SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Discrete Dynamical Systems, Chaotic Behavior and Fractal Geometry

av

Johan Lord

2017 - No 33

Discrete Dynamical Systems, Chaotic Behavior and Fractal Geometry

Johan Lord

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Yishao Zhou

2017

Abstract

In this thesis we give an introduction to the dynamics of difference equations and the complex behaviour that sometimes arise out of seemingly simple equations. We start off with a short historical introduction and a comparison with differential equations. We introduce the concept of dynamics of difference equations, where we define and explain concepts such as: orbit, fix points, periodic points, and discuss the notion of stability, as well as state and prove criteria for determining stability of periodic points. We then slowly introduce non-linear dynamics through the example of population models and the logistic map, and we also discuss the theory of bifurcations. We give a short historical introduction to chaotic dynamics, and after making the necessary definitions, we give our definition of chaotic behaviour. Different definitions of chaotic behaviour are discussed, mainly the one due to Devaney, and we briefly address the various ambiguities regarding definitions in this rather recent field of research. After introducing a possible quantification of chaotic behaviour, through the concept of Lyapunov exponents, we move from the dynamics to the geometric aspects of chaotic systems, via fractal geometry. Classical notions of dimension are discussed via e.g. the Lebesgue covering dimension, and with a few examples of fractals, we give some intuition for how and why these classical ideas may be extended to something called fractal dimension. We then give a thorough explanation of different measures of fractal dimension, and apply these ideas to chaotic attractors of dynamical systems, in the form of Renyi dimension. Results of the authors own numerical estimations of the dimension of well known chaotic attractors are presented, and we tie together the dynamics with the geometry side of things with a discussion of the Kaplan-Yorke conjecture. Lastly we give a few concluding remarks and a brief discussion of potential applications to number theory.

Acknowledgements

I would like to thank my supervisor Yishao Zhou for her support, and for always being available on sometimes very short notice. I would also like to thank Torbjörn Tambour for his support throughout the whole project, and also for taking such interest in my work, even though it is somewhat outside his field of research.

Contents

1 Introduction

This thesis is written as an introduction to the study of discrete dynamical systems and their dynamics. These are one or more difference equations, also often called recurrence relations, that form a system of equations in the same sense one or more differential equations would form a system. In that sense they are the discrete counterpart of the continuous systems, with which the reader might already be familiar. These equations form a set of rules that govern how a point or state is affected as time progresses. We use the word 'time' here, as this is a very common case, because many of the systems we study come from biological or physical applications, where time is ever present. For this reason, the systems are often referred to as *continuous time systems* and *discrete time systems*. For example, the position of a stone being hurled from a catapult, may be described by a set of equations where time certainly plays an important role, and this would most likely benefit from a continuous time model. However, when describing for example growth of a population, it could definitely be of interest to let the independent variable be the number of generations, in case it would most likely be modelled by a discrete time system. We can of course never really work with infinitely small time intervals in real life, so continuous time models often end up being approximated by discrete systems after all, for example when running numerical simulations on a computer. The system could then be viewed as taking snapshots of it, at sufficiently short intervals. The reader can think of a film of the stone from before; it looks like it is moving through the air continuously, but we know it consists of many still images, with time intervals of $\frac{1}{25}s$ between each one.

It is important to remember that these are all mere applications of the mathematics, and although very useful, they are just a few from an infinite set of applications. It is therefore just as important to study the mathematics itself, as it is as real as the physical phenomenon for which it can be used to study. We should therefore think not so much of time progressing, as of moving through the solution space. This may sound very abstract, and that is the point, but it is not stranger than moving through time (perhaps less so in a way, and for all we know, that is what time really is). As an example, take the difference equation $x_{n+1} = 2x_n$. This just says to take an initial condition, x_0, and double it in order to get x_1; double this to get x_2; double that to get x_3, and so on. It is clear that these rules give rise to a sequence, namely to $(x_n)_{n=0}^{\infty}$. A solution to this equation is then a sequence for which these rules apply. It is clear that such a solution depends on x_0, for if $x_0 = 0$, then $(x_n) = (0, 0, 0, ...)$, but if $x_0 = 1$, then $(x_n) = (1, 2, 4, 8, ...) = 2^n, n = 1, 2, 3, ...$ These solutions can be seen as infinite dimensional vectors in the vector space of solutions, and moving through "time" could then just be seen as moving along these vectors. So, especially for discrete systems, the word 'time' just refers to a specific point in the solution sequence. We will also see that this, in turn, just corresponds to a

specific number of iterations of a function applied to the initial value. For now, the point is that the reader should not feel forced to think of time, when describing the progression of a system from a set of rules. We will sometimes still use the word 'time' in this sense, or we may just say that the system progresses.

As we explore different systems further on, we will see that remarkable complexity will come from seemingly simple systems. We will encounter systems, even among these simple ones, that undergo such extreme changes that they become, in a sense, unpredictable. As we are still dealing with a deterministic system, we do not mean this in a literal sense, but even the smallest perturbations to the initial conditions, will give rise to huge changes further on. In practice, this means that the long term behaviour becomes near unpredictable. This type of behaviour is called chaotic, and as a field of study, it is considered by most to be very much in its infancy. Before we go on to study dynamics, we will take a look at the possible origins of difference equations, in a brief historical section.

1.1 A Little History

In the third section of the book "Liber Abaci" (which roughly translates to "The Book of Calculations"), published in 1202, and written by Leonardo Pisano (nicknamed Fibonacci), a problem was proposed. The now famous problem concerned the evolution of a population of rabbits. If we assume that a pair of rabbits always give birth to two more rabbits each month, and that it takes just under two months for a rabbit to mature, then how will this population evolve? Taking time to be discrete (counting in months and pairs), we see that if we begin with one pair, there won't be any more until two months later, when there will be two pair. This new pair won't give birth until after two months, during which the first pair will have given birth to two new pair, and so on[1]. This give rise to the sequence that bears Leonardo Pisanos nickname, namely the Fibonacci sequence $(1, 1, 2, 3, 5, 8, ...)$. Each term of the sequence is the sum of the two preceding terms, and thus we can state this as the following difference equation:

$$x_{n+2} = x_{n+1} + x_n. \tag{1}$$

Equation (1) is an example of a second order difference equation, and the Fibonacci sequence $(1, 1, 2, 3, 5, 8, 13, ...)$ is a solution to this equation. If we asked for the 111th term, this would be a cumbersome task, as we would have to find it by recursively applying the rules in (1). Thankfully there are algebraic methods for solving these equations, which in practice results in a formula of the form $x_n = f(n)$.

The reader might have noticed that we started the Fibonacci sequence with 1

[1]This is of course a rather unrealistic model since we would probably have to consider a lot more factors, for this to accurately model a real population. Also the model would not likely be linear, as we shall discuss in a later section.

rather than 0, which certainly seems more common these days. This is because Fibonacci himself started at 1. Understanding where the problem comes from, a pair of rabbits does not just jump into existence from nothing do they? Also, it really does not matter from what number we start. The nature of the equation is what's important, not what part of a solution we decide to write down. We may as well describe it for negative n. By manipulating (1) we get $x_n = x_{n+2} - x_{n+1}$, which for $x_0 = 1, x_1 = 1$ would yield $(...13, -8, 5, -3, 2, -1, 1, 0, 1, 1)$, for $n = -1, ..., -8$. As we see, this mirrors the previous sequence for positive n, but with alternating sign.

1.2 Continuous vs. Discrete

Before looking at the equations and their solutions, it may be a good idea to have some intuition of the difference between the continuous and the discrete case. We will try to supply this by briefly investigating how we could approximate the one with the other.

As explained in the introduction, difference equations is the discrete counterpart of differential equations (which are equations involving differentials), as they are equations involving differences. Differential equations explain how a system evolves continuously, by describing how one or more of a functions derivatives change. To give some intuition of this, we make the following definition:

$$\Delta x_n := x_{n+1} - x_n \Rightarrow x_{n+1} = x_n + \Delta x_n. \tag{2}$$

So we get the next value in the solution by adding Δx_n to the previous term, which makes sense.

A first order differential equation is usually of the form

$$\frac{dy}{dx} = f(x, y(x)). \tag{3}$$

With a little informal algebra of differentials, we may just as well view this as

$$dy = dx f(x, y(x)),$$

where dy is simply the change in y. Similarly, if this was a discrete system, starting with y_0, one would just like in (2) get the next term by adding the change, i.e., $y_1 = y_0 + dx f(x_0, y_0)$. In general we would have the formula

$$y_{n+1} = y_n + dx f(x_n, y_n). \tag{4}$$

Exchanging dx for h, and making the definition $x_n = x_0 + nh$, the reader may recognize (4) as Euler's method for approximating a solution to an initial value problem, often referred to as 'Forward Euler'. This is a difference equation

that can be seen as an approximation of (3). The reason why this is only an approximation, is of course that in the continuous case, we are dealing with infinitely small changes, while in the case of (4), we would have to settle for dx to be finite. In this case we would usually write it as Δx instead. We will see more similarities between the two cases in later sections.

1.3 Solutions and Iterated Maps

By a **difference equation** we mean an equation of the form

$$x_n = f(x_{n-1}, x_{n-2}, ..., x_{n-k}, n), \tag{5}$$

where the function f is usually a function from $\mathbb{R}^{k+1}$ to $\mathbb{R}$, often called the recursion function. When this is linear, (5) is called a linear difference equation of order k [2], and the equation takes the form

$$x_n = g_1(n)x_{n-1} + g_2(n)x_{n-2} + ... + g_k(n)x_{n-k} + h(n), \tag{6}$$

where the coefficients $g_1, ..., g_k, h$ are complex valued functions.

Just as with differential equations, the problem of finding a solution is usually stated as an initial value problem, i.e. to find a sequence that solves (6), given k initial values $x_0, ..., x_{k-1}$. What it means, in practice, is finding a general closed formula for x_n. The methods for finding such solutions differ for different types of equations, and even though the theory for difference equations are somewhat simpler than for differential equations, there are still a lot to say on this subject. There is however one important and helpful fact, that may seem obvious, about equations of the form of (6), and that is that every initial value problem for a finite difference equation has a unique solution. This is evident from the fact that $x_n = f(x_{n-1})$ is uniquely determined by x_{n-1}, and since we start with x_0 which then determines the whole solution.

This is one of the big differences between differential equations and difference equations, that in the continuous case we are not always guaranteed a solution. In the case of differential equations, we are often resorted to approximating solutions, using for example Euler's method, described in the previous section.

A solution to a differential equation is a function on $\mathbb{R}^n$, while a solution to a difference equation is a function on $\mathbb{N}$, i.e. a sequence.

For the most part, we will restrict ourselves to the special case when $g1, ..., gk$ are constant functions, i.e. they do not depend on n (autonomous case), and when $h(n) = 0$ (homogeneous case). In this case the equation is called autonomous and homogeneous.

[2] Actually this is a finite difference equation as we could consider an equation of infinite order. We choose to focus on finite ones however.

Instead of defining x_n as a function of preceding terms, we can of course use the fact that each term is determined by repeatedly applying f to the initial condition x_0. This is then referred to as an iterated map, and by the n:th iterate of the function f, we mean the n:th power of f under function composition. We write this as f^n, and we let $f^0 := Id$, where Id denotes the identity function. This means that we have

$$f^n = \underbrace{f \circ f \circ \ldots \circ f}_{n \text{ times}}.$$

For a solution to the one dimensional system $x_n = f(x_{n-1})$, starting with the initial condition x_0, we then have

$$\begin{aligned} x_0 &= Id(x_0) = f^0(x_0), \\ x_1 &= f(x_0) = f^1(x_0), \\ x_2 &= (f \circ f)(x_0) = f^2(x_0), \\ &\vdots \\ x_n &= f^n(x_0). \end{aligned} \tag{7}$$

It therefore makes sense to just write our discrete dynamical system as (7). Starting with x_0, the solution sequence then takes the form $(f^0(x_0), f^1(x_0), f^2(x_0), \ldots)$, which can be shortened to $(f^n(x_0))_{n \geq 0}$.

If the equation would be of a higher order, we would have to expand this to higher dimensions. Let the k:th order equation be given by

$$x_{1n} = f(x_{1n-1}, x_{1n-2}, ..., x_{1n-k}).$$

We can then rewrite this as a system of first order equations in the following way

$$\begin{cases} x_{1n} = f(x_{1n-1}, x_{2n-1}, ..., x_{kn-1}) \\ x_{2n} = x_{1n-1} \\ \vdots \\ x_{kn} = x_{k-1n-1}. \end{cases}$$

Then by letting $\bar{\Psi}(x_1, x_2, ..., x_k) = (f(x_1, x_2, ..., x_k), x_1, x_2, ..., x_{k-1})$ and $\bar{x} = (x_1, x_2, ..., x_k)$, we get a formula of the same form as (7), namely

$$\bar{x}_n = \bar{\Psi}^n(\bar{x}_0), \tag{8}$$

where each state $\bar{x}_i$ is a k-dimensional vector.

For example, rewriting the equation $x_n = x_{n-1} + x_{n-2}$ in this way, we get

$$\begin{cases} x_n = x_{n-1} + y_{n-1} \\ y_n = x_{n-1}. \end{cases}$$

Thus we can say that $F(x, y) = (x + y, x)$ describes the system and we see that

$$(F^n(1,0))_{n=0}^{\infty} = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 3 \end{pmatrix}, ... \right).$$

If we read off the sequence $(y_n)_{n=0}^{\infty}$ we recognize this as the ordinary Fibonacci sequence.

We will frequently be referring to the recursion function as the function describing the system. We may also refer to a system as 'a system described by f', or 'a system governed by f'. Also, for the remainder of this thesis, we will not distinguish k-dimensional vectors from 1-dimensional ones by using $\bar{x}$ to denote the former. It should always be either explicitly stated when defining the systems, or be clear from context, what dimension the states are in.

2 Dynamical Properties

Dynamical properties usually indicate properties of a system unchanged as the system progresses. For instance, this incorporates the properties of being fixed, periodic, and stable for points as well as for orbits. We shall also see that it is of great interest to see how one system, with a change in one or more of the parameters, could suddenly change properties completely, and thus become a different system. This phenomenon is called 'bifurcation', and will be studied in Section 2.2.2. Let us now look at the behaviour of different points under iteration of a function f. This is essentially what we mean by the dynamics of a system.

2.1 Orbits, Fix Points and Periodic Points

For a system described by a function f, we call the sequence of points

$$(x_0, f(x_0), f^2(x_0), \ldots, f^n(x_0), \ldots)$$

the **orbit** of x_0 under f, and when it is clear what function is describing the system, we denote the orbit of x_0 by $O(x_0)$. Sometimes we also write out the orbit as $(x_0, x_1, x_2, ...)$. Some comments on notation is needed here. The phrase 'orbit of x_0', refers to the sequence $(f^n(x_0))_{n=0}^{\infty}$, sometimes also written $(f^n(x_0))_{n\geq 0}$. However, sometimes it is convenient to refer to the set of distinct

points of $O(x_0)$, and usually this is understood from context. Thus we may refer to the sequence $O(x_0)$ as a set if it is clear from context what we mean. Otherwise we will make the distinction explicitly. For the set of points of $O(x_0)$ we will write $\{f^n(x_0)\}_{n\geq 0}$ and for the sequence we will write $(f^n(x_0))_{n\geq 0}$. As a sequence, the orbit is infinite, while the set of points $\{f^n(x_0)\}_{n\geq 0}$ can be finite if it is periodic (explained below), or infinite if it is not. As an example, the orbit of $x_0 = 1$ under iteration of $f(x) = -x$ is $((-1)^{n+1})_{n\geq 0}$, but $\{f^n(x_0)\}_{n\geq 0} = \{-1, 1\}$.

Some set of points exhibit properties of special interest. The first of these sets is the set of fix points of f.

Definition 2.1 *A point x_s is called a **fix point** of a function f, (sometimes referred to as a **stationary state** of $x_n = f(x_{n-1})$) if $f(x_s) = x_s$.*

We often denote such a point by x_s, but in general it should be explicitly stated if a point is a fix point. A rather intuitive corollary to this is that if x_s is a fix point of f, then x_s is a fix point of f^n. As an example, the set of fix points for

$$f(x) = -3x + 2 \tag{9}$$

is $\{\frac{1}{2}\}$, since $f(\frac{1}{2}) = \frac{1}{2}$, and there is no other solution to $f(x) = x$. The fix points of $Id : \mathbb{R} \to \mathbb{R}$ however, is $\mathbb{R}$. We understand that finding fix points is just a matter of solving the equation $f(x) = x$.

A generalization of this concept is the notion of periodic points. This means that you get back to a previous point after a number of iterations, and the sequence repeats itself.

Definition 2.2 *A point x is called a **periodic point**, with period n, of a function f, if x is a fix point of f^n, $n \in \mathbb{N}$. The smallest positive such n is called the **prime period** of x.*

By this definition a fix point is just a period-1 point. In the above example, $x = \frac{1}{2}$ is a period 1 point of (9). It is of course also a period 1489 point. Its prime period however is 1. For the map $f(x) = -x$, $x = 1$ is a period 2 point, since $f^2(1) = f(f(1)) = f(-1) = 1$. The period 2 points of $f(x) = -x$ is $\mathbb{R}$, but the only period 1 point is $\{0\}$. Often, when referring to a point of prime period n, we usually just call it n-periodic. One could also talk about a point x as being **eventually n-periodic**, if $f^n(x)$ eventually becomes n-periodic as n increases. We also say that a point x is **forward asymptotic** to a point p with prime period k, if $\lim_{n\to\infty} f^{nk}(x) = p$. The **stable set** of p, denoted $W^s(p)$, is then all points that are forward asymptotic to p.

Just as with points, we will also refer to a whole orbit, where $|O(x_0)| = n$, as n-periodic. For orbits, we make the following definitions:

Definition 2.3 *A point p is called a **limit point** of $O(x_0)$, if there exists a subsequence $(x_{n_k})_{k\geq 0}$ in $O(x_0)$, such that $x_{n_k} \to p$, as $k \to \infty$. Further, we call the set of all limit points of $O(x_0)$ the **limit set** of the orbit of x_0, and we denote it $L(x_0)$.*

For example, if x_s is a fix point, then $\{f^n(x_s)\}_{n\geq 0} = \{x_s\}$ and $L(x_s) = \{x_s\}$. It is also easy to see that if $\{f^n(x_0)\}_{n\geq 0} = \{x_0, x_1, ..., x_n\}$, i.e. $O(x_0)$ is $(n+1)$-periodic, then $L(x_0) = \{x_0, x_1, ..., x_n\}$ has $(n+1)$ points.

It is important to remember that a limit point of a sequence is not defined in the same way that one usually defines limit points of sets. For a set A, in any topological space, one usually defines a limit point $p \in A$, as a point such that every neighbourhood of p contains a point of A different from p.

Definition 2.4 *For an orbit $(f^n(x_0))$, we will say that it is **asymptotically stationary**, if $L(x_0) = \{x_s\}$, for some fix point x_s, we call it **asymptotically periodic** if $L(x_0)$ is finite, and **eventually stationary**, if $f(x_n) = x_n$ for some $n \geq 1$, and **eventually periodic**, if $f^k(x_n) = x_n$ for some $k > 1$.*

Intuitively, one can think of an asymptotically stationary or asymptotically periodic orbit, as an orbit converging to a fix point x_s, or to a periodic orbit respectively, but not necessarily attaining the values at those points. It is clear that eventually stationary and eventually periodic implies asymptotically stationary and asymptotically periodic, respectively. However, let $f(x) = x(1-x)$, then $O(\frac{1}{2}) = (\frac{1}{2}, \frac{1}{4}, \frac{3}{16}, ...)$ would be an example of an orbit that is asymptotically stationary, since $L(\frac{1}{2}) = \{0\}$ and 0 is a fix point of f. It is, however, not eventually stationary, since there is no n such that $f(x_n) = x_n$.

One question, that may arise from the definition above, is what happens if an orbit is not asymptotically periodic. We make the following definition:

Definition 2.5 *We call $O(x_0)$ **aperiodic** if $L(x_0)$ is not finite.*

In this case the set $\{f^n(x_0)\}_{n\geq 0}$ is infinite, since the orbit never settles down into an periodic orbit. Further it is not asymptotically periodic either, since $L(x_0)$ is infinite. Informally but intuitively, one can think of the orbit as never settling down into a predictable behaviour. It is not hard to realize that this property is of key importance in the study of chaotic behaviour. Another feature of aperiodicity is that, since all points in the orbit are distinct, the orbit actually visits every open neighbourhood of each of the points of $L(x_0)$. Thus, the points of $L(x_0)$ are also limit points of the set $\{f^n(x_0)\}_{n\geq 0}$. In later sections we will see that aperiodicity is not enough for a map to be called chaotic, and we will develop more properties needed in the sections to come.

Before going further, we will first look at a few of the introduced concepts in the simple case of linear and affine mappings.

2.1.1 Linear and Affine Mappings

By a linear one dimensional mapping $f : \mathbb{R} \to \mathbb{R}$, we mean one of the form $f(x) = ax, a \in \mathbb{R} \setminus \{0\}$. Finding fix points to such mappings is of course very easy, and there are a few conclusions one may draw about these mappings and their fix points. First we note that 0 is always a fix point, and that it is the only fix point since it is the only solution to $ax = x$; unless $a = 1$ because then $f = Id$, and every point is a fix point.

Since the solution to $x_n = ax_{n-1}$ is described by $x_n = a^n x_0$, we see that if $|a| < 1$, then

$$\lim_{n\to\infty} a^n = 0.$$

Therefore, no matter where we start, the solution will always tend to 0, i.e $\lim_{n\to\infty}\{x_n\} = 0$. So, all points $x \in \mathbb{R}$ are asymptotically 1-periodic, or equivalently, asymptotically stationary. We also see that $W^s(0) = \mathbb{R}$. However, if $|a| > 1$, then

$$\lim_{n\to\infty} a^n = \infty,$$

and $\lim_{n\to\infty}\{x_n\} = \infty$. We then say that $W^s(\infty) = \mathbb{R}\backslash\{0\}$ (since if we start at 0, we stay there).

If we translate the map, so that its graph does not intersect the origin, we see that we are not always guaranteed a fix point, since the equation $ax+b = x$ does not have a solution if $a = 1, b \neq 0$. Mappings where $b \neq 0$ are not called linear, and they are instead called **affine** (i.e mappings that have a graph consisting of a line, which of course also include linear maps). Affine mappings where $a \neq 1$ has one fix point, namely the solution to $ax + b = x$. Generalizing the above discussion, we may state the following:

Theorem 2.1 *For any affine mapping $f(x) = ax + b$ we have that if $a = 1, b = 0$, then every point is a fix point. If $a = 1, b \neq 0$, then the mapping has no fix points. If $|a| < 1$, then $O(x_0)$ converges to the unique fix point p, which is the solution of $ax + b = x$. If $|a| > 1$ and $x_0 \neq p$, then $O(x_0)$ diverges away from p. If $a = -1$, then every point except p has prime period 2.*

We will now develop our collection of properties of orbits and periodic points even further, by looking at the concept of stability of periodic points.

2.1.2 Stability of Periodic Points

As we saw in the previous example of an affine map f, if $|a| < 1$ the orbit of f tends to the fix point, while it diverges away from the fix point if $|a| > 1$. Thus there seems to be some aspects of stability connected with a fix point. We make the following definitions:

Definition 2.6 *Let x_s be a fix point of f and let $N_\alpha(p)$ denote a neighbourhood of radius α around the point p. We then say*

i) x_s *is* ***stable****, if for each $\epsilon > 0$ there exists a $\delta > 0$ such that*

$$f^n(N_\delta(x_s)) \subset N_\epsilon(x_s), \quad \forall n \in \mathbb{N}$$

ii) x_s *is* ***unstable*** *if it is not stable.*

iii) x_s *is* ***asymptotically stable*** *or* ***attracting****, if x_s is stable and there exist a neighbourhood $N_r(x_s)$ such that $x \in N_r(x_s) \Rightarrow \lim_{n\to\infty} f^n(x) = x_s$.*

iv) x_s *is* ***repelling*** *if there exist a neighbourhood $N_r(x_s)$ such that $x_0 \in N_r(x_s)\backslash\{x_s\} \Rightarrow f^n(x_0) \notin N_r(x_s)$ for sufficiently large n.*

Informally speaking, a fix point x_s of a map f is stable if points within a neighbourhood $N_\delta(x_s)$ of x_s stay close to x_s under iteration of f. To be precise, we would call x_s **locally stable** if the neighbourhood $N_\delta(x_s)$ is finite. Otherwise we would call it **globally stable**, in which case points in $O(x_0)$ can not get arbitrarily far away from x_s, no matter where we start. Since a globally stable fix point is of course also locally stable, we will often just refer to locally stable as stable if not otherwise stated. The most common term for an asymptotically stable fix point, and the one we will use, is 'attracting' since $O(x_0)$ is "moving" closer and closer to x_s and is thus "attracted" by it. In case (iv), $O(x_0)$ instead diverges away from the fix point, as in the affine case where $|a| > 1$, it is therefore called a **repelling** fix point, and is thus unstable (a fix point can however be unstable without being repelling).

To arrive at a criteria for stability, we consider the difference equation

$$x_n = f(x_{n-1}), \tag{10}$$

where $f : S \to S \subset \mathbb{R}$ is differentiable. By Taylor expansion around the fix point x_s, we get that

$$f(x_{n-1}) = f(x_s) + f'(x_s)(x_{n-1} - x_s) + \mathcal{O}((x_{n-1} - x_s)^2) \Leftrightarrow$$

$$\Leftrightarrow x_n - x_s = f'(x_s)(x_{n-1} - x_s) + \mathcal{O}((x_{n-1} - x_s)^2).$$

Thus the linear system of differences

$$y_n = f'(x_s)y_{n-1},$$

where $y_n = x_n - x_s$, approximates (10) when x_n is close to x_s. We then know from our earlier discussion that $\lim_{n\to\infty}\{y_n\} = 0$ if $|f'(x_s)| < 1$, and that

$\lim_{n\to\infty}\{y_n\} = \infty$ if $|f'(x_s)| > 1$. Now obviously, if x_s is attracting, then $\{y_n\} \to 0$, as $n \to \infty$.

To state and prove the criteria for stability of fix points of differentiable one dimensional mappings, we first need the following lemma:

Lemma 2.1 [3] *If $f : S \to S$ is differentiable at $p \in S \subseteq \mathbb{R}$ and $|f'(p)| < 1$, then there is a positive number $a < 1$ and a neighbourhood $N_r(p)$ such that, for all $x \in N_r(p)$,*

$$|f(x) - f(p)| \leq a|x - p|.$$

Similarly, if $|f'(p)| > 1$, then there is a positive number $a > 1$ and a neighbourhood $N_s(p)$, such that for all $x \in N_r(p)$,

$$|f(x) - f(p)| \geq a|x - p|.$$

Proof. Let $\lim_{x\to p} \dfrac{f(x) - f(p)}{x - p} = f'(p)$ and suppose $|f'(p)| < 1$. We can then choose $a \in (0, 1)$ such that $f'(p) \in (-a, a)$. By the definition of limit, we then have that there exists a neighbourhood $N_r(p)$, such that
$x \in N_r(p), x \neq p \Rightarrow f'(p) \in (-a, a)$. Thus
$x \in N_r(p), x \neq p \Rightarrow -a < \dfrac{f(x) - f(p)}{x - p} < a \Leftrightarrow \left|\dfrac{f(x) - f(p)}{x - p}\right| < a$. So, for all $x \in N_r(p)$, we have $|f(x) - f(p)| \leq a|x - p|$.

Similarly, if $|f'(p)| > 1$, then we can instead pick $a > 1$ such that $1 < a < |f'(p)|$. By an analogous argument we may deduce that there exists a neighbourhood $N_s(p)$ such that $x \in N_s(p), x \neq p \Rightarrow \left|\dfrac{f(x) - f(p)}{x - p}\right| > a$. So, for all $x \in N_s(p)$, we have $|f(x) - f(p)| \geq a|x - p|$. □

Lemma 2.1 can be used to prove the following theorem about fix points of differentiable maps.

Theorem 2.2 *Suppose x_s is a fix point of the map $f : S \to S$, where $S \subset \mathbb{R}$, and that f is differentiable at x_s with $|f'(x_s)| \neq 1$, then*

i) $|f'(x_s)| < 1 \Rightarrow x_s$ is stable, and there exist a neighbourhood $N_r(x_s)$ such that $x_0 \in N_r(x_s) \Rightarrow \lim_{n\to\infty} f^n(x_0) = x_s$. Thus x_s is an attracting fix point.

ii) $|f'(x_s)| > 1 \Rightarrow x_s$ is unstable, and there exist a neighbourhood $N_r(x_s)$ such that $x_0 \in N_r(x_s)\backslash\{x_s\} \Rightarrow f^n(x_0) \notin N_r(x_s)$ for sufficiently large n. Thus x_s is a repelling fix point.

[3] This lemma and the following theorem, are modified and slightly extended versions of Lemma 5.2.2 and Theorem 5.2.1 in Banks [2]

Proof. Let x_s be a fix point of f, and assume first that $|f'(x_s)| < 1$. From Lemma 2.1 and the fact that $f(x_s) = x_s$, we then get that there exists a positive $a < 1$, and a neighbourhood $N_r(x_s)$, such that $x \in N_r(x_s) \Rightarrow |f(x) - x_s| \leq a|x - x_s|$. Since $x_n = f(x_{n-1})$, $a \in (0, 1)$ and we assume that $x_0 \in N_r(x_s)$, we get that

$$\begin{aligned} |x_1 - x_s| \leq a|x_0 - x_s| \Rightarrow x_1 \in N_r(x_s) \\ \Rightarrow |x_2 - x_s| \leq a|x_1 - x_s| \leq a^2|x_0 - x_s| \Rightarrow x_2 \in N_r(x_s) \\ \Rightarrow |x_3 - x_s| \leq a|x_2 - x_s| \leq a^3|x_0 - x_s| \Rightarrow x_3 \in N_r(x_s) \\ \ldots \Rightarrow |x_n - x_s| \leq a|x_{n-1} - x_s| \leq a^n|x_0 - x_s|. \end{aligned}$$

Since $a^n \to 0$, as $n \to \infty$ and $|x_0 - x_s|$ is a constant, $x_n \to x_s$, as $n \to \infty$ and $f^n(N_r(x_s) \subseteq N_r(x_s)$. So x_s is stable and $x_0 \in N_r(x_s) \Rightarrow \lim_{n\to\infty} f^n(x_0) = x_s$.

Suppose instead that $|f'(x_s)| > 1$ but that x_s is stable. Then there exists a neighbourhood $N_\epsilon(x_s)$ such that $x \in N_\epsilon(x_s) \Rightarrow |f(x) - x_s| \geq a|x - x_s|$ by the second part of Lemma 2.1. By the definition of stable there also exists a neighbourhood $N_\delta(x_s)$ such that $f^n(N_\delta(x_s)) \subseteq N_\epsilon(x_s)$ for all $n \in \mathbb{Z}_+$. So $x_0 \in N_\delta \Rightarrow f^n(x_0) \in N_\epsilon(x_s)$ for all $n \in \mathbb{Z}_+$. By a similar reasoning as before we get that

$$\begin{aligned} |x_2 - x_s| \geq a|x_1 - x_s| \\ |x_3 - x_s| \geq a^2|x_1 - x_s| \\ \vdots \\ |x_k - x_s| \geq a^{k-1}|x_1 - x_s|, \end{aligned}$$

and since $a > 1$, $a^{k-1} \to \infty$, as $k \to \infty$. So, for sufficiently large k, $f^k(x_0) \notin N_\epsilon(x_s)$. So $f^n(N_\delta(x_s)) \not\subseteq N_\epsilon(x_s)$, which contradicts the fact that x_s was stable, and hence x_s is unstable. Also, x_s is an repelling fix point. □

In the case when $|f'(x_s)| = 1$, x_s is usually called **indifferent**, and it can be attracting, repelling or actually both. Thus we can not say much about the stability of an indifferent fix point by analysing the first derivative; for this we would need more sophisticated methods.

Since a point p is a n-periodic point of f, if it is a fix point of f^n, we also see that for a periodic point p of prime period n, we have that

$$|(f^n)'(p)| < 1 \Rightarrow \quad p \text{ is an attracting fix point of } f^n,$$

$$|(f^n)'(p)| > 1 \Rightarrow \quad p \text{ is repelling fix point of } f^n.$$

A similar argument, as the one using Taylor expansion for (10), works even in higher dimensions. So, for a function $F : S \to S \subset \mathbb{R}^n$, the n-dimensional

system given by

$$x_n = F(x_{n-1}) \tag{11}$$

still has an approximation

$$y_n = F'(x_s)y_{n-1},$$

where $y_n = x_n - x_s$, and where $F'(x_s)$ is now the Jacobian of F evaluated at x_s. So we have approximated the system (11) by this linear system, where $F'(x_s)$ is a matrix, let's call it J. We then have that $y_n = J^n y_0$. We also know from linear algebra that if J is diagonalizable, then we can write $J^n = P\Lambda^n P^{-1}$, where Λ is a diagonal matrix with the eigenvalues of J on its diagonal. Now, if $\lambda_i, i = 1, ..., n$ are the eigenvalues of J, and $|\lambda_i| < 1$ for $i = 1, ..., n$, then $\mathbf{y}^n \to 0$, as $n \to \infty$, but if $|\lambda_i| > 1$ for at least one i, then Λ^n will blow up and $\mathbf{y}^n \to \infty$, as $n \to \infty$.

For a square matrix A that is not diagonalizable, there is still always a block diagonal matrix Θ, called the Jordan canonical form of A, such that $A = T\Theta T^{-1}$. So, we still have that $A^n = T\Theta^n T^{-1}$. The matrix Θ has the following form:

$$\Theta = \begin{pmatrix} \Theta_1 & & \\ & \ddots & \\ & & \Theta_p \end{pmatrix},$$

where each block Θ_i is a square matrix of the form

$$\Theta_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix},$$

where λ_i is the i:th eigenvalue of A.

The power of a block diagonal matrix $B := B_1 \oplus \cdots \oplus B_m$ is the direct sum of the powers of the blocks, i.e $B^k = B_1^k \oplus \cdots \oplus B_m^k$. From a more general formula, for applying a matrix function to a Jordan block, that we won't state here, one can derive the following formula for the n:th power of a $m \times m$ Jordan block Θ_i:

$$\Theta_i^n = \begin{pmatrix} \lambda_i^n & \binom{n}{1}\lambda_i^{n-1} & \binom{n}{2}\lambda_i^{n-2} & \cdots & \binom{n}{m-1}\lambda_i^{n-m+1} \\ 0 & \lambda_i^n & \binom{n}{1}\lambda_i^{n-1} & \cdots & \binom{n}{m-2}\lambda_i^{n-m+2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_i^n & \binom{n}{1}\lambda_i^{n-1} \\ 0 & 0 & \cdots & 0 & \lambda_i^n \end{pmatrix}$$

To find out what happens to Θ^n for large n, we look at the limit of the elements of the j:th super diagonal. We first note that

$$\begin{aligned}\binom{n}{j}\lambda_i^{n-j} &= \frac{n(n-1)(n-2)\dots(n-k+1)}{k!}\lambda_i^{n-j} \\ &= \frac{(1-\frac{1}{n})(1-\frac{2}{n})\dots(1-\frac{k-1}{n})n^{k+2}}{k!}\lambda_i^{n-j},\end{aligned}$$

where we get the last equality by multiplying with $\frac{n^{k+2}}{n^{k+2}}$. For large n we then have that

$$\binom{n}{j}\lambda^{n-j} \approx \frac{n^{k+2}}{k!}\lambda^{n-j}.$$

From this we derive that

$$\begin{aligned}\lim_{n\to\infty}\binom{n}{j}\lambda^{n-j} &= 0, \quad \text{for} |\lambda| < 1 \\ \lim_{n\to\infty}\binom{n}{j}\lambda^{n-j} &= \infty, \quad \text{for} |\lambda| > 1.\end{aligned} \tag{12}$$

A useful terminology about matrices, that is often used, is the **spectrum** of a matrix. The spectrum of a matrix A, denoted σ_A, is just the set of its eigenvalues. We also often talk about a matrix **spectral radius**. We make the following definition:

Definition 2.7 *The spectral radius of a matrix A is denoted $\rho(A)$, and we define it as:*

$$\rho(A) = max\{|a| : a \in \sigma_A\}.$$

With the above discussion in mind we now state, without a formal proof, a generalization of the earlier theorem about the stability of fix points of one-dimensional maps.

Theorem 2.3 [4]

Let x_s be a fix point of the continuous function $f : S \to S$, where $S \subset \mathbb{R}^m$, and assume that f is differentiable in a neighbourhood of x_s with continuous derivative at x_s. Let $J = f'(x_s)$ be the Jacobian of f evaluated at x_s. Then

i) x_s *is an attracting fix point if* $\rho(J) < 1$,

ii) x_s *is an unstable fix point if* $\rho(J) > 1$,

iii) x_s *is a repelling fix point if* $min\{|a| : a \in \sigma_J\} > 1$.

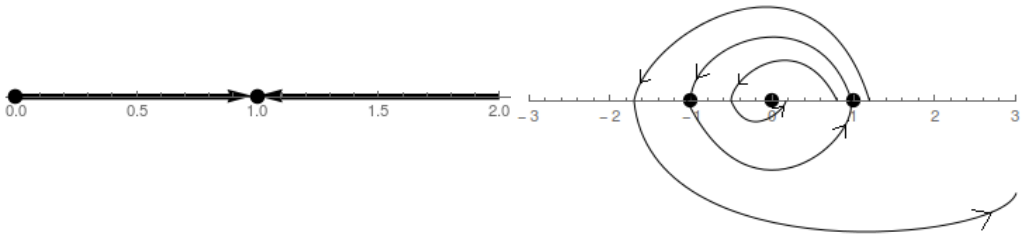
This of course generalizes Theorem 2.2, since the Jacobian of a one dimensional map $f : S \to S \subseteq \mathbb{R}$ is just the 1×1-matrix $\left(\frac{df}{dx}\right)$. Thus, in that case we just get $\rho(J) = |\frac{df}{dx}|$.

Before moving on to non-linear maps, as a side note, it can be a good idea to introduce some graphical aids. As things get more complicated it usually helps to illustrate them to get a more complete picture.

2.1.3 Graphical Tools

One common way to illustrate the evolution of a dynamical system is by plotting a so called **phase portrait**. This is just a graph of the states of the system, i.e a diagram showing where the rules takes a specific point, and where that point goes on from there, and so on. This is more commonly used for continuous systems, as there is a natural vector field associated with the system, that you can also plot. This gives you a very nice presentation of the evolution of different starting positions. For a discrete system, however, the next state of the system can be very far away from the current one, and since we do not get smooth curves describing the evolution of an initial condition, the phase portrait would get quite messy. We still use the idea at times when it suits us however. For example, say we wanted to illustrate how the one dimensional system represented by $f(x) = \sqrt{x}$ evolves when starting from different points. We could simply start from x_0, and then plot the sequence $f^n(x_0), n = 0, 1, 2, ...$ This would tell us where the points are, but not where to go from a specific point. Instead we indicate this by arrows. We usually also mark fix points by larger dots. If all points within a given interval converges (diverges) to (away from) a fix point, we represent this by an arrow covering the interval and pointing towards (away from) the fix point. When the orbit does not converge (diverge) monotonically, i.e when it jumps between two sides of a fix point on the real line, as

[4] This theorem is somewhat of a compound of several theorems that can be found in different literature on the subject. For a proof of part (i) see Theorem 5.3.3 in [25], for part (ii) see Theorem 10.4.2 in [44], and for part (iii) see Theorem 5.4.1 in [25].



(a) $f(x) = \sqrt{x}$ (b) $g(x) = -x^3$

Figure 1: Example of phase portraits for two one dimensional maps. The large dots indicate fix points, and the arrows indicate where a point is taken under iteration. Thick arrows indicate that all points in the interval covered by the arrow is taken in the direction of the arrow.

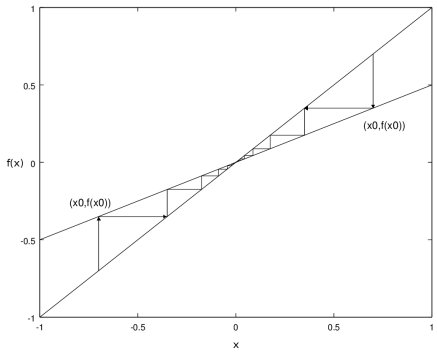
for $f(x) = -x^3$, the arrows may need to be bent in order to illustrate the orbits. In these cases the phase portrait could become a bit messy. We are still dealing with a one dimensional system, so the arrows are of no other significance than to show the order in which the points occur under iteration of f. We can see two examples of phase portraits in Figure 1.

A perhaps more descriptive way of illustrating the orbit of a point x_0 under iteration of a function f is by plotting $f(x)$ and Id, together with the sequence

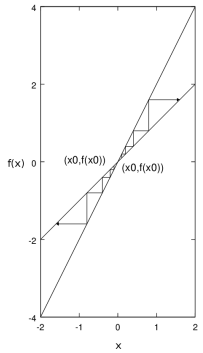
$$((x_0, x_0), (x_0, f(x_0)), (f(x_0), f(x_0)), (f(x_0), f^2(x_0)), ..., (f^{n-1}(x_0), f^n(x_0)))$$

for a satisfyingly large n, and join the points of that sequence together by arrows. The resulting image is called a **cobweb**. You simply start at x_0 and "move" vertically to the graph of $f(x)$, now move horizontally until you get to the graph of $Id(x)$, you are then at $x = f(x_0)$. Then move vertically again, to the graph of $f(x)$, at which point you are at $y = f^2(x_0)$, and so on. This is a very intuitive way of following the iterations of x_0 under f. As we can see in Figure 2, it also clearly illustrates the characteristics of the fix points as attracting or repelling. The n-periodic points are also easy to spot, as they show up as square shaped cycles, seen in Figure 2c. In Figure 2d, things look a bit more messy. We do not seem to have any fix points or periodic points, even though this is of course not entirely clear just from the picture. This is an example of a chaotic system, and it is something we will explore in further detail in Section 3.

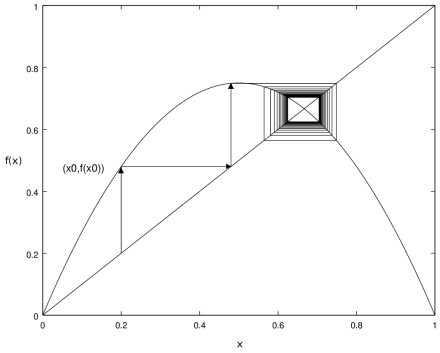
We will now start to explore the dynamics of non-linear maps, and we will see that things get very complicated even for seemingly simple maps.



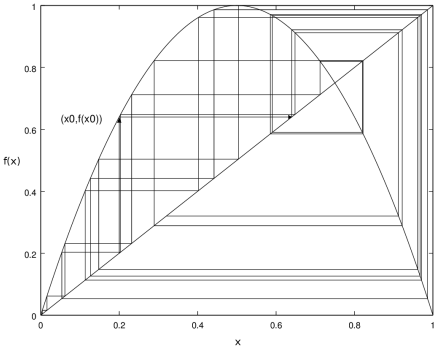
(a) Cobwebbing of $f(x) = \frac{1}{2}x$ starting at -0.7 and 0.7.



(b) Cobwebbing of $f(x) = 2x$ starting at -0.1 and 0.1



(c) Cobwebbing of $f(x) = 3x(1-x)$ starting at $x_0 = 0.2$



(d) Cobwebbing of $f(x) = 4x(1-x)$ starting at $x_0 = 0.2$

Figure 2: Different examples of cobwebbing on a map f, as an illustration of the orbit of a point x_0 under iteration of f.

2.2 Non-Linear Dynamics

We have already seen a lot of theory that apply to the case where the function that describes a dynamical system is not necessarily linear or affine. In this section we will discuss some more specialized theory suited for non-linear discrete dynamical systems. We will start off with some motivation for a non-linear theory, with an application to population dynamics. We will also introduce the concept of a family of systems and bifurcations, as well as fractals and fractal dimension. This will hopefully give us enough background to tackle the section about chaotic dynamics.

2.2.1 Population Models and the Logistic Map

One very simple model for population growth would be to assume that for each generation the number of individuals grows proportionally to the present generation, i.e. that

$$p_{n+1} = qp_n,$$

where p_n is the number of individuals after n time units, and q is some growth factor. The solution to this difference equation is $p_n = q^n x_0$. Now say that such a population of 100 individuals grows by 50% each year, i.e. $q = 1.5$, then the number of individuals after 10 years would be $x_{10} = 5767$, and after 100 years it would be on the scale of 10^{19}.

In Section 1.1 we gave an example of a second order difference equation, namely the Fibonacci recurrence. We also explained how this was initially a simplified model of the evolution of a population of rabbits. There are many reasons that the model is a simplified one. One of them is of course that there are probably many more external factors in play, such as other predators and/or sickness and starvation, resulting in a death rate. A more obvious reason (albeit due to the first reason) is that after just 5 years we would have 1548008755920 rabbits. That is about 209 times earth's population (of 2016). After 40 years the number of rabbits would be close to 10^{100}, which is well over the estimated number of atoms in the universe (usually estimated to somewhere on the scale of 10^{80}). Obviously something has to limit growth for these models to make sense.

Let us describe a different way of modelling population growth. As the population of a certain species increase we can assume that, at a certain point, it will be harder for each individual to survive due to competition for food, etc. Let the number of individuals after n time units be x_n, and let r be some growth factor. We want $(x_n)_{n\geq 0}$ to start decreasing when the population has become big enough, say at size K, usually called the **carrying capacity**. This behaviour can be modeled by the letting the rate of growth x_{n+1}/x_n be

$$\frac{x_{n+1}}{x_n} = r(K - x_n) \Rightarrow x_{n+1} = x_n r(K - x_n). \tag{13}$$

Without loss of generality we can assume that $K = 1$ (rather than describing the number of individuals we are then describing the percentage of the carrying capacity population). The difference equation (13) can then be described by

$$f(x) = rx(1 - x). \tag{14}$$

This is the so called **logistic map**. This map is the discrete version of the logistic equation, which is a model for population growth that was first published by Pierre François Verhulst, in 1845 [41]. It has been known for a long time that the logistic map is capable of displaying very complicated dynamics. It was even suggested by John Von Neumann, in the late 1940s, that it could be used as a random number generator. However, it was probably first popularized in 1976, by the Australian physicist Robert May, in a paper called *Simple mathematical models with very complicated dynamics*[28]. Because of its simple nature yet complicated dynamics, it is very well suited for demonstration, and we will thus return to this map several times throughout this thesis.

2.2.2 Family of Maps and Bifurcations

Sometimes we want to study how the behaviour of a system changes with a change in one or more of the parameters. We will therefore introduce the concept of a **family of maps**. We say that the set $\{f(a, x)|a \in \mathbb{R}^n\}$ is a n-**parameter family of maps**. When $a \in \mathbb{R}$, we often write this map as f_a, or even leave out the parameter in the function handle completely, as in the example of $f(x) = r\cos(x)$. Since this is not a paper on function theory, and it should always be clear from context what parameters, if any, we are changing, we will use the notation best suited for the case at hand. Quite often, we will omit the parameters from the function handle. Exceptions to this will be made if very convenient, or perhaps to emphasize the fact that we are talking about a family of maps, and not just one of the maps in a family.

The logistic map is part of a family of maps also called the logistic family, or sometimes the quadratic family of maps, i.e. the family $f_r(x) = rx(1 - x)$, for $r \in \mathbb{R}$. Most interesting dynamics of these maps seem to occur when $r \in [0, 4]$. Before going further and studying bifurcations, let us first look at some of the properties we are already familiar with in the example of $f(x) = rx(1 - x)$.

Since $f'(x) = r - 2rx$, we see that f attains a maximum at $x = 1/2$, and clearly $f(0) = f(1) = 0$. So, for $r \in [0, 4]$, f can be viewed as a map $f : I = [0, 1] \to I$. This is how we will view it now, and also how it was used in modelling since, as was mentioned above, $f(x)$ then represents a percentage. Solving $f(x) = rx(1 - x) = x$ for x yields $x_1 = 0, x_2 = \dfrac{r - 1}{r}, r \neq 0$. So, f has two fix points, but we see that $x_2 \in I$ only when $r \geq 1$ (and distinct from x_1 when $r > 1$). When $r < 1$, we see that x_1 is the only fix point in I, and that this is attracting since $|f'(0)| = |r| < 1$. So x_1 first becomes repelling when

$r > 1$, and another fix point x_2 appears, which at this stage is attracting since $f'(x_2) = f'(\frac{r-1}{r}) = r - 2r(\frac{r-1}{r}) = 2 - r$. So x_1 is repelling and x_2 is attracting. Continuing to increase r, we see that x_2 stays attracting until $r > 3$, when it becomes repelling.

So for $r \in (1, 3)$, the dynamics of f in I is well understood and simple. All points in I, except for 0 and 1, are forward asymptotic to x_2, and hence $W^s(x_2) = (0, 1)$.

Now, one may ask, what happens when $r > 3$ and both x_1 and x_2 become repelling? Up until now, we have only been looking at fix points of f, and ignored the fact that there may be periodic points of higher periodicity present. Remember from earlier, that a n-periodic point of f is a fix point of f^n. So finding the 2-periodic points is a matter of solving

$$\begin{aligned} f^2(x) = x &\Leftrightarrow r(rx(1-x))(1-rx(1-x)) = x \\ &\Leftrightarrow -r^3x^4 + 2r^3x^3 - r^3x^2 - r^2x^2 + r^2x = x. \end{aligned}$$

Solving this yields the four solutions

$$x_1 = 0, \quad x_2 = \frac{r-1}{r},$$
$$x_3 = \frac{r^2 + r - r\sqrt{r^2 - 2r - 3}}{2r^2}, \quad x_4 = \frac{r^2 + r + r\sqrt{r^2 - 2r - 3}}{2r^2},$$

of which we can of course see that x_1 and x_2 are the fix points we already had. Hence, x_3 and x_4 are the two points of prime period 2. Solving $r^2 - 2r - 3 = 0$ yields the two solutions $r_1 = -1, r_2 = 3$. So, we see that $r^2 - 2r - 3$ is a parabola with zeros at -1 and 3. This means that the points x_3 and x_4 only exist in I for $r > 3$, and at $r = 3$, we see that

$$\begin{aligned} \frac{(3)^2 + (3) - (3)\sqrt{(3)^2 - 2(3) - 3}}{2(3)^2} &= \frac{(3)^2 + (3) - (3)\sqrt{(3)^2 - 2(3) - 3}}{2(3)^2} \\ &= \frac{(3)^2 + (3)}{2(3)^2} = \frac{3-1}{3} = \frac{2}{3}, \end{aligned}$$

so that $x_2 = x_3 = x_4$, when $r = 3$. Some tedious but straightforward calculations also gives $(f^2)'(x_3) = (f^2)'(x_4) = -r^2+2r+4$, and solving $|r^2+2r+4| = 1$ gives us $r_1 = -1, r_2 = 3, r_3 = 1 - \sqrt{6}, r_4 = 1 + \sqrt{6}$. We see then that x_3 and x_4 stay attracting until $r = 1 + \sqrt{6}$.

So, to conclude this example: as we increased r, $x_1 = 0$ became repelling and another fix point, x_2 was born at $r = 1$, with $x_1 = x_2$ when $r = 1$. We then continued to increase r until x_2 became repelling at $r = 3$, at which two new attracting periodic points, x_3, x_4 of period 2 (or fix points of f^2) were born, with $x_2 = x_3 = x_4$, at $r = 3$. This continued up to $r = 1 + \sqrt{6}$ after which

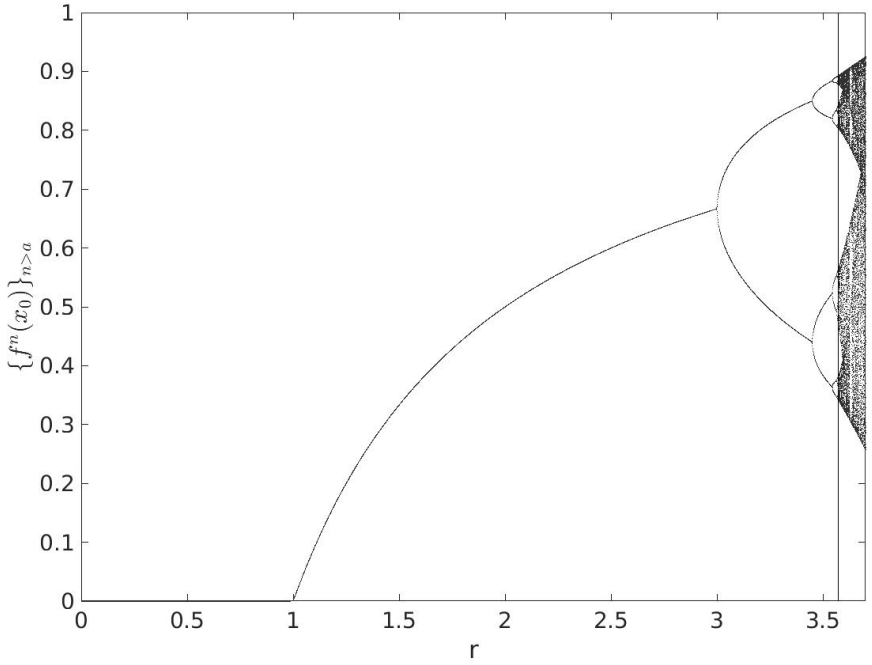


Figure 3: A so called bifurcation diagram of the logistic map, where we can see the periodic points on the y-axis and the parameter r on the x-axis. This shows how the dynamics change when we change the parameter.

we have points of period 4. This can all be illustrated using something called a **bifurcation diagram**, or sometimes, perhaps more accurately, called an **orbit diagram**. This is the attracting periodic points of a system, plotted against some parameter. More precisely, we plot the parameter r against $\{f^n(x_0)\}_{n\geq k}^N$, for a large N. If we let k be sufficiently large, the orbit will have settled down into a periodic cycle (if one exists of course), so the y-axis will present the attracting periodic points. In this way we can see how the dynamics changes when we change the parameter. An example of such a diagram can be seen in Figure 3. There we can also see that there seems to be another point, just to the right of $r = 3.5$, where we seem to have points of period 8. We have also marked a specific point further to the right with a line, after which we cannot seem to read out any periodicity at all. We will return to this value later in Section 4.4.

We have already gone ahead of ourselves a bit, and thus we will now introduce a concept that this last example gave a taste of, namely bifurcations.

As we saw in the previous section, the dynamics of a system can change dramatically when we introduce a change in one of its parameters. To be very precise we are not really looking at the same system as soon as we change a parameter (remember the notion of a family of maps), but merely different maps in a family. For a lot of these maps however, the dynamics will be exactly the

same. We would call these systems **topologically conjugate** or **topologically equivalent**. We will not go into much detail of topological conjugacy here, but as we will use the concept later, we will give a brief introduction starting with the definition.

Definition 2.8 *Let X and Y be topological spaces. Two continuous functions $f : X \to X$, and $g : Y \to Y$ are called topologically conjugate, if there exists a homeomorphic function (i.e. a bijective continuous function with a continuous inverse) $h : X \to Y$ such that $h \circ f = g \circ h$.*

We may consult the following commutative diagram for some clarity:

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ \downarrow{\scriptstyle h} & & \downarrow{\scriptstyle h} \\ Y & \xrightarrow{g} & Y \end{array}$$

Definition 2.8 is done for general topological spaces, but for us we may just aswell think of X and Y as subsets of $\mathbb{R}^n$. Essentially (but informally) this means that, as far as the dynamics goes, the systems described by f and g are equivalent. Everything we know about the dynamics of one of them, we automatically know about the other (e.g. in terms of fix points and their stabilities). For example, we note that for $x_0 \in X$ we have that

$$h \circ f = g \circ h \Rightarrow h \circ f \circ h^{-1} = g \Rightarrow g^n = h \circ f^n \circ h^{-1} \Rightarrow h \circ f^n = g^n \circ h,$$

which leads us to

$$h(O(x_0)) = \{h(f^n(x_0))\}_{n\geq 0} = \{g^n(h(x_0))\}_{n\geq 0} = O(h(x_0)).$$

Further, since if $f^k(x_0) = x_0$, then $g^k(h(x_0)) = h(f^k(x_0)) = h(x_0)$, we see that h takes orbits to orbits and periodic points to periodic points. In a similar fashion, one can go further to show that periodic orbits go to periodic orbits of the same period, and essentially that every dynamical feature is carried over by h.

As we saw in the previous section, the dynamics may change when we pass certain values of the parameters. Thus passing a point in the parameter space could make us go from one set of topologically equivalent systems to another set of equivalent systems, but that are no longer equivalent to the ones in the first set. The point in the parameter space where this happens is called a **bifurcation point**[5]. As we may understand from Theorem 2.2 and 2.3, for one dimensional

[5]This idea can actually be used to make a quite general definition of bifurcation point, in loose terms, as a point in parameter space whose every neighbourhood contain topologically different maps.

systems, this typically happens when a change in a parameter makes $|(f^n)'(x_s)|$ pass 1, at which point the stability of x_s changes. For higher dimensional systems, we instead look at the spectral radius $\rho(J)$ of the jacobian $J = (f^n)'(x_s)$. There are many ways in which the dynamics of different systems may change and we will only make the concept precise here in terms of one dimensional systems.

The above description gives some intuition for the concept, but in order to make a more precise definition we first need another concept. Let J be an interval and let $f(a, x)$ be real valued maps, $f : S \subseteq \mathbb{R} \to \mathbb{R}$, for each parameter value a (as was mentioned earlier, we often just refer to 'the map' f, instead of the family of maps). Let V be a subset of J, and let x_s be a fix point of $f(a, x)$, for every a in $V \subseteq J$. From the previous section it is hopefully clear that x_s may be dependent on a. We can therefore see x_s as a function $x_s : V \to \mathbb{R}$, of a. We may also define fix points of $f^n(a, x)$ (periodic points) in an analogous fashion. Let us therefore rather say that $x_s(a)$ is a fix point of $f^n(a, x)$ for some $n \geq 1$. For example, remember that $\dfrac{r-1}{r}$ was a fix point of $f(x) = rx(1-x)$ for $r \in (1, 3)$, thus $x_s(r) = \dfrac{r-1}{r}$, with $V = (1, 3)$. As we have seen earlier, we may now plot the function $x_s(a)$ against a, and get a bifurcation diagram. Usually we also refer to the graph of x_s in V, as a **branch** of fix points of f^n, fore some $n \geq 1$.

Having built up the intuition, let us now define what we mean by a bifurcation point.

Definition 2.9 *Let $f(a, x)$ be real valued, differentiable maps, with continuous derivatives, for each $a \in J$. We call $p \in J$ a **bifurcation point** of f, if there are continuous functions $x_i : V_i \to \mathbb{R}$, for some non-empty $V_i \subseteq J, i = 1, 2$ or $i = 1, 2, 3$, and whose graphs are branches of fixed points of f^n, for some $n \geq 1$, such that*

$$x_i(p) = x_j(p), x_i(a) \neq x_j(a), a \neq p, i \neq j.$$

where p could be one of the endpoints of V_i, for some i.

The bifurcation diagram in Figure 3 illustrate this rather well. As we saw before, this is called a bifurcation diagram of the logistic map $f(x) = rx(1-x)$ [6]. Let us first look at what happens at $r = 1$. Consider the branches $x_1(a) = 0$ and $x_2(a) = \dfrac{r-1}{r}$. We have already seen, since $f'(x_1) = r$ and $f'(x_2) = 2-r$, that when $r < 1$ on I, x_1 is attracting and x_2 is repelling (however, it's not even in I, if we necessarily look at $f : I \to I$). The two branches meet at $r = 1$, when $x_1(a) = x_2(a) = 0$, but $x_1(r) \neq x_2(r)$, when $r \neq 1$. Thus, letting

[6] We will not describe the details of numerically generating this diagram. However, the code can be found in the appendix B.2

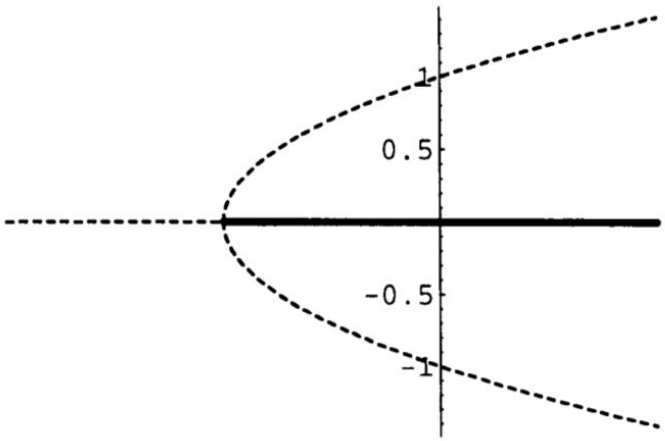


Figure 4: For the map $f(x) = x^3 - ax$, there is a bifurcation point (pitchfork bifurcation) at $a = -1$, but there is no exchange of stability between the branches. The branch $x_s = 0$ just becomes stable as a crosses -1 from the left.

$J_1 = J_2 = (a, b)$, for $a < 1 < b$, we see that $r = 1$ is a bifurcation point, call it p_1. We also noted that when passing p_1, x_1 became repelling, but that instead x_2 became attracting. We often say that there is an **exchange of stability** at p_1, or that x_1 passes its stability to x_2. Another interesting point was when $r = 3$. At this point, as we saw in the previous section, three branches meet, namely $x_2(r), x_3(r), x_4(r)$, but they are different from each other when $r \neq 3$, hence $r = 3$ is another bifurcation point, call it p_2. At p_2 there is also an exchange of stability between x_2, x_3 and x_4 in that x_2 passes on its stability to x_3 and x_4. Then x_3 and x_4 continue to be attracting until $r = 1 + \sqrt{6}$.

As the example of $f(x) = x^3 - ax$, seen in Figure 4, shows, there does not have to be an exchange of stability between branches of periodic points at a bifurcation point.

3 Chaotic Dynamics

In this section, we will give a very brief history of chaotic dynamics as a field of study, and build up the tools needed to give a, hopefully, comprehensible definition of chaotic behaviour. We will end by discussing an attempt to quantify chaos in terms of a measure of the rate of divergence of nearby orbits.

3.1 Short History and Introduction

The study of dynamical systems is a wide field of study, and the branch of it called chaotic dynamics (sometimes just referred to as chaos theory, especially in popular science) is a very recent development. Most people seem to agree that

chaotic dynamics, as a field on its own, started with the American mathematician and meteorologist Edward Lorentz[7].

In 1961, as Lorentz was working on an simplified model for atmospheric convection, he rounded off a few values to save time, and noticed a complete change in the new solutions. His discovery showed that even a very small change in the initial conditions could have a massive effect on the long term solutions. We may want to remind ourselves that these are all deterministic processes, and we like to think that a deterministic processes is, by definition, easy to predict. If we knew the initial conditions of a physical system exactly (and had a perfect model of it), we could predict its future exactly. However, for some systems, if we are off even slightly (which may always be the case in physical experiments), we are not even able to predict its relatively near future to a satisfying degree. This was the sort of system that Lorentz were working on when modelling the weather, perhaps without realizing it. This is a phenomenon called 'sensitivity to initial conditions'. This key component is often referred to as the **butterfly effect**, a phrase thought to have been coined by Edward Lorentz, however the concept of it appeared in Ray Bradbury's novel *A sound of thunder* from 1952 [8]. Lorentz also described this phenomenon in a paper from 1969, via the metaphor of a seagull flapping its wings [31].

One may think that such extreme sensitivity on initial conditions could only be exhibited by very complicated systems, but recall the logistic map from before. As we described in Section 2.2.1, this was even part of a paper named *Simple mathematical models with very complicated dynamics*[28], by Robert May. In Figure 5 below, one may see a demonstration of the butterfly effect using the logistic map. In this example, two different orbits $O(x_0)$ and $O(\tilde{x_0})$, under $f(x) = rx(1-x)$, with $x_0 = 0.1, \tilde{x_0} = 0.1 + 10^{-16}$, was generated using numerical methods. To give some perspective, say that these points were part of some physical experiment. Now, if x_0 were to represent the exact value, and $\tilde{x_0}$ our approximation, and that the values gave a measure of say mass in kg, we would only have been off by a tenth of the mass of an E.coli bacterium. Still only about 48 iterations were needed to give a significant difference in the results.

Later on, we will also see that sensitivity to initial conditions is part of even simpler systems, and that it is far from enough to characterize something as beeing chaotic. Before trying to define chaotic behaviour mathematically, we will first discuss a few definitions that we will need in order to do so.

[7]The essence of a chaotic system, namely its sensitivity to initial conditions and thus apparent randomness, was however noticed much earlier by a number of scientists, e.g James Clerk Maxwell in 1860 and Henri Poincaré in 1890 [43].

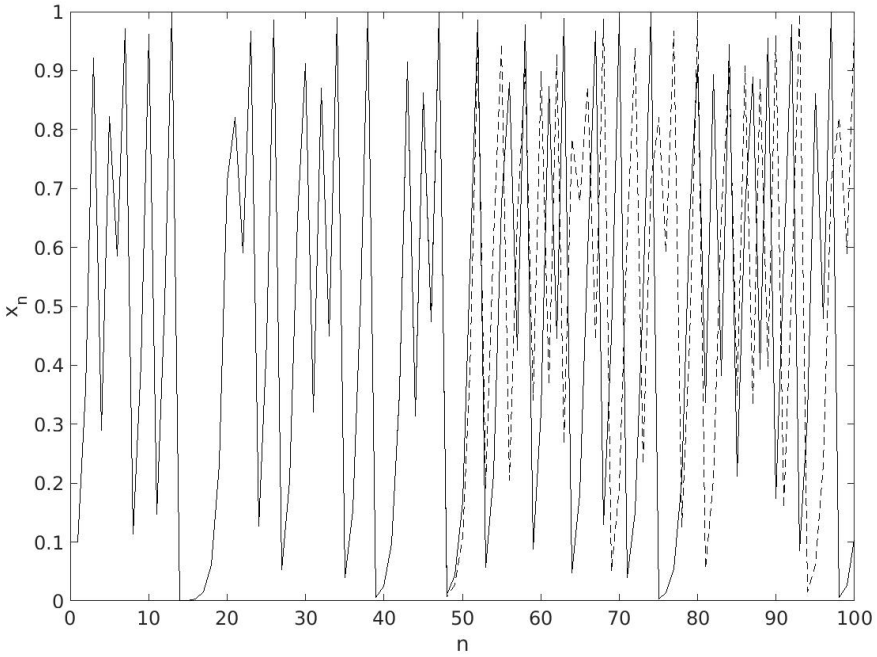


Figure 5: Illustration of the 'Butterfly effect' using the logistic map $f = rx(1-x)$, with $r = 4$. The values of two orbits, of length 100, are plotted against the number of iterations n. The first orbit starts at $x_0 = 0.1$ (solid line), and the other at $x_0 + 10^{-16}$ (dashed line). For about the first 48 iterations, the two orbits seem to coincide, but after that, one can clearly see a divergence as the dashed line become visible.

3.2 Stable and Unstable Orbits

We already know what it means for fix points and periodic points to be stable or unstable, and we will now discuss a similar notion for whole orbits. The intuition should be that we need something to push orbits apart, i.e. if two orbits start very close to each other, we do not want them to end up close to each other after enough iterations (not necessarily at least). In Section 2.1 we discussed aperiodicity, and it seemed that this was a key ingredient of chaotic behaviour. This was due to the fact that an aperiodic orbit never settles down into a periodic one, and in a way, thus never repeating its behaviour. However, a system could exhibit this behaviour while still not having a sensitive dependence on initial conditions, as the following example[8] shows:

Example 3.1 *Recall from linear algebra that the counter clockwise rotation by θ radians is given by the transformation*

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos(\theta) - y\sin(\theta) \\ x\sin(\theta) + y\cos(\theta) \end{pmatrix}$$

Let $\mathbb{S}^1$ be the unit circle, $x_0 \in \mathbb{S}^1$, and let $f(x,y) = (x\cos(1) - y\sin(1), x\sin(1) + y\cos(1))$. Then f is a counter clockwise rotation of the plane of 1 radian around the origin and we will now show that every open set of $\mathbb{S}^1$ contains points of $O(x_o)$.

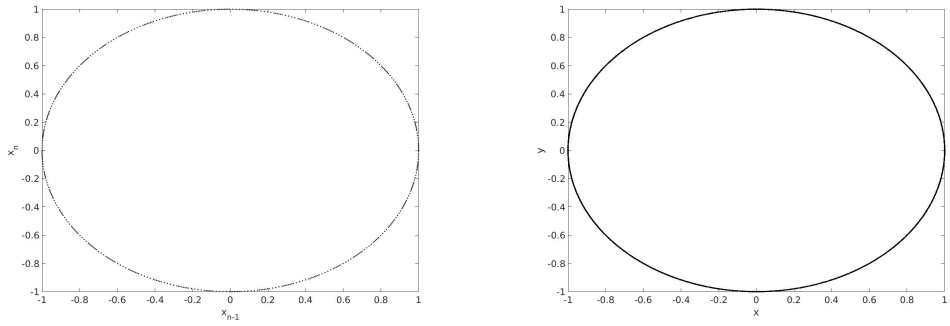
Assume there is an arc A of positive length of $\mathbb{S}^1$, such that $A \cap O(x_0) = \emptyset$, then there exists some open arc A_0 containing A, such that $A_0 \cap O(x_0) = \emptyset$ but the endpoints of A_0 are points of $O(x_0)$. Now, let $A_1 = f(A_0)$. If $A_1 = A_0$ then $1 = 2\pi$, which is false. Further, if $A_0 \neq A_1$, $A_0 \cap A_1 \neq \emptyset$, then A_0 would contain one of the endpoints of A_1, which belong to $O(x_0)$. This thus contradicts the assumption that $A_0 \cap O(x_0) = \emptyset$. So $A_0 \cap A_1 = \emptyset$. By an analogous argument, we would have $f(A_1) \cap A_0 = \emptyset$ and further $f^n(A_0) \cap A_0 = \emptyset$, for $n = 1, 2, ..., k$. Since the length of A was positive, the length of A_0 is positive and the length of its image under f^n is the same for all n. Thus, eventually, the length of all of $A_0, A_1, ..., A_k$ for some k, will exceed 2π, thus contradicting the fact that they are all disjoint arcs of $\mathbb{S}^1$.

Then, since every arc of $\mathbb{S}^1$ around an arbitrary point $z \in \mathbb{S}^1$, would contain a point of $O(x_0)$, every point $z \in \mathbb{S}^1$ is a limit point of $O(x_0)$. So $L(x_0)$ is infinite and thus $O(x_0)$ is aperiodic.

In Figure 6, one can also find some numerical verification of this fact, showing that the orbit of a point under f seems to trace out the entire unit circle.

Since f is an ordinary rotation of points in $\mathbb{R}^2$, one realizes that two orbits, starting close to each other, will remain close forever. Thus it does not feel right

[8] Example 3.1 is taken from [25], however the proof has been modified slightly. Many of the definitions in this section are also borrowed from [25]



(a) (b)

Figure 6: Plots of the first 500 (6a) and 30000 (6b) iterations of $x_0 = (1, 0)$ under the map $f(x, y) = (x\cos(\theta) - y\sin(\theta), x\sin(\theta) + y\cos(\theta))$, for $\theta = 1$, i.e. the map that rotates points in the plane one radian counter-clockwise. It seems that f maps x_0 into every open arc of the unit circle.

to characterize f as being chaotic in any sense. //

We now introduce a concept that will give us one of the properties missing in the above example. First, let us state what we mean by an invariant set. A set $X \subseteq \mathbb{R}^n$ is called **invariant** under f if $f(X) \subseteq X$. As an example, $I = [0, 1]$ is invariant under $f(x) = x(1 - x)$, since we know from before that every orbit starting outside of I will diverge to negative infinity, but we also know that 0 is attracting every point in I, so $W^s(0) = [0, 1]$. Simply put, if any orbit under f, that starts in X, stays in X, then X is invariant under f.

Definition 3.1 *An orbit $O(x_0)$ is called **stable**, if for every $\epsilon > 0$ there exists $\delta > 0$ such that $||x_0 - y_0|| \leq \delta \Rightarrow ||f^n(x_0) - f^n(y_0)|| \leq \epsilon$, for every $n \geq 1$. If $O(x_0)$ is not stable it is called **unstable**. If $x_0 \in X$ for some invariant set X, and if we require $y_0 \in X$, then we say that $O(x_0)$ is stable/unstable in X.*

Intuitively, this means that no matter how close to $O(x_0)$ we want some other orbit $O(y_0)$ to remain, we can provide a neighbourhood U of x_0, so that $O(y_0)$ will remain closer than that forever, provided $y_0 \in U$. Again, as a simple example, $O(0)$ of $f(x) = x(1 - x)$ is unstable, since any orbit starting at $x < 0$ diverges away from it. The orbit $O(0)$ is however stable in I.

Thus, if an orbit is unstable in some closed and bounded set X, we can see this as having encapsulated the orbit in X, while still having instability, i.e. nearby orbits will push away from it. As the reader may now suspect, if we combine this with our previous concept of aperiodicity, we arrive at something behaving the way we may expect from a chaotic system.

Definition 3.2 *Let $f : S \to \mathbb{R}^n$, be continuous for some open $S \subset \mathbb{R}$, let $X \subset S$ be a closed and bounded invariant set and let $x_0 \in X$. Then $O(x_0)$ is called* ***dense*** *in X, if $L(x_0) = X$.*

Recall from Example 3.1 above, that every point $z \in \mathbb{S}^1$, which is clearly invariant, was a limit point of $O(x_0)$, for a point $x_0 \in \mathbb{S}^1$. Thus $L(x_0) = \mathbb{S}^1$. So the orbit of any point in $\mathbb{S}^1$ is dense in $\mathbb{S}^1$.

As we saw earlier, the reason that f in Example 3.1 failed to behave like something we would call chaotic, was that nearby orbits would stay close forever. We are now ready to move on and give our first definition of chaotic behaviour, based on the concept we have introduced.

3.3 Defining Chaos

As a field of mathematics, chaotic dynamics could be considered a rather recent one. This pose a few problems, mainly that there are still a lot of ambiguities regarding definitions. It takes time for a theory like this to develop a solid basis on which to stand. Because of this, there are actually many different definitions of chaotic behaviour, and we will only touch on a few of them here. The one most suitable for our purposes is simply going to be the one we have built up intuition for. Definition 3.3 below is due to Martelli, Dang and Seph [26], and it combines precisely the two ideas we have previously discussed [9].

Definition 3.3 *Let $f : S \to \mathbb{R}^n$, be continuous in $X \subset S$, for some closed and bounded invariant set X and open $S \subset \mathbb{R}^n$. We call f* ***chaotic*** *in X, if there exists a point $x_0 \in X$ such that*

i.) $O(x_0)$ is dense in X,

ii.) $O(x_0)$ is unstable in X.

Another important concept that often comes up when studying chaotic systems, and that we will see is even part of another common definition of chaos, is the notion of **topological transitivity**.

Definition 3.4 *The continuous map $f : S \to S$, is called transitive in S, if for every pair of non empty open sets U and V in S, there is a non-negative integer n such that $f^n(U) \cap V \neq \emptyset$*

This means that points in any open set in S eventually moves (under f) into any other open set. It should be clear that this is deeply connected with our notion of dense orbits. As a matter of fact, in a paper by Stephen Silverman, from 1992 [35], it is proven that if S is *perfect*, *separable* and *second category*, then the

[9] Definition 3.3 is taken from [25]

notions of transitivity and the existence of a dense orbit are equivalent. In the chapter about fractal geometry we will explain what perfect means, but the other notions are out of scope for this paper. Right now, it is enough to know that the sets we are considering, for example closed and bounded subsets of $\mathbb{R}^n$, fulfill these requirements. For the sake of completeness however, we state and prove the following special case, which will be enough for our purposes:

Theorem 3.1 *Let X be compact, and let $f : X \to X$ be a continuous function. If f is topologically transitive in X, then there exists $x_0 \in X$ such that $O(x_0)$ is dense in X.*

Proof. Let $r_i = \frac{1}{i}$, and let $\{B_j^{r_i}\}_{j=1}^k$ be a finite cover of X, of closed balls of radius r_i. Let $U_1 \subset X$ be a closed subset of X, then there is some $n_{11} \in \mathbb{N}$, such that $V_{11} = f^{n_{11}}(U_1) \cap B_1^{r_1} \neq \emptyset$, by transitivity since $f^{n_{11}}(U_1^\circ) \cap B_1^{r_1} \neq \emptyset$. Then there is a n_{12} such that $V_{12} = f^{n_{12}}(V_{11}) \cap B_2^{r_1} \neq \emptyset$, and so on to n_{1k}, such that $V_{1k} = f^{n_{1k}}(V_{1k-1}) \cap B_k^{r_1} \neq \emptyset$. Let $U_2 = f^{-(n_{11}+\ldots+n_{1k})}(V_{1k}) \cap U_1$. Do the same for a cover $\{B_j^{r_2}\}_{j=1}^k$ to obtain $V_{21} = f^{n_{21}}(U_1) \cap B_1^{r_2} \neq \emptyset, \ldots, V_{2k} = f^{n_{2k}}(V_{2k-1}) \cap B_k^{r_2} \neq \emptyset$. Then let $U_3 = f^{-(n_{21}+n_{22}+\ldots+n_{2k})}(V_{2k}) \cap U_2$, and so on to get $U_{m+1} = f^{-(n_{m1}+\ldots+n_{mk})}(V_{mk}) \cap U_m$, for a cover $\{B_j^{r_m}\}_{j=1}^k$. We then have that $U_1 \supset U_2 \supset \ldots \supset U_{m+1}$. Letting $m \to \infty$ we get that

$$W = \bigcap_{i=1}^{\infty} U_i \neq \emptyset,$$

by Cantor's intersection theorem, since each U_i is compact by the continuity of f.

If so, there exists some $x_0 \in W$, and then for each $i > 0$ and $j \in \{1, 2, \ldots, k\}$ $\{f^n(x_0)\}_{n\geq 0} \cap B_j^{r_i} \neq \emptyset$ for $B_j^{r_i} \in \{B_j^{r_i}\}_{j=1}^k$. Since every neighbourhood of any point $p \in X$ will contain an open ball B of some cover $\{B_j^{r_i}\}_{j=1}^k$ for sufficiently large i, any neighbourhood of any point $p \in X$ will contain a point of $O(x_0)$, and thus $\{f^n(x_0)\}_{n\geq 0}$ is dense in X. Then every point of X is a limit point of $O(x_0)$, so $L(x_0) = X$. □

In [25, Theorem 6.2.1], Martelli states the following theorem without proof, under the same assumptions as in Definition 3.3, that can be applied to prove the presence of chaos for certain maps.

Theorem 3.2 *Assume that for every $x \in X$ and every $r > 0$, there exists $n \geq 1$ such that $F^n(B(x, r) \cap X) = X$. Then there exists $x_0 \in X$ such that $L(x_0) = X$.*

Proof. Since the assumption in Theorem 3.2 obviously implies topological transitivity of F, the proof of this theorem is automatic from the proof of 3.1.□

Let us now put Theorem 3.2 to use, in proving chaotic behaviour of a one dimensional system.

Theorem 3.3 *$f(x) = 2|x| - 1$ is chaotic in $[-1, 1]$.*

Proof. First of all, $[-1, 1]$ is closed and bounded, and it is also easy to see that $f([-1, 1]) = [-1, 1]$, so $[-1, 1]$ is invariant under f. We also see that $f'(x) = 2$, for $x \in (0, 1]$ and $f'(x) = -2$ for $x \in [-1, 0)$. Either way, $|f'(x)| > 1$ for all $x \neq 0$. So every orbit of f will be unstable.

If $a, b > 0$, we have $f((a, b)) = (2|a| - 1, 2|b| - 1)$, and if $a, b < 0$, we get $f((a, b)) = (2|b| - 1, 2|a| - 1)$. So, if (a, b) does not contain 0, we have that

$$|f((a, b))| = |2|a| - 2|b|| = 2|a - b|$$

and hence the length of an open interval is doubled with each iteration of f. Let $U = N_\epsilon(x) \cap [-1, 1]$, for some $x \in [-1, 1]$ and neighbourhood $N_\epsilon(x)$ of x, $\epsilon > 0$. Then, if U or any of $f^1(U), ..., f^{k-1}(U)$ does not contain 0, and the length is doubled at each iteration, at some point we will have $0 \in f^k(U)$. Now, since $f(0) = -1, f(-1) = 1$ and $f(1) = 1$ we get that $1 \in f^{k+l}, l = 2, 3, 4....$ Again, as with U, if V or any of $f^1(V), ..., f^{m-1}(V)$ does not contain 0, and the length is doubled at each iteration, at some point we will have $0 \in f^m(V)$. So, $f^{m+1}(U) = [-1, 1]$ and therefore $f^{k+m+3} = [-1, 1]$ and thus for every $x \in [-1, 1]$, and every neighbourhood of x, there exists n such that $f^n(N_\epsilon(x) \cap [-1, 1]) = [-1, 1]$. From Theorem 3.2 we may conclude that there exists x_0 $[-1, 1]$ such that $L(x_0) = [-1, 1]$.

Hence by Definition 3.3, f is chaotic in $[-1, 1]$. □

Now, let $I = [-1, 1]$, $f : [0, 1] \to [0, 1], f(x) = 4x(1 - x)$, which we recognize from before as the logistic map with $r = 4$. Let further $g : I \to I, g(x) = 2x^2 - 1$ and $h : I \to I, h(x) = 2|x| - 1$. In [25] it is shown that h is topologically conjugate to g via the conjugacy $\phi : I \to I, \phi(x) = \frac{2}{\pi} \arcsin x$. It is also easy to see that f is conjugate to g via $\psi : I \to [0, 1], \psi(x) = \frac{1 - x}{2}$ since, for $\psi^{-1}(x) = 1 - 2x$, we have

$$\psi^{-1} \circ f \circ \psi(x) = 1 - 2(4(\frac{1 - x}{2})(1 - \frac{1 - x}{2}))$$
$$= 1 - 2(2(1 - x) - (1 - x)^2) = 1 - 2(1 - x^2) = 2x^2 - 1 = g(x).$$

So $g = \psi^{-1} \circ f \circ \psi \Rightarrow \psi \circ g = f \circ \psi$, and f and g are conjugate. If we now argue for why g and h are conjugate via ϕ, we could directly say that f and h are conjugate. Instead we notice that, since $h \circ \phi = \phi \circ g \Rightarrow g = \phi^{-1} \circ h\phi$, we get

$$f \circ \psi = \psi \circ g = \psi \circ \phi^{-1} \circ h\phi \Rightarrow f \circ (\psi\phi^{-1}) = (\psi\phi^{-1}) \circ h,$$

which in turn would show that f and h are conjugate via $\psi \circ \phi^{-1} : I \to [0,1], \psi\phi^{-1}(x) = \dfrac{1-\sin\frac{\pi}{2}x}{2}$. We can then just check to see if $\psi \circ \phi^{-1}$ is a conjugacy between f and h. We see that

$$\begin{aligned} f \circ (\psi \circ \phi^{-1})(x) &= 4\Big(\frac{1-\sin(\frac{\pi}{2}x)}{2} - \frac{1-\sin^2(\frac{\pi}{2}x)}{2}\Big) \\ &= 4\Big(\frac{1-\sin(\frac{\pi}{2}x)}{2} - \frac{1-2\sin(\frac{\pi}{2}x)+\sin(\frac{\pi}{2}x)}{4}\Big) \\ &= 2 - 2\sin(\frac{\pi}{2}x) - 1 + 2\sin(\frac{\pi}{2}x) - \sin^2(\frac{\pi}{2}x) \\ &= 1 - \sin^2(\frac{\pi}{2}x) = \cos^2(\frac{\pi}{2}x). \end{aligned}$$

We also see that

$$\begin{aligned} (\psi \circ \phi^{-1}) \circ h(x) &= \frac{1-\sin(\frac{\pi}{2}(2|x|-1))}{2} \\ = \frac{1-\sin(\pi|x|-\frac{\pi}{2})}{2} &= \frac{1-\cos(\pi|x|-\pi)}{2} = \frac{1+\cos(\pi|x|)}{2} \\ &= \cos^2(\frac{\pi}{2}|x|) = \cos^2(\frac{\pi}{2}x). \end{aligned}$$

So $f \circ (\psi \circ \phi^{-1}) = (\psi \circ \phi^{-1}) \circ h$, and thus $(\psi \circ \phi^{-1})$ is a conjugacy between f and h. Further, it is shown in [25, Therorem 6.2.2 and 6.2.3] that if $f_1 : X \to X$ and $f_2 : Y \to Y$ are conjugate maps, then f_1 is chaotic in X if and only if f_2 is chaotic in Y. This is somewhat intuitive since chaotic behaviour is a property related directly to the dynamics of a system, and as we mentioned earlier, everything we know about the dynamics of f_1, we also know about f_2. Now, since we have shown that h is chaotic in I, and that f is conjugate to h, we may draw the conclusion that f is chaotic in $[0,1]$. This is of course something we have suspected all along, but now we had the theory to actually prove it according to a more precise definition.

This shows how conjugacy can be used to show that certain maps are chaotic, by finding conjugacys to other maps that we already know are chaotic. We will end this section by discussing some of the other definitions of chaotic behaviour that exists.

3.3.1 On Other Definitions

Informally, a chaotic system is a dynamical system that exhibits chaotic behaviour in some region of its phase space. Chaotic behaviour, in turn, has many definitions. One of the most popular seems to be the one by the American mathematician Robert L. Devaney, which is the following:

Definition 3.5 *Let $f : U \to U \subset \mathbb{R}^n$ be continuous. Then f is said to be* ***chaotic in U*** *if*

1. *f is topologically transitive in U,*
2. *Periodic points are dense in U,*
3. *f has sensitive dependence on initial conditions.*

Devaney calls the three properties 1,2,3; *indecomposability*, an *element of regularity* and *unpredictability* respectively. We have already discussed topological transitivity in the previous section. This was a property that made sure that every open neighbourhood eventually got mapped into any other. This property is often referred to as mixing property for obvious reasons. Devaney calls it indecomposability since the system cannot be broken down into several subsystems that does not interact with each other. We will return to the concept of a dense set when we discuss fractals in Section 4, but 2. means that every open subset of U contains a periodic point. This is what Devaney therefore calls an *element of regularity*. We have already touched on the subject of sensitive dependence on initial conditions, but not properly defined it. This is perhaps one of the most intuitive parts of the definition, since it was even part of why chaotic dynamics was more or less discovered, as discussed in Section 3.1. The definition is pretty much straight forward, and it is usually stated as follows:

Definition 3.6 *The function $f : S \to S$ has* ***sensitive dependence on initial conditions*** *in S, if there exists a real number $r > 0$ such that for any $x \in S$, and any neighbourhood N of x, there exists $y \in N$ and $m > 0$ such that $|f^m(x) - f^m(y)| > r$.*

We may note that it is not necessary that $N \subseteq S$. We therefore also define sensitive dependence on initial conditions **with respect to** S, by adding the restriction in Definition 3.6, that $y \in N \cap S$. Informally speaking, if no matter how close to every initial condition x_0 we start, that orbit will eventually get far away from the orbit of x_0, then the system exhibits sensitive dependence. As has been mentioned before, this is a key component in a chaotic system, often referred to as the 'butterfly effect'. However, it is not a sufficient condition, as can easily be seen from the simple system governed by e.g. $f = 5x$. The orbits of any two points, no matter how close, will eventually get separated far enough. However, one may argue that in such an example, f has the whole of $\mathbb{R}$ to "move around" in. Some, especially physicists, may still argue that in a bounded invariant set X, this is still the simplest and most important feature, and therefore take sensitive dependence on initial conditions in X as a definition of chaos. This is actually referred to as **the experimentalists definition of chaos** in [26]. However, as Martelli, Dang and Seph shows in [26], there are maps that we would not want to classify as chaotic that still shows sensitive

dependence on initial conditions even in bounded invariant sets. They take $F : D \to D, F(\rho, \theta) = (\rho, \theta + \rho)$, where $D = \{x \in \mathbb{R}^2 || x| \leq 2\}$ as an example. They show that $C_\rho = \{x \in \mathbb{R}^2 || x| = \rho\}$ is invariant, and that F shows sensitive dependence on initial conditions there, but at the same time F is just a rotation in C_ρ and it does not seem appropriate to classify it as chaotic.

It has been shown by J. Banks et al. in 1992 [1] that, in Definition 3.5, 1 and 2 implies 3. Now, as we mentioned in Section 3.3, Silverman had shown that under the given circumstances, transitivity and existence of a dense orbit are equivalent. In [26], Martelli, Dang and Seph also shows that, under the assumptions of Definition 3.3, f has sensitive dependence on initial conditions with respect to X, if and only if $O(x_0)$ is unstable in X. This means that the two definitions, Definition 3.3 and Definition 3.5, are closely related since, at least for a large class of functions on closed and bounded subsets of $\mathbb{R}^n$, 1 and 2 of Definition 3.3 are equivalent to 1 and 3 of Definition 3.5. It has also been proven by M. Vellekoop and R. Berglund [40] that on an interval, transitivity implies sensitivity and dense periodic orbits, and hence is enough for chaos.

From the above discussion we see that the only notion that Devaney's definition adds, in comparison with Definition 3.3, is the denseness of periodic points. This is actually similar to part of another common definition of chaos, due to T.Y. Li and J.A. Yorke., who also coined the term 'chaos' in their paper *Period three implies chaos* [21] from 1975. In their paper, they proved that for an interval I, and a continuous function $f : I \to I$, if f has a periodic orbit of period 3, then f has a periodic orbit of every period. Further there is an uncountable set $S \subset I$, such that for every $x \in S$ $O(x)$ is aperiodic and unstable, which was the two ingredients that eventually lead us to Definition 3.3. So one possible definition of chaos is to say that a map $f : I \to I$ is **chaotic in the Li-Yorke sense**, if f has a periodic point of period 3 in I. This is then a very experimentally nice definition, since all that needs to be done is to find a point of period three, but a drawback is of course that the definition is only applicable on intervals [10].

So in conclusion, the two main properties that all sensible definitions of chaos wants to capture is the ones of mixing, or as Devaney calls it, indecomposability, and of sensitivity. These two properties were illustrated in Figure 5, where we saw two orbits starting very close to each other. After only about 48 iterations we saw that the two orbits had separated quite a lot and that they did no longer approximate each other even slightly. In Figure 6 we saw a demonstration of transitivity, as the orbit seems to visit every open arc of the unit circle. In a way, we may intuitively but informally see these two properties, of mixing and sensitivity, as embodying the concepts of geometry and dynamics respectively.

Before ending the section on chaos, we will introduce a dynamical concept that quantifies the notion of sensitive dependence, and gives us a way to try and

[10] An extension to $\mathbb{R}^n$ was however made by F.R. Marotto in 1978 [24], using a concept he called a *snapback repeller*.

measure how sensitive a system is to initial conditions. In the later half of this thesis, we will instead spend some time trying understand the geometry side of things. Later we will try and tie these two concepts together when defining a measure of dimension (geometry) in terms of this new measure of sensitivity (dynamics).

3.3.2 Lyapunov Exponents

Let us look at a rather intuitive concept that is closely related to the sensitive dependence on initial conditions, which we need for chaotic behaviour. Let us consider two different orbits, starting very close to each other. One can imagine that if no matter how close to another orbit we start, the two different orbits will diverge away from each other exponentially fast, but still be confined to a bounded region, we have a good chance of that system behaving chaotically. Also, if you think about the three notions *no intersections of orbits*, *orbits confined to a bounded region*, and *exponential divergence of nearby orbits*, as the always present features of chaotic systems, it is relatively easy to see that chaos cannot emerge in a continuous system in less than three dimensions. This can actually be realized by just thinking about how this would work, drawing on a piece of paper. One can think of this as the system needing that extra dimension to wrap its orbits around themselves. One makes this idea of exponential divergence of orbits precise by introducing something called a **Lyapunov exponent**. Let $\delta(t)$ denote the distance between two orbits at time t, and say that two orbits start close to each other, at a distance $\delta(0)$. If the distance $\delta(t)$ between the two orbits satisfies

$$\delta(t) = \delta(0)e^{\lambda t}, \tag{15}$$

then λ is called the **Lyapunov exponent** for the orbits. Figure 7 illustrates this. We also talk about the average Lyapunov exponent, where we consider a large amount of orbits. Most authors actually just call this 'the Lyapunov exponent' for the system [16].

Let us consider a one-dimensional system described by a function f, a starting point x_0, and another point y_0 very close to x_0, say at a distance ϵ of x_0. Let us first note that, for an iterated function f^n we get its derivative as

$$\begin{aligned}(f^n(x_0))' &= f'(f^{n-1}(x_0))f'(f^{n-2}(x_0))f'(f^{n-3}(x_0))\dots f'(f(x_0))f'(x_0)\\ &= f'(x_0)f'(x_1)f'(x_2)f'(x_3)\dots f'(x_{n-1}),\end{aligned}$$

just by using the chain rule for derivatives.

We will now see what happens to the distance $|f^n(x_0) - f^n(y_0)|$, as the system progresses, i.e. what happens to the distance between the two orbits starting close to each other. From the definition of the derivative, and for a

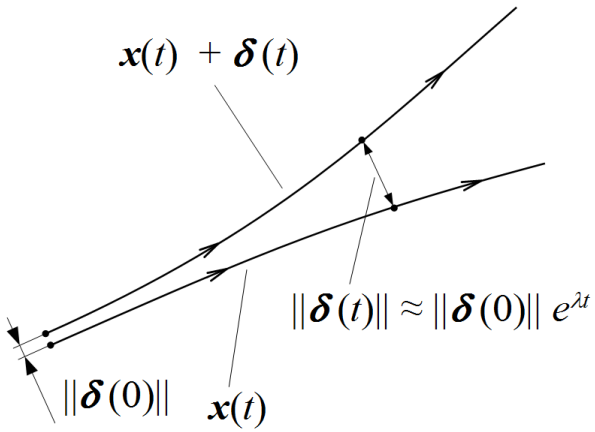


Figure 7: Illustration of the Lyapunov exponent as a measure of the rate of divergence of two trajectories. The two orbits, represented by $x(t)$ and $x(t) + \delta(t)$, start off at a distance of $\delta(0)$ from each other. The distance between them grow/shrink exponentially on average.

small ϵ, we then have that

$$|f^n(x_0) - f^n(y_0)| \approx |(f^n(x_0))'||x_0 - y_0| \tag{16}$$

$$= |f'(x_0)||f'(x_1)||f'(x_2)||f'(x_3)|| \ldots |f'(x_{n-1})|\epsilon, \tag{17}$$

from which we see that, as $n \to \infty$, the distance between the two orbits will either remain constant, grow or shrink exponentially on an average. As we expect the distance to grow according to an exponential law, we write

$$\frac{|f^n(x_0) - f^n(y_0)|}{\epsilon} = e^{\tilde{\lambda} n}$$

(compare this to (17)), and to get a quantitative measure of the rate of divergence we solve for $\tilde{\lambda}$, for a very small ϵ, to get

$$\begin{aligned} \tilde{\lambda} &= \frac{1}{n}\Big(\ln\Big(\frac{|f^n(x_0) - f^n(y_0)|}{\epsilon}\Big)\Big) \\ &\approx \frac{1}{n}\Big(\ln\left(|f'(x_0)||f'(x_1)||f'(x_2)| \ldots |f'(x_{n-1})|\right)\Big) \\ &= \frac{1}{n}\Big(\ln\left(|f'(x_0)| + \ln|f'(x_1)| + \ldots + \ln|f'(x_{n-1})|\right)\Big), \end{aligned} \tag{18}$$

where $\tilde{\lambda}$, or perhaps more accurately $\tilde{\lambda}(x_0)$, would be the the approximated Lyapunov exponent for the finite orbit. As we would not want to restrict ourselves

to finite orbits of length n, it makes sense to define the Lyapunov exponent for $O(x_0)$ as

$$\lambda(x_0) = \lim_{n\to\infty} \frac{1}{n}\Big(\ln\left(|f'(x_0)| + \ln|f'(x_1)| + \ldots + \ln|f'(x_{n-1})|\right)\Big). \qquad (19)$$

It is also convenient to consider the so called **Lyapunov numbers**. In the above example, the Lyapunov number of the orbit of x_0, denoted $\Lambda(x_0)$[11], would be the average growth factor of the separation, where we would use a geometric mean, i.e.

$$\Lambda(x_0) = \lim_{n\to\infty} (|f'(x_0)||f'(x_1)|\ldots|f'(x_n)|)^{\frac{1}{n}}. \qquad (20)$$

This is perhaps a more intuitive measure as e.g. $\Lambda(x_0) = 2$ would mean that the separation of the orbits grow, on an average, by a factor 2 at each iteration. So for exponential divergence of orbits we simply need $\Lambda(x_0) > 1$. Of course we realize that the Lyapunov exponent is just the logarithm of the Lyapunov number, i.e. $\lambda(x_0) = \ln(\Lambda(x_0))$, and that $\lambda(x_0) > 0$ precisely when $\Lambda(x_0) > 1$. By applying the natural logarithm to (20), and using simple logarithm laws, we also see that we get (19).

This also gives us a rather convenient way to numerically approximate the average Lyapunov exponent for a system. We will discuss the notion of an attractor in Section 4.4, but let's just say that we wanted to measure the average Lyapunov exponent of a system on an invariant and attracting set A. We first generate a large finite orbit, starting from a random point p_1 close to or in A. We omit a portion from the beginning of the generated orbit, to make sure the points we are left with are actually in A (this works since A is attracting). We then use (18) to calculate $\tilde{\lambda}(p_1)$. We then do the same thing for another random point p_2, to get $\tilde{\lambda}(p_2)$, and so on, until we have $\tilde{\lambda}(p_m)$. We then simply let the approximated average Lyapunov exponent be $\frac{1}{m}\sum_{i=1}^{m}\tilde{\lambda}(p_i)$. In Figure 8, we can see the result of such a numerical approximation done in the example of the logistic map. Here we have calculated the average Lyapunov exponent for several different values of r, and presented the results along with the bifurcation diagram that shows the regions of r, in which the map behaves chaotic. In this way, one can clearly see the correspondence between a positive Lyapunov exponent and chaotic behaviour.

In the case of higher dimensional maps, the role of the derivative is as usual played by the Jacobian. We may remember from linear algebra that for an n-dimensional linear map M the eigenvalues of M can be seen as a factor of how much the transformation M streches in the direction of the corresponding eigenvector. Thus we could define the Lyapunov numbers for M as the eigenvalues of M. For a, not necessarily linear n-dimensional map Ω, we can, as in (17),

[11] Unfortunately this is most commonly denoted $L(x_0)$. Since we have reserved this notation for the *limit set* of an orbit, we denote it $\Lambda(x_0)$ instead.

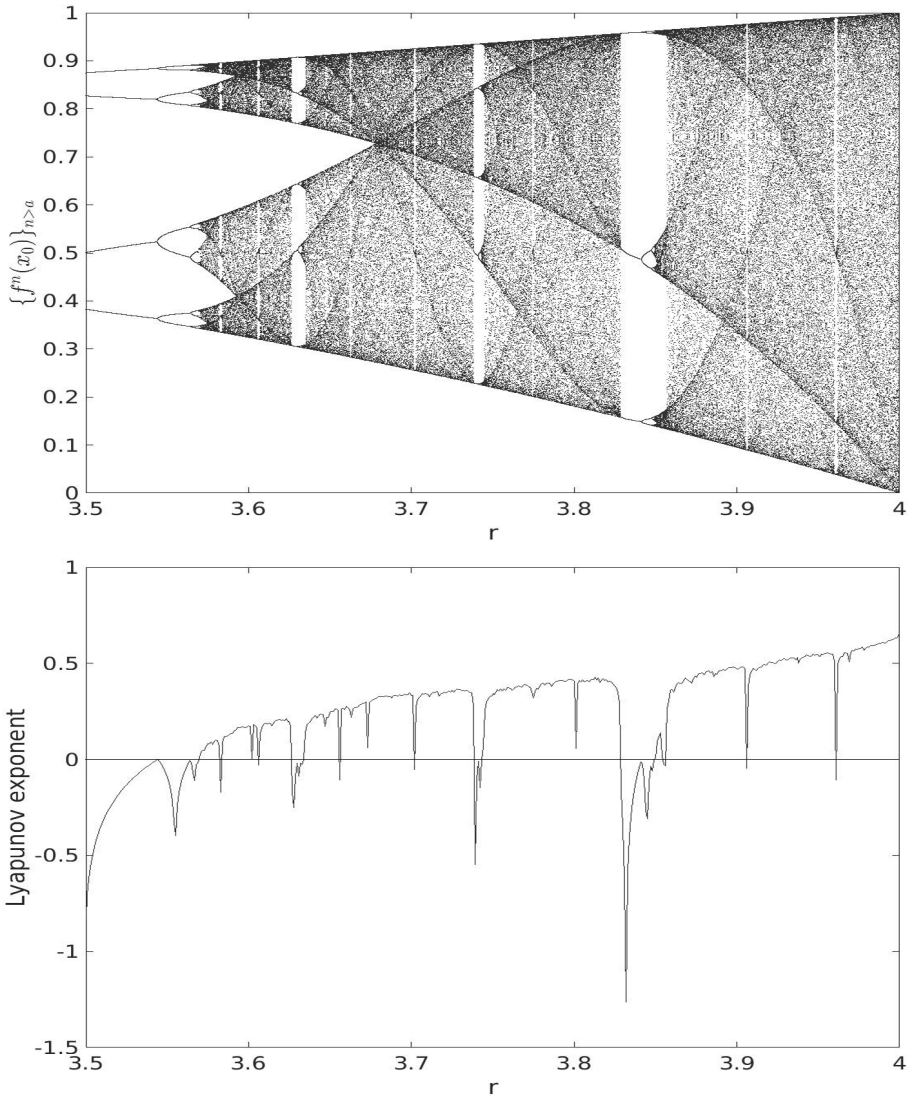


Figure 8: Illustration of the correspondence between a positive Lyapunov exponent and chaos. In the figure we can see that the so called chaotic bands of the bifurcation diagram seem to correspond well with a positive Lyapunov exponent.

make the linear approximation

$$\Omega^n(x_0) - \Omega^n(y_0) \approx (\Omega^n(x_0))'(x_0 - y_0)$$

in a neighbourhood of x_0, and look at the behaviour of the Jacobian $(\Omega^n(x_0))'$. In a similar fashion as for one dimensional maps, but instead now looking at the eigenvalues of a matrix, we thus define the Lyapunov numbers $\Lambda_1(x_0), \Lambda_2(x_0), \ldots, \Lambda_n(x_0)$ as the n eigenvalues of

$$\lim_{n\to\infty} \left((\Omega(x_0))'(\Omega(x_1))'(\Omega(x_2))' \ldots (\Omega(x_n))' \right)^{\frac{1}{n}}.$$

Thus a two-dimensional system would have two Lyapunov exponents associated with it, one for each dimension. This of course makes sense, since the orbits could now separate in two directions. One usually refers to the set of Lyapunov exponents as the **Lyapunov spectrum**. Also, when referring to the average Lyapunov exponents or Lyapunov numbers for a whole system, as opposed to just an orbit, we usually denote them λ_i and Λ_i respectively.

We will now leave the dynamics behind, to discuss the more geometrically related subject of dimension. As it turns out, the geometry of many invariant sets, on which maps behave chaotically, is not that simple. Again, this can perhaps be realized by thinking of the three notions of chaotic behaviour mentioned in the first paragraph of this section. Sets with these properties often have a geometry which is called fractal, and the dimension of such sets are referred to as a fractal dimension. This is the rather counter-intuitive idea of a non-integer dimension. As we said in the end of Section 3.3.1, we will return to the Lyapunov exponents in Section 4.5.2, when we tie together the dynamics with the geometry, suggesting that the dimension of an attractor (geometry) can be expressed in terms of the Lyapunov exponents (dynamics).

4 Fractal Dimension and Strange Attractors

The notion of dimension probably seems clear to most people. We have no trouble imagining up to at least three dimensions, since this is what we experience every day. When we pass three it may be trickier, but even then our imagination can help us in a number of ways. We may think of four dimensional objects as three dimensional ones existing in time or having different temperatures at different points, etc. If we let go of our need to imagine the object existing in front of us in space, we could just as well deal with objects having a thousand dimensions.

Informally speaking, and according to classical definitions, the dimension of an object is the minimum number of coordinates needed to specify a point within it. Thus a circle is one dimensional rather than two, since it could be

parametrized using just one parameter, namely the angle. Therefore the dimension of an object is not necessarily the same as the dimension of the space it is embedded in (e.g a circle is embedded in two and a sphere in three dimensions). As we shall see, for some geometrical shapes, this way of thinking about dimension breaks down completely, and as so often happens in mathematics, we come up with new ways of defining and thinking about it.

We will start off by explaining one of the perhaps most natural ways of making the concept of dimension precise, and that is that of the Lebesgue covering dimension or topological dimension. We will then introduce certain geometrical objects called fractals, give a few examples of them, and discuss their properties. This will hopefully build up enough intuition for why we may want to extend the concept of dimension. These extended definitions will coincide, i.e they will yield an equivalent measure of the dimension, with the topological dimension of 'ordinary' geometrical objects, but for fractals they will yield a dimension that can even be non-integer.

We shall introduce a few different ways of measuring this so called fractal dimension, and see how this can be applied to the field of chaotic dynamics.

4.1 Lebesgue covering dimension

One way of defining the dimension of an object, or more precisely of a topological space, is the so called **topological dimension**, also known as **Lebesgue covering dimension**. To define it we first need the following.

Definition 4.1 *A collection $\mathscr{A}$ of subsets of the space X is said to have **order** $m+1$ if some point of X lies in $m+1$ elements of $\mathscr{A}$, and no point of X lies in more than $m+1$ elements of $\mathscr{A}$.*

Intuitively one could imagine a number of two dimensional sets in the plane being placed in a "pile". Some of these sets may intersect each other, and the order of this "pile" is the largest number of sets sharing a piece of the plane. Using this concept, we can now make the following definition.

Definition 4.2 *A space X is said to be **finite dimensional** if there is some integer m such that for every open cover $\mathscr{A}$ of X, there is an refinement $\mathscr{B}$ of $\mathscr{A}$ of order at most $m+1$. The smallest such integer m is called the **topological dimension** of X. We denote it by $\dim X$.*

We are thus looking for a way to cover a set with the least amount of overlapping, and in such a cover we look for the maximum number of sets overlapping each other, and we call this the topological dimension. In Figure 9, we can see covers of a circle and a disk. We may refine these covers further, but it is hopefully clear that any such refinement will have the same maximum amount of intersecting sets as before. It will be impossible to cover the disk with a "chain" of sets only having at most two intersecting sets. You need atleast three sets

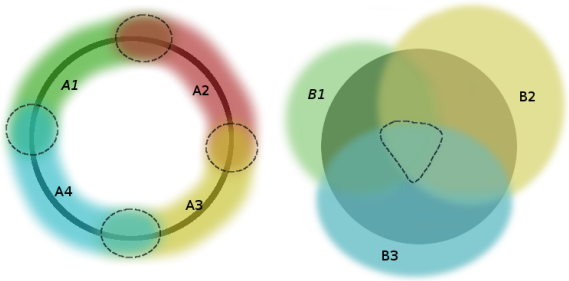


Figure 9: Illustration of topological dimension by covering of a circle (left), and a disk (right), with open covers $\{A1, A2, A3, A4\}$, and $\{B1, B2, B3\}$ respectively. We see that the largest number of intersecting elements, i.e the order, of the cover is two and three respectively. The dashed lines mark the intersections determining the order of the cover.

having a common intersection and thus, according to Definition 4.2, the topological dimension of the disk is $3 - 1 = 2$ (while it is 1 for the circle as seen in Figure 9).

The topological dimension of ”ordinary” geometrical shapes corresponds with the notion of the minimum number of coordinates needed to specify a point in it that we mentioned earlier. We will now introduce other objects where this notion seems to fail.

4.2 Fractals

The word ’fractal’ (from the Latin word ’fractus’, meaning broken) was coined in 1975 by the Polish-French mathematician Benoît B. Mandelbrot. Mandelbrot himself described them as, *”[B]eautiful, damn hard, increasingly unuseful. That’s fractals.”* However, since then, many have tried to formally define what a fractal actually is. Still to this day, we lack a well accepted universal definition. Most sets we call fractals seems to have a number of properties in common however. First and foremost, they seem to be consisting of parts that can be constructed from simple rules via an iterative process that creates a repeating pattern, which is seen at every scale. Also, many fractals exhibits something called self-similarity, which means that the whole set is similar to a part of it. We will now give a few examples of sets which we will call fractals, without making a formal definition of the term itself.

4.2.1 The Cantor Set

In 1883, the German mathematician Georg Cantor published a paper [3], where he introduced what is now known as the Cantor ternary set or the Cantor middle thirds set. The term ’The Cantor Set’ usually refers to this set, but this was just

an example in Cantor's paper. In general we call any perfect and nowhere dense set a Cantor set. Remember that a set is **perfect** if it is closed and has no isolated points. No isolated points means, for a set S, that given a point $x \in S$, every neighbourhood of x contains a point $y \neq x$ such that $y \in S$. The intuition for a nowhere dense set may be trickier, even though we may remember what dense means.

Definition 4.3 *A subset S of a topological space X is called **nowhere dense** in X if the interior of the closure of S is empty.*

One could also say that it is a set that is not dense in any non-empty, open subset of X. In a more loose sense it is a set whose elements are not tightly packed together. As an example, $\mathbb{Z}$ is nowhere dense in $\mathbb{R}$, since the closure of $\mathbb{Z}$ is just $\mathbb{Z}$ itself, and it does not have any interior points. However, $K = \mathbb{Z} \cup \Big((0,1) \cap \mathbb{Q} \Big)$ is not nowhere dense, since $\mathbb{Q}$ is dense in $\mathbb{R}$, and thus K is dense in $[0,1]$. For us, it is enough to realize that a nowhere dense set in $\mathbb{R}$ can not contain any non-degenerate intervals.

Thus we have two properties for a Cantor set in $\mathbb{R}$ that seem to contradict each other. Nowhere dense seems imply that points are separated from each other, while perfect seems to imply that the points are infinitely close together.

Let us now demonstrate the construction of a Cantor set via the middle thirds Cantor set.

We construct the middle thirds Cantor set iteratively by staring from the closed interval $C_0 = [0,1]$. We then remove the middle third part of C_0, i.e $\big(\frac{1}{3}, \frac{2}{3}\big)$. Thus leaving $C_1 = \big[0, \frac{1}{3}\big] \cup \big[\frac{2}{3}, 1\big]$. From each of these two intervals we then delete its middle third, namely $\big(\frac{1}{9}, \frac{2}{9}\big)$ and $\big(\frac{7}{9}, \frac{8}{9}\big)$, which gives us four new segments $C_2 = \big[0, \frac{1}{9}\big] \cup \big[\frac{2}{9}, \frac{1}{3}\big] \cup \big[\frac{2}{3}, \frac{7}{9}\big] \cup \big[\frac{8}{9}, 1\big]$, and so on. The open sets we remove from C_{n-1} to get C_n has the general form

$$\Big(\frac{3k+1}{3^n}, \frac{3k+2}{3^n}\Big),$$

where $k \in \{0, 1, 2, ..., 3^{n-1} - 1\}$.

Remark 1 *Note that k counts the number of possible thirds, even the ones previously removed. As we are taking the union of these, that does not matter.*

The closed formula for the Cantor middle thirds set then becomes

$$C = [0,1] \setminus \bigcup_{n=1}^{\infty} \bigcup_{k=0}^{3^{n-1}-1} \Big(\frac{3k+1}{3^n}, \frac{3k+2}{3^n}\Big). \tag{21}$$

As we see from (21), we are dealing with a limiting set as $n \to \infty$. The Cantor set is what is left over after iterating the above described process infinitely many steps. In other words, C_∞ is the Cantor set. This is best illustrated by Figure 10.

Notice also how many length units we remove from $[0,1]$. The first part is $\frac{1}{3}$, the second is $\frac{2}{9}$, and the n:th is $\frac{2^n}{3^{n+1}}$ units. In total we thus remove

$$\sum_{n=0}^{\infty} \frac{2^n}{3^{n+1}} = \frac{1}{3}\Big(\frac{1}{1-\frac{2}{3}}\Big) = 1$$

units of length.

So then one may ask, "What is left?". One thing we can notice is that, since we only remove open intervals, the endpoints all remain. So at least C_∞ is non-empty. We will explore this further using expansion in a different base.

Our usual way of writing decimal numbers in base-10 holds within it some information that we may not think of as useful. We may ask ourselves the question, "Where in $[0,1]$ is the number 0.3456 located?". The base-10 expansion then in fact tells us that it is in the fourth ($0.\mathbf{3}456$) tenth (base-10) from the left, in the fifth ($0.3\mathbf{4}56$) tenth of the previous fourth tenth, and in the sixth ($0.34\mathbf{5}6$) tenth of that, and so on (remember that we of course start from 0 indicating the first tenth). If we instead expand in base three, we get the same thing but in thirds, which will be useful for us here.

In base three, the point 0.012 will for example be in the last third of the second third of the first third of $[0,1]$. Now let us imagine the points that are left in the actual Cantor set expanded in base-3. Since we always remove the middle third of all the intervals left over at any stage, we first remove numbers of the form $0.1...$ (the middle of $[0,1]$), then the ones of the form $0.01...$ and $0.21...$, then $0.001...$, $0.021...$, $0.201...$ and $0.221...$, and so on. We end up deleting all points with 1's in their base-3 expansion. In fact, we only delete the points that has a 1 in its base-3 expansion, and we are thus left with all the other points.

Remark 2 *Note that endpoints like* 1 *and* $\frac{1}{3} = (0.1)_3$ *are never removed. However, these could equivalently be written as* $0.222...$ *and* $0.0222...$ *respectively (just as* $1 = 0.999...$ *in base-10).*

Now suppose we could make a list L of all the points in the Cantor middle thirds set. We would have something like $L = \{0.a_{11}a_{12}a_{13}..., 0.a_{21}a_{22}a_{23}..., ...\}$, where all numbers are in base-3. Now let $x^* = 0.b_1b_2b_3...$, where $b_i = 0$ if $a_{ii} = 2$, and $b_i = 2$ if $a_{ii} = 0$. Then $x^* \in C_\infty$, since C_∞ contain all points with no 1's, but $x^* \notin L$. We have constructed a number in C_∞ that is not on our list. Thus C_∞ contains uncountably many points! As the reader may notice, this is just Cantors famous diagonal argument.

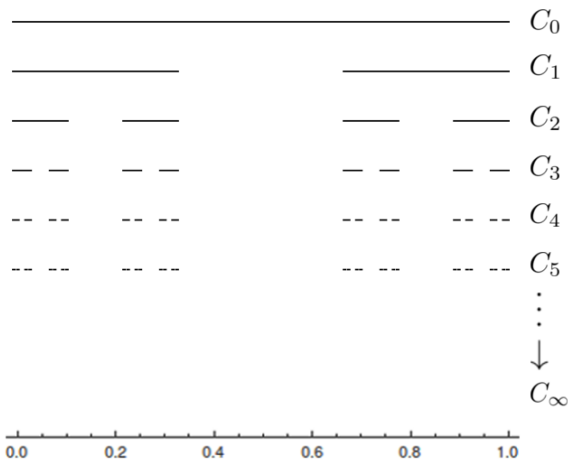


Figure 10: Demonstrating the construction of the Cantor middle thirds set. In each step a new set C_n is created by removing the middle third of each of the segments in C_{n-1}. The Cantor middle thirds set is the set we get at infinity, C_∞.

In conclusion, the Cantor middle thirds set is constructed from removing infinitely many intervals from $[0, 1]$, it is closed (since any intersection of closed sets is closed), has zero measure, but contain infinitely many points and is nowhere dense in $\mathbb{R}$. In fact C_∞ contains as many points as $[0, 1]$, but still contains no non-degenerate intervals.

4.2.2 The Von Koch Curve

The Von Koch curve is a fractal described by the Swedish mathematician Helge Von Koch in a paper from 1904. We start off by looking at the construction of it in a similar fashion as for the Cantor middle thirds set.

To construct the Von Koch curve we start with a line segment K_0. Let K_{ij} denote the j:th straight line segment of K_i from the left. We then do the following three steps to get to K_1:

1. Divide each of $K_{ij}, j \in \{1, ..., 4^i\}$ into three equal parts.

2. For each j, create an equilateral triangle pointing upwards using the middle third of K_{ij} as a base.

3. Remove the base(s) of the triangle(s) created in step 2.

Do these steps again, starting with K_1 to get to K_2, and so on iteratively with K_l to get to K_{l+1}. The construction is illustrated in Figure 11.

The reason why we haven't drawn C_∞ or K_∞ in Figure 10 or Figure 11 respectively, is of course that we can't. These sets are objects that you would

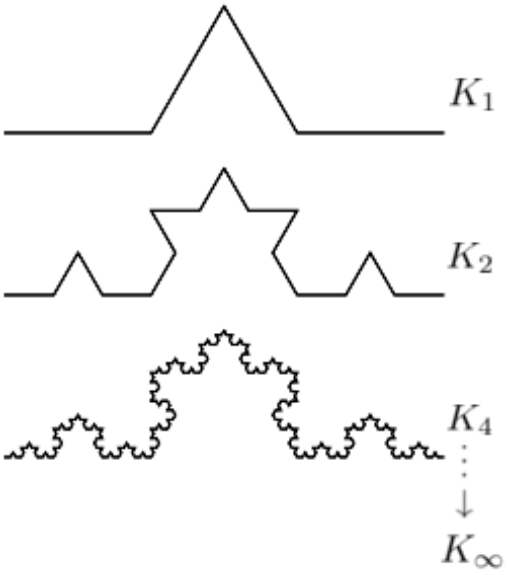


Figure 11: Demonstrating the construction of the Koch curve. In each step a new set K_i is created by creating an equilateral triangle out of the middle third of each line segment created by the previous step, and then removing its base. The Koch curve is the limiting set K_∞.

have to imagine, if you can. This is one of the most important advantages of mathematics, it can describe it precisely.

Let us now look at the length of K_∞. Let L_i denote the length of the straight line segments of K_i. Since the triangles being created from each segment of K_i are equilateral, with one third of the previous segment as its base and the base is discarded, we end up with four new line segments. Each of these then have the length $L_i = 4\frac{1}{3}L_{i-1}$. It is then easy to see that in general we get

$$L_i = (\frac{4}{3})^n L_0 \to \infty, \quad \text{as } i \to \infty.$$

So the Koch curve has infinite length. It is also self similar, meaning that you can take any part of it and scale it up to match the whole thing. Since any two different points on K_∞ could be seen as lying in two different line segments at some stage of the construction, an infinitely long curve will be created between them. Thus any two points on K_∞ are infinitely far away from each other on K_∞. This makes it hard to classify the Koch curve as being 1-dimensional. How do we even specify the position for an arbitrary point on it? However, it does not seem justified to classify it as being 2-dimensional either, since it is a curve after all. So what is it?

Let us now introduce different ways of making sense of the dimensions of

these sets.

4.3 Fractal Dimension

Because of the intricate geometrical nature of the fractals introduced in the previous section (we can't even specify a point on K_∞), it seems that our notion of dimension of a set as being the minimum number of coordinates needed to specify a point in it, may not be as rigid as we thought. Constructions made by Giuseppe Peano [30] in 1890, and by David Hilbert [15] in 1891, i.e the Peano curve and Hilbert curve respectively, further substantiates these doubts. These are examples of space filling curves. Normally we think of a curve as something one dimensional, as it can be parametrized using exactly one parameter. What Peano and Hilbert showed was that there are curves that visits all points of a square, and more specifically that the unit square $[0,1] \times [0,1]$ (2-dimensional) could be parametrized by one parameter, ranging over $[0,1]$ (1-dimensional).

Another important paper on the subject, that has to be mentioned, is the one by Benoit Mandelbrot [23] from 1967. In his paper, Mandelbrot discuss the idea of measuring a coast line using a measuring stick (or a compass if you like). Say you choose a length of s units for the stick, and you start placing it around the coastline, you will then get a length of say L units. Now, suppose next you choose a shorter stick, i.e. you decrease s. What happens to L? The result is that L will increase due to all the creases and irregularities in the coastline that the stick of length s missed. If we choose an even smaller s, then L will again increase. Thus there seem to be a relation between the scale of the individual measurements and the length of the coastline. The suggestion is that they are related by a power law $L \propto s^d$, where d is the **fractal dimension** of the object which we are measuring. Since we have that

$$L = Cs^d \iff \log(L) = \log(Cs^d) \iff \log(L) = d\log(s) + \log(C)$$

for some constant C, experimentally one may do such a measurement, and then plot $\log(L)$ against $\log(s)$ to hopefully get points almost on a line with slope d. In practice one would do a least squares approximation to fit a line through the points, and take the slope of this line to be d. Say that $C = 1$, so that $L = s^d \Leftrightarrow d = \dfrac{\log(L)}{\log(s)}$. Intuitively one may see this as a ratio describing how the object change in detail with a change in scale.

Mandelbrot mentions in a note to his paper that, *"The concept of dimension is elusive and very complex, and is far from exhausted by simple considerations of the kind used in this paper. Different definitions frequently yield different results, and the field abounds in paradoxes. However, the Hausdorff-Besicovitch and the capacity dimensions, when computed for random self-similar figures, have so far yielded the same value as the similarity dimension."*[23] This sum-

marizes our suspicions that dimension might not have been as intuitive of a concept as we thought. To clarify what was meant by the last sentence in the quotation above, we will further study the three concepts that Mandelbrot mentioned.

4.3.1 Hausdorff-Besicovitch Dimension

The Hausdorff-Besicovitch dimension (introduced in 1918), after Felix Hausdorff (1868-1942) and Abram Samoilovitch Besicovitch (1891-1970), is probably the oldest and perhaps most important definition of dimension to the field of fractal geometry[12]. It generalizes the concept of dimension of a vector space, and it nicely captures the idea of dimension as a number that indicates the amount of space a set occupies near each of its points [32].

The Hausdorff-Besicovitch dimension uses another concept called the Hausdorff measure, that in turn generalizes the concepts of length, area, and volume, etc. A complete treatment of Hausdorff-Besicovitch dimension is out of scope for this article, but we will take a look at the main definitions and give a short intuition for them.

Definition 4.4 *Let (X, ρ) be a metric space, S be any subset of X, and $\delta > 0$ be a real number. Further define*

$$H^d_\delta(S) := \inf\Big\{ \sum_i^\infty (\operatorname{diam} U_i)^d : \bigcup_{i=1}^\infty U_i \supseteq S, \operatorname{diam} U_i < \delta \Big\}.$$

Then

$$\mathcal{H}^d(S) := \lim_{\delta \to 0} \mathcal{H}^d_\delta(S)$$

*is called the **d-dimensional Hausdorff measure** of S.*

So we look at all covers $\{U_i\}$ of a set S with sets U_i of diameter at most δ. For each such cover we then look at the d:th powers of the diameters. Now if S has a topological dimension of d, then $(\operatorname{diam} U_i)^d$ is an approximation of the d-dimensional volume of a little piece of S, i.e. U_i. If we then sum up those volumes for all sets in a cover of S, then we should get at least the volume of S. Surely it will be a lot bigger for many covers, hence we take the infimum of these sums. If d is however larger than the topological dimension of S then these sums will get arbitrarily close to 0 (e.g. if S is a disk but we have $d = 3$, then since a cube of side δ is a smaller portion of a ball than a square of side δ is of a disk, when we take smaller and smaller δ and we only add up over sets that cover the disk, we are going to get smaller and smaller sums.) If we then

[12]Hausdorff did not use the term fractal as this was something coined by Mandelbrot. Mandelbrot, however, used the concept of Hausdorff-Besicovitch dimension as a basis for further study of the geometrical objects he called fractals.

look at the values of $d \geq 0$ where $\mathcal{H}^d(S) = 0$, and take the infimum of these, we should get the dimension of S. This is just before $\mathcal{H}^d(S)$, as a function of d, switches over to 0, i.e. when we pass the threshold where d becomes too big. We therefore define the **Hausdorff-Besicovitch dimension** as

$$\dim_H(S) := \inf\{d \geq 0 : H^d(S) = 0\}.$$

It can be shown for subsets of $\mathbb{R}^n$ that n-dimensional Hausdorff measure coincides, to within a constant multiple, with the usual n-dimensional volume [5], i.e. $H^0(A)$ is the number of points in A, while $H^1(B)$ is the length of the curve B, and so on. For further and more rigorous treatment of the Hausdorff-Besicovitch dimension, see [5].

Hausdorff's definition is mathematically nice and hugely important in theory, however, as Peitgen puts it, it is impractical to use even in elementary examples and nearly impossible to estimate in practical applications [32].

4.3.2 On the Definition of Fractal

Remembering the definition of topological dimension from before, the essential property of a fractal set is that its Hausdorff-Besicovitch dimension exceeds its topological dimension. Objects with this feature are referred to as having a **fractal geometry**, or as being fractals, and their possibly approximated Hausdorff dimension is referred to as their fractal dimension. This suggests for example that coastlines are often fractals, since we are essentially measuring 1-dimensional curves to begin with, but usually get dimensions $d \in (1, 2)$. The fractal dimension of the curves of Peano and Hilbert from before has also been approximated, having a fractal dimension of 2 even though being 1-dimensional curves in terms of topological dimension. This of course reflects their space-filling nature.

Since Mandelbrot, the fractal dimension of different coastlines has been approximated by several people over the years. Other than the already mentioned Britain (1.25), a couple of example measurements include Ireland of 1.22 ± 0.02 [17], and Norway of 1.52 [7]. This then suggests that the coast of Norway has a somewhat more complicated structure than Ireland. That would certainly correspond well with observation since Norway has a very convoluted coastline with lots of bays and fjords.

As we can see from the quotation of Mandelbrot in Section 4.3, there are other definitions with the same idea in mind that mostly yields the same results as Hausdorff-Beisicovitch dimension. Some of these are also much easier to work with even though being somewhat ad hoc at times. For example, the experimental approach discussed earlier about measuring coastlines aims at approximating the Hausdorff-Beisicovitch dimension. In the next two sections we will be discussing the two other concepts that Mandelbrot mention.

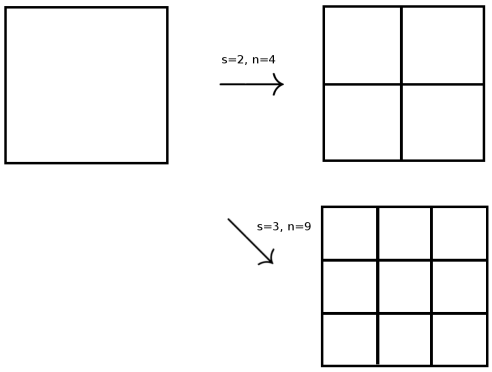


Figure 12: Similarity dimension of the square as a ratio .

4.3.3 Similarity Dimension

If an object can be seen as a collection of sets, each of which is similar to the whole set, we call the set **self-similar**. Self-similar then refers to a an object where you can take any of the parts that make up the object, and scale it uniformly to get the whole set. Another related property is self-affinity, where the scaling would not be uniform, i.e not the same in all directions. To retrieve a measure for the dimension in terms of a ratio discussed in the previous section for self similar objects, one could do like follows:
Scale down the object A by a certain amount s, and see how s is related to the number of scaled down versions n it takes to make up A, using the power law $n = s^d$. We would then define d to be the ***similarity dimension*** *of A.*
Take the unit interval $[0, 1]$ for example. Let I_s denote $[0, 1]$ scaled down by s, e.g. $I_2 = [0, \frac{1}{2}]$. One would need 2 copies of I_2 to make up the whole of I_1. In general you would need s copies of I_s to make up I_1. We then have that $n = s$ so that $s = s^d$, and thus $d = 1$, which makes sense. For the unit square, when $s = 2$, then $n = 4$ or $4 = 2^d \Rightarrow d = 2$, as seen from Figure 12 which also includes an example when $s = 3$. Thus the similarity dimension is 2 for the square. In general we solve for d as we saw before and get $d = \log(n)/\log(s)$. One can easily see the connection to the example of measuring coastlines from before. Instead of choosing a smaller size of the measuring stick, we pick a smaller "stick" in the shape of the object we are measuring, and then count the length as the number of "sticks" we need in order to cover the object.

We have seen then that similarity dimension makes sense for "ordinary" self-similar objects. Let us now look at the Cantor set and the Koch curve from before.

Example 4.1 ***The Cantor Set****: At each stage in the construction C_k is composed of two copies of C_{k-1} scaled down by a factor 3, e.g. in Figure 10 the left and right part of C_2 is C_1 scaled down by 3, the left and right part of C_4 is C_3 scaled down by 3, and so on. We get that $s = 3, n = 2$ thus giving us $2 = 3^d \Leftrightarrow d = \log_3(2)/\log_3(3) = \log(2)/\log(3) \approx 0.63$.* //

Example 4.2 ***The Koch curve****: At each stage in the construction K_k is composed of four copies of K_{k-1} scaled down by a factor 3, e.g. in Figure 11 K_2 is composed of four copies of K_1 scaled down by 3. We get that $s = 3, n = 4$ thus giving us $4 = 3^d \Leftrightarrow d = \log_3(4)/\log_3(3) = \log(4)/\log(3) \approx 1.26$.* //

4.3.4 Minkowski–Bouligand dimension

The similarity dimension is of course very convenient for self-similar sets. However, for a wider collection of sets we need to generalize the concept even further. As with the previous definitions, we are interested in the way the measure of a set at a scale ϵ varies as we vary ϵ. One popular definition is the so called Minkowski–Bouligand dimension, also known as **capacity** (capacitary dimension in Mandelbrot's note above), **box dimension** or **box-counting dimension**[13]. It is computationally easy to use and has become one of the most frequently used definitions of dimension [5]. As before, we study how the measurement at a small scale varies as the scale varies. Hopefully the reader has started to become familiar with the power law idea *Measurement* $\propto$ *scale*^{d}. We make the following definition.

Definition 4.5 *Let S be a non-empty bounded subset of $\mathbb{R}^n$, and let $N_\delta(S)$ be the minimum number of sets of diameter at most δ which cover S. Then*

$$\dim_B S := \lim_{\delta \to 0} \frac{\log N_\delta(S)}{-\log(\delta)}$$

is called the ***Minkowski-Bouligand dimension*** *or* ***box dimension*** *of S, if it exists.*

One may also define upper and lower limits accordingly, and call them upper and lower box dimensions respectively.

The Minkowski-Bouligand dimension is in many ways similar to the similarity dimension, but with the crucial difference that we are not forced to deal with self-similar sets any more. To give an example of a fractal that is not self-similar, we give the following construction and calculation [38].

[13] Many of these measures of dimension has a variety of different names and sometimes different, albeit equivalent, ways to define them. This paper tries to focus on the names that create least confusion with other definitions, and that pays respect to the mathematicians who discovered them.

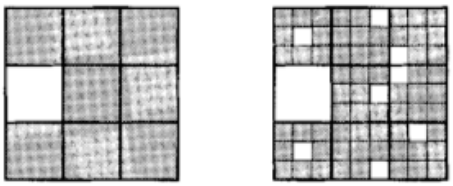


Figure 13: Illustrating the construction of a non-self-similar fractal. The construction resembles the Menger carpet, but the sections to be removed are determined at random. *Image source:[38]*

Example 4.3 *Start with the unit square, divide it into nine equal squares (three by three), select one of the nine squares at random and delete it. Repeat this process for each of the remaining eight squares, and so on. The process is illustrated by Figure 13. Now we see that* S_1 *is covered by* $N_{\frac{1}{3}}(S_1) = 8$ *squares of side* $\delta = \frac{1}{3}$. *Further we get* $N_{\frac{1}{9}}(S_2) = 16 = 8^2$. *In general we get* $N_{(\frac{1}{3})^n}(S_n) = 8^n$. *Then, since* $N_\delta(S_n) = 8^n$, *we can use the sequence* $(\delta_n) = ((\frac{1}{3})^n)$ *and Definition 4.5 to get that*

$$\dim_B(S_\infty) = \lim_{\delta \to 0} \frac{\log(8^n)}{-\log((\frac{1}{3})^n)} = \lim_{n \to \infty} = \frac{n\log(8)}{-n\log(\frac{1}{3})} = \frac{\log(8)}{\log(3)} \approx 1.893 /\!/$$

Before we move on to discuss the concept of Renyi dimension, we first look at the concept of a strange attractor.

4.4 Strange Attractors

Unfortunately there is not even a universally accepted definition of attractor [38, p. 324], but informally we may think of an attractor as a set A in phase space that attracts nearby trajectories. It is also reasonable to require that A is invariant under the function f that governs the system, i.e. $f(A) \subset A$. Further it should be the minimal set for which the previous applies, otherwise there would be some other set $B \subset A$ that could just as well be seen as the attractor. It seems reasonable to make the following definition:

Definition 4.6 *Let* $f : U \subset \mathbb{R}^n \to U$. *A closed and bounded set* $A \subset U$ *is an* ***attractor*** *if* $f(A) = A$, *and there exists* $r > 0$ *such that* $d(x_0, A) \leq r \Rightarrow d(x_n, A) \to 0, \quad as\, n \to \infty$, *where* $d(x, S) = \inf\{|x - y| \Big| y \in S\}$ *is the distance from a point to a set.*

This means that if we start within a distance r from the attractor A (usually referred to as within the **basin of attraction**), that orbit will eventually end up in A. Further, since $f(A) = A$, this is the smallest such set. Many authors also require that the orbits starting within the basin of attraction, are dense in A. Numerically, one can visualize an attractor by starting sufficiently close to it, omitting the transient points and look at a very large, but thus finite, orbit. This should thus "fill up" the attractor. As we saw in Section 3.3, the logistic map was chaotic in $[0, 1]$ for certain values of r. Seen as a map $f : \mathbb{R} \to \mathbb{R}$, we realize that even though $[0, 1]$ is invariant, and orbits starting within $[0, 1]$, were dense in it, it is not an attractor since no matter how close to $[0, 1]$, outside of it, we start, the orbit will diverge away to $-\infty$, as we have seen before. So a map can of course be chaotic in an invariant set that is not necessarily an attractor. It is however very common for the attractors, on which maps behave chaotically, to have a strange geometry, as we gave some intuition for in Section 3.3.1. This is perhaps why they got the name 'strange attractors'.

The definition of a **strange attractor** is not very universal, and there is still some disagreement on this topic. The first use of the term 'strange attractor' is usually attributed to David Ruelle and Floris Takens, in an article on turbulence from 1971 [33]. Originally it was meant to reflect a geometrical property of an object, i.e an attractor was strange if it had a fractal geometry. This is sometimes still used, but these days it seems more common to refer to attractors as strange if they exhibit sensitive dependence on initial conditions. This is because the geometrical aspect of being fractal is usually regarded as less important than the dynamical property of exhibiting sensitive dependence on initial conditions [38].

Let us look at two examples of attractors that have been called strange. We are already familiar with the logistic map $f(x) = rx(1 - x)$. If we look at the bifurcation diagram, we may view this as a graph of the attractor for a specific value of r. So each value of r gives a different attractor. As we can see, for values of r below approximately 3.57 (a more accurate approximation is 3.5699456) the attractor is just a set of periodic points. As we increase r, the number of periodic points increase by a factor two. This was what we called a period doubling bifurcation in Section 2.2.2. If we start after the first bifurcation point $r = 1$, which we called p_1, and continue to increase r, we saw that we got $p_2 = 3$ and $p_3 = 1 + \sqrt{6}$. As we can see in Figure 14, this continues, and the next value is actually $p_4 \approx 5.4409$. We thus get a sequence $(p_n)_{n\geq 2}$ such that every p_n is a period doubling bifurcation point. The sequence $(p_n)_{n\geq 2}$ converges to 3.5699456, which is often called the **onset of chaos**. This type of sequence of bifurcation points is often called a **bifurcation cascade**, and in specific, when all bifurcations are of the period doubling type, a **period doubling cascade**[14]. This process should remind ourselves of the creation of the

[14]This is something that M.J. Feigenbaum noted was a common, so called, route to chaos in many systems. It has therefore been called the *period doubling route to chaos* .

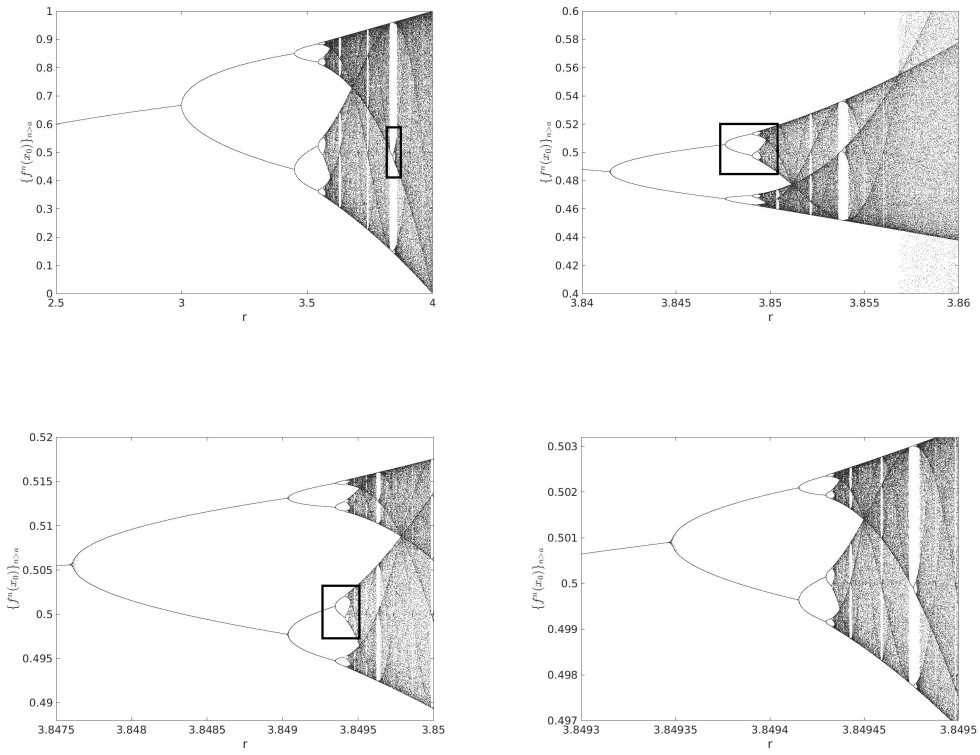


Figure 14: Results of zooming in on the bifurcation diagram for the logistic map. The boxed regions of the image on the left are blown up to the right. So in (b) we see the boxed region of (a), and in (c) we see the boxed region of (b), and in (d) we see the boxed region of (c).

Cantor Middle Thirds set, and it turns out, as we shall see in Section 4.5.1, that they both have a similar value for its fractal dimension. As a matter of fact, the attractor for the logistic map can be shown to be a general Cantor set, for certain values of r. If we zoom in on the bifurcation diagram of the logistic map, we can also see, in Figure 14, the self similar pattern emerging on smaller and smaller scales.

Another map that has a strange attractor, is the Hénon map. This was introduced by the French mathematician Michel Hénon, as a simplified, lower dimensional model of the Lorentz systems, which Edward Lorentz studied when discovering chaotic dynamics. This is a two-dimensional map defined by $h(x, y) = (y + 1 - ax^2, bx)$, where a and b are parameters. In his paper from 1976 [14], Hénon argues that the systems converges to a strange attractor for parameter values $(a, b) = (1.4, 0.3)$. This is one of the simplest models that rather elegantly illustrates one very common feature of chaotic maps, namely stretching and folding. In Figure 15, we can see what happens when we apply the Hénon map to a square. We can see the square being squashed, folded, and

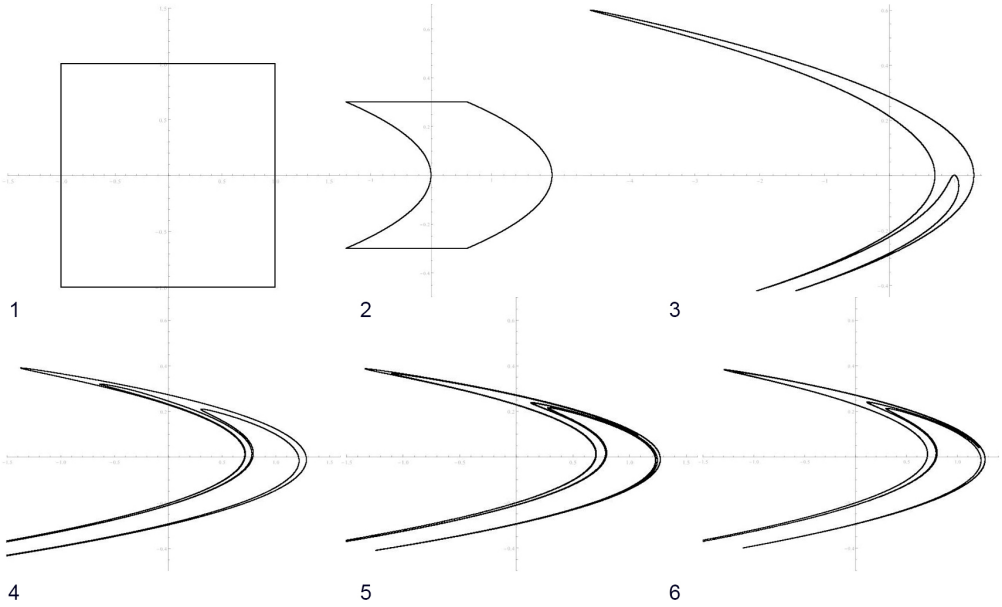


Figure 15: This shows the formation of the so called Hénon attractor. We see five iterations of the Hénon map on a square (left to right, up, down), demonstrating the folding and streching effect of the map.

stretched out, and so on. What is slowly forming is the so called Hénon attractor. This results in a very detailed micro-structure that shows similarity on smaller and smaller scales, as seen in Figure 16.

4.5 Measuring the Dimension of Strange Attractors

When trying to numerically approximate the fractal dimension of an attractor, the most straightforward way would probably be to generate a large set of points from the orbit $O(x_0)$, and select a subset U of this set by skipping a number of points from the beginning. This makes sure that the points in our set is actually located on the attractor (this method was already touched upon in Section 3.3.2, and in the construction of the bifurcation diagram). We will sometimes, when clear from context what we mean, refer to this obviously finite set as the orbit, even though it is not literally the whole orbit, since that would be infinite. Continue by partitioning the d-dimensional phase space in equal sized d-dimensional squares of side length ϵ_i, and count the number $\bar{N}(\epsilon_i)$ of squares containing at least one point of U (those squares will constitute a cover of the attractor). Following the procedure suggested in the introduction to this section, take $\bar{N}(\epsilon_i)$ as the measure of your attractor and plot $\ln\left(\bar{N}(\epsilon_i)\right)$ against $\ln\left(1/\epsilon_i\right)$ for many small values of ϵ. This will hopefully give a nice fit with a straight line, and we may then take the slope of this line as an approximation of the Minkowski-Bouligand dimension $\dim_B$, or box dimension for short.

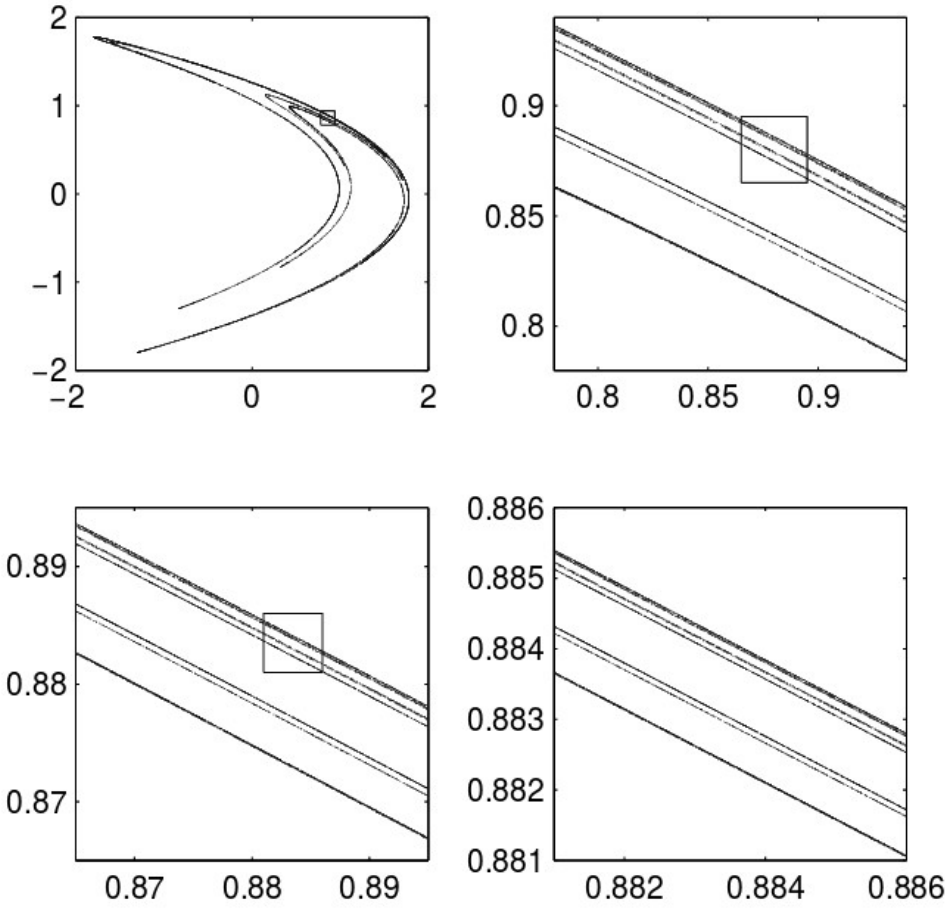


Figure 16: By zooming in on the Hénon attractor (above to the left) we see that the same type of pattern emerge at smaller and smaller scales, demonstrating the self similarity of the attractor. *Image source: [22]*

The above mentioned procedure is of course a direct analogue of what we were discussing for measuring coast lines before. However, there is at least one possible problem with this method, and it has to do with the fact that a box could contain one or a very large number of points. As a matter of fact, when ϵ is small enough, most boxes will contain a very small amount of points, and only a few of the boxes will contain the majority of the orbit. Actually there may be a large amount of boxes that are not counted, since they are empty, but that would contain points if the generated orbit were made just a little bit longer. This especially poses a problem when the points of the orbit tend to clutter together on the attractor, i.e. filling up the attractor faster in some areas than others. We will now illustrate this problem, with an example.

Consider the map defined as

$$\Gamma(x, y) = (2x + y, x + y)(\mod 1).$$

This is known as Arnold's cat map, and it is a map $\Gamma : \mathbb{T}^2 \to \mathbb{T}^2$, which is perhaps more intuitively clear if we think of the torus $\mathbb{T}^2$ as the quotient space $\mathbb{R}^2/\mathbb{Z}^2$. It is named after the Russian mathematician Vladimir Igorevich Arnold (12 June 1937 – 3 June 2010), who first described the effects of the map in the 1960:s, by applying it to a picture of a cat. For this map, the whole torus is an attractor, but what happens if we introduce a small perturbation? Let us look at[15]

$$\Gamma_*(x, y) = (2x + y, x + y + \delta \cos(2\pi x))(\mod 1).$$

In Figure 17 we see three different length orbits, for $\delta = 0.001, \delta = 0.08$ and $\delta = 0.23$, plotted on a torus[16]. We see, for $\delta = 0.001$, that the orbits slowly start to fill up $\mathbb{T}^2$, rather uniformly. As a matter of fact, Yakov Sinai has shown that the whole torus is still an attractor for small δ [36]. As δ increases however, points are clustering together at certain areas, and the torus fills up faster in these areas than in other. In Figure 17f, for $\delta = 0.23$, we see that it is no longer even clear that the torus is an attractor. We may actually have a smaller attracting set made up of the bands, seen swirling around the torus in the plot. If we look at this plot, on the square and zoom in on the bands, as seen in Figure 18, we also see that the pattern repeats itself on smaller scales. As was discussed in the previous paragraph, using box counting, there will therefore be fewer boxes

[15]We will not go into details and specifics of the dynamics and the attracting sets of this family of maps. This is just used here to illustrate some points about different ways to measure dimension. For more info about these type of maps, called Anosov diffeomorphisms, we refer to [36]

[16]For estethic purposes we have plotted this on a somewhat larger torus, of major radius 2 instead of 1. The purpose of plotting this on a torus in the first place, is merely of an illustrative nature. This way, one can clearly see the orbits 'swirling' around the torus instead of having to imagining the square as the torus. For most purposes however, one may just as well look at plots on the square, as in 18.

actually accounting for most of the points in the orbit. Since there are potentially a lot of empty boxes that would eventually fill up, if the orbit was made longer, we tend to underestimate $\bar{N}(\epsilon_i)$, and thus in turn, underestimate $\dim_B$. One may also argue that it makes sense to distinguish between boxes that the orbit spends a lot of time in, and the ones that it hardly ever visits. We will therefore now look at a measure of dimension that does precisely this. In Figure 18, we have zoomed in on some of the bands (here depicted in the unit square for simplicity). As we zoom in, we can see that the bands seem to consist of several similar bands. As a matter of fact, this pattern continues on smaller and smaller scales, and may suggest a fractal micro-structure. However, as we mentioned before, it is not entirely clear that we have a smaller attracting set (which may be fractal), or if the entire torus is still an attractor, and we just need to make the orbit longer.

4.5.1 Renyi Dimension

As we just saw, there are some possible drawbacks with using the box-dimension, as it does not distinguish between the different boxes needed to cover the attractor. There is a generalized notion of dimension, often referred to as the **Renyi dimension** or sometimes the **generalized dimension**, that does this by weighting each box B_i according to its so called natural measure $\mu(B_i)$. We will not go into details of the definition of μ here, but instead just informally state that $\mu(B_i)$ can be seen as the limit, as $T \to \infty$, of the fraction of points, of an orbit, in B_i, during the interval $(0, T)$ to the total number of points in the orbit. Thus, one may equally think of $\mu(B_i)$ as the probability of a random point being in B_i. We then define the order q Renyi dimension D_q as

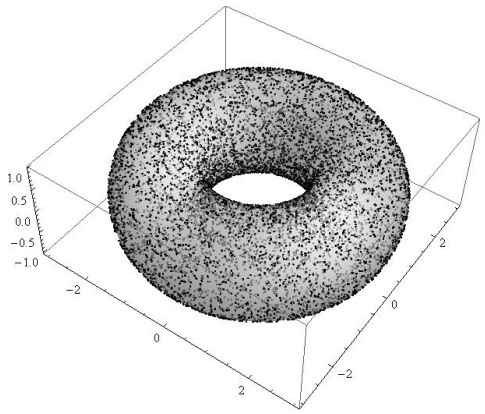
$$D_q = \lim_{\epsilon \to 0} \frac{1}{1-q} \frac{\ln\left(\sum_j (\mu(B_i))^q\right)}{\ln\left(\frac{1}{\epsilon}\right)}. \tag{22}$$

where q gives the strength of the weighting, i.e. the higher value of q, the stronger the weighting of the boxes with larger measure $\mu(B_i)$, compared to the ones with smaller measure. We note that setting $q = 0$, we retrieve the box dimension, i.e. $D_0 = \dim_B$.

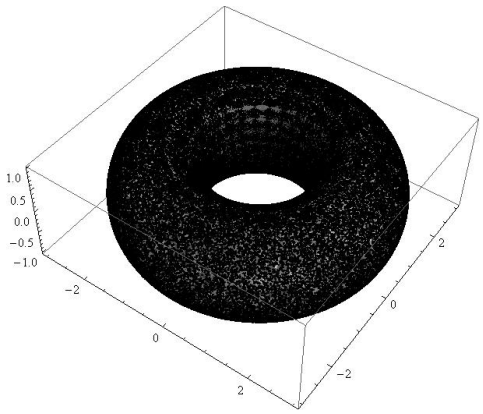
Taking the limit as $q \to 1$ in (22), we may note, since $\sum_i \mu(B_i) = 1$, that

$$\ln\left(\sum_i \mu(B_i)^q\right) \to ln(1) = 0 \text{ and}$$

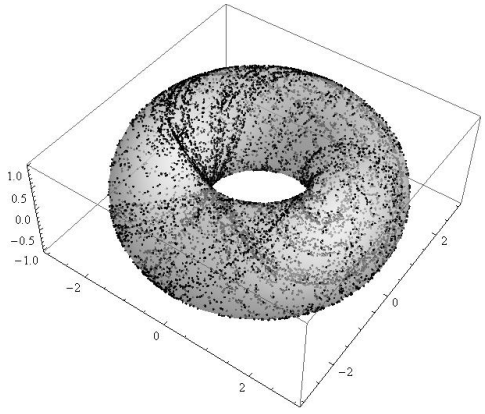
$$(1-q)\ln\left(\frac{1}{\epsilon}\right) \to 0,$$



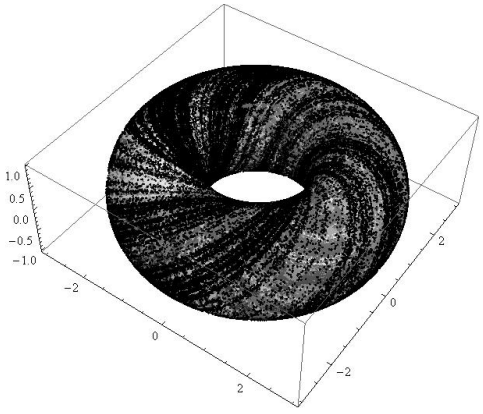
(a) $\delta = 0.0001, N = 10000$



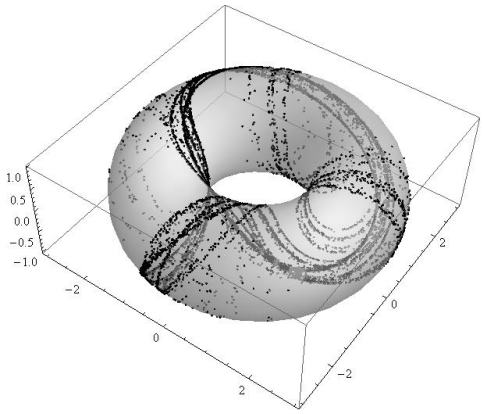
(b) $\delta = 0.0001, N = 150000$



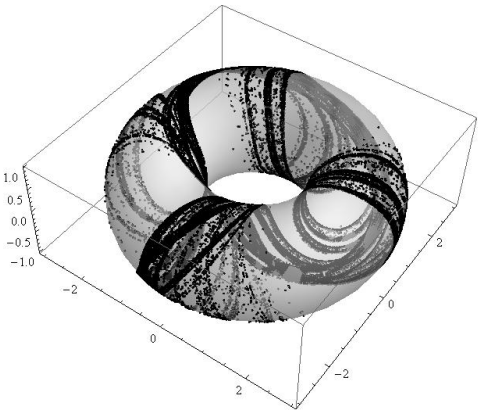
(c) $\delta = 0.1, N = 10000$



(d) $\delta = 0.1, N = 150000$

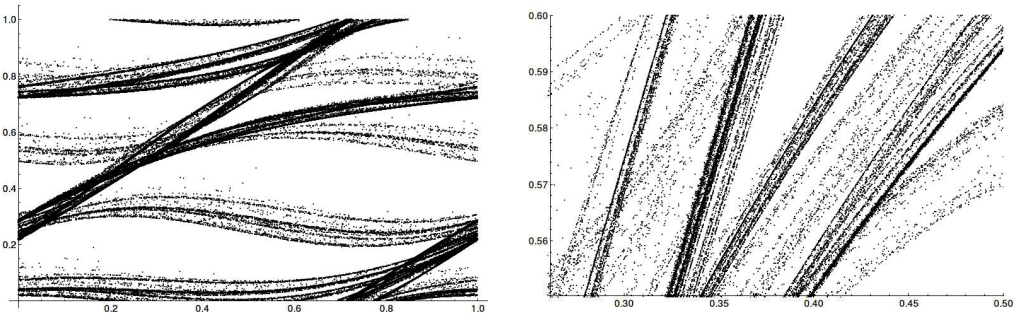


(e) $\delta = 0.23, N = 10000$



(f) $\delta = 0.23, N = 150000$

Figure 17: N iterations of Γ_* starting from $(0.1, 0.1)$ projected onto a torus, for three different values of δ. For a small δ, numerical results suggests that the whole torus is an attractor. When δ is increased, patterns emerge, showing that the orbit spend more time in certain areas than other. It is not clear whether, for a larger δ, we have a smaller attracting subset of the torus, or if the torus would be filled if the orbit was made long enough.



(a) $(x, y) \in I \times I$ (b) $x \in [0.26, 0.5], y \in [0.55, 0.6]$

Figure 18: 200000 and 600000 iterations of Γ_* starting from $(0.1, 0.1)$, plotted in the unit square $I \times I$, for $\delta = 0.23$. Figure 18b shows some of the bands magnified, which reveals a similar micro-structure of new bands.

and therefore we may apply l'Hospital's rule, to arrive at

$$\lim_{\epsilon \to 0} \lim_{q \to 1} \frac{1}{1-q} \frac{\ln\left(\sum_j (\mu(B_i))^q\right)}{\ln\left(\frac{1}{\epsilon}\right)} = \lim_{\epsilon \to 0} \lim_{q \to 1} \frac{\sum_i \left(\mu(B_i)^q \ln\left(\mu(B_i)\right)\right)}{\sum_i \left(\mu(B_i)^q\right) \ln(\epsilon)}$$
$$= \lim_{\epsilon \to 0} \frac{\sum_i \left(\mu(B_i) \ln\left(\mu(B_i)\right)\right)}{\ln(\epsilon)} := D_1.$$

The limit D_1 is usually called the **information dimension**. It is also worth noting that D_q is a non-increasing function in q [29].

In 1983 Peter Grassberger and Itamar Procaccia introduced the correlation dimension in their paper *Measuring the Strangeness of Strange Attractors* [10], as a method for estimating the fractal dimension of so called strange attractors. This dimension actually corresponds to D_2 above, and it has a convenient numerical algorithm, often referred to as the Grassberger-Procaccia algorithm. The idea is similar to the previously presented definitions, but here we look at pairwise distances instead of looking at a cover of boxes. The main advantage of this is that it allows one too look at very small scales. If the same small scales would to be used in a box-counting algorithm, we would have the problem discussed above, with many boxes being considered empty [39].

We start by counting the number of pairs of points in the orbit, or more specifically of the first N points in it, that are closer together than ϵ. We then divide by the total number of pairs

$$\binom{N}{2} = \frac{N!}{2(N-2)!} = \frac{N(N-1)(N-2)!}{2(N-2)!} = \frac{N(N-1)}{2},$$

to get a ratio that describes the fraction of pairs closer to each other than ϵ. We

thus define the **correlation integral** as

$$C(\epsilon) = \lim_{N\to\infty} \frac{2}{N(N-1)} \sum_{i<j=1}^{N} \Theta(\epsilon - \rho(x_i, x_j)), \tag{23}$$

where Θ is the Heaviside step function, i.e.

$$\Theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0, \end{cases},$$

and $\rho(x, y)$ is the metric used in the space that embeds A (usually the standard euclidean metric). We then define the **correlation dimension** for A (or to be precise, for the orbit) as

$$\dim_C = \lim_{\epsilon\to 0} \frac{ln(C(\epsilon))}{\ln(\epsilon)}. \tag{24}$$

Of course it is impossible for us to deal with these limits in practice, so what we do is that we approximate $C(\epsilon)$ by summing over all points in a large orbit we generated numerically, i.e. for a very big N. We then do this for many different values of ϵ and plot $\ln(C(\epsilon))$ against $\ln(\epsilon)$. If the points are related by a power law, there should be a linear correlation between the two, meaning that we should be able to find a close fit of a line through the points using the method of least squares.

Remark 3 *As we can see from* (23)*, for large enough ϵ we get $C(\epsilon) = 1$. This means that* (24) *won't supply us with any useful information. If on the other hand ϵ is small enough, $C(\epsilon)$ will be close to or equal to zero, and* (24) *will not supply us with useful information. In practice we must thus tailor our choice of ϵ to avoid this.*

In Figure 19a and 19b we can see an approximation of the correlation dimension using the above method for two well known attractors to discrete dynamical systems. In Figure 19a we generated points using the logistic map $f(x) = rx(1-x)$ for $r = 3.56995$, starting from $x_0 = 0.5$, and in Figure 19b we looked at the Hénon map $h(x, y) = (y + 1 - ax^2, bx)$ with $a = 1.4$, $b = 0.3$ starting from $x_0 = (0.1, 0.1)$. We found an approximated correlation dimension of 0.50117 for the logistic attractor and 1.2057 for the Hénon attractor. We may compare this result to the values of Grassberger and Procaccia of 0.500 ± 0.005 and 1.21 ± 0.01 [17] respectively [10]. We have also numerically approximated the information dimension (D_1) of Arnold's cat map, with a small perturbation,

[17] The results of the dimension of the Hénon attractor are actually underestimated and a better method for approximating the dimension, using embedding dimensions, are discussed in [10]

i.e. Γ_* from the previous section. We have both looked at this for small values of δ, where the whole torus is an attractor, and for larger values of δ, where we seem to have a smaller attractor. The results of this can be seen in Figure 19c and 19d respectively.

4.5.2 Lyapunov Dimension, and the Kaplan-Yorke Conjecture

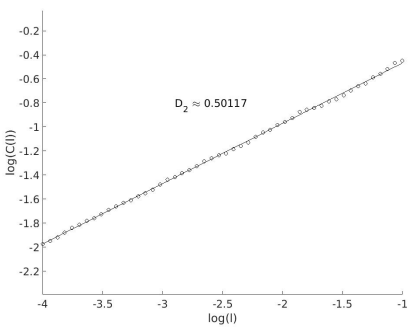
Before we conclude this thesis, we will now, as promised in the end of Section 3.3.2, in part return to the dynamics. We will introduce a conjecture that in a way tries to bind together the notions of geometry and dynamics. It does so by suggesting that an attractors dimension (geometry) could be expressed in terms of the systems Lyapunov exponents (dynamics). The intuition for, and presentation of this conjecture, known as the **Kaplan-Yorke Conjecture**, will be rather brief and will merely be used as a concluding subject, demonstrating the connections between the different areas that we have been discussing.

We begin by building some intuition. Imagine covering an attractor A, for a 2-dimensional system, governed by a function f, with boxes of side δ. Let, as in the section on box-dimension, $N(\delta)$ denote the number of such boxes needed to cover the attractor (purely for convenience, we depart from the previously used notation $N_\delta(A)$ that included the attractor A in the notation). Then study how an arbitrary box B is deformed by f^n, i.e. after n iterations. As we saw in Section 3.3.2, when discussing the Lyapunov numbers, the stretching factors were given by the eigenvalues of the Jacobian $(f^n(x_0))'$. Suppose that the Lyapunov numbers are the same for almost every starting point x_0 within the basin of attraction [6], so we can say that Λ_1, Λ_2 are the Lyapunov numbers for the whole attractor. Then, for a sufficiently small δ, the box B will be transformed into a parallelogram $f^n(B)$ with length $\Lambda_1^n\delta$ and width $\Lambda_2^n\delta$. We then look at how many boxes of side $\Lambda_2^n\delta$ it takes to cover $f^n(B)$, which of course will be $\left(\frac{\Lambda_1}{\Lambda_2}\right)^n$. It is therefore reasonable to say that

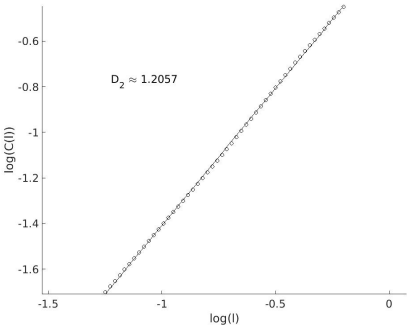
$$N(\Lambda_2^n\delta) \approx \left(\frac{\Lambda_1}{\Lambda_2}\right)^n N(\delta). \tag{25}$$

We now remember, from the discussion of the box dimension, that

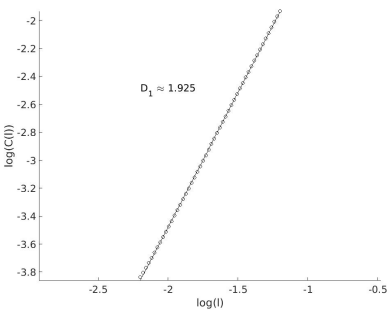
$$N(\delta) \propto \left(\frac{1}{\delta}\right)^{\dim_B}, \tag{26}$$



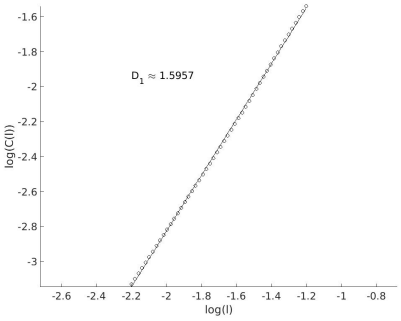
(a) Logistic Map



(b) Hénon Map



(c) Arnold's Cat Map



(d) Perturbed Arnold's Cat Map

Figure 19: Results of numerical approximations of the correlation dimension D_2, and the information dimension D_1, for three well known attractors. In (a) we see the result of 30000 iterations of the logistic map starting from $x_0 = 0.5$ and with $r = 3.56995$. This yielded $D_2 \approx 0.50117$, and in (b), the Hénon map with $a = 1.4, b = 0.3$ and 15000 iterations which gave $D_2 = 1.2057$. In (c) and (d) we see a significant difference when introducing the perturbation. For $\delta = 0$ the map is just the usual Arnold's cat map and the whole torus should be an attractor. The approximated dimension is therefore close to 2, $D_1 \approx 1.925$. When a large perturbation is introduced the dimension has gone down significantly to 1.5957.

and substituting this into (25), for some proportionality constant c, we get

$$
\begin{aligned}
c\Big(\frac{1}{\Lambda_2^n\delta}\Big)^{\dim_B} &= c\Big(\frac{\Lambda_1}{\Lambda_2}\Big)^n\Big(\frac{1}{\delta}\Big)^{\dim_B} \\
\Rightarrow \dim_B \ln\Big(\frac{1}{\Lambda_2^n\delta}\Big) &= n\ln\Big(\frac{\Lambda_1}{\Lambda_2}\Big) + \dim_B \ln\Big(\frac{1}{\delta}\Big) \\
\Rightarrow \dim_B \Big(\ln\big(\frac{1}{\Lambda_2^n\delta}\big) - \ln\big(\frac{1}{\delta}\big)\Big) &= n\ln\big(\frac{\Lambda_1}{\Lambda_2}\big) \\
\Rightarrow \dim_B = -\frac{\ln\big(\frac{\Lambda_1}{\Lambda_2}\big)}{\ln(\Lambda_2)} &= 1 - \frac{\ln(\Lambda_1)}{\ln(\Lambda_2)} = 1 + \frac{\ln(\Lambda_1)}{\ln\big(\frac{1}{\Lambda_2}\big)}.
\end{aligned}
$$

This is then the so called, **Lyapunov dimension**, or sometimes also called **Kaplan-Yorke dimension** of A, i.e.

$$
\dim_L(A) = 1 + \frac{\ln(\Lambda_1)}{\ln\big(\frac{1}{\Lambda_2}\big)},
$$

where Λ_1, Λ_2 are the Lyapunov numbers of the attractor.

One may generalize these arguments to higher dimensions, and the more general Lyapunov dimension is usually defined using the Lyapunov exponents λ_i instead of the Lyapunov numbers Λ_i (remember that $\lambda_i = \ln(\Lambda_i)$). The Lyapunov dimension, that was introduced by J. Kaplan and J. Yorke in 1978 [19], is defined for an attractor to a n-dimensional system, as

$$
\dim_L = k + \frac{\sum_{i=1}^{k}\lambda_i}{|\lambda_{k+1}|}, \tag{27}
$$

where the Lyapunov exponents are ordered $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and k is the index for which $\sum_{i=1}^{k}\lambda_i > 0$ and $\sum_{i=1}^{k+1}\lambda_i < 0$. The Kaplan-Yorke conjecture says that, for typical attractors, $dim_L = D_1$, where we may remember D_1 as the information dimension from before [29]. Using a Mathematica function called LCEsD [34] (Lyapunov Characteristic Exponent Discrete), developed by Marco Sandri, we obtain the Lyapunov spectrum for the Hénon map with $a = 1.4, b = 0.3$ as $\lambda_1 = 0.418, \lambda_2 = -1.622$. Substituting this into (27), we get $\dim_L \approx 1.258$. As we can see, this is considerably higher than the value obtained from numerical approximations of the information dimension in the previous section. Grassberger and Procaccia noted this in their paper ([10]) and suggested a more accurate way of approximating the dimension using different embedding dimensions.

There are much more to say on the subject, but we will end the discussion

here. For more on the Lyapunov dimension, see [6] and [19].

5 Concluding Remarks and Future Research

The fields of chaotic dynamics and fractal geometry are both in their infancy, and this is especially evident when we look at definitions. It takes time for these concepts to find a solid ground, and for the mathematical world to find definitions that people are satisfied with. Therefore you can find many different definitions of notions such as 'attractor', 'strange attractor', 'fractal', and 'chaotic'. This is also perhaps much because of the interdisciplinary nature of both fields (especially chaotic dynamics). Most articles one will find on chaotic dynamics will likely come from researchers within physics, or mathematical physics, but as far as we know, chaos is a purely mathematical phenomenon. However, it seems there are many phenomena in nature that exhibits a behaviour that could be closely described by chaotic dynamics, and thus interest about such models come from all sorts of fields that study natural phenomena.

It is true that attractors of chaotic systems seem to have fractal geometry, and it was because of this that the term 'strange' was somewhat reserved for these. Nevertheless, both fractal chaotic attractors, fractal non-chaotic attractors [12], and even non-fractal chaotic attractors [37] exists. One could argue that the term 'strange' should then refer to the geometrical aspects of the attractor rather than the dynamical, and thus argue that a strange attractor can exist without it exhibiting chaotic dynamics. We could then instead be talking about strange chaotic, non-strange chaotic, strange non-chaotic (and of course non-strange non-chaotic) attractors.

Perhaps since chaotic dynamics, as a field of study, was more or less born out from physics, it seems that it has, to a large extent, stayed there. There seems to be much fewer (or at least lesser known) attempts to apply these theories to problems within the field of mathematics itself. Of course, application of chaotic dynamics could be considered as soon as we have something behaving seemingly randomly, and that we do not understand in a better way. A particular area within mathematics that comes to mind for such applications, is the field of number theory. As an example, we seem to have a pretty poor grasp of how and why the prime numbers are distributed the way they are, and they seem to pop up more or less at random. However, plotting them in certain ways reveil patterns in an otherwise seemingly random distribution, such as the Ulam spiral [9]. Many of these patterns may be better explained by theories different from the one of dynamical systems, but at least the Ulam spiral are connected with many, still unanswered, questions in number theory. Another place where chaotic dynamics could perhaps be applied to number theory, is in the case of the Collatz conjecture. The Collatz conjecture states that, for the system governed

by $f : \mathbb{N}_{>0} \to \mathbb{N}_{>0}$, given by

$$f(x) = \begin{cases} x/2 & \text{if } x \equiv 0 \pmod 2 \\ 3x+1 & \text{if } x \equiv 1 \pmod 2. \end{cases}$$

$L(x_0) = \{1\}$ for all $x_0 \in \mathbb{N}_{>0}$, i.e. no matter what point x we start at, the orbit will always converge down to 1.

Experimental evidence suggests that the conjecture is true, but it has yet to be proven. This problem, also often referred to as the '$3n+1$-problem', has been referred to as especially difficult by many famous mathematicians. The Hungarian mathematician Paul Erdős even went as far as saying: *"Mathematics may not be ready for such problems."*[13].

In a paper from 1999 [20], this problem was extended to the complex plane by S. Letherman, D. Schleicher and R. Wood, to reveil, among other things, what has been named the Collatz fractal, seen in Figure 20. In a similar fashion, it would be interesting to see what the dynamics would look like if we extend the Collatz function f, e.g. to a field of p-adic numbers. Studying various properties and invariants of the systems we get from expanding f for different p, may reveil information useful in trying to prove the conjecture. This is, however, very much based on pure intuition and nothing else. In conclusion, since chaotic dynamics is such a recent development, and since it seems to have been somewhat stuck in the hands of physicists for a long time, there may be reasons to believe that many applications (and perhaps breakthroughs) within pure mathematics are soon to come.

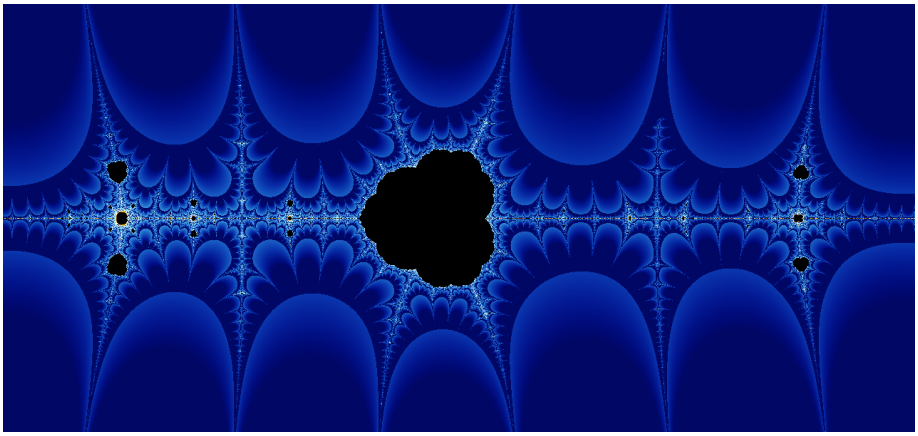


Figure 20: A so called escape-time coloured plot of an extension of the Collatz function to $\mathbb{C}$. The extension is given by the function $f : \mathbb{C} \to \mathbb{C}, f(z) = \frac{1}{4}(2 + 7z - (2 + 5z)\cos(\pi z))$, and regions in the complex plane are colored after how fast those values diverge away to infinity under the iteration of f. Black regions are starting values whose orbits remain bounded, while other colours indicate divergence. The image is centred at $z = 0$ and the long spiky horizontal line that goes through the whole image is $\mathbb{R}$. *Image source: [18]*

References

[1] J. Banks, J. Brooks, G. Cairns, G. Davies, P. Stacey, *On Devaney's Definition of Chaos*, The American Mathematical Monthly, Vol. 99, No. 4 (1992), pp. 332-334.

[2] J. Banks, V. Dragan, A. Jones, *Chaos: A Mathematical Introduction*, Cambridge University Press, (2003), Print.

[3] G. Cantor, *Über Unendliche, Lineare Punktmannichfaltigkeiten V*, Mathematische Annalen 21, (1883), pp. 545-591.

[4] C. Danforth, *Chaos in an Atmosphere Hanging on a Wall*, `http://mpe2013.org/2013/03/17/chaos-in-an-atmosphere-hanging-on-a-wall`, *Mathematics of Planet Earth 2013*, Retrieved 1 Mars 2017.

[5] K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, Chichester, England, Wiley, (2003), Print.

[6] J. D. Farmer, E. Ott, J. A. Yorke, *The Dimension of Chaotic Attractors*, Physica 7D, North Holland Publishing, (1983), pp. 153-180

[7] J. Feder, *Fractals*, New York, Plenum Press, (1988).

[8] F. Flam, *The Physics of Ray Bradbury's "A Sound of Thunder"*, Philadelphia Inquirer, (2015).

[9] M. Gardner, *Mathematical Games: The Remarkable Lore of the Prime Number*, Scientific American 210, (1964), pp. 120–128

[10] P. Grassberger, I. Procaccia, *Measuring the strangeness of strange attractors*, Physica D: Nonlinear Phenomena, (1983).

[11] P. Grassberger, *Grassberger-Procaccia algorithm*, Scholarpedia, 2(5):3043, (2007).

[12] C. Grebogi et al., *Strange attractors that are not chaotic*, Phys D 13(1-2), (1984), pp. 261-268.

[13] R. K. Guy, *E17: Permutation Sequences / Unsolved problems in number theory* (3rd ed.), Springer-Verlag, (2004), pp. 336–7.

[14] M. Hénon, *A two-dimensional mapping with a strange attractor*, Communications in Mathematical Physics 50, (1976), pp. 69-77.

[15] D. Hilbert, *Über die stetige Abbildung einer Linie auf ein Flächenstück*, Mathematische Annalen 38, (1891), pp. 459-460.

[16] R. C. Hilborn, *Chaos and Nonlinear Dynamics, An Introduction for Scientists and Engineers, Second Edition*, Oxford University Press, (2000), Print

[17] S. Hutzler, *Fractal Ireland*, Science Spin 58, (2013), pp. 19-20

[18] N. Johnston, *The Collatz Conjecture as a Fractal*, Webpage:`http://www.njohnston.ca/2009/06/the-collatz-conjecture-as-a-fractal/`, (2009).

[19] J. Kaplan, J. Yorke, *Chaotic behaviour of multidimensional difference equations*, Functional Differential Equations and the Approximation of Fixed Points, Lecture Notes in Mathematics vol. 730, Berlin, Springer, (1978), p.228.

[20] S. Letherman, D. Schleicher, R. Wood, *The 3n+1-Problem and Holomorphic Dynamics*, Experimental Mathematics 8 Issue 3, (1999), pp.241-251.

[21] T.Y. Li, J.A. Yorke, *Period Three Implies Chaos*, American Mathematical Monthly 82, 985 (1975).

[22] Y.Y. Lu, *Introduction to Dynamical Systems and Chaos: Lecture notes*, Dept. of Mathematics, Univ. of Hong Kong.

[23] B. Mandelbrot, *How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension*, Science 156.3775, (1967), pp. 636-638.

[24] F.R. Marotto, *Snap-back repellers imply chaos in* $\mathbb{R}^n$, Journal of Mathematical Analysis and Applications 63, (1978), pp.199–223

[25] M. Martelli, *Introduction to Discrete Dynamical Systems and Chaos*, Wiley, (1999), Print.

[26] M. Martelli, M. Dang, T. Seph, *Defining Chaos*, Mathematics Magazine, Vol. 71, No.2 (1998), pp. 112-122.

[27] J. R. Munkres, *Topology*, Saddle River, Prentice Hall, (2000).

[28] R. M. May, *Simple mathematical models with very complicated dynamics*, Nature 261(5560), pp. 459-467, (1976).

[29] E. Ott, *Attractor dimensions*, Scholarpedia, 3(3):2110, (2008).

[30] G. Peano, *Sur une courbe, qui remplit toute une aire plane*, Mathematische Annalen 36, (1890), pp. 157-160.

[31] T. Palmer, *Edward Norton Lorentz*, Physics Today, (2008).

[32] H.-O. Peitgen, H. Jürgens, D. Saupe, *Chaos and Fractals: New Frontiers of Science*, New York, Springer-Verlag, (1992), Print.

[33] D. Ruelle, F. Takens, *On the nature of turbulence*, Communications in Mathematical Physics 20, (1971), pp.167-192.

[34] M. Sandri, *Numerical Calculation of Lyapunov Exponents*, The Mathematica Journal Vol. 6 Issue 3, (1996)

[35] S. Silverman, *On Maps With Dense Orbits and the Definition of Chaos*, Rocky Mountain Journal of Mathematics Vol 22, Number 1, (1992).

[36] Y. G. Sinai, *Gibbs Measure in Ergodic Theory*, Russian Math. Surveys 4, (1972), pp. 21-64.

[37] J. Starrett, *Non-strange chaotic attractors equivalent to their templates*, Dynamical Systems Vol. 27, Issue.2, (2012).

[38] S. H. Strogatz, *Nonlinear dynamics and Chaos: with applications to physics, biology, chemistry, and engineering*, Reading MA, Addison-Wesley Pub, (1994).

[39] J. Theiler, *Estimating the Fractal Dimension of Chaotic Time Series*, The Lincoln Laboratory Journal Vol. 3, Num. 1, (1990).

[40] M. Vellekoop, R. Berglund, *On Intervals, Transitivity = Chaos*, American Mathematical Monthly 101, (1994), pp.353-355.

[41] P. F. Verhulst, *Recherches mathématiques sur la loi d'accroissement de la population*, Nouv. mém. de l'Academie Royale des Sci. et Belles-Lettres de Bruxelles 18, (1845), pp. 1-41.

[42] S. Wiggins, *Chaotic transport in dynamical systems*, New York, Springer-Verlag, (1992), Print.

[43] S. Wolfram, *A New Kind of Science*, IL, Wolfram Media Champaign, (2002)

[44] Cull, Paul; Flahive, Mary; Robson, Robby (2005). Difference Equations: From Rabbits to Chaos. ISBN 978-0-387-23233-1.

Appendices

A Classification of Bifurcation Points

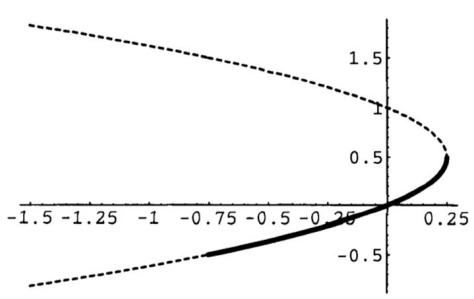
There are several characterizations of bifurcation points that may be of special interest, and therefore it may be of use to classify them accordingly. We will only mention a few types here, and we will not go into too much detail. The important message is just that there are many types of bifurcation points with different properties, and that this is an area of study in itself called bifurcation theory.

Above, we have already seen examples of two different types of bifurcation points. The first one, p_1, is called a **transcritical** bifurcation point. The characteristics of this type is that two different branches ($x_1(r)$ and $x_2(r)$ in our case) exists in an open neighbourhood of the bifurcation point they meet at. A **supercritical (subcritical) pitch fork** bifurcation is when three branches meet at a bifurcation point p, two of them existing only in $[p, b)$ ($(a, p]$), for some $b > p$ ($a < p$), and the other one existing in an open interval of p, but all of them of the same period. If we look at the point $a = 1$ in Figure 21c, it is easy to see where the name comes from. Another type of bifurcation point is the so called **supercritical (subcritical) fold** bifurcation. This is when two branches only exists in an interval $[p, b)$ ($(a, p]$), and meet at p, as seen in Figure 21b.

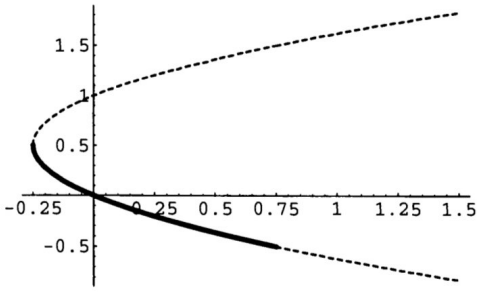
When the period of the points in one or two branches is double the period of the other one, the bifurcation is called a **period doubling** bifurcation. This is clearly what happens at p_2 above. This looks very similar to a pitch fork bifurcation. The difference is that if the branch to the left of the bifurcation point has period p, the two branches to the right of it has period $2p$. This type of bifurcation is of special interest in the study of chaotic dynamical systems, as we shall see further on.

B Matlab Code

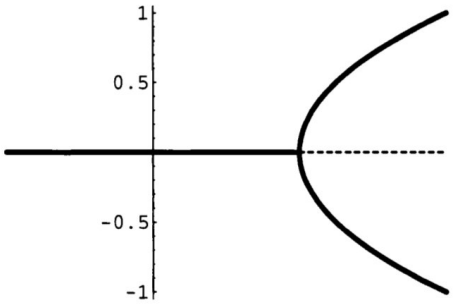
This section includes some of the scripts needed to generate the most important figures and results in this thesis. Some of the most simple scripts have been omitted. Since the focus has been on the mathematics, few considerations regarding complexity of the algorithms has been taken. Since, to a great extent, Matlab's own functions has been used, and these seems to be rather optimized, this is hopefully not of great importance. The scripts run in under approximately one minute, even for a rather large number of iterations, on a fairly standard modern PC (of 2017). Some of the scripts could also have been incorporated into one, more general function, that takes a function handle of a map of any dimension, with parameters and initial condition, and e.g. returns the correlation dimension. Again, programming and efficiency has not been the main focus



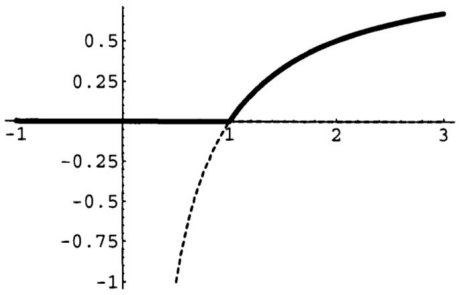
(a) Subcritical fold



(b) Supercritical fold



(c) Supercritical Pitchfork



(d) Transcritical

Figure 21: Examples of some different types of bifurcation points. In (a) we have a Subcritical fold at $a = \frac{1}{4}$ for $f(x) = x^2 + a$, In (b) we have a Supercritical fold at $a = -\frac{1}{4}$ for $f(x) = x^2 - a$, In (c) we have a Supercritical Pitchfork at $a = 1$ for $f(x) = ax - x^3$, and in (d) we have a Transcritical bifurcation at $r = 1$ for the logistic map $f(x) = rx(1 - x)$.

here, therefore the scripts are somewhat ad hoc.

B.1 Illustrating the 'Butterfly effect' using the logistic map

```
%#!/usr/bin/octave -qf

% This is a script to illustrate the butterfly effect of a chaotic map.
% The map used to illustrate this is the logistic map, f.
% An extremely small change in initial conditions, 1e-15, still makes
% a significant difference after quite few iterations.

clear all

timestart = cputime;

% Parameters
N = 100; % number of iterations (10000 good test)
r = 4;
x0 = 0.1
y0 = x0+10^(-16);
n = 1:N;

% The logistic map
f = @(p,r)r.*p.*(1-p);

% Allocating memory
p1 = zeros(N,1);
p2 = zeros(N,1);

% Generating two orbits
p1(1) = x0;
p2(1) = y0;
for i=2:N
    p1(i) = f(p1(i-1),r);
    p2(i) = f(p2(i-1),r);
end

plot(n,p1,'k'), hold on
plot(n,p2,'--k')
set(gca,'fontsize',18)
xlabel('n')
ylabel('x_n')
```

B.2 Bifurcation diagram of the logistic map

```
%#!/usr/bin/octave -qf
```

```
% Creates a bifurcation diagram of the logistic map. One may use the
% different parameters below to obtain different zoom levels, or play
% around on their own. For the plots to look good, one has to adjust both
% N, omit, and res (the resolution of r) when zooming.

clear all

% Parameters zoom1:
N = 900; % number of iterations (500 good test)
omit = 875; % number of transient points to omit (100 good test)
res = 0.0001; % step size of r values (0.001 good test)
r = 2.5:res:4;
YaxisPlotRange = [0 1];

% Parameters zoom2:
%N = 1800;%N = 1600;
%omit = 1770;%omit = 1580;
%res = 0.000001;%res = 0.0000008;
%r = 3.84:res:3.86;
%YaxisPlotRange = [0.4 0.6];

% Parameters zoom3:
%N = 3500;
%omit = 3470;
%res = 0.00000005;
%r = 3.8475:res:3.85;
%YaxisPlotRange = [0.488 0.52];

% Parameters zoom4:
%N = 7500;
%omit = 7100;
%res = 0.00000001;
%r = 3.8493:res:3.8495;
%YaxisPlotRange = [0.497 0.5032];

% Map specific parameters
x0 = 0.3; % initial condition

% Start CPU time
timestart = cputime;

% Allocating memory
p = zeros(length(r),N-omit);
tmp = zeros(1,N);

i = 1;
for R = r(1):res:r(end)
    for j = 2:N
        tmp(1) = x0;
        tmp(j) = R.*tmp(j-1).*(1-tmp(j-1));
    end
```

```
    p(i,:) = tmp(omit+1:end)';
    i = i+1;
end

% Stop CPU time
timeend = cputime-timestart

% Plotting
TeXylabel = ['$\{f^n(x_0)\}_{n>a}$'];
plot(r,p(:,:),'.','Color',[0.2 0.2 0.2],'MarkerSize',0.01)
set(gca,'fontsize', 21)
axis([r(1) r(end) YaxisPlotRange(1) YaxisPlotRange(2)])
xlabel('r')
ylabel(TeXylabel, 'Interpreter', 'Latex', 'FontSize', 24)
```

B.3 Lyapunov exponent for the logistic map

```
%#!/usr/bin/octave -qf

clear all

% Parameters:
N = 10000; % number of iterations, length of orbit
omit = 500; % number of points to omit from the start of the orbit. (100 good test
rrange = [3.5 4]; % range that r varies over
rres = 0.001; % increments of r

% Allocating memory
lyapunov_exponents=zeros(1,length(rrange(1):rres:rrange(2)));

% The logistic map
f = @(p,r)r.*p.*(1-p);

timestart = cputime;
j = 1;
for r = rrange(1):rres:rrange(2)
  s = 0;
  p = rand; % Pick a random starting point in (0,1)
  for i = 1:N % Generate orbit of length N
    p = f(p,r);
    if i > omit % Omit the first omit nr. of points
      s = s + log(abs(r.*(1-p)-r.*p)); % Accumulate the logs of the derivatives
    end
  end
  lyapunov_exponents(j)=s./N; % Assign the average lyapunov exponent to each eleme
  j = j + 1;
end
timeend = cputime-timestart

% Plots
```

```
r = rrange(1):rres:rrange(2);
plot(r, lyapunov_exponents,'Color',[0.2 0.2 0.2]);
set(gca,'fontsize', 21);
referenceline = refline([0 0]);
referenceline.Color = 'k';
axis([rrange(1) rrange(2) -1.5 1])
xlabel('r');
ylabel('Lyapunov exponent');
```

B.4 Correlation dimension for the logistic map

```
%#!/usr/bin/octave -qf

% Calculating the correlation dimension of the logistic map. This script is
% tailored to use a one dimensional map, just change f to use another map.
% The same procedure can easily be extended to higher dimensions by making
% the appropriate changes to this script. An example of such an extension
% is correlation_dimension_henon.m.

clear all

timestart = cputime;

% Parameters:
N = 30000; % number of iterations (10000 good test)
omit = 1000; % number of points to omit from the start of the orbit. (500 good tes
res = 50; % resolution, the number of l values. (50 good test)

% Map specific parameters
x0 = 0.5; % initial condition
R = 3.56995; % parameter

% The logistic map
f = @(p,r)r.*p.*(1-p);

% Allocating memory
p = zeros(N,1);
distMatrix = zeros(N-omit,N-omit);
Cl = zeros(res,1);

% Generating the orbit p
p(1) = x0;
for i=2:N
    p(i) = f(p(i-1),R);
end

% Generating distance matrix distMatrix from p, omitting <omit>
% number of points from the start of the orbit
distMatrix = pdist(p(omit+1:end));
distMatrix = squareform(distMatrix);
```

```
% Counting pair of points with distance smaller than l
l = (logspace(-4,-1,res))';
k = 1;
for i = 1:res
    Cl(k) = (1/2).*nnz(distMatrix(:) < l(i));
    k = k + 1;
end

% Generating approx. correlation integral
Cl = (2/(N^2-N)).*Cl;

% Preparing for log plot and least squares curve fitting
Cl = log10(Cl);
l = [ones(res,1) log10(l)];

% Least squares curve fitting
b = l\Cl
lsY = b(2)*l(:,2)+b(1);

% Cpu time
timeend = cputime-timestart

% Plotting

% Label showing slope d
txt1 = ['D_2 \approx ',num2str(b(2))];
% Positioning of label
xlbl = ((max(l(:,2))+min(l(:,2))))/2-0.4;
ylbl = ((max(lsY)+min(lsY)))/2+0.4;
% Plots
scatter(l(:,2),Cl,'k')
hold on
set(gca,'fontsize',18)
plot(l(:,2),lsY,'k')
axis equal
xlabel('log(l)')
ylabel('log(C(l))')
text(xlbl,ylbl,txt1,'HorizontalAlignment','left','fontsize',18);
%print -dpng corr_logistic.png
```

B.5 Correlation dimension for the Hénon map

```
%#!/usr/bin/octave -qf

% Calculating the correlation dimension of the Henon map. This script is
% tailored to use a two dimensional map, just change h to use another map.
% The same procedure can easily be extended to higher dimensions by making
% the appropriate changes to this script
```

```
clear all

timestart = cputime;

% Parameters:
N = 15000; % number of iterations (10000 good test)
omit = 1000; % number of points to omit from the start of the orbit. (500 good tes
res = 50; % resolution, the number of l values. (50 good test)

% Map specific parameters
x0 = [0.1 0.1]; % initial condition
a = 1.4; b = 0.3;

% The Henon map
h = @(x,y,a,b)[y+1-a.*x.^2; b.* x];

% Allocating memory
p = zeros(N,2);
distMatrix = zeros(N-omit,N-omit);
Cl = zeros(res,1);

% Generating the orbit p
p(1,:) = x0;
for i=2:N
    p(i,:) = h(p(i-1,1),p(i-1,2),a,b);
end

% Generating distance matrix distMatrix from p, omitting <omit>
% number of points from the start of the orbit
distMatrix = pdist(p(omit+1:end,:));
distMatrix = squareform(distMatrix);

% Counting nr of pairs with distance smaller than l
%l = (linspace(2.*10^(-2),0.5,res))';
% Using logspace instead for equal spacing between log(l) values
l = (logspace(-1.25,-0.2,res))';%(logspace(-1.25,-0.2,res))';%-1.7,-0.3
k = 1;
for i = 1:res
    Cl(k) = (1/2).*nnz(distMatrix(:) < l(i));
    k = k + 1;
end

% Creating approx. correlation integral
Cl = (2/(N^2-N)).*Cl;

% Preparing for log plot and least squares curve fitting
Cl = log10(Cl);
l = [ones(res,1) log10(l)];

% Least squares curve fitting
b = l\Cl
lsY = b(2)*l(:,2)+b(1);
```

```
% Cpu time
timeend = cputime-timestart

% Plotting

% Plot the attractor
%plot(p(:,1),p(:,2),'k.','MarkerSize', 4)
% Label showing slope d
txt1 = ['D_2 \approx ',num2str(b(2))];
% Positioning of label
xlbl = (max(l(:,2))+min(l(:,2)))/2-0.5;
ylbl = (max(lsY)+min(lsY))/2+0.3;
% Plots
figure
scatter(l(:,2),Cl,'k')
hold on
set(gca,'fontsize',18)
plot(l(:,2),lsY,'k')
axis equal
xlabel('log(l)')
ylabel('log(C(l))')
text(xlbl,ylbl,txt1,'HorizontalAlignment','left','fontsize',18);
%print -dpng corr_henon.png
```

B.6 Information dimension of the perturbed Arnold's cat map

```
%#!/usr/bin/octave -qf

% Calculating the information dimension of the perturbed Arnold's cat map.
% This script is tailored to use a two dimensional map, just change f or h
% to use another map. The same procedure can easily be extended to higher
% dimensions by making the appropriate changes to this script.

clear all

timestart = cputime;

% Parameters:
N = 30000; % number of iterations (1000 good test)
omit = 1000; % number of points to omit from the start of the orbit. (100 good tes
res = 50; % resolution, the number of l values. (50 good test)

% Map specific parameters
x0 = [0.1 0.1]; % initial condition
delta = 0.235;
a = 1.4; b = 0.3;

% The Henon map
h = @(x,y,a,b)[y+1-a.*x.^2; b.* x];
```

```
% Arnold's cat map, perturbed
f = @(x,y,delta)[mod(2.*x+y,1); mod(x+y+delta.*cos(2.*pi.*x),1)];

% Allocating memory
p = zeros(N,2);
Cl = zeros(res,1);
n = zeros(N-omit,res);

% Generating radii-vector l; using logspace for equal spacing
% between log(l) values.
l = (logspace(-2.2,-1.2,res))';%(logspace(-2,-0.4,res))'

% Generating the orbit p
p(1,:) = x0;
for i=2:N
    p(i,:) = f(p(i-1,1),p(i-1,2),delta);%h(p(i-1,1),p(i-1,2),a,b);%
end

% Generating matrix n, where n(k,i) are the number of points
% within a l(i) neighborhood of the point p(k,:). Iterating
% through all points of the orbit p and all radii l.
for i = 1:res
    Idx = rangesearch(p(omit+1:end,:),p(omit+1:end,:),l(i));
    for k = 1:N-omit
        n(k,i) = length(Idx{k});
    end
end

% Generating approximate 'Information integral' Cl based on
% formula for D_1, derived from applying l'Hospital to general formula
% for Renyi-dimension.
for r = 1:res
    Cl(r) = exp((1./(N-omit)).*sum(log(n(:,r)./((N-omit)-1))));
end

% Preparing for log plot and least squares curve fitting
Cl = log10(Cl);
l = [ones(res,1) log10(l)];

% Least squares curve fitting
b = l\Cl
lsY = b(2)*l(:,2)+b(1);

% Cpu time
timeend = cputime-timestart

% Plotting
% Plot the attractor
%plot(p(:,1),p(:,2),'k.','MarkerSize', 4)
%print -dpng arnoldcat_perturbed_phase.png
% Label showing slope d
```

```
txt1 = ['D_1 \approx ',num2str(b(2))];
% Positioning of label
xlbl = (max(l(:,2))+min(l(:,2)))/2-0.5;
ylbl = (max(lsY)+min(lsY))/2+0.4;
% Plots
figure
scatter(l(:,2),Cl,'k')
hold on
set(gca,'fontsize',18)
plot(l(:,2),lsY,'k')
axis equal
xlabel('log(l)')
ylabel('log(C(l))')
text(xlbl,ylbl,txt1,'HorizontalAlignment','left','fontsize',18);
%print -dpng arnoldcat_perturbed.png
```