Juan Pablo Vigneaux

Lectures on Mathematical Chaos

A first course in dynamical systems

August 5, 2024

California Institute of Technology

Preface

These are lecture notes for the course *Ma 4/104 Introduction to Mathematical Chaos* at the California Institute of Technology.

The material in Part I comes mostly from [3], whereas Part II has been heavily influenced by [1].

For the formatting, I have used the SVMono class by Springer Verlag (which is free of copyright as far as I know). most of the graphics have been generated by hand or with Mathematica.

Pasadena, August 5, 2024, *Juan Pablo Vigneaux*

Contents

List of abbreviations and symbols

i.e. *id est*, Latin for “that is”.
:= The term on the left is defined to be the term on the right. It can also be used in the opposite direction, =:.
$\mathbb{N}$ The natural numbers $\{0, 1, 2, \ldots\}$.[1]
$\mathbb{N}^*$ The nonnegative integers $\{1, 2, 3, \ldots\}$.
$\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ Respectively: the integers, the rational numbers, the real numbers, the complex numbers.
K The Cantor middle-thirds set.
Σ The space of binary sequences $\{0, 1\}^{\mathbb{N}}$.

[1] For definitions, see the Wiki article about Peano axioms.

Chapter 1
Introduction

A ***dynamical system*** is a mathematical model that prescribes how certain time-dependent quantities x, y, z, ... evolve in time.

When time is regarded as a continuous independent variable, whose values are real numbers, the aforementioned models are typically differential equations.

Example 1.1 Consider the ordinary differential equation $\dot{x} = ax$ (remember that $\dot{x} := \frac{\mathrm{d}x}{\mathrm{d}t}$). The unique solution is of the form $x(t) = Ae^{at}$ (how do we know it is unique?); the leading constant A is determined by an *initial value* of x, for instance, $x(0) = A$.

Example 1.2 Newton's law

$$\text{force} = \text{mass} \times \text{acceleration} \tag{1.1}$$

is basically a recipe to produce differential equations.

For instance, the force exerted by a spring that has been extended or contracted by a distance x is $-kx$ (opposed to the extension/contraction), hence the movement of a body of mass m attached to the spring is governed by

$$m\ddot{x} = -kx. \tag{1.2}$$

This model is known as the *harmonic oscillator*.

Recalling that if $f(t) = \sin(t)$ then $\ddot{f}(t) = -\sin(t)$, one may start with the guess

$$x(t) = A\sin(\omega t + \phi), \tag{1.3}$$

where A, ω and ϕ are constants to be determined, known respectively as the *amplitude*, *frequency* and *phase*. If we substitute $x(t)$ in (1.2), we immediately see that $\omega = \sqrt{k/m}$. Moreover, initial conditions $x(0) = x_0 \in \mathbb{R}$ and $\dot{x}(0) = v_0 \in \mathbb{R}$ determine the value of A and ϕ through the system of equations

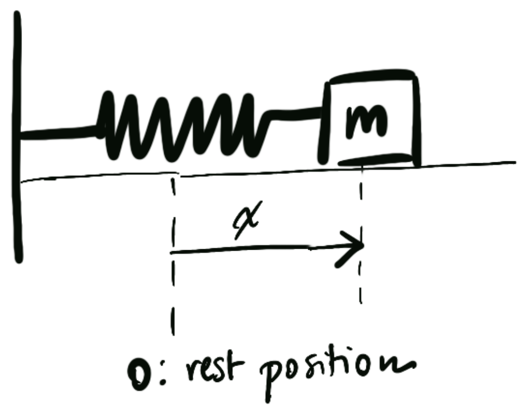

Fig. 1.1 An idealized spring. The horizontal distance between the rest position and the center of mass of the body attached at its end is x.

$$x_0 = A\sin\phi \tag{1.4}$$

$$v_0 = A\sqrt{\frac{k}{m}}\cos\phi. \tag{1.5}$$

Example 1.3 Similarly, Newton used (1.1) and a formula for the gravitational force between two bodies to calculate the orbit of a planet around a sun, which turns out to be an ellipse. This was consistent with Kepler's observations.

The solution (1.3) of the harmonic oscillator, or the elliptical orbits of a planet calculated by Newton, are such that if the initial state of the system is known (in this case, the position and velocity of the bodies at a certain time t_0), then we can determine precisely the position and velocities of all the bodies at any time $t > t_0$. This determination is robust: if the initial condition is perturbed slightly, the parameters are also perturbed slightly (they depend continuously on the initial condition) and the solution changes in a controlled way. Hence these systems are *predictable*. Moreover, they exhibit a high degree of *regularity*, describing a well-defined trajectory that repeats over time: the oscillator moves back and forth; the planet repeatedly describes the same ellipse.

Already Newton wondered if this was still the case for a system of three bodies mutually attracted by the gravitational force. Towards the end of the XIX century, it was proved, with differential Galois theory, that it is impossible to write a closed formula describing the solutions akin to (1.3). In old terminology: most differential equations have no "integral". However, one can still hope to determine the qualitative behavior of the orbits (e.g. that they are ellipses, etc). Henri Poincaré established that the solutions were highly unpredictable and irregular also in the qualitative sense.

Towards the end of the course, we shall be able to understand the main ideas introduced by Poincaré. Nowadays, with computers, we can simulate the orbit of a tiny mass interacting gravitationally with two larger bodies. A result of this soimulation is shown in Figure 1. In this solution, the two larger bodies move circularly around one another and the figure shows the trajectory of the small mass in a rotating coordinate system where the two planets are located at the extremes of the central segment.

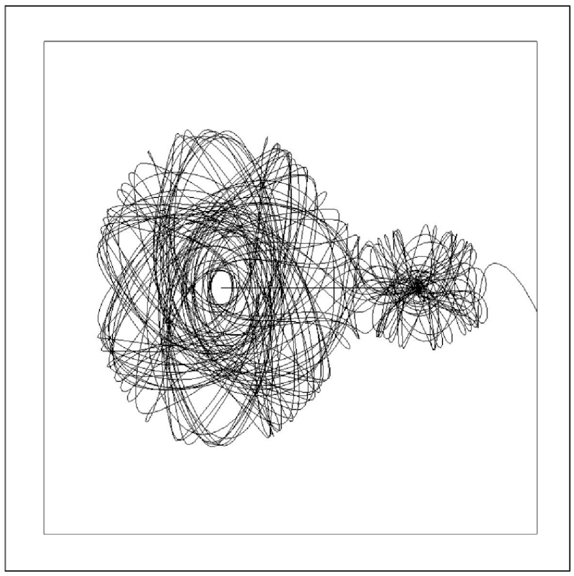

Fig. 1.2 Source: Alligood, Sauer, and Yorke: Chaos: An Introduction to Dynamical Systems, Springer, 1966.

1.1 The Lorenz attractor

The mathematician turned metheorologist E. Lorenz (MIT) wanted to study the statistical methods used back to predict the weather. He knew that the phenomenon was nonlinear, hence in principle not very apt to simple linear methods like regression that were popular in his time.

In order to prove his point, he produced a simplified "toy model" given by the nonlinear system of equations

$$\dot{x} = -\sigma x + \sigma y \tag{1.6}$$

$$\dot{y} = rx - y - xz \tag{1.7}$$

$$\dot{z} = xy - bz \tag{1.8}$$

with parameters $\sigma = 10$, $r = 28$ and $b = 8/3$, which he solved numerically.

At some point, Lorenz wanted to look closer at a particular solution, so he restarted the numerical algorithm introducing by hand, as initial condition, the printed value of some of the iterations (which had been truncated by the computer). To his surprise, the iterates that he obtained from that initial condition diverged very significantly from what he had obtained in his initial run of the algorithm. He concluded that the system was highly sensitive to initial conditions.

He also noticed that the trajectories are attracted to a peculiar set, which looks like a butterfly. This set does not look like a curve or a surface; it is an example of a *fractal*. The definition of fractal is not evident and it will be one of the things discussed in this course. Lorenz's butterfly is an example of *strange attractor*.

1.2 Discrete-time systems

For the sake of simplicity, for the most part of this course, we shall focus on models where time is a discrete parameter $t = 0, 1, 2, 3, \ldots$. The evolution of the values $x_n = x(n)$ is then given by a recursive relation $x_{n+1} = F(x_n)$. It will become clear that these systems, despite their simple appearance, might exhibit extremely complex chaotic behavior.

Example 1.4 The simplest model of growth for a population of bacteria is

$$x_{n+1} = 2x_n, \tag{1.9}$$

that is, the bateria duplicate between time n and $n + 1$. Similarly, in finance, the application of an interest rate of 1% every month translates into

$$x_{n+1} = 1.01x_n, \tag{1.10}$$

where x_n is the money after n months have passed (with an initial deposit of x_0).

In both cases, we have

$$x_{n+1} = rx_n \quad \text{for some } r \in (0, \infty), \tag{1.11}$$

and with initial condition $x_0 \geq 0$. Then

$$x_n = rx_{n-1} = r^2 x_{n-2} = \cdots = r^n x_0, \tag{1.12}$$

so it is clear that $x_n \to 0$ as $n \to \infty$ if $r < 1$, and $x_n \to \infty$ as $n \to \infty$ if $r > 1$. Moreover, if $r = 1$ the sequence $(x_n)_{n\geq 0}$ is constant. So the analysis of this model is fairly simple.

Example 1.5 Other choices of the function F can lead to an extremely rich and complex behavior. Consider for instance another model of population growth,

$$\tilde{x}_{n+1} = \tilde{r}\tilde{x}_n(M - \tilde{x}_n), \tag{1.13}$$

for some positive parameters $\tilde{r}$ and M. In this model, there is a maximal *carrying capacity* M and growth slows down as $x \to M$.

Making the change of coordinates $x_n = \tilde{x}_n/M$ and $r = \tilde{r}M$, we can put it in the form

$$x_{n+1} = rx_n(1 - x_n), \tag{1.14}$$

with initial condition $x_0 \in (0, 1)$.

Remark that $F(x) = rx(1 - x)$ has a maximum at $x_* = 1/2$, with $f(x_*) = r/4$. So if $r \in [0, 4]$, the map f has images in $[0, 1]$. So we can consider interations of $f : [0, 1] \to [0, 1]$: this means, $F(F(x)) =: F^2(x)$, $F(F(F(x))) =: F^3(x)$, etc.

In order to analyze what happens under iterations, one can perform a graphical analysis. A useful tool is the *Cobweb diagram*, explained in Figure 1.3.

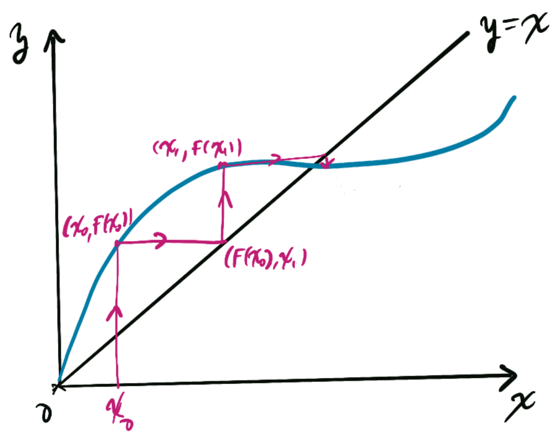

Fig. 1.3 The Cobweb diagram is a representation of an initial point x_0 and its iterates $x_i = F^i(x_0)$ under a function F.

Part I
One-dimensional systems

Chapter 2
Orbits

2.1 Iteration

For the moment, we focus on discrete-time dynamical systems given by successive iterations of a single function $x \mapsto f(x)$, which possibly depends on certain ***parameters***.

Example 2.1 Some functions that will play an important role in this course are:

1. $Q_c(x) = x^2 + c$ for $c \in \mathbb{R}$;
2. $F_\lambda(x) = \lambda x(1-x)$ for $\lambda \in [0,4]$, and
3. $S_\mu(x) = \mu \sin(x)$ for $\mu \in \mathbb{R}$.

The parameters are the numbers c, λ, and μ.

To ***iterate*** means to evaluate the function over and over, using the output of the previous application as input of the following one. We write $F(F(x)) =: F^2(x)$ and in general

$$F^n(x) = \underbrace{F \circ F \circ \cdots \circ F}_{n \text{ times}}(x) \tag{2.1}$$

$$= \underbrace{(F(F(\cdots(F}_{n \text{ times}}(x)\cdots). \tag{2.2}$$

! Warning

In the context of this course, $F^n(x)$ will never mean "raise $F(x)$ to the n^{th} power.

Example 2.2 For $F(x) = x^2 + 1$, one has $F^2(x) = (x^2+1)^2 + 1$ and

$$F^3(x) = ((x^2+1)^2+1)^2 + 1 = x^8 + 4x^6 + 8x^4 + 8x^2 + 5. \tag{2.3}$$

2.2 Orbits

Definition

Let X be a set (for instance, $\mathbb{R}$) and $F : X \to X$ a function whose range is contained in the domain of definition of the function, that is, such that $\text{Ran}(f) \subset \text{Dom}(f)$. Then, given $x_0 \in \text{Dom}(f)$, the ***orbit of*** x_0 ***under*** F is defined to be the sequence $(x_0, x_1, x_2, \ldots)$ given by $x_i = F(x_{i-1})$ for all $i \geq 1$. Sometimes we call x_0 the ***seed*** of the orbit.

Exercise 2.1 Compute the first elements of the orbit of 256 under $F(x) = \sqrt{x}$.

In general, a computer is needed to find numerically the elements of the orbit. But as we will see, the computer might lie.

Some types of orbits

1. **Fixed point**: a point x_0 that satisfies $F(x_0) = x_0$ is called a fixed point.
   For such x_0, the orbit is a constant sequence $(x_0, x_0, x_0, \ldots)$. They can be found by solving the equation
   $$F(x) = x. \tag{2.4}$$
   Exercise 2.2 Find the fixed points of x^2.

   Geometrically, fixed points correspond to the intersections of the graphs of F and $y = x$.
2. **Periodic orbit** or **cycle**: We call x_0 a periodic point of period k if $F^k(x_0) = x_0$ and if k is the *smallest* such positive integer.
   The orbit of such an x_0 consists of exactly k different values that repeat,
   $$(x_0, x_1, \ldots, \underbrace{x_{k-1}}_{F^{k-1}(x_0)}, \underbrace{x_0}_{F^k(x_0)}, x_1, \ldots) \tag{2.5}$$
   We talk about ***periodic*** k***-point***, ***periodic*** k***-orbit*** or k***-cycle***. All the points in the orbit have period k as well.

   Remark 2.1 Every period k-point is a fixed point of F^k, but the converse is not true.
   In particular, a fixed point of F is also a fixed point of F^k for every k.
   More generally, a fixed point of F^m is a fixed point of F^k if m divides k (denoted $m \mid k$).
3. **Eventually fixed** or **eventually periodic orbits**: If x_0 is itself not fixed or periodic, but some point in its orbit is.

   Example 2.3 If $F(x) = x^2 + 1$ and $x_0 = \sqrt{2}$, then $F(x_0) = 1$, $F^2(x_0) = F(1) = 0$, $F^3(x_0) = F(0) = 1$, etc.

4. **Other orbits**: We saw in the introduction, for instance, that $F_4(x) = 4x(1-x)$ has "dense" orbits that visit every subinterval of $[0,1]$, no matter how small.

Example: The doubling function

Let $[0,1)$ denote the set $\{ x \in \mathbb{R} \mid 0 \leq x < 1 \}$. Consider the function $D : [0,1) \to [0,1)$ given by

$$D(x) = \begin{cases} 2x & \text{if } 0 \leq x < 1/2 \\ 2x - 1 & \text{if } 1/2 \leq x < 1 \end{cases}. \tag{2.6}$$

This means that we keep only the numbers after the decimal point. Another way of writing this is $D(x) = 2x \mod 1$.

Remark that

1. 0 is a fixed point.
2. $1/3$ and $2/3$ lie in a 2-cycle.
3. $1/5$ is the seed of a 4-cycle:

$$\left(\frac{1}{5}, \frac{2}{5}, \frac{4}{5}, \frac{3}{5}, \frac{1}{5}, \dots \right). \tag{2.7}$$

4. $1/9$ lies in a 6-cycle:

$$\left(\frac{1}{9}, \frac{2}{9}, \frac{4}{9}, \frac{8}{9}, \frac{7}{9}, \frac{5}{9}, \frac{1}{9}, \dots \right). \tag{2.8}$$

But we shall see that "most" orbits are not periodic, eventually periodic or even convergent to a cycle.

2.3 Graphical analysis

A tool to study orbit analysis is the cobweb diagram introduced in Figure 1.3.

Example 2.4 $F(x) = \cos(x)$ and $x_0 = -\frac{\pi}{2} + 0.1$. Figure 2.1 shows convergence to a fixed point.

We see convergence to a fixed point.

Example 2.5 $Q_{-1}(x) = x^2 - 1$ and $x_0 = -0.1$. Figure 2.2 shows convergence to the 2-cycle given by 0 and -1.

Sometimes, one can do a complete orbit analysis just from a Cobweb diagram.

Example 2.6 $F(x) = x^3$.

We see that 0, 1 and -1 are fixed points. Also $\lim_n F^n(x_0) = 0$ if $|x_0| < 1$ and $\lim_n F^n(x_0) = \pm\infty$ if $|x_0| > 1$.

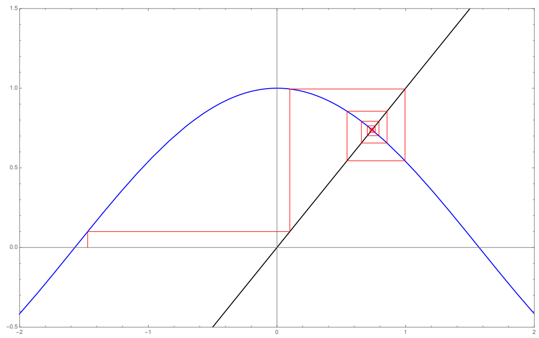

Fig. 2.1 Cobweb diagram of the first 200 iterates of $F(x) = \cos(x)$ with seed $x_0 = -\frac{\pi}{2} + 0.1$.

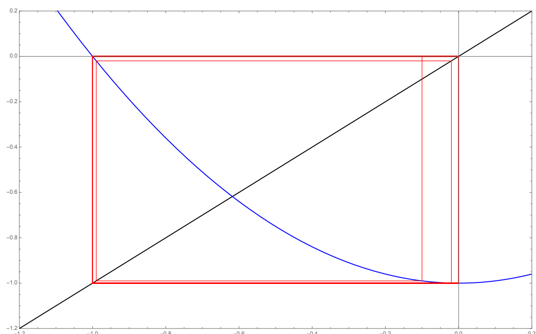

Fig. 2.2 Cobweb diagram of the first 200 iterates of $Q_{-1}(x) = x^2 - 1$ with seed $x_0 = -0.1$ (after some iterations the it is impossible to distinguish these iterates from 0 and -1 in the picture).

2.4 Two results from single-variable calculus

Theorem 2.1 (Intermediate value theorem, IVT) *Suppose $F : [a, b] \to \mathbb{R}$ is continuous and hat y_0 lies between $F(a)$ and $F(b)$. Then there exists an $x_0 \in [a, b]$ such that $F(x_0) = y_0$.*

Theorem 2.2 (Mean value theorem, MVT) *Suppose F is a differentiable function on the interval $[a, b]$. Then there exists $c \in (a, b)$ such that*

$$F'(c) = \frac{F(b) - F(a)}{b - a}. \tag{2.9}$$

2.5 Fixed points

We can find fixed points solving the equation $F(x) = x$, but sometimes this is difficult. In some situations, it is enough to establish the *existence* of a fixed point by means of the following corollary of the intermediate value theorem (IVT).

Corollary 2.1 (Fixed point theorem) *Suppose $F : [a, b] \to [a, b]$ is continuous. Then there exists a fixed point of F in $[a, b]$.*

Proof Apply the IVT to $H(x) = F(x) - x$. Remark that $H(a) = F(a) - a \geq 0$ and $H(b) = F(b) - b \leq 0$, so 0 lies between $H(a)$ and $H(b)$.

We saw in the numerical experiments that some fixed points attract orbits of other points, whereas some repel nearer points.

Definition 2.1 A ***neighborhood*** of a point $x_0 \in \mathbb{R}$ is an open interval I that contains x_0.

Definition 2.2 Let x_0 be a fixed point of a map F.

We say that x_0 is (locally) ***attracting*** if there exists a neighborhood I of x_0 such that if $x \in I$, then $F^n(x) \in I$ for all $n \in \mathbb{N}$ and $\lim_n F^n(x) = x_0$.

We say that x_0 is (locally) ***repelling*** if there exists a neighborhood I of x_0 such that if $x \in I$ and $x \neq x_0$, then there is an integer $n > 0$ such that $F^n(x) \notin I$.

Example 2.7 In example 2.6, 0 is attracting and 1 is repelling.

For simplicity, let us consider first the linear function $L_\alpha(x) = \alpha x$, for $\alpha \in \mathbb{R}$. 0 is a fixed point of this map.

- When $|\alpha| < 1$, 0 is an attracting fixed point.
- When $|\alpha| > 1$, 0 is a repelling fixed point.

We shall see that this holds more generally, at least near a fixed point, since $F'(x_0)$ approximates the tangent of F at x_0,

Definition 2.3 Let x_0 be a fixed point of the map F. We say that x_0 is a ***stable*** fixed point if $|F'(x_0)| < 1$, and that it is an ***unstable*** fixed point if $|F'(x_0)| > 1$. If $|F'(x_0)| = 1$, the fixed point is called ***neutral***.

Theorem 2.3 *Let F be a continuously differentiable function. If x_0 is a stable fixed point for F, then x_0 is locally attracting.*

Proof Choose λ such that $|F'(x_0)| < \lambda < 1$.

By continuity of F', for any $\epsilon > 0$, we can choose $\delta = \delta(\epsilon) > 0$ such that if $|x - x_0| < \delta$, then $|F'(x) - F'(x_0)| < \epsilon$.

Remark that $|F'(x) - F'(x_0)| < \epsilon$ implies in particular $|F'(x)| < |F'(x_0)| + \epsilon$. Let us choose ϵ_0 such that $|F'(x_0)| + \epsilon_0 < \lambda$, and set $I = (x_0 - \delta(\epsilon_0), x_0 + \delta(\epsilon_0))$. By Theorem 2.2 (MVT), for every $x \in I \setminus \{x_0\}$,

$$\frac{|F(x) - F(x_0)|}{|x - x_0|} = |F'(c)| < \lambda. \tag{2.10}$$

where c denotes a point between x and x_0.

Equation (2.10) says, equivalently, that F is a contractive map: for all $x \in I \setminus \{x_0\}$

$$|F(x) - F(x_0)| < \lambda |x - x_0| \tag{2.11}$$

From (2.11), we also conclude that $F(x)$ also belongs to I. So we can apply (2.11) twice:

$$\begin{aligned} |F^2(x) - x_0| &= |F^2(x) - F^2(x_0)| \\ &< \lambda |F(x) - F(x_0)| \\ &< \lambda^2 |x - x_0|, \end{aligned}$$

Similarly, for all $n \geq 1$,

$$|F^n(x) - x_0| < \lambda^n |x - x_0|, \tag{2.12}$$

hence $|F`n(x) - x_0| \to 0$ as $n \to \infty$. □

Remark that the last part of the proof shows *exponential convergence* of $F^n(x)$ to x_0.

Analogously:

Theorem 2.4 *Let F be a continuously differentiable function. If x_0 is an unstable fixed point for F, then x_0 is locally repelling.*

The proof of this one is left as an exception.

? Question

What are the chances of finding a repelling fixed point numerically?

In the case of a neutral fixed point, we cannot determine its character based purely on the derivative.

Example 2.8 For $F(x) = x$, all points are fixed.

Example 2.9 For $F(x) = -x$, only 0 is fixed. All other points have period 2.

Example 2.10 $F(x) = x - x^2$.

Exercise 2.3 When $F(x) = x - x^3$, 0 attracts the neighborhood $|x| < 1$, but when $F(x) = x + x^3$, 0 repels this neighborhood.

2.6 Periodic points

Periodic points can also be consistently classified as attracting, repelling or neutral, based on the following fact.

Theorem 2.5 *If* $(x_0, ..., x_{n-1})$ *is an n-cycle for a differentiable function F, then*

$$(F^n)'(x_0) = \cdots = (F^n)'(x_{n-1}). \tag{2.13}$$

Proof If x_0 is a period-n point, then the chain rule of the derivatives implies that

$$(F^n)'(x_0) = F'(\underbrace{x_{n-1}}_{F^{n-1}(x_0)})F'(x_{n-2})\cdots F'(x_0). \tag{2.14}$$

If x_0 is replaced by another x_i on the left-hand side, the same principle can be applied, and one will find the same factors on the right-hand side, just in a different order. □

Hence we say that a period-n point is attracting or repelling if itself or any of its iterates is, respectively, an attracting or repelling fixed point of F^n.

Example 2.11 $F(x) = x^2 - 1$. The derivative of $F^2(x) = x^4 - 2x^2$ at $x = 0$ vanishes, hence 0 is an attracting period-2 point. This explains the figure we saw above.

Chapter 3
Bifurcations

3.1 Bifurcations of the quadratic map

Let us analyze the maps $Q_c(x) = x^2 + c$ for different values of c. Our goal is to understand the qualitative changes in the dynamics of Q_c as c varies. The family Q_c is simple enough that many of the computations can be performed explicitly with relative ease.

First of all, the fixed points of Q_c are found by solving $x^2 + c = x$. This is a quadratic equation, with roots

$$p_+ = \frac{1}{2}(1 + \sqrt{1 - 4c}), \quad p_- = \frac{1}{2}(1 - \sqrt{1 - 4c}). \tag{3.1}$$

There are real solutions only if $1 - 4c \geq 0$, that is, $c \leq 1/4$.

One can see by the graphical method that if $c > 1/4$ there are no fixed points and all the orbits diverge.

If $c = 1/4$ there is exactly one fixed point, and if $c < 1/4$ there are two different fixed points, p_+ and p_-.

Moreover, since $Q_c'(x) = 2x$,

$$Q_c'(p_+) = 1 + \sqrt{1 - 4c}, \quad Q_c'(p_-) = 1 - \sqrt{1 - 4c}, \tag{3.2}$$

hence $p_+ = p_-$ is neutral when $c = 1/4$, and when $c < 1/4$,

- p_+ is unstable/repelling
- p_- is stable/attracting, as long as $|Q_c'(p_-)| < 1$, which is equivalent to $c \in (-\frac{3}{4}, \frac{1}{4})$.

What we have just described represents a first kind of bifurcation, that we are going to call *tangent bifurcation*:

1. There is no fixed point when $c > 1/4$ (all orbits tend to infinity).
2. If $c = 1/4$ there is a single fixed point that is neutral.

3. For $c < 1/4$, there are two different fixed points p_+ and p_-: the former is always repelling and the latter is attracting if $c \in (\frac{3}{4}, \frac{1}{4})$.

This situation is pictured in Figure 3.1.

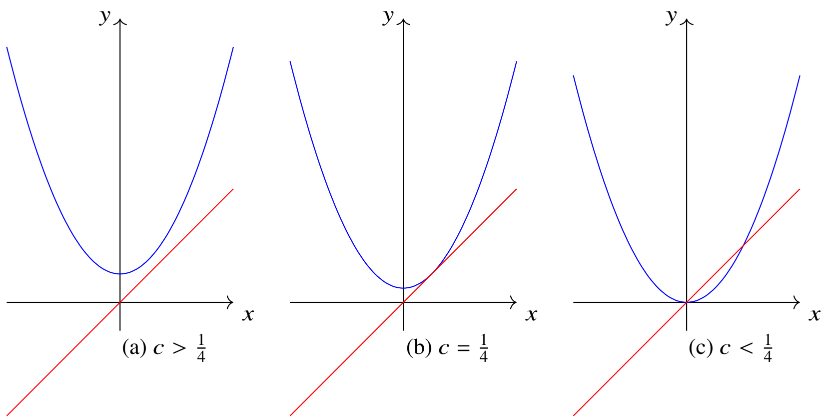

Fig. 3.1 Plots of $f(x) = x^2 + c$ and $g(x) = x$ for different values of c. As c decreases, a fixed point appears when $c = 1/4$, when the graph of g is tangent to the graph of f.

Exercise 3.1 Show that

- If $x < -p_+$ or $x > p_+$ then $Q_c^n(x) \to \infty$.
- $-p_+$ is eventually fixed.
- When $c \in (-\frac{3}{4}, \frac{1}{4})$, $Q_c^n(x) \to p_-$ for every $x \in (-p_+, p_+)$ (easier case: $0 \leq c < 1/4$).

At $c = -3/4$, one had $Q_c'(p_-) = -1$, and p_- becomes a repelling fixed point for $c < -3/4$. But also a period-2 orbit appears for these values of c. To see this, let us consider the equation

$$Q_c^2(x) = x \quad \Leftrightarrow \quad p(x) := x^4 + 2cx^2 - x + c^2 = 0. \tag{3.3}$$

We know that p_+ and p_- are roots of this equation, so $p(x)$ is divisible by

$$(x - p_-)(x - p_+) = x^2 - x + c. \tag{3.4}$$

Performing the division, we conclude that

$$\frac{x^4 + 2cx^2 - x + c^2}{x^2 - x + c} = x^2 + x + c + 1. \tag{3.5}$$

The roots of the resulting polynomial are

$$q_\pm = \frac{1}{2}(-1 \pm \sqrt{-4c - 3}). \tag{3.6}$$

The solutions are real if and only if $c \le -3/4$.

So at $c = -3/4$ we find a second type of bifurcation, that we are going to call *priod-doubling*:

1. For $-3/4 < c < 1/4$, Q_c has an attracting fixed point and no 2-cycles.
2. For $c = -3/4$, Q_c has a neutral fixed point at $p_- = q_\pm$ and no 2-cycles.
3. For $-5/4 < c < 3/4$, Q_c has repelling fixe points and an attracting 2-cycle made by q_+ and q_-.

3.2 Types of bifurcations

Let $F_\lambda(x)$ be a one-parameter family of functions $F_\lambda : \mathbb{R} \to \mathbb{R}$, with $\lambda \in \mathbb{R}$.

A ***bifurcation*** is a change in the dynamical behavior of F_λ when λ crosses some particular value.

Definition 3.1 A family F_λ undergoes a ***tangent bifurcation*** (also known as ***saddle-node bifurcation***) at λ_0 if there is an interval $I \subset \mathbb{R}$ and an $\epsilon > 0$ such that

1. For $\lambda_0 - \epsilon < \lambda < \lambda_0$, F_λ has no fixed points in I.
2. For $\lambda = \lambda_0$, F_λ has one fixed point.
3. For $\lambda_0 < \lambda < \lambda_0 + \epsilon$, F_λ has two fixed points, one attracting and one repelling.

Remark that this is again a local definition. We obtain the same kind of bifurcation when we invert all the inequalities in Definition 3.1.

The definition above can also be applied to iterates F_λ^n: in that case, an n-cycle appears at λ_0 (i.e. a fixed point of F_λ^n), and there is an attracting n-cycle and a repelling n-cycle that appear when $\lambda > \lambda_0$.

The tangent bifurcation typically occurs when the graph of F_{λ_0} has a quadratic tangency with $y = x$ at a point (x_0, x_0), which means that $F'_{\lambda_0}(x_0) = 1$ and $F''_{\lambda_0}(x_0) \neq 0$ (i.e. there is local curvature: F_λ is locally concave or convex). See again Figure 3.1-(b).

Example 3.1 As we saw above, the family Q_c has a tangent bifurcation at $c = 1/4$.

A ***bifurcation diagram*** plots the parameter on the horizontal axis and the value of the fixed point of the vertical axis. A continuous line denote the attracting fixed points, and a broken line the repelling ones. For the bifurcation diagram of Q_c around $c = 1/4$, see Figure 3.2.

Definition 3.2 F_λ undergoes a ***period-doubling bifurcation*** at $\lambda = \lambda_0$ if there is an open interval I and $\epsilon > 0$ such that

1. For each $\lambda \in (\lambda_0 - \epsilon, \lambda_0 + \epsilon)$, there is a unique fixed point p_λ of F_λ in I.
2. For $\lambda_0 - \epsilon < \lambda \le \lambda_0$, F_λ has no cycle of period 2 in I and p_λ is attracting (respectively, repelling).
3. When $\lambda_0 < \lambda \le \lambda_0 + \epsilon$, there exists a unique 2-cycle made by q_λ^1 and q_λ^2 is I, with $F_\lambda(q_\lambda^1) = q_\lambda^2$. The 2-cycle is attracting (resp. repelling).

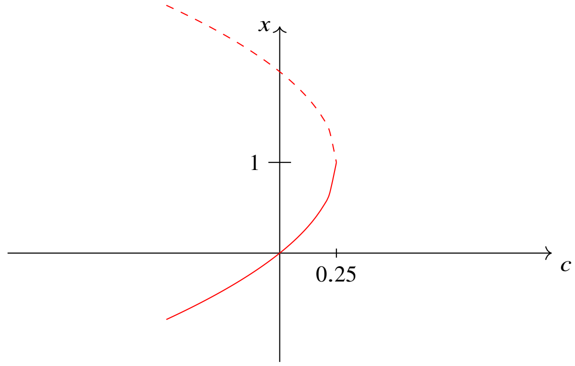

Fig. 3.2 Plots of the unstable fixed point $q_+ = 1 + \sqrt{1+4c}$ (dashed line) and the stable fixed point $q_- = 1 - \sqrt{1-4c}$ (solid line) for c between -1 and 1.

4. As $\lambda \to \lambda_0^+$, $q_\lambda^i \to p_\lambda$ for $i = 1, 2$.

The period-doubling bifurcation occurs when $F'_{\lambda_0}(p_{\lambda_0}) = -1$, that is, when the tangent at the fixed point is perpendicular to the diagonal. In this case, in virtue of ..., $(F^2_{\lambda_0})'(p_{\lambda_0}) = 1$, so $y = x$ is tangent to F^2. When λ is slightly bigger than λ_0, to situations are possible

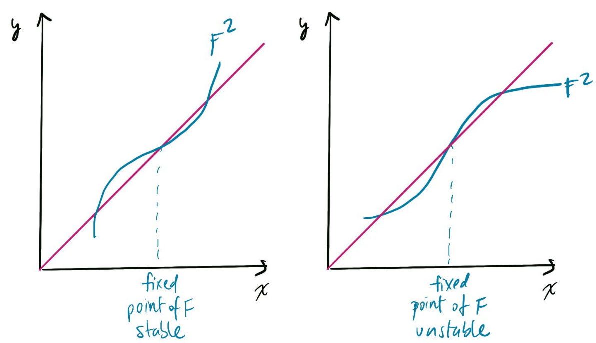

Fig. 3.3 Graphs of F^2_λ (blue) and $y = x$ (fuchsia) for λ slightly bigger than λ_0.

The left panel corresponds to p_λ being attracting and the 2-cycle repelling; the right panel to p_λ being repelling and the 2-cycle attracting.

Example 3.2 The family $Q_c(x) = x^2 + c$ undergoes a period-doubling bifurcation at $c = -3/4$, in which an attracting fixed point becomes repelling when an attracting 2-cycle emerges.

Example 3.3 Another example is given by $F_\lambda(x) = \lambda x - x^3$. Remark that $x_0 = 0$ is a fixed point, which is attracting when $-1 < \lambda < 1$, neutral when $\lambda = -1$ and repelling when $\lambda < -1$.

To find the 2-cycles, remark first that F is odd, that is, $F_\lambda(-x) = -F_\lambda(x)$. This implies that any x_0 that solves $F_\lambda(x) = -x$ is a fixed point of F_λ^2, because

$$F_\lambda^2(x_0) = F_\lambda(F_\lambda(x_0)) = F_\lambda(-x_0) = -F_\lambda(x_0) = -(-x_0) = x_0. \tag{3.7}$$

The equation $F_\lambda(x) = -x$ can be written as $x((\lambda+1)+x^2) = 0$, and its solutions are 0 and $\pm\sqrt{\lambda+1}$. So we have indeed a nontrivial 2-cycle when $\lambda < -1$.

A computation shows that $|F_\lambda'(\pm\sqrt{\lambda+1})| = |\lambda - 3(\lambda+1)^2|$. The reader may verify that this is indeed strictly greater than 1 when $\lambda < -1$.

Exercise 3.2 To conform completely to Definition 3.2, prove that in some interval containing 0 and $\pm\sqrt{\lambda+1}$ there is no other 2-cycle.

3.3 More on period doubling

Any iterate F_λ^n of F_λ may undergo a tangent of period-doubling bifurcation as in Definitions 3.1 and 3.2. In case of a period-doubling bifurcation, the fixed point p_λ of F_λ^n might be a period-n point of F_λ, and the 2-cycle of F_λ^n that emerges represents a new period $2n$-cycle of F_λ.

This might be illustrated with the familiar example of Q_c. Figure 3.4 plots Q_c, $y = x$ and a square with vertices (p_+, p_+) and $(-p_+, p_+)$ for different values of c, whereas Figure 3.5 plots Q_c^2 together with $y = x$ and a square with vertexes (p_-, p_-) and (s, p_-), where s is the point to the left of p_- that satisfies $Q_c^2(s) = p_-$. Between panels (a) and (b) in Figure3.4 a tangent bifurcation occurs. The sequence (b)-(c)-(d) in Figure 3.4 corresponds to a period-doubling bifurcation; the corresponding panels (a)-(b)-(c) in Figure 3.5 shows what happens with Q_c^2 for those values of c; we see the emergence of two new fixed points. Observe that locally (inside the highlighted square) this emergence of a fixed point of Q_c^2 can be seen as a tangent bifurcation.

Similarly, in Figure 3.5-(c), the fixed point q_- of Q_c^2 becomes neutral, and Q_c^2 undergoes a period-doubling bifurcation i.e. a 4-cycle of Q_c appears. A similar reasoning could then be applied to Q_c^4 for different values of c. We obtain in this way a "period-doubling cascade".

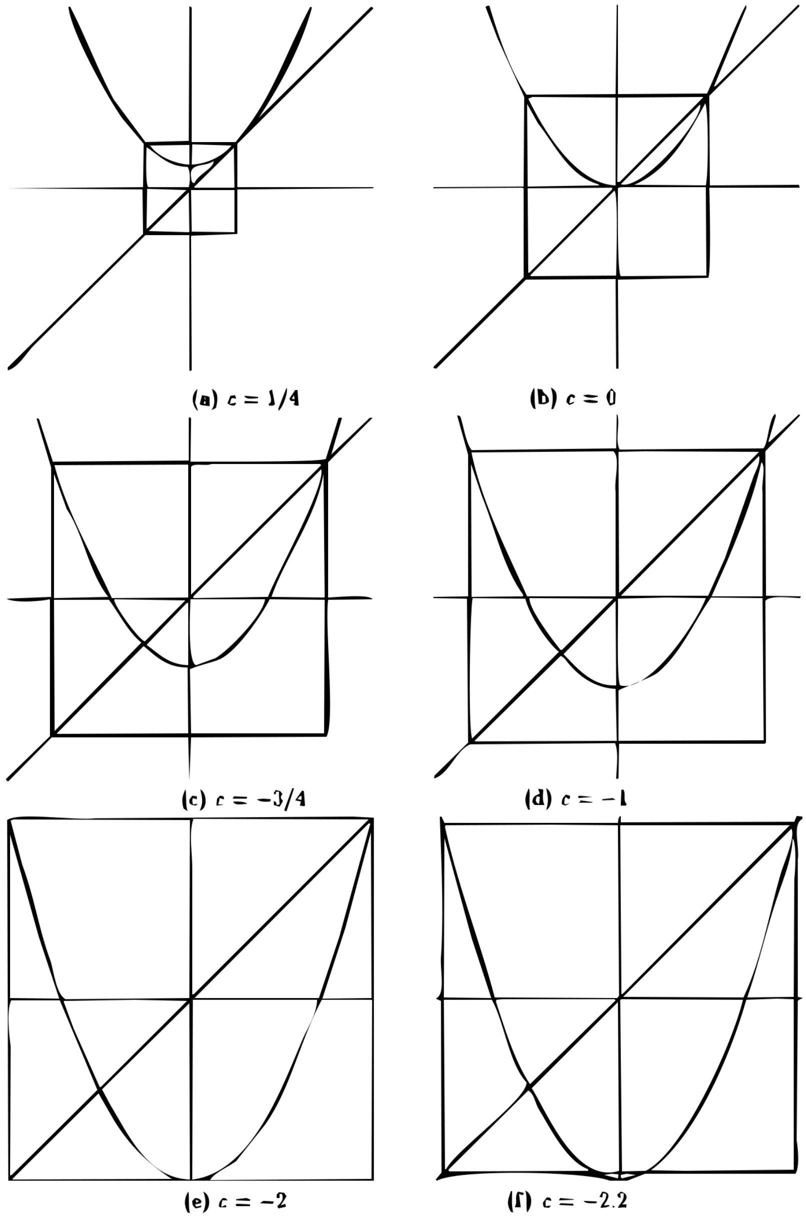

Fig. 3.4 Graphs of Q_c for different values of c. Taken from [3].

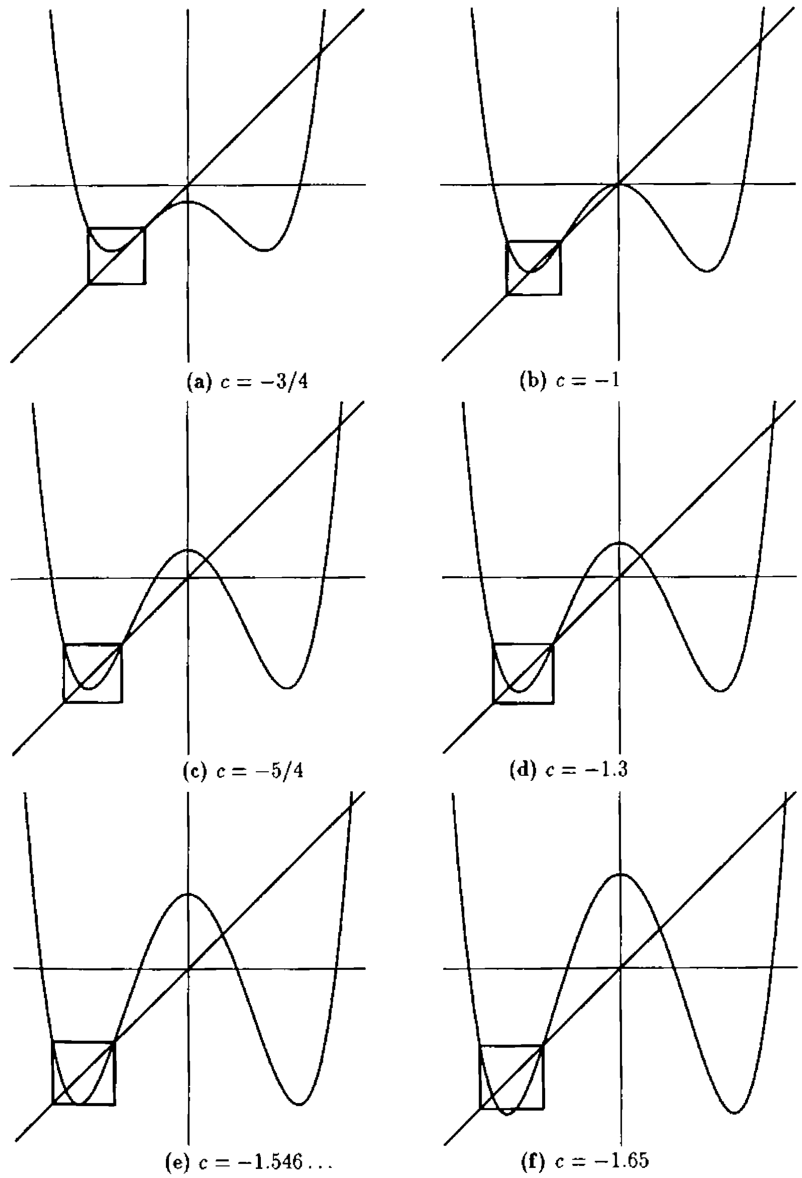

Fig. 3.5 Graphs of Q_c^2 for different values of c; remark that these values of c are not as in Figure 3.4. Taken from [3].

Chapter 4
Periods

4.1 Periodic points of Q_{-2}

[...]

We see in this example that sometimes periodic points of many different cycles might appear all at the same time. The result in the following section represents an even more surprising version of this.

4.2 Period three implies all periods

We are going to prove the following remarkable theorem.

Theorem 4.1 *Suppose $F : \mathbb{R} \to \mathbb{R}$ is continuous. If F has a periodic point of period 3, then F also has periodic points of all other periods.*

This result was published by Li and Yorke [5] in 1975. Back then there was a significant interest in chaos theory originated by Lorenz's research. Because of the Cold War, communications between the U.S.S.R. and the U.S.A. were very difficult, and translations limited, so Li and Yorke were unaware that their result was a particular case of a more general theorem published by Sarkovskii in 1964 [6].

To formulate Sharkovsky's theorem, let us introduce an unusual total order $\lhd$ of the positive integers, given by

$$
\begin{aligned}
&3 \lhd 5 \lhd 7 \lhd 9 \lhd 11 \lhd \cdots \\
&2 \cdot 3 \lhd 2 \cdot 5 \lhd 2 \cdot 7 \lhd 2 \cdot 9 \lhd 2 \cdot 11 \lhd \cdots \\
&2^2 \cdot 3 \lhd 2^2 \cdot 5 \lhd 2^2 \cdot 7 \lhd 2^2 \cdot 9 \lhd 2^2 \cdot 11 \lhd \cdots \\
&\cdots \\
&\cdots \lhd 2^n \lhd 2^{n-1} \lhd \cdots 2^3 \lhd 2^2 \lhd 2 \lhd 1.
\end{aligned}
$$

Theorem 4.2 (Sharkovsky) *Assume that F is a continuous map on an interval and has a period p orbit. If $p \lhd q$, then F has a period q orbit too.*

The proof of Sharkovsky's theorem is quite involved, but most of its fundamental ideas are already present in the proof of Theorem 4.1. hence we shall limit ourselves to the proof of this simpler result. The interested reader might consult [1, p. 135] for a guided exercise that proves Theorem 4.2.

Theorem 4.1's proof depends on two fundamental lemmas.

Lemma 4.1 *Let $I = [a, b]$ and $J = [c, d]$. Suppose $I \subset J$. If $F(I) \supset J$, then F has a fixed point in I.*

Proof Let $H(x) = F(x) - x$. A fixed point is a root of H.

Since I is closed and bounded, F attains its maximum at minimum at points $p_{\min}$ and $p_{\max}$ in I, respectively.

If $p_{\min} < p_{\max}$, we have $[p_{\min}, p_{\max}] \subset J \subset [F(p_{\min}), F(p_{\max})]$, hence $H(p_{\min}) \leq 0$ and $H(p_{\max}) \geq 0$. We conclude by the IVT that H has a root in $[p_{\min}, p_{\max}]$.

If $p_{\max} < p_{\min}$, one establishes similarly that $H(p_{\max}) \leq 0$ and $H(p_{\min}) \geq 0$. □

We say that an interval I of the real line is ***compact*** if it is closed and bounded, that is, if $I = [a, b]$ with a and b finite.

Lemma 4.2 *Let I be an interval and $F : I \to \mathbb{R}$ a continuous function. For any compact interval $J \subset F(I)$, there is a compact interval $I' \subset I$ such that $F(I') = J$.*

Proof We write first $J = [G(p), G(q)]$, where $p, q \in I$. We have to distinguish two cases: $p < q$ and $q < p$.

Suppose $p < q$. Set

$$r = \sup\{\, x \in [p, q] \mid G(x) = G(p) \,\}. \tag{4.1}$$

Remark that this is the supremum of a nonempty bounded set (at least it contains p), so r is finite, and $r \in I$ because I is closed. Moreover, $G(r) = G(p)$ by continuity of F. Then set

$$s = \inf\{\, x \in [r, q] \mid G(x) = G(q) \,\}. \tag{4.2}$$

For similar reasons $s \in I$ and $G(s) = G(q)$.

Set $I' = [r, s]$. If $G(I')$ contains an element z that is not in J, then either $z < G(p)$ or $z > G(q)$. In the former case, we would conclude that there is a preimage of p that is greater than r, contradicting the definition of r, and in the latter that there is a preimage of q smaller than s, also contradicting its definition.

The case $q < p$ is analogous and we leave it as an exercise. □

Proof (of Theorem 4.1) Suppose F has a period 3 cycle $(a, b, c, a, b, c, \ldots)$, which means that $F(a) = b$ and $F(b) = c$. Without loss of generality, let a be the leftmost point in $\mathbb{R}$. We distinguish again two cases: $a < b < c$ and $a < c < b$.

Let us suppose $a < b < c$. Set $I_0 = [a, b]$ and $I_1 = [b, c]$. Then $F(I_0) \supset I_1$, in virtue of the IVT. Similarly, $F(I_1) \supset I_0 \cup I_1$.

1. *Existence of a fixed point (period 1):* We simply remark that $F(I_1) \supset I_1$, hence the existence of a fixed point follows from Lemma 4.1.
2. *Existence of a 2-cycle:* Because

$$F^2(I_0) \supset F(I_1) \supset I_0 \cup I_1 \supset I_0, \tag{4.3}$$

there is a fixed point of F^2 in I_0 by Lemma 4.1; unfortunately we cannot guarantee that this point has period 2. But we can refine the argument as follows: since $F(I_0) \supset I_1$, Lemma 4.2 ensures the existence of a compact interval $A \subset I_0$ such that $F(A) = I_1$; but $F(I_1) \supset I_0 \supset A$, that is, $F^2(A) \supset A$, so there must be a fixed point x_0 of F^2 in A, again by Lemma 4.1. We know x_0 is not a fixed point, because $x_0 \in I_0$ but $F(x_0) \in I_1$, and the only point that I_0 and I_1 have in common has period 3.
3. *Existence of a point of period $n > 3$:* We have not used yet that $F(I_1) \supset I_1$. By Lemma 4.2, there is a compact interval $A_1 \subset I_1$ such that $F(A_1) = I_1$. Since $F(A_1) \supset A_1$, we can now define $A_1 \subset A_1$ such that $F(A_2) = A_1$. Recursively, we introduce sets

$$A_{n-2} \subset A_{n-3} \subset \cdots A_1 \subset I_1 \tag{4.4}$$

such that $F(A_i) = A_{i-1}$ for $i = 2, ..., n-2$.
Remark that $F(I_0) \supset I_1 \supset A_{n-2}$, so there exists A_{n-1}, compact subinterval of I_0, such that $F(A_{n-1}) = A_{n-2}$.
Finally, because $F(I_1) \supset I_0 \cup I_1 \supset A_{n-1}$, there is $A_n \subset I_1$, a compact interval, such that $F(A_n) = A_{n-1}$.
Because of our choices, $F^n(A_n) = I_1$, so $F^n(A_n) \supset A_n$, which implies that F^n has a fixed point x_1 in A_n. This point must have period n, because $F(x_1) \in I_0$ but $F^n(x_1) \in I_1$ for $n = 2, 3, ..., n$. □

4.3 Transition graphs

The proof of Theorem 4.1 can be turned into a general method to establish the existence of periodic points of a given period.

Consider a continuous function $F : I \to \mathbb{R}$ defined on a compact interval I. A ***partition*** of I is a finite collection of compact intervals $\{I_i\}$ such that $\bigcup_i I_i = I$ and $I_i \cap I_j$ is empty or contains a single point whenever $i \neq j$.

A ***transition graph*** is a graph whose vertices are the intervals I_i and that has an arrow $I_i \to I_j$ iff $I_j \subset F(I_i)$, that is if F can take a point from I_i to I_j.

Exercise 4.1 Prove that if $S_1 \to S_k \to S_{k-1} \to \cdots \to S_2 \to S_1$ is an allowable path in the graph, then F^k has a fixed point in S_1.

Exercise 4.2 Prove that if $S_1 \to S_k \to S_{k-1} \to \cdots \to S_2 \to S_1$ has no repeated pattern, the fixed point above must have period k.

Chapter 5
The Cantor set

5.1 A motivating example

Let us consider now the function $Q_c(x) = x^2 + c$ when $c < -2$. We know that any seed outside the interval $I = [-p_+, p_+]$ has an orbit that diverges to $+\infty$. But what happens with the points in I?

Let us introduce

$$A_1 = \{ x \in I \mid Q_c(x) \in I^c \}, \tag{5.1}$$

the points whose first iterate falls outside I. It is clear that these points also diverge to $+\infty$. The set A_1 is an open interval around 0.

Similarly, the points in

$$A_2 = \{ x \in I \setminus A_1 \mid Q(x) \in I^c \} \tag{5.2}$$

also have divergent orbits. This set is the union of two open intervals, one to the right and one to the left of A_1.

We may define

$$\Lambda = \{ x \in I \mid Q^n(x) \in I \text{ for all } n \in \mathbb{N} \}, \tag{5.3}$$

the set of points of I whose orbits remain in I. What does this set look like?

Λ can be constructed recursively, first removing A_1 (an open interval) from I, then $A_2 = Q_c^{-1}(A_1)$ (two open intervals) from $I \setminus A_1$, and so on (removing each time $A_n = Q_c^{-1}(A_{n-1})$).

Is the resulting set empty? No: it contains at least p_+ (which is a fixed point), as well as $-p_+$ and the endpoints of each interval A_i, $i \in \mathbb{N}$ (all these points are eventually fixed). We shall see that Λ contains much more: it is an uncountable set.

5.2 Cantor middle-thirds set

To isolate the set-theoretic aspects of the construction evoked in the previous section, we consider here a more idealized recursive construction, that does not make explicit reference to a continuous map like Q_c. It is known as the ***Cantor middle-thirds set***, and we shall see that it appears over and over in dynamical systems, sometimes in disguise. It is also a fundamental example of a ***fractal*** or self-similar set.

The Cantor middle-thirds set is constructed via the following "recipe":

Step 0: Start with the interval $K_0 = [0, 1]$

Step 1: Remove the open interval $(1/3, 2/3)$ to obtain $K_1 = [0, 1/3] \cup [2/3, 1]$.

Step 2: Remove the middle-thirds of the remaining two intervals, that is, $(1/9, 2/9)$ from $[0, 1/3]$ and $(7/9, 8/9)$ from $[2/3, 1]$, to obtain $K_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$.

...

Step n: Introduce a set K_n by removing the middle-thirds of the 2^n closed intervals (maximal by inclusion) that form K_{n-1}.

...

Remark that $K_{n+1} \subset K_n$ for all n, so the set $\tilde{K}$ that results from performing all the steps can be expressed as $\bigcap_{n\in\mathbb{N}} K_n$. The set $\tilde{K}$ is nonempty: it contains, at least, all the *triadic points* i.e. those of the forms $m/3^n$ for any $n \in \mathbb{N}$ and $m \in \{0, 1, 2, ..., 3^n\}$.

Although this description of the Cantor middle-thirds set is very graphical, it is not easy to work with it.

Ternary expansion

The ternary expansion of a point $x \in [0, 1]$ is

$$x = \frac{c_1}{3} + \frac{c_2}{3^2} + \cdots + \frac{c_n}{3^n} + \cdots \tag{5.4}$$

where $c = (c_1, c_2, c_3, ...)$ is an element of $\{0, 1, 2\}^{\mathbb{N}}$, that is, a sequence whose terms c_i are 0, 1 or 2 for any $i \in \mathbb{N}$. More compactly, we write $x = 0._3 c_1 c_2 c_3$; the decorated dot $._3$ emphasizes we are talking about the ternary expansion (as opposed to the usual *decimal* expansion).

We shall use repeatedly that whenever $|a| < 1$,

$$\sum_{i=k}^{\infty} a^i = \frac{a^k}{1-a}. \tag{5.5}$$

This implies, in particular, that for any choice of c,

$$0 \le \sum_{i=1}^{\infty} \frac{c_i}{3^i} \le \sum_{i=1}^{\infty} \frac{2}{3^i} = \frac{\frac{2}{3}}{1-\frac{1}{3}} = 1. \tag{5.6}$$

hence indeed (5.4) represents a number in $[0, 1]$.

Example:

$$0._3 0202... = 2\sum_{i=1}^{\infty}\left(\frac{1}{9}\right)^i = \frac{1}{4}. \tag{5.7}$$

Remark that triadic points have two different ternary expansions. For instance,

$$\frac{1}{3} = 0._3 100000.... = 0._3 022222... \tag{5.8}$$

because

$$\sum_{i=2}^{\infty}\frac{2}{3^i} = \frac{2}{9}\sum_{i=0}^{\infty}\left(\frac{1}{3}\right)^i = \frac{1}{3}. \tag{5.9}$$

Similarly,

$$\frac{8}{9} = 0._3 22000... = 0.212222... \tag{5.10}$$

because

$$\frac{2}{3}+\frac{1}{9}+\sum_{i=3}^{\infty}\frac{2}{3^i} = \frac{8}{9} = \frac{2}{3}+\frac{2}{9} \tag{5.11}$$

What happens in these cases is analogous to the equality 0.9999... = 1 in decimal representation. This kind of ambiguity occurs when $x = 0._3 c_1 c_2 c_3 0000\cdots$, which is equivalent to $x = p/3^n$ for some p, n.

Remark that the points of the form $0._3 0 c_2 c_3\cdots$ all fall in $[0, 1/3]$, with $1/3$ corresponding to $c_i = 2$ for all $i \geq 2$. Similarly,

$$[1/3, 2/3] = \{\, x = 0._3 1 c_2 c_3\cdots \mid c_i \in \{0, 1, 2\}\,\}$$
$$[2/3, 1] = \{\, x = 0._3 2 c_2 c_3\cdots \mid c_i \in \{0, 1, 2\}\,\}$$

Again, $1 = 0._3 22222\cdots$.

Cantor originally characterized the middle-thirds set as follows.

Definition 5.1 The ***Cantor middle-thirds set*** is the set of those points $x_c \in [0, 1]$ whose ternary expansion is

$$x_c = \frac{c_1}{3}+\frac{c_2}{3^2}+\cdots+\frac{c_n}{3^n}+\cdots \tag{5.12}$$

for some sequence $c = (c_1, c_2, c_3, ...) \in \{0, 2\}^{\mathbb{N}}$.

Exercise 5.1 Prove that every real number in K (that is once we remove the possibility of having 1) has a unique ternary expansion (5.12).

Let us establish that this indeed corresponds to the set K described at the beginning of this section. First, the sets

$$C_0 = \{ x = 0._3 d_1 d_2 d_3 \cdots \in [0,1] \mid d_1 = 0 \} = [0, 1/3] \tag{5.13}$$
$$C_2 = \{ x = 0._3 d_1 d_2 d_3 \cdots \in [0,1] \mid d_1 = 2 \} = [2/3, 1] \tag{5.14}$$

are exactly the two intervals that remain after Step 1 of the construction of K is performed. Similarly, the sets

$$C_{i,j} = \{ x = 0._3 d_1 d_2 d_3 \cdots \in [0,1] \mid d_1 = i, d_2 = j \} \tag{5.15}$$

for $i, j \in \{0, 2\}$ are the four intervals that remain after Step 2 is performed. In general, given a finite word $(c_1, c_2, ..., c_n) \in \{0, 2\}^n$, we can introduce the *cyclindrical set*

$$C_{c_1,...,c_n} = \{ x = 0._3 d_1 d_2 d_3 \cdots \in [0,1] \mid d_i = c_i \text{ for each } i = 1, ..., n \}. \tag{5.16}$$

These sets, which are pairwise disjoint, result from performing Step n is the construction of K described at the beginning of this section. Therefore,

$$K = \bigcap_{n\in\mathbb{N}} \left(\bigcup_{(c_1,...,c_n)\in\{0,2\}^n} C_{c_1,...,c_n} \right). \tag{5.17}$$

If we (cleverly) change the order of the unions and intersections, we obtain

$$K = \bigcup_{(c_1,c_2,...)\in\{0,2\}^{\mathbb{N}}} \left(\bigcap_{n\in\mathbb{N}} C_{c_1,...,c_n} \right); \tag{5.18}$$

observe that here we have $\{0, 2\}^{\mathbb{N}}$ instead of $\{0, 2\}^n$. The key fact now is that given an infinite word $c = (c_1, c_2, ...) \in \{0, 2\}^{\mathbb{N}}$, the intersection

$$\bigcap_{n\in\mathbb{N}} C_{c_1,...,c_n} \tag{5.19}$$

consists of a single point (why?), which necessarily is $x_c = 0._3 c_1 c_2 c_3 \cdots$.

5.3 Uncountability

An advantage of Definition 5.1 is that we can "easily" proof the following.

Proposition 5.1 *The set K is uncountable i.e. there is no bijection between K and $\mathbb{N}$.*

Proof A bijection between K and $\mathbb{N}$ consists in an enumeration $x_1, x_2, x_3, ...$ of all the points of K. We prove that this is impossible by contradiction. If such enumeration exists, we can make a rectangular array with the ternary expansions of each $x \in K$:

$$
\begin{aligned}
x_1 &= 0._3 c_1^1 c_2^2 c_3^3 \cdots \\
x_2 &= 0._3 c_1^2 c_2^2 c_3^2 \cdots \\
x_3 &= 0._3 c_1^3 c_2^3 c_3^3 \cdots \\
&\vdots
\end{aligned}
$$

Define $\bar{x} = 0._3 d_1 d_2 d_3 \cdots$ such that

$$
d_i = \begin{cases} 0 & \text{if } c_i^i = 2 \\ 2 & \text{if } c_i^i = 0 \end{cases}. \tag{5.20}
$$

Then $\bar{x} \neq x_i$ for all $i \in \mathbb{N}$ (from Exercise 5.1 we know that the expansion of x_i is unique). But this is a contradiction, since our sequence $(x_i)_{i \in \mathbb{N}}$ was supposed to contain all the points in K. □

The argument in the proof, known as *Cantor's diagonalization argument*, is arguably one of the most important in the history of mathematics.

Exercise 5.2 Prove that

$$
K - K := \{\, p - q \mid p, q \in K \,\} \tag{5.21}
$$

equals $[0, 1]$. (Hint: By induction in the length of the ternary expansions.)

Chapter 6
Symbolic dynamics

6.1 The space of binary sequences

The considerations in Section 5.2 mean that we can study the Cantor middle-third set through the lens of the ***symbolic space***

$$\Sigma := \{0,1\}^{\mathbb{N}} := \{\, (s_0, s_1, s_2, \ldots) \mid s_i = 0 \text{ or } 1 \,\}, \tag{6.1}$$

i.e. the space of infinite binary sequences. Here the symbols 0 and 1 do not play any special role: we could equally well use $\top$ and $\perp$, or A and B.

To simplify notation, we sometimes write $s_0 s_1 s_2 \cdots$ instead of $(s_0, s_1, s_2, \ldots)$.

It is important to remark that Σ is not a subset of some Euclidean space, hence it is not so easy to visualize. However, we shall see that we can still use some of our geometric intuitions with it; for this, we must introduce a notion of distance between two points. Unavoidably, the only way of doing this is by introducing a higher level of abstraction.

Definition 6.1 Let $s = (s_0, s_1, s_2, \ldots)$ and $t = (t_0, t_1, t_2, \ldots)$ be two points of Σ. The ***distance*** $d(s,t)$ between s and t is given by

$$d(s,t) = \sum_{i=0}^{\infty} \frac{|s_i - t_i|}{2^i}. \tag{6.2}$$

Example 6.1 If $s = (1,1,1,\ldots)$ and $t = (0,1,0,1,\ldots)$, then

$$d(s,t) = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^5} + \cdots = \frac{1}{2}\sum_{i\geq 0}\left(\frac{1}{4}\right)^i = \frac{2}{3}. \tag{6.3}$$

The distance d satisfies the following properties, which are familiar properties of the distance $|x-y|$ between two points of the real line or, more generally, of the distance between two points in Euclidean space.

Proposition 6.1 *1. Positivity: For all $s, t \in \Sigma$, $d(s,t) \geq 0$.*

2. *Nondegeneracy:* $d(s,t) = 0$ *if and only if* $s = t$.
3. *Symmetry: For all* $s, t \in \Sigma$, $d(s,t) = d(t,s)$.
4. *Triangular inequality: For all* $s, t, u \in \Sigma$,

$$d(s,u) \leq d(s,t) + d(t,s)$$

Proof 1. Clearly, because $d(s,t)$ is a sum of nonnegative terms.
2. $d(s,t) = 0$ iff for all $i \geq 0$, $|s_i - t_i| = 0$, which in turn is equivalent to $s = t$.
3. This holds because for each i, $|s_i - t_i| = |t_i - s_i|$.
4. This is a consequence of the triangular inequality for the absolute value: for each i,

$$|s_i - u_i| \leq |s_i - t_i| + |t_i - u_i|.$$

Definition 6.2 A set X with a function $d : X \times X \to \mathbb{R}$ that satisfies the properties 1-4 in Proposition 6.1 is called a ***metric space*** and d is called a ***metric***.

So Proposition 6.1 proves that (Σ, d) is a metric space. The metric d gives an intrinsic notion of "closeness" between two points.

Theorem 6.1 (Proximity theorem) *Let* $s, t \in \Sigma$. *If* $s_i = t_i$ *for* $i \in \{0, ..., n\}$, *then* $d(s,t) \leq \frac{1}{2^n}$. *Conversely, if* $d(s,t) < \frac{1}{2^n}$, *then* $s_i = t_i$ *for* $i \in \{0, ..., n\}$

Proof If $s_i = t_i$ for $i \in \{0, ..., n\}$, then $d(s,t) \leq \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \frac{1}{2^n}$. To prove the second statement, suppose $s_i \neq t_i$ for some $j \leq n$; in this case $d(s,t) \geq 2^{-j} \geq 2^{-n}$, contradicting $d(s,t) < 2^{-n}$. □

6.2 Shift map

Definition 6.3 The map $\sigma : \Sigma \to \Sigma$ given by

$$\sigma(s_0, s_1, s_2, ...) = (s_1, s_2, s_3, ...) \tag{6.4}$$

is called the ***shift map***.

The iterations of σ define a discrete-time dynamical system.
The shift map is easy to iterate: one just has to drop more entries. Hence

$$\sigma^n(s_0, s_1, s_2, s_3, ...) = (s_n, s_{n+1}, s_{n+2}, ...). \tag{6.5}$$

It is also easy to find periodic points: they consist in any pattern that repeats. So

$$\mathbf{s} = \overline{s_0 s_1 \cdots s_{n-1}} = (s_0, ..., s_{n-1}, s_0, ..., s_{n-1}, s_0, ...) \tag{6.6}$$

is a fixed point of σ^n, which is of period n if $s_0 \cdots s_{n-1}$ has no repeated pattern.
hence the fixed points of σ are the words $\overline{0}$ and $\overline{1}$, its only 2-cycle is made of $\overline{01}$ and $\overline{10}$. It has two 3-cycles, given by

$$\overline{001} \mapsto \overline{010} \mapsto \overline{100110} \mapsto \overline{101} \mapsto \overline{011}. \tag{6.7}$$

The explicit determination of periodic points is almost trivial; this is in stark contrast with the functions we saw before, like Q_c.

6.3 Continuity

There is a general definition of continuity for maps between metric spaces.

Definition 6.4 Let $F : X \to Y$ be a map between metric spaces (X, d_X) and (Y, d_Y). We say that F is ***continuous*** at $x_0 \in X$ if for all $\epsilon > 0$ there exists a $\delta > 0$ such that if $d_X(x_0, x) < \delta$ then $d_Y(F(x_0), F(x)) < \epsilon$.

This is just like the usual definition of continuity, except that "closeness" in the real line is measured by the absolute value of the difference of two numbers, whereas in general it is measured by the distance functions d_X, d_Y, etc.

We can apply this more general definition of continuity to maps $F : \Sigma \to \Sigma$, so in particular to the shift map.

Proposition 6.2 *The shift map σ is continuous.*

Proof Consider $x_0 \in \Sigma, x_0 = (x_0^0, x_1^0, x_2^0, ...)$. Given $\epsilon > 0$, chose n such that $2^{-n} < \epsilon$.

By the Proximity Theorem, if $d(x_0, s) < 2^{-(n+1)}$, then $x_0^i = s_i$ for $i \in \{0, ..., n+1\}$. But then $\sigma(x_0)_i = \sigma(s)_i$ for $i \in \{0, ..., n\}$, which in turn implies that $d(\sigma(x_0), \sigma(s)) \leq 2^{-n} < \epsilon$, again by the Proximity Theorem.

This means that we can choose $\delta = 2^{-(n+1)}$. □

6.4 Symbolic coding

Let us go back to the study of Q_c for $c < -2$. Set $I = [-p_+, p_+]$. We introduced above the set

$$\Lambda = \{ x \in I_0 \mid Q_c^n(x) \in I \text{ for all } n \in \mathbb{N}^* \}. \tag{6.8}$$

This set has a complicated structure, but we saw that we can construct it iteratively, removing first a set

$$A_1 = \{ x \in I \mid Q_c(x) \notin I \}. \tag{6.9}$$

to obtain a set $\Lambda_1 = I \setminus A_1 = I_0 \cup I_1$, where I_0 is a compact subinterval of $(-\infty, 0)$, and I_1 a compact subinterval of $(0, \infty)$.

The next stage is to remove from Λ_1 the set

$$A_2 = \{ x \in \Lambda_1 \mid Q_c(x) \notin \Lambda_1 \}, \tag{6.10}$$

to obtain $\Lambda_2 = \Lambda_1 \setminus A_2$. In this operation, we remove an open interval from I_0 and an open interval from I_1. Because Q_c establishes a monotonically increasing (resp.decreasing) bijection between I_1 (resp. I_0) and I, we can identify the remaining left or right fragment of I_i according to where it is mapped by F, either I_0 or I_1.

More generally, we can introduce a map $h : \Lambda \to \Sigma$ given by

$$h(x) = (s_0, s_1, s_2, ...) \quad \text{where} \quad \begin{cases} s_j = 0 & \text{if } Q_c^j(x) \in I_0 \\ s_j = 1 & \text{if } Q_c^j(x) \in I_1 \end{cases}. \tag{6.11}$$

For instance, $h(p_+) = (1, 1, 1, ...)$, $h(-p_+) = (0, 1, 1, ...)$, $h(x) = (0, ...)$ if $x \in I_0$ i.e. to the left of A_1, $h(x) = (1, ...)$ if $x \in I_1$, and

[picture]

The map h is known as the ***coding map***; we shall see that it encodes every point of Λ as a binary sequence and, in fact, it established a homeomorphism between Λ and Σ.[1] In particular, points that are close in Λ are mapped to sequence that are close in Σ.

Not only the coding map allows us to identify Λ as a copy of the Cantor set, but also gives a symbolic representation of the map Q_c. We are going to see that the square

$$\begin{array}{ccc} \Lambda & \xrightarrow{Q_c} & \Lambda \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h} \\ \Sigma & \xrightarrow{\sigma} & \Sigma \end{array}, \tag{6.12}$$

by which we mean that $h \circ Q_c = \sigma \circ h$. We say that Q_c and Σ are ***(topologically) conjugate***.

6.5 Conjugacy

Proposition 6.3 *For every $x \in \Lambda$, the equality $h \circ Q_c(x) = \sigma \circ h(x)$ holds.*

Proof For any $x \in \Lambda$, $h(x) = (s_0, s_1, s_2, ...)$ iff $x \in I_{s_0}$ and $Q_c(x) \in I_{s_1}$,..., $Q_c^n(x) \in I_{s_n}$,...

Hence $h(Q_c(x)) = (s_1, s_2, s_3, ...) = \sigma(h(x))$. □

From Proposition (6.3), it also follows that

$$h \circ Q_c^n = \sigma \circ h \circ Q_c^{n-1} = ... = \sigma^n \circ h. \tag{6.13}$$

This can be pictured in the diagram

[1] A ***homeomorphism*** is a continuous invertible map with a continuous inverse.

$$\begin{array}{ccccccccc} \Lambda & \xrightarrow{Q_c} & \Lambda & \xrightarrow{Q_c} & \cdots & \xrightarrow{Q_c} & \Lambda & \xrightarrow{Q_c} & \Lambda \\ {\scriptstyle h}\downarrow & & & & & & & & \downarrow{\scriptstyle h} \\ \Sigma & \xrightarrow{\sigma} & \Sigma & \xrightarrow{\sigma} & \cdots & \xrightarrow{\sigma} & \Sigma & \xrightarrow{\sigma} & \Sigma \end{array}, \tag{6.14}$$

Proposition 6.4 *If s is a periodic point of σ, then $h^{-1}(s)$ is a periodic point of Q_c with the same period.*

Proof Let s be a period-n point of σ. Then $\sigma^n(s) = s$ and $\sigma^k(s) \neq s$ for $k = 1, \ldots, n-1$.

Because of (6.13), for any positive integer m,

$$\sigma^m(s) = \sigma^m \circ h(h^{-1}(s)) = h(Q_c^m(h^{-1}(s))). \tag{6.15}$$

We conclude that $Q_c^n(h^{-1}(s)) = h^{-1}(s)$ and $Q_c^k(h^{-1}(s)) \neq h^{-1}(s)$ for $k = 1, \ldots, n-1$. □

Not only periodicity is preserved by conjugation. For any $c < -2$, the coding map h and its inverse h^{-1} are continuous, so that "closeness" is preserved under it. This implies, for instance, that $\sigma^n(s) \to s_0$ if and only if $Q^n(h^{-1}(s)) \to h^{-1}(s_0)$, so the property of being an attracting fixed point is preserved under coding.

To keep things simple, we restrict ourselves to the case $c < -(5+2\sqrt{5})/4$; in this case, $|Q_c'(x)| > 1$ for all $x \in I \setminus A_1$.

Proposition 6.5 *If $c < -(5+2\sqrt{5})/4$, then $h : \Lambda \to \Sigma$ is an homeomorphism.*

Proof Recall that the coding map $h : \Lambda \to \Sigma$, $x \mapsto h(x) = (s_0, \ldots, s_n, \ldots)$ is such that $s_n = a$ if and only if $Q_c^n(x) \in I_a$ (for $a = 0, 1$).

We have to prove that

1. h is invertible,
2. h is continuous, and
3. the inverse h^{-1} is continuous. □

To prove that h is invertible, let us establish that the inverse image $h^{-1}(s)$ of an arbitrary word $s = (s_0, s_1, \ldots) \in \Sigma$ is a singleton. Define

$$I_{s_0, s_1, \ldots, s_n} = \{\, x \in I \mid x \in I_{s_0}, \ldots, Q_c^n(x) \in I_{s_n} \,\} \tag{6.16}$$

$$= I_{s_0} \cap Q_c^{-1}(I_{s_1}) \cap \cdots \cap Q_c^{-n}(I_{s_n}), \tag{6.17}$$

where $Q_c^{-n}(J) = \{\, x \in I \mid Q_c^n(x) \in J \,\}$. One has

$$Q_c^{-1}(A \cap B) = Q_c^{-1}(A) \cap Q_c^{-1}(B) \tag{6.18}$$

thus

$$I_{s_0,s_1,\ldots,s_n} = I_{s_0} \cap Q_c^{-1}(I_{s_1} \cap Q_c^{-1}(I_{s_1}) \cap \ldots \cap Q_c^{-(n-1)}(I_{s_n})) \tag{6.19}$$
$$= I_{s_0} \cap Q_c^{-1}(I_{s_1,\ldots,s_n}). \tag{6.20}$$

We prove by induction in n that $I_{s_0,\ldots,s_n}$ is a closed interval contained in $I_{s_0,\ldots,s_{n-1}}$ and that

$$|I_{s_0,\ldots,s_n}| < \eta^{-n} K, \tag{6.21}$$

where η is a lower-bound for $|Q_c'(x)|$ on $I_0 \cup I_1$, and $K = \max(|I_0|, |I_1|)$.

Let us consider the case $n = 1$. Remark that I_{s_0} equals I_0 or I_1, whereas $Q_c^{-1}(I_1)$ consists in the union of two closed intervals, one in I_0. Hence $I_{S_0,s_1} = I_{s_0} \cap Q_c^{-1}(I_{s_1})$ is a closed interval too, contained in I_{s_0}. Let us denote this interval by $[a, b]$.

The function Q_c is strictly decreasing on I_0 and strictly increasing on I_1. So if $s_0 = 0$, $Q_c(a)$ and $Q_c(b)$ are the endpoints of I_{s_1}. By the MVT, for some y between a and b,

$$\eta < |Q_c'(y)| = \frac{|Q_c(b) - Q_c(a)|}{|b - a|}, \tag{6.22}$$

therefore

$$|I_0 \cap Q_c^{-1}(I_{s_1})| < \eta^{-1}|I_{s_1}| < \eta^{-1}K. \tag{6.23}$$

The case $s_1 = 1$ follows similarly.

We leave the general case as an exercise.

We thus have that

$$h^{-1}(s) = \bigcap_{n\in\mathbb{N}} I_{s_0,\ldots,s_n} = \bigcap_{n\in\mathbb{N}} [l_n, r_n], \tag{6.24}$$

The intersection is also a compact interval, $[r, l]$, where $r = \lim r_n$ and $l = \lim l_n$. This compact interval must be a point (i.e. $r = l$) because of (6.21).

To prove that h is continuous at $x_0 \in \Lambda$, we have to show that for every $\epsilon > 0$, there exists a $\delta > 0$ such that $|x - x_0| < \delta$, then $d(h(x), h(x_0)) < \epsilon$. Remark that x and x_0 are both in Λ (the function h is defined on Λ).

The closed intervals in $C_n = \{ I_{s_0 \cdots s_n} \mid (s_0, \ldots, s_n) \in \{0, 1\}^n \}$ are pairwise disjoint[2] and their diameters go to zero. Hence any two intervals in C_n are at a positive distance of each other[3] (otherwise, they would intersect at a common point), so the infimum A_n of all those pairwise distances is also strictly greater than zero (because C_n is finite; in other words, the infimum is just a minimum).

Consider an arbitrary $\epsilon > 0$ and $x_0 \in X$. Denote by $s^0 = (s_0^0, s_1^0, s_2^0, \cdots)$ the image $h(x_0)$ of x_0. Choose N such that $2^{-N} < \epsilon$.

We pick $\delta = A_N/2$. In this way, the condition $|x - x_0| < \delta$ implies that x also belongs to $I_{s_0^0 \cdots s_N^0}$. For both x and x_0 belong to $I_{s_0^0 \cdots s_N^0}$, the images $h(x)$ and $h(x_0)$ coincide in the first $N+1$ entries. By the Proximity Theorem, $d(h(x), h(x_0)) \leq 2^{-N}$.

---

[2] This means that for every $I, I' \in C$, we have $I \cap I' = \emptyset$

[3] The distance between to subsets A and B of a metric space (X, d) is $\inf\{d(a, b)\, a \in A, b \in B\}$. If $X = \mathbb{R}$, equipped with the usual distance $d(x, y) = |y - x|$, then the distance between $[a, b]$ and another closed interval $[c, d]$ to the right of it is $c - b$.

Hence we have that $|x - x_0| < \delta = A_N/2$ implies $d(h(x), h(x_0)) \leq 2^{-N} < \epsilon$. Thus h is continuous.

Chapter 7
Chaos

In the literature, there are many nonequivalent definitions of chaos, but we are going to work with one introduced by Devaney. For him, a chaotic dynamical system comprises three essential features:

1. unpredictability,
2. undecomposability,
3. regularity.

The unpredictability refers to the sensitive dependence on initial conditions that we illustrated in the Introduction; it is formally defined in Section 7.3. The undecomposability refers to the fact that we cannot divide the system into two independent subsystems, because if we start in the neighborhood of any point we eventually visit any neighborhood of any other point; that corresponds to the notion of *transitivity* in Section 7.2. Finally, the regularity will be given by periodic points that *dense*, that is, arbitrarily close to any point in the domain.

7.1 Density

The definition of chaos depends on the topological concept of ***density***, which is the subject of this section.

Definition 7.1 Let (X, d) be a metric space and $Y \subset X$. We say that Y is ***dense*** in X if for all $x \in X$ and all $\epsilon > 0$, there exists a $y \in Y$ such that $d(x, y) < \epsilon$.

So we say that Y is dense in X if it is always possible to find a point $y \in Y$ arbitrarily close to any $x \in X$.

Example 7.1 The rationals $\mathbb{Q}$ are dense in the reals $\mathbb{R}$. If $x \in \mathbb{R}$ has decimal expansion $a_0 \cdots a_n.b_1 \cdots b_n \cdots$, the truncated number $y = a_0 \cdots a_n.b_1 \cdots b_n$, which is rational, is such that $|x - y| \leq 10^{-n}$. As $n \to \infty$, this becomes arbitrarily small.

Remark 7.1 The rationals are countable, let $(q_n)_{n\in\mathbb{N}}$ be an enumeration. For any $\epsilon \in (0,1)$, we evidently have $\mathbb{Q} \subset \bigcup_{n\in\mathbb{N}}(q_n - \epsilon^{n+1}, q_n + \epsilon^n)$, and the length of this union of intervals is bounded by $\sum_{n\in\mathbb{N}} \epsilon^n = \frac{\epsilon}{1-\epsilon}$, which converges to 0 as $\epsilon \to 0$. In other words, the rationals are contained in sets of arbitrarily small length, so they must have vanishing length (or "measure zero").[1]

Proposition 7.1 *Periodic points are dense in* Σ.

Proposition 7.2 *There is a point whose orbit is dense in* Σ.

Proof Consider the word $\hat{s}$ formed by the juxtaposition of all binary blocks of length 1, followed by all binary words of length 2, then those of length 3, and so on. That is

$$\hat{s} = (\underbrace{0,\ 1,}_{\text{length 1}}\ \underbrace{0,0,\ 0,1,\ 1,0,\ 1,1,}_{\text{length 2}}\ldots) \tag{7.2}$$

Let $s = (s_0, s_1, \ldots) \in \Sigma$ and $\epsilon > 0$. Choose n such that $2^{-n} < \epsilon$. There exists a block $(s_1, \ldots, s_n)$ in $\hat{s}$, starting at the index $k \in \mathbb{N}$. Then $\sigma^k(\hat{s})$ and s coincide in the first $n+1$ components, so by the Proximity theorem, $d(\sigma^k(\hat{s}), s) \leq 2^{-n} < \epsilon$. ☐

Remark 7.2 How big is k? Suppose $n = 101$. In the worst case (the block $(s_0, \ldots, s_n)$ does not appearing as concatenation of smaller blocks), we have that

$$k \geq \sum_{i=1}^{100} i2^i \geq 2^{102} \approx 5.07 \times 10^{30}, \tag{7.3}$$

which has the same order of magnitude as the mass of the sun in kilograms.

Remark 7.3 $\hat{s}$ is not periodic.

7.2 Transitivity

Definition 7.2 Let (X, d) be a metric space. A dynamical system $F : X \to X$ is ***transitive*** if for any $x, y \in X$ and $\epsilon > 0$, there is a third point z and $k \in \mathbb{N}$ such that $d(x, z) < \epsilon$ and $d(F^k(z), y)) < \epsilon$, i.e. z is within ϵ of x and its orbit comes within ϵ of y.

An ***open ball*** centered at $x \in X$ of radious $r > 0$ is the set $B(x; r) = \{\, y \in X \mid d(y, x) < \epsilon \,\}$. The condition in Definition 7.2 can be expressed as $F^k(B(x;\epsilon)) \cap B(y;\epsilon) \neq \emptyset$.

[1] Formally, the Lebesgue measure of a set A is

$$\lambda(A) = \inf\{\, \sum_{i=0}^{\infty} |b_i - a_i| \mid \{(a_i, b_i)\}_{i\in\mathbb{N}} \text{ covers } A \,\}. \tag{7.1}$$

Proof Let $z' \in F^k(B(x;\epsilon)) \cap B(y;\epsilon)$. Then $d(z', y) < \epsilon$ and $z' = F^k(z)$ for some $z \in B(x; epsilon)$. □

As a preparation for the next proof, we prove the following lemma.

Lemma 7.1 *Let (X, d) be a metric space, $x \in X$ an arbitrary point, and $\epsilon > 0$. For any $y \in B(x;\epsilon)$ there exists an ϵ' such that $B(y;\epsilon') \subset B(x;\epsilon)$.*

Proof Because $\epsilon - d(x, y) > 0$, we can choose an ϵ' such that $0 < \epsilon' < \epsilon - d(x, y)$.

If z is any point in $B(y, \epsilon')$ then, by definition, $d(y, z) < \epsilon'$. In turn, because of the triangular inequality,

$$d(x, z) \leq d(x, y) + d(y, z) < \epsilon, \tag{7.4}$$

so $z \in B(x;\epsilon)$. This shows that $B(y;\epsilon') \subset B(x, \epsilon)$. □

Proposition 7.3 *Let (X, d) be a metric space and $F : X \to X$ a function. Suppose that every open ball in X has an infinite number of points. If there is a point $\hat{x}$ whose orbit is dense in X, then $F : X \to X$ is transitive.*

Proof Let $\hat{x}$ be the point with dense orbit. Then, for any $x, y \in X$ and $\epsilon > 0$, there are $m, n \in \mathbb{N}$, $F^m(\hat{x}) \in B(x;\epsilon)$ and $F^n(\hat{x}) \in B(y,\epsilon)$. If $n \geq m$ we are done: we can take $z = F^m(\hat{x}) \in B(x;\epsilon)$ and $k = n - m$, because $F^k(z) = F^n(\hat{x}) \in B(y;\epsilon)$. But if $n < m$, we still need to prove that there is an iterate of $F^m(x)$ that belongs to $V = B(y;\epsilon)$.

Let $n_1, ..., n_s$ be the indexes in $[0, m]$ such that $F^{n_i}(\hat{x}) \in V$. We remove from V the subset $W = \{F^{n_1}(\hat{x}), ..., F^{n_s}(\hat{x})\}$. Because V is infinite, $V \setminus W$ must be nonempty. A point $w \in V \setminus W$ is at distance D_i of $F^{n_i}(\hat{x})$, for $i = 1, ..., s$. Set $D = \min_i D_i$ and $\epsilon' < D$ small enough, such that $B(w, \epsilon') \subset V$ (using Lemma 7.1). By construction $B(w, \epsilon') \cap W = 0$, but being an open ball$B(w, \epsilon')$ must contain some iterate of $\hat{x}$, $F^N(\hat{x})$, and N is necessarily greater or equal than m. □

A more general condition is introduced in [2]. The same reference proves that, under fairly general assumptions, the converse is also true.

Corollary 7.1 *The shift is transitive.*

7.3 Sensitive dependence

Definition 7.3 Let (X, d) be a metric space. A dynamical system $F : X \to X$ ***depends sensitively on initial conditions*** if there exists $\beta > 0$ such that for all $x \in X$ and $\epsilon > 0$, there is $y \in B(x, \epsilon)$ and $k \in \mathbb{N}$ such that $d(F^k(x), F^k(y)) \geq \beta$.

In other words, given any point $x \in X$, one can find an arbitrarily close point $y \in X$ such that the iterates $F^k(x)$ and $F^k(y)$ are at a prescribed distance β (irrespective of how close x and y where).

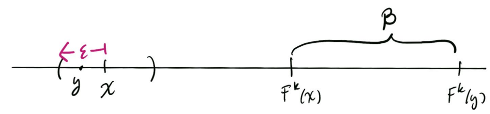

Fig. 7.1 Situation in Definition 7.3

Proposition 7.4 *The shift exhibits sensitive dependence on initial conditions.*

Proof Take $\beta = 1$. For any $s = (s_i)_{i\in\mathbb{N}} \in \Sigma$ and $\epsilon > 2^{-N} > 0$, the point $t = (s_0, ..., s_n, 1 - s_{n+1}, s_{n+2}, s_{n+3}, ...)$ belongs to $B(s, \epsilon)$ i.e. $d(s, t) \leq 2^{-N} < \epsilon$, in virtue of the Proximity Theorem. But it is also clear that

$$|\sigma^{N+1}(s) - \sigma^{N+1}(t)| = 1 = \beta. \tag{7.5}$$

7.4 Chaos

Definition 7.4 (Devaney's chaos) A dynamical system $F : X \to X$ is ***chaotic*** if

1. Periodic orbits of F are dense in X.
2. F is transitive.
3. F depends sensitively on initial conditions.

It is possible to prove that the first and second properties in this definition, taken together, imply the third one [**?**].

Theorem 7.1 *The shift map $\sigma : \Sigma \to \Sigma$ is a chaotic dynamical system.*

Fact: chaos is preserved under topological conjugation.

Corollary 7.2 *The quadratic map $Q_c : \Lambda \to \Lambda$ is chaotic when $c < -(5 + 2\sqrt{5})/4$.*

Proof We proved in the previous chapter that there is a topological conjugation between the shift map $\sigma : \Sigma \to \Sigma$ and $Q_c : \Lambda \to \Lambda$. And the shift map is chaotic. □

Part II
Two-dimensional systems

Chapter 8
Stability

In the first part, we analyzed, among other things, iterations of functions $F : X \to X$ where X is a subset of the real line. When F was continuously differentiable, we were able to characterize the behavior of fixed points, periodic orbits, etc. in terms of the derivative F'. We shall see now that we can do something similar when $X \subset \mathbb{R}^2$ (and more generally of $\mathbb{R}^n$); the Jacobian matrix takes the role of F'.

8.1 Some basic definitions

We work with the vector space $\mathbb{R}^n$ of n-tuples $(x_1, ..., x_n)$ of real numbers. This vector space, when equipped with the ***norm*** $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$, is called ***Euclidean space***. Remark that the norm satisfies the properties:

1. $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$,
2. $\|\lambda\mathbf{x}\| = |\lambda| \, \|\mathbf{x}\|$ for any $\mathbf{x} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$,
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

From these properties, it is easy to prove that $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ defines a metric on $\mathbb{R}^n$. In particular, we can formalize the notion of "closeness" through open balls ϵ-balls $B(\mathbf{x}; \epsilon)$. In more sophisticated terms: the metric defines a topology.

Given a function $F : \mathbb{R}^n \to \mathbb{R}^n$, a fixed points $\mathbf{p}$ of F is an element $\mathbf{p} \in \mathbb{R}^n$ such that $F(\mathbf{p}) = \mathbf{p}$.

Definition 8.1 Let $\mathbf{p}$ be a fixed point of $F : \mathbb{R}^n \to \mathbb{R}^n$.

1. If there exists $\epsilon > 0$ such that for all $\mathbf{x} \in B(\mathbf{p}; \epsilon) \setminus \{p\}$,

$$\lim_{k\to\infty} F^k(\mathbf{x}) \to \mathbf{p} = F(\mathbf{p}),$$

   then $\mathbf{p}$ is a ***sink*** or ***attractor***.
2. If there exists $\epsilon > 0$ such that for all $\mathbf{x} \in B(\mathbf{p}; \epsilon) \setminus \{p\}$, there exists a $k = k(v) \in \mathbb{N}$ such that $f^k(\mathbf{x}) \notin B(\mathbf{p}; \epsilon)$, then $\mathbf{p}$ is a ***source*** or ***repeller***.

In the case $n = 1$ we essentially recover the definitions of attracting and repelling fixed points that we introduced in Chapter 2.

8.2 Linear maps

The simplest case that we can consider are linear maps.

Definition 8.2 A map $F : \mathbb{R}^n \to \mathbb{R}^n$ is linear if for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$,

$$F(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha F(\mathbf{x}) + \beta F(\mathbf{y}).$$

We denote by $\mathbf{e}_i$ the vector that has a 1 as i^{th}-component and 0 elsewhere. The ordered list $(\mathbf{e}_1, ..., \mathbf{e}_n)$ forms a basis of $\mathbb{R}^n$, known as the canonical basis.

If we know that $F(\mathbf{e}_i) = (a_{1,1}, a_{2,1}, ..., a_{n,1})$, then linearity implies that

$$F\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 F(\mathbf{e}_1) + \cdots + x_n F(\mathbf{e}_n) \tag{8.1}$$

$$= \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \tag{8.2}$$

So the linear map F can be represented as multiplication by a matrix.

More generally, a basis of $\mathbb{R}^n$ is an ordered list $(\mathbf{b}_1, ..., \mathbf{b}_n)$ of linearly independent vectors. The images $F(\mathbf{b}_i)$ completely determine the linear map F, because any $\mathbf{x} \in \mathbb{R}^n$ can be expressed as a linear combination $x'_1\mathbf{b}_1 + \cdots x'_n\mathbf{b}_n$ in a unique way, and

$$F(\mathbf{x}) = \sum_{i=1}^{n} x'_i F(\mathbf{b}_i) = \big(F(\mathbf{b}_1) \; \cdots \; F(\mathbf{b}_n)\big) \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}, \tag{8.3}$$

where $\big(F(\mathbf{b}_1) \cdots F(\mathbf{b}_n)\big)$ is a matrix that has the $F(\mathbf{b}_i)$ as columns.

Definition 8.3 We say that λ is an ***eigenvalue*** of the matrix A if there exists a vector $\mathbf{v} \neq 0$ such that $A\mathbf{v} = \lambda v$. In this case $\mathbf{v}$ is called an ***eigenvector***.

If eigenvectors $(\mathbf{v}_1, ..., \mathbf{v}_n)$ with associated eigenvalues $(\lambda_1, ..., \lambda_n)$ form a basis of $\mathbb{R}^n$, we say that A is ***diagonalizable***. In this case, we can express any $\mathbf{x}$, uniquely, as a linear combination of eigenvectors, $\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{v}_i$, and then

$$F(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i F(\mathbf{v}_i) = \sum_{i=1}^{n} \alpha_i \lambda_i \mathbf{v}_i. \tag{8.4}$$

In other words, F is represented by the diagonal matrix $D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$.

Example 8.1 Let

$$A = \begin{pmatrix} 5 & 1 \\ -12 & -2 \end{pmatrix}. \tag{8.5}$$

1. Find the eigenvalues of A. What are the corresponding eigenvectors?
2. Write $\mathbf{x}_0 = (-2, 3)$ as a linear combination of eigenvectors.
3. Compute $F^n(\mathbf{x}_0)$ as a linear combination of the same eigenvectors.

Solution 8.1 The characteristic polynomial of A is

$$p(\lambda) = \det(A - \lambda I) = (5 - \lambda)(-2 - \lambda) + 12 \tag{8.6}$$

$$= \lambda^2 - 3\lambda + 2 = (\lambda - 2)(\lambda - 1). \tag{8.7}$$

The eigenvalues are its roots: 2 and 1. To find an eigenvector corresponding to 2, we determine

$$\ker(A - 2I) = \ker \begin{pmatrix} 3 & 1 \\ -12 & -4 \end{pmatrix}.$$

So an eigenvector is $\mathbf{v}_1 = (-1, 3)$. Remark that any scalar multiple of it is also an eigenvector.

Similarly, because

$$\ker(A - I) = \ker \begin{pmatrix} 4 & 1 \\ -12 & -3 \end{pmatrix} = \ker \begin{pmatrix} 4 & 1 \\ 0 & 0 \end{pmatrix},$$

$\mathbf{v}_2 = (-1, 4)$ is an eigenvector associated to 1.

How do we write $x = (x_1, x_2)$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$? We have to solve the system $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 = \mathbf{x}$, which matricially corresponds to

$$\begin{pmatrix} -1 & 1 \\ 3 & -4 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{8.8}$$

We can interpret this as follows: the matrix $G = \begin{pmatrix} -1 & 1 \\ 3 & -4 \end{pmatrix}$ converts a vector in the basis $B' = (\mathbf{v}_1, \mathbf{v}_2)$ into a vector in the canonical basis i.e. the usual coordinates/components of $\mathbb{R}^n$. The inverse G^{-1} turns the vector (x_1, x_2) in the canonical basis into the vector (α_1, α_2) of coordinates in the B' basis.

Because $G^{-1} = \begin{pmatrix} -4 & -1 \\ -3 & -1 \end{pmatrix}$, we conclude that for $x = (-2, 3)$ one has $(\alpha_1, \alpha_2) = (5, 3)$.

Once we have expressed $\mathbf{x} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2$, remark that $F(\mathbf{x}) = \alpha_1 F(\mathbf{v}_1) + \alpha_2 F(\mathbf{v}_2) = \alpha_1 2\mathbf{v}_1 + \alpha_2 1\mathbf{v}_2$, and more generally,

$$F^n(\mathbf{x}) = \alpha_1 2^n \mathbf{v}_1 + \alpha_2 \mathbf{v}_2. \tag{8.9}$$

When a linear map is diagonalizable, the analysis of the dynamics is relatively simple: if an eigenvalue $|\lambda_i| > 1$, the corresponding eigenvector $\mathbf{v}_i$ defines an *expanding direction*; similarly, if $|\lambda_i| < 1$, the corresponding eigenvector defines a *contracting direction*. When the eigenvalue is negative the expansion or contraction is accompanied by a reflection with respect to the plane with normal vector $\mathbf{v}_i$ that passes through the origin.

Proposition 8.1 *Let A be a diagonalizable matrix of size $n \times n$. Remark that 0 is the only point fixed by $F(\mathbf{x}) = A\mathbf{x}$. Let $(\lambda_1, ..., \lambda_n)$ be the eigenvectors of A. If $|\lambda_i| > 1$ for all $i = 1, ..., n$, then 0 is a source. If $|\lambda_i| < 1$ for all $i = 1, ..., n$, then 0 is a sink.*

We see that even if all the eigenvalues have modulus different from 1 (this condition is called ***hyperbolicity***), there are more possibilities. In general, if $(\mathbf{v}_1, ..., \mathbf{v}_n)$ are the eigenvalues associated with $(\lambda_1, ..., \lambda_n)$, we can define

$$E_s = \text{span}\{\, v_i \mid |\lambda_i| < 1 \,\} \quad \text{and} \quad E_u = \text{span}\{\, v_i \mid |\lambda_i| > 1 \,\}. \tag{8.10}$$

The space $\mathbb{R}^n$ decomposes as a direct sum $E_s \oplus E_u$, which means that any vector $\mathbf{x}$ can be written in a unique way as a sum $\mathbf{x}_s + \mathbf{s}_u$ with $x_s \in E_s$ and $x_u \in E_u$. It is easy to see that $F^{-n}(\mathbf{x}_u) \to 0$ and $F^n(\mathbf{x}_s) \to 0$.

Example 8.2 A symmetric matrix has real eigenvalues and its eigenspaces are orthogonal.

8.3 Standard forms

Not all linear maps are diagonalizable, although we'll mainly focus on those. We focus here on the case $n = 2$.

Proposition 8.2 *Let A be a 2×2 matrix. There exists a real invertible matrix G such that $G^{-1}AG$ assumes one of the following forms:*

1. $C = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$ *for some $\beta \neq 0$,*
2. $D = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$,
3. $E = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$.

Remark that $\det(A - \lambda I)$ is in this case a polynomial of degree 2. The three cases above correspond, respectively, to two conjugate roots $\alpha \pm i\beta$, a repeated real root, or two different real roots λ and μ.

Since we have $A = GEG^{-1}$ (resp. C or D instead of E), we can see G as a change-of-basis matrix (into a basis made by the columns on G), in which the matrix

A has a simpler representation E (resp. C or D). We saw the usefulness of this in Example 8.1.

Proof

Exercise 8.1 Describe the dynamics given by

1.
$$L_1(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & -1/2 \end{pmatrix} \mathbf{x}.$$

2.
$$L_2(\mathbf{x}) = \begin{pmatrix} 0 & -1/2 \\ 1/2 & 0 \end{pmatrix} \mathbf{x}.$$

8.4 Stable Manifold Theorem

Recall that a *curve* γ (a.k.a. 1-dimensional manifold) in $\mathbb{R}^2$ is a subset of points that "locally" looks like $\mathbb{R}^1$. More formally, this means that every $\mathbf{x} \in \gamma$ has a neighborhood U such that there is a diffeomorphism $f : (-1, 1) \to U \cap \gamma$.

The following example illustrates the concept of stable and unstable curves (manifolds) [1, Ex. 2.21].

Example 8.3 Let $F(x, y) = (x/2, 2y - 7x^2)$. The point $\mathbf{p} = (0, 0)$ is fixed. Locally, the map f behaves like

$$Df(0,0) = \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix}, \tag{8.11}$$

which implies that $\mathbf{p}$ is a saddle point. In this nonlinear case, there are stable and unstable *curves* that pass through $\mathbf{p}$. The eigenvector $\mathbf{v}_{0.5}$ is tangent to the stable curve at $\mathbf{p}$, whereas $\mathbf{v}_2$ is tangent to the unstable curve at $\mathbf{p}$.

Exercise 8.2 1. Find F^{-1}.
2. Show that $S = \{\, x, 4x^2) \mid x \in \mathbb{R} \,\}$ is invariant under F, that is if $\mathbf{v} \in S$, then $F(\mathbf{v})$ and $F^{-1}(\mathbf{v})$ too.
3. Show that each point in S converges to 0 under F.
4. Show that no point outside F goes to zero under F.

The set S above is a parametric curve. It is a *stable* manifold of the fixed point $\mathbf{p}$, because it is invariant under F and $F^n(\mathbf{x}) \to \mathbf{p}$ for all $\mathbf{x} \in S$. Similarly, there is an unstable curve $U = \{\, (0, 2y) \mid y \in \mathbb{R} \,\}$ which is invariant under F and such that $F^{-n}(\mathbf{x}) \to \mathbf{p}$ for all $\mathbf{x} \in U$.

Although for a general map one cannot find explicit parametrizations of the stable and unstable curves, it is possible to prove that such curves exist and are well-defined.

Theorem 8.1 (Stable manifold theorem) *Suppose $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a diffeomorphism with a sadddle point at $\mathbf{p}$. Then there exists $\epsilon > 0$ and a smooth curve $\gamma_s : (-1, 1) \to \mathbb{R}^2$ (resp. γ_u) such that (the $\bullet$ stands for s and u):*

1. $\gamma_\bullet(0) = \mathbf{p}$*,*
2. $\gamma'_\bullet(0) \neq 0$ *for all* $t \in (-1, 1)$*,*
3. $\gamma'_s(0)$ *is a stable (resp.* γ'_u *is an unstable) eigenvector of* $DF(\mathbf{p})$*,*
4. $\gamma_\bullet$ *is* F *invariant,*
5. for all t*,* $F^n(\gamma_s(t)) \to \mathbf{p}$ *(resp.* $F^{-n}(\gamma_u(t)) \to \mathbf{p}$*) as* $n \to \infty$ *.*
6. If $|F^n(q) - p| < \epsilon$ *(resp.* $|F^{-n}(q) - p| < \epsilon$*) for all* $n \geq 0$*, then* $q = \gamma_s(t)$ *(resp.* $q = \gamma_u(t)$*) for some* t

Proof

Definition 8.4 The ***unstable manifold*** at $\mathbf{p}$, denoted $W^u(\mathbf{p})$, is given by

$$W^u(\mathbf{p}) = \bigcup_{n>0} F^n(\operatorname{im} \gamma_u). \tag{8.12}$$

Similarly, the ***stable manifold*** $W^s(\mathbf{p})$ at $\mathbf{p}$ is given by

$$W^s(\mathbf{p}) = \bigcup_{n>0} F^{-n}(\operatorname{im} \gamma_s). \tag{8.13}$$

Remark that both $W^u(\mathbf{p})$ and $W^s(\mathbf{p})$ contain infinitely many orbits.

The stable and unstable manifolds do not have self-intersections, being always locally isomorphic to a line segment. However, the stable and unstable manifolds can intersect each other. A point in the intersection of $W^s(\mathbf{p})$ and $W^u(\mathbf{p})$ is called a ***homoclinic point***.

Because the stable and unstable manifold are invariant under F, if $\mathbf{x}$ is homoclinic, then also $F^k(\mathbf{x})$ is homoclinic for any $k \in \mathbb{Z}$, which means that there is, in generalm an *infinite number of intersections* of $W^s(\mathbf{p})$ and $W^u(\mathbf{p})$.

Smale showed that, in this case, there is an *hyperbolic horseshoe*. Topologically, one cna introduce a rectangular neighborhood R os $\mathbf{p}$. Then $F^{-l}(R)$ is an elongated region in the stable direction, whereas $F^k(R)$ is an elongated region in the unstable direction. When there is an homoclinic point both regions must intersect for some value of k and l. In the next section we study an idealized version of the map F^{k+l} from $S = F^{-l}(R)$ to $F^{-l}(R)$.

Chapter 9
Chaos again

9.1 Smale's horseshoe

We follow [4, pp. 368ss].

We start with a "stadium" D made of a square S with semicircles D_1 and D_2 attached above and below.

We consider a map F that first linearly contracts S in the horizontal direction by a factor of δ and linearly expands it in the vertical direction by a factor of $1/\delta$, then "curling" it to form a horseshoe:

The map F is one-to-one but not onto. Outside S the map F is not linear, but we observe that $F(D_1) \cup F(D_2) \subset D_1$.

The image of F is S consists of two vertical strips V_0 (left) and V_1 (right). The set $F^{-1}(V_L \cup V_R)$ consists of two horizontal strips of height δ, a strip H_0 (below) that is mapped to V_0, and a strip H_1 (above) that is mapped to V_1. The linearity is such that horizontal (resp. vertical) lines in H_0 and H_1 are mapped to horizontal (resp. vertical) in V_0 and V_1.

We also assume that there exists a unique $d \in D_1$ such that for all $y \in D_1 \cup D_2$, $F^n(y) \to d$. In this case, if for any $\mathbf{x} \in S$ there exists an iterate $F^k(\mathbf{x})$ that belongs to $D_1 \cup D_2$, for $k \in \mathbb{N}$, then $F^n(\mathbf{x}) \to d$ and $n \to \infty$.

The interesting points are therefore those that never leave S:

$$\Lambda_+ = \{\, \mathbf{x} \in S \mid F^n(\mathbf{x}) \in S \text{ for } n \in \mathbb{N} \,\}. \tag{9.1}$$

The reader should be suspecting that this set shares many properties with the space Σ of sequences that we have studied in previous chapters.

The points in Λ_* must belong to $H_0 \cup H_1$ (to have $F(x) \in S$). But since $F^2(x) \in S$ too, we need $F(F(x)) \in V_L \cup V_R \subset S$, so $F(x) \in H_0 \cup H_1$ too i.e. $x \in F^{-1}(H_0 \cup H_1)$. The preimage $F^{-1}(H_0)$ is made of two horizontal lines,one contained in H_0 and other contained in H_1, and mapped respectively to $H_0 \cap V_0$ and the other to $H_0 \cap V_1$. Reasoning similarly for the other iterates, we conclude that

$$\Lambda_+ = \bigcap_{n\geq 0} F^{-n}(H_0 \cup H_1) \tag{9.2}$$

By recursion, one proves that $F^{-n}(H_0 \cup H_1)$ consists of 2^{n+1} horizontal strips of height δ^{n+1}, contained in $F^{-(n-1)}(H_0 \cup H_1)$ (for $n \geq 1$).

We conclude that the set Λ_+ is of the form (line segment)×(Cantor set). In fact, we can introduce a coding map

$$h_+ : \Lambda_+ \longrightarrow \Sigma, \quad \mathbf{x} \longrightarrow (s_i)_{i\geq 0}$$

such that

$$s_i(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in F^{-i}(H_0) \\ 1 & \text{if } \mathbf{x} \in F^{-i}(H_1) \end{cases} \tag{9.3}$$

but this coding does not specify a unique point, but rather a whole horizontal segment (of points that are necessarily mapped to the same thing as $n \to \infty$). The map h_+ prescribes the location H_0 or H_1 of $F^n(\mathbf{x})$ for any point $\mathbf{x} \in S$.

Although F is not invertible, we can attempt something similar to describe a *backward orbit* $\{F^{-n}(x)\}_{n\geq 1}$. A point in S has a well-defined backward orbit only if it lies in the image of F^n for every n; we call the set of such points Λ_-. Hence have

$$\Lambda_- = \bigcap_{n\geq 1} F^{n}(H_0 \cup H_1). \tag{9.4}$$

Here $F(H_0 \cup H_1)$ is made of two vertical strips $V_0 \cup V_1$ of width δ; $F^2(H_0 \cup H_1) = F(V_0 \cup V_1)$ consists of four vertical strips of width δ^2 and, in general $F^n(H_0 \cup H_1)$ consists of 2^n vertical strips of with δ^n. Hence Λ_+ is of the form (Cantor set)×(line segment).

We can introduce a coding

$$h_- : \Lambda_- \longrightarrow \Sigma, \quad \mathbf{x} \mapsto (\cdots, s_{-3}, s_{-2}, s_{-1}),$$

where

$$s_{-i}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in F^{i}(H_0) \\ 1 & \text{if } \mathbf{x} \in F^{i}(H_1) \end{cases} \tag{9.5}$$

The points in $\Lambda = \Lambda_+ \cap \Lambda_-$ have an entire orbit, both forward and backward, in S. There is a coding map

$$h : \Lambda \longrightarrow \Sigma_\pm, \quad \mathbf{x} \longrightarrow (s_i(\mathbf{x}))_{i\in\mathbb{Z}} \tag{9.6}$$

with values in the set of bi-infinite sequences

$$\Sigma_\pm = \left\{ (s_i)_{i\in\mathbb{Z}} \,\middle|\, s_i \in \{0,1\} \right\}, \tag{9.7}$$

which is equipped with the metric

$$d(\mathbf{s}, \mathbf{t}) = \sum_{k \in \mathbb{Z}} \frac{|s_i - t_i|}{2^{|k|}}. \tag{9.8}$$

For this metric we also have a "proximity theorem": two words are close if they coincide on a finite block $(s_{-k}, ..., s_k)$.

The map h establishes an homeomorphism between Λ and $\Sigma_{\pm}$. Moreover, one has the conjugation $\sigma \circ h = h \circ F$. Hence one can establish the chaoticity of the horseshoe map working with $\Sigma_{\pm}$ pretty much as we did for the quadratic map.

References

1. K.T. Alligood, T.D. Sauer, and J.A. Yorke. *Chaos: An Introduction to Dynamical Systems*. Textbooks in Mathematical Sciences. Springer New York, 2012.
2. Nedim Değirmenci and Şahin Koçak. Existence of a dense orbit and topological transitivity: when are they equivalent? *Acta Mathematica Hungarica*, 99(3):185–187, 2003.
3. R.L. Devaney. *A First Course In Chaotic Dynamical Systems: Theory And Experiment*. Studies in Nonlinearity. Avalon Publishing, 1992.
4. Morris W. Hirsh, Robert L. Devaney, and Stephen Smale. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press, 2013.
5. Tien-Yien Li and James A. Yorke. Period three implies chaos. *The American Mathematical Monthly*, 82(10):985–992, 1975.
6. Oleksandr Mykolayovych Sharkovsky. Existence of cycles of continuous transformation of the real line in itself. *Ukr. Mat. Zh*, 26:61–71, 1964.