MAT733 - HS2018

Dynamical Systems and Ergodic Theory

Part I: Examples of dynamical systems

Prof. Dr. Corinna Ulcigrai

October 10, 2018

Chapter 1

Examples of dynamical systems

1.1 Introduction

Dynamical systems is an exciting and very active field in pure and applied mathematics, that involves tools and techniques from many areas such as analyses, geometry and number theory and has applications in many fields as physics, astronomy, biology, meterology, economics.

The adjective *dynamical* refers to the fact that the systems we are interested in is *evolving in time*. In applied dynamics the *systems* studied could be for example a box containing molecules of gas in physics, a species population in biology, the financial market in economics, the wind currents in metereology. In pure mathematics, a dynamical system can be obtained by iterating a function or letting evolve in time the solution of equation.

Discrete dynamical systems are systems for which the time evolves in discrete units. For example, we could record the number of individuals of a population every year and analyze the growth year by year. The time is parametrized by a discrete variable n which assumes integer values: we will denote natural numbers by $\mathbb{N}$ and integer numbers by $\mathbb{Z}$. In a *continuous* dynamical system the time variable changes continuously and it is given a real number t. We will denote real numbers by $\mathbb{R}$.

Our main examples of *discrete dynamical systems* are obtained by iterating a map. Let X be a space. For example, X could be the unit interval $[0,1]$, the unit square $[0,1]\times[0,1]$, a circle (but also the surface of a doughnut or a Cantor set). Let $f : X \to X$ be a map. We can think as f as the map which gives the time evolution of the points of X. If $x \in X$, consider the iterates $x, f(x), f(f(x)), \dots$.

Notation 1.1.1. *For $n > 0$ we denote by $f^n(x)$ the n^{th} iterate of f at x, i.e. $f \circ f \circ \cdots \circ f$, n times.*[1] *In particular, $f^1 = f$ and by convention f^0 : is the identity map, which will be denoted by Id ($Id(x) = x$ for all $x \in X$).*

We can think of $f^n(x)$ as the status of the point x at time n. We call forward *orbit*[2] the evolution of a point x.

Definition 1.1.1. *We denote by $\mathcal{O}^+_f(x)$ the forward orbit of a point $x \in X$ under iterates of the map f, i.e.*

$$\begin{aligned}\mathcal{O}^+_f(x) &:= \{x, f(x), f^2(x), \dots, f^n(x), \dots\}\\ &= \{f^n(x), \quad n \in \mathbb{N}\}.\end{aligned}$$

[1] Do not confuse this notation with the n^{th} derivative, which will be denoted by $f^{(n)}$, or by the n^{th} power, which will not be used!

[2] The name orbit cames from astronomy. The first dynamical system studied were indeed the solar system, where trajectory of a point (in this case a planet or a star) is an orbit.

This gives an example of a discrete dynamical system parametrized by $n \in \mathbb{N}$.

Example 1.1.1. *Let $X = [0,1]$ be the unit interval. Let $f : X \to X$ be the map $f(x) = 4x(1-x)$. For example*

$$\mathcal{O}_f\left(\frac{1}{3}\right) = \left\{\frac{1}{3}, \frac{4}{3}\left(1-\frac{1}{3}\right) = \frac{8}{9}, 4\cdot\frac{8}{9}\left(1-\frac{8}{9}\right) = \frac{32}{81}, \dots\right\}.$$

Example 1.1.2. *Let X be a circle of radius 1. An example of map $f : X \to X$ is the (clockwise) rotation by an angle $2\pi\alpha$, which maps each point on the circle to the point obtained by rotating clockwise by an angle $2\pi\alpha$.*

If f is invertible, we have a well defined inverse $f^{-1} : X \to X$ and we can also consider backwards iterates $f^{-1}(x), f^{-2}(x), \dots$.

Notation 1.1.2. *If f is invertible and $n < 0$, we denote by $f^n(x)$ the n^{th} iterate of f^{-1} at x, i.e. $f^{-1} \circ f^{-1} \circ \cdots \circ f^{-1}$, n times. Remark that even if f is not invertible, we will often write $f^{-1}(A)$ where $A \subset X$ to denote the set of premimges of A, i.e. the set of $x \in X$ such that $f(x) \in A$.*

Definition 1.1.2. *If f is invertible, we denote by $\mathcal{O}_f(x)$ the (full) orbit of a point $x \in X$ under forward and backward iterates of f, i.e.*

$$\begin{aligned}\mathcal{O}_f(x) &:= \{\dots, f^{-k}(x), \dots, f^{-2}(x), f^{-1}(x), x, f(x), f^2(x), \dots, f^k(x), \dots\}\\ &= \{f^k(x), \quad k \in \mathbb{Z}\}.\end{aligned}$$

In this case, we have an example of discrete dynamical system in which we are interested in both past and future and the time is indexed by $\mathbb{Z}$.

Even if the rule of evolution is deterministic, the long term behavior of the system is often "'*chaotic*"'. For example, even if two points x, y are very close, there exists a large n such that $f^n(x)$ and $f^n(y)$ are far apart. This property (which we will express formally later) is known as *sensistive dependence of initial conditions.* There are various mathematical definitions of *chaos*, but they all include sensitive dependence of initial conditions. Different branches of dynamical systems, in particular topological dynamics and ergodic theory, provide tools to quantify *how chaotic* is a systems and to predict the asymptotic behaviour. We will see that often even if one cannot predict the behaviour of each single orbit (since even if deterministic it is too complicated), one can predict the *average* behaviour.

The main objective in dynamical systems is to understand the behaviour of all (or almost all) the orbits. Orbits can be fairly complicated even if the map is quite simple. A first basic question is whether orbits are finite or infinite. Even if the index run through a infinite set (as $\mathbb{N}$ or $\mathbb{Z}$) it could happen that $\mathcal{O}_f(x)$ is finite, for example if the points in the orbit repeat each other. This is the simplest type of orbit.

Definition 1.1.3. *A point $x \in X$ is* periodic *if there exists $n \in \mathbb{N}\backslash\{0\}$, such that $f^n(x) = x$. If $n = 1$, so that we have $f(x) = x$, we say that x is a* fixed point. *More in general, if $f^n(x) = x$ we say that x is periodic of periodic n or that n is a period for x. In particular, $f^{n+j}(x) = f^j(x)$ for all $j \geq 0$.*

Example 1.1.3. *In example 1.1.1, the point $x = 3/4$ is a fixed point, since $f(3/4) = 4 \cdot 3/4(1-3/4) = 3/4$.*

Example 1.1.4. *In example 1.3, $\alpha = 1/4$, i.e. we consider the rotation by $\pi/2$, all points are periodic with period 4 and all orbits consist of four points: the initial points are the points obtained rotating it by $\pi/2$, π and $3\pi/2$.*

Definition 1.1.4. *If x is a periodic point, the* minimal period *of x is the minimum integer $n \geq 1$ such that $f^n(x) = x$.*

In particular, if n is the minimal period of x, the points $f(x), \dots, f^{n-1}(x)$ are all different than x. Be aware that in some textbook the *period* of a periodic point x means the *minimal period*.

Definition 1.1.5. *A point $x \in X$ is* preperiodic *if there exists $k, n \in N$ such that $f^{n+k}(x) = f^k(x)$. In this case $f^{n+j}(f^k(x)) = f^j(f^k(x))$ for all $j \in \mathbb{N}$.*

Exercise 1.1.1. *Show that if f is invertible every preperiodic point is periodic.*

Examples of questions that are investigated in dynamical systems are:

Q1 Are there fixed points? Are there periodic points?

Q2 Are periodic points dense?

Q3 Is there an orbit which is dense, i.e. an orbit which gets arbitrarily close to any other point in X?

Q4 Are all orbits dense?

We will answer these questions for the first examples in the next lectures. More in general, these properties are studied in *topological dynamics*[3], see Chapter 2.

If an orbit is dense, it visits every part of the space. A further natural question is how much time it spends in each part of the space. For example, let $A \subset X$ be a subset of the space. We can count the number of visits of a segment $\{x, f(x), \dots, f^n(x)\}$ of the orbit $\mathcal{O}_f^+(x)$ to the set A. If χ_A denotes the characteristic function of A, i.e.

$$\chi_A(x) = \begin{cases} 1 & \text{if} x \ \in A \\ 0 & \text{if } x \notin A \end{cases},$$

let us count the number of visits

$$Card\{0 \leq k < n, \quad \text{such that } f^k(x) \in X\} = \sum_{k=0}^{n-1} \chi_A(f^k(x))$$

and divide by n to get the *frequency* of visits in time n:

$$\frac{1}{n} \sum_{k=0}^{n-1} \chi_A(f^k(x)). \tag{1.1}$$

Intuitively, orbit $\mathcal{O}_f^+(x)$ is *equidistributed* if the frequency in (1.1) is getting closer and closer, as n increases, to the *volume* of A (or the lenght, or the area, ...)[4]. This means that the orbit asymptotically spends in each part of the space a time proportional to the volume.

Q1 Are orbits equidistributed?

This last question is a main quesition in ergodic theory[5] A priory, not even the existence of a limit of the frequency (1.1) is guaranteed. One of the main theorems that we will see in Chapter 4, the Birkhoff ergodic theorem, will show that for *almost all* points the limit exists and guarantee that if the system is enough chaotic (more precisely, *ergodic*), than the frequency converge to the expected limit. As we will see, questions related to equidistributions have many connections and applications in number theory.

[3]Topological Dynamics is a branch of dynamics that investigate the properties of *continous* maps.

[4]More in general, we will have a measure on X (lenght, area and volume are all examples of measures) which is preserved by the map f and we will ask if the frequency tends to the measure of A. See Chapter 4.

[5]Ergodic Theory is a branch of dynamics which investigate the chaotic properties of maps which preserves a *measure*.

Course Outline

The course will provide an introduction to subject of dynamical systems, from a pure-mathematical point of view. The first part of the course will be driven by examples. We will present many fundamental examples of dynamical systems, such as

- rotations of the circle;
- quadratic maps;
- the doubling map;
- the baker map;
- automorphisms of the torus;
- the Gauss map and continued fractions.

Driven by the examples, we will introduce some of the phenomena and main concepts which one is interested in studying.

In the second part of the course, we will formalize these concepts. We will then develop the mathematical background and cover the basic definitions and some fundamental theorems and results in three areas of dynamical systems:

- Topological Dynamics
- Symbolic Dynamics
- Ergodic Theory

Furthermore, towards the end of the course, after a few preliminaries about the upper half plane and hyperbolic geometry, we will also introduce two very famous continuous time dynamical systems, namely the geodesic and horocycle flow and prove their ergodicity (a classical and beautiful result which dates back to Hopf).

In the course we will encounter several mathematical definitions of properties which encode *chaotic features* of a dynamical systems. In particular, we will define *sensitive dependence on initial conditions* (the mathematical formulation of the "butterfly effect"), *chaotic* topological dynamical systems (according to Devanay), *ergodicity* (a key concept in ergodic theory which dates back to Boltzmann ergodic hypothesis), *mixing* (a stronger property of equidistribution of *sets*), ... Hence a recurring theme of this course is on *how to describe chaotic behaviour mathematically.*

During the course we will also mention some applications both to other areas of mathematics, such as number theory, and to problems as data storage and Internet search engines. We will see that the algorithm for the widely used search engine google is based on an idea which uses the theory of dynamical systems.

References

We will follow the first four Chapters of [1]. Many of the arguments are also covered in more depth in [2] (which is very elementary and assumes no background) and in [3], which is more abstract and recommended for a more theoretical approach. For the last part, on geodesic and horocycle flow, we will follow [4] (Chapter 9).

[1] M. Brin, G. Stuck *Introduction to Dynamical Systems*, Cambdrige University Press.

[2] B. Hasselblatt, A. Katok *A First Course in Dynamics*, Cambdrige University Press.

[3] M. Brin, G. Stuck *Dynamical Systems and Ergodic Theory*, London Mathematical Society.

[4] M. Einsiedler, T. Ward *Ergodic Theory with a view towards Number Theory*, Springer.

Extra: Continuous Dynamical Systems

A continous dynamical system can be given by a 1−parameter family of maps $f_t : X \to X$ where $t \in \mathbb{R}$. The main example is given by solutions of a differential equation. Let $X \subset \mathbb{R}^n$ be a space, $g : X \to \mathbb{R}^n$ a function, $x_0 \in X$ an initial condition and

$$\begin{cases} \dot{x}(t) = g(x) \\ x(0) = x_0 \end{cases} \tag{1.2}$$

be a differential equation. If the solution $x(x_0, t)$ is well defined, unique and exists for all t and all initial conditions $x_0 \in X$, if we set $f_t(x_0) := x(x_0, t)$ we have an example of a continous dynamical system. In this case, an orbit is given by the trajectory described by the solution:

Definition 1.1.6. *If $\{f_t\}_{t\in\mathbb{R}}$ is a continuous dynamical system, we denote*

$$\mathcal{O}_{f_t}(x) := \{f_t(x), \quad t \in \mathbb{R}\}.$$

More in general, a 1-parameter family $\{f_t\}_{t\in\mathbb{R}}$ is called a *flow* if f_0 is the identity map and for all $t, s \in \mathbb{R}$ we have $f_{t+s} = f_t \circ f_s$, i.e.

$$f_{t+s}(x) = f_t(f_s(x)) = f_s(f_t(x)), \quad \text{for all } x \in X.$$

Extra: Dynamical systems as actions

A more formal way to define a dynamical system is the following, using the notion of *action.*

Let X be a space and G group (as $\mathbb{Z}$ or $\mathbb{R}$ or $\mathbb{R}^d$) or a semigroup (as $\mathbb{N}$).

Definition 1.1.7. *An* action *of G on X is a map $\psi : G \times X \to X$ such that, if we write $\psi(g, x) = \psi_g(x)$ we have*

(1) If e id the identity element of G, $\psi_e : X \to X$ is the identity map;

(2) For all $g_1, g_2 \in G$ we have $\psi_{g_1} \circ \psi_{g_2} = \psi_{g_1 g_2}$.[6]

A discrete dynamical systems is then defined as an action of the group $\mathbb{Z}$ or of the semigroup $\mathbb{N}$. A continous dynamical system is an action of $\mathbb{R}$. There are more complicated dynamical systems defined for example by actions of other groups (for example $\mathbb{R}^d$).

Exercise 1.1.2. *Prove that the iterates of a map $f : X \to X$ give an action of $\mathbb{N}$ on X. The action $\mathbb{N} \times X \to X$ is given by*

$$(n, x) \to f^n(x).$$

Prove that if f is invertible, one has an action of $\mathbb{Z}$.

Exercise 1.1.3. *Prove that the solutions of a differential equation as (1.2) (assuming that for all points $x_0 \in X$ the solutions are unique and well defined for all times) give an action of $\mathbb{R}$ on X.*

1.2 Quadratic maps: attracting and repelling fixed points

Let $X = I = [0, 1]$ be the unit interval. Let $f : I \to I$ be the map

$$f(x) = \frac{5}{2}x(1 - x).$$

[6] If X has an additional structure (for example X is a topological space or X is a measured space), we can ask the additional requirement that for each $g \in G$, $\psi_g : X \to X$ preserves the structure of X (for example ψ_g is a continous map if X is a topological space or ψ_g preserves the measure. We will see more precisely these definitions in Chapters 2 and 4.

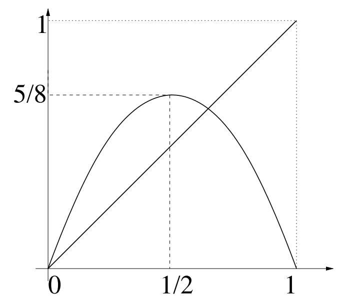


Figure 1.1: The graph of the quadratic map $f(x) = \frac{5}{2}x(1-x)$.

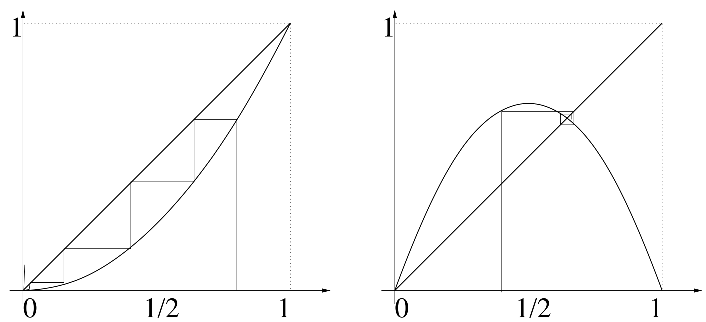


Figure 1.2: Examples of graphical analyses.

The graph of f can be drawn noticing that it is a parabola, $f(0) = f(1) = 0$ and that the derivative $f'(x) = 5/2 - 5x$ is zero at $x = 1/2$ for which $f(1/2) = 5/8$. In particular, f maps $[0,1]$ to $[0,1]$. See Figure 1.2.

Fixed points of f are solutions of the equation $f(x) = x$. In this case, solutions of $5/2x - 5/2x^2 = x$ or equivalenty $x(3/2 - 5/2x) = 0$ are only $x = 0$ and $x = 3/5$.

Remark 1.2.1. *Graphically, fixed points are given by considering the intersections of the graph of f with the diagonal $\Delta = \{(x,x), \quad x \in X\}$, which is the graph of the Id map, and taking their horizontal components.*

To have an idea of the empirical behaviour of an orbit $\mathcal{O}_f(x)$ one can use the following graphical method.

Graphical Analyses

- Draw the graph of f and the diagonal $\Delta = \{(x,x), \quad x \in X\}$;
- Start from $(x,0)$. Move vertically up until you intersect the graph of f at $(x, f(x))$;
- Move horizontally until you hit the diagonal Δ, at $(f(x), f(x))$; the horizontal projection is now $f(x)$;
- Move vertically to hit the graph, and then again horizontally to hit the diagonal;
- Repeat the step above.

At step $n \geq 1$ one hits the graph at $(f^{n-1}(x), f^n(x))$ and the diagonal at $(f^n(x), f^n(x))$. Thus the horizontal projections of the points obtained give the orbit $\mathcal{O}_f(x)$.

Exercise 1.2.1. *Draw the behaviour of the orbits of* $f(x) = 5/2x(1-x)$ *near* 0 *and near* 3/5.

Exercise 1.2.2. *Use the graphical analyses to find fixed points and study the behaviour or orbits nearby for the following functions:*

(1) $g(x) = x - x^2$ *for* $0 \le x \le 1$*;*

(2) $g(x) = 2x - x^2$ *for* $0 \le x \le 1$*;*

(3) $g(x) = -x^3$ *for* $-\infty \le x \le \infty$*;*

Given a ball $U := B(x, \epsilon) = \{y | \ d(x, y) < \epsilon\}$, let $\overline{U} := \overline{B(x, \epsilon)}$ be the closed ball $\{y | \ d(x, y) \le \epsilon\}$.

Definition 1.2.1. *We say that a fixed point* x *is an* attracting *fixed point if there exists a ball* $U := B(x, \epsilon)$ *around* x *such that*

$$f(\overline{U}) \subset U, \quad \text{and} \quad \cap_{n \in \mathbb{N}} f^n(U) = \{x\}.$$

We say that a fixed point x *is an* repelling *fixed point if there exists a ball* $U := B(x, \epsilon)$ *around* x *such that*[7]

$$\overline{U} \subset f(U), \quad \text{and} \quad \cap_{n \in \mathbb{N}} f^{-n}(U) = \{x\}.$$

Exercise 1.2.3. *Show that if* f *is invertible,* x *is an attracting fixed point if and only if it is a repelling fixed point for* f^{-1} *and viceversa.*

When X is an interval in $\mathbb{R}$ there is an easy criterium to determine whether a fixed point is attracting or repelling.

Theorem 1.2.1. *Let* $X \subset \mathbb{R}$ *be an interval and let* $f : X \to X$ *be a differentiable function with continuous derivative. Let* $x = f(x)$ *be a fixed point.*

(1) If $|f'(x)| < 1$*, then* x *is an attracting fixed point. More precisely, we will find and open ball* U *such that* $f(\overline{U}) \subset U$ *and for all* $y \in U$ *we have*

$$\lim_{n \to \infty} f^n(y) = x.$$

(2) If $|f'(x)| > 1$*, then* x *is a repelling fixed point.*

Remark 1.2.2. *Remark that if* $|f'(x)| = 1$ *it is not possible to determine just from this information whether the fixed point is repelling or attracting.*

Proof. Let us prove (1). Since f' is continuous and $|f'(x)| < 1$, there exist an $\epsilon > 0$ such that for all $y \in [x - \epsilon, x + \epsilon] = \overline{B(x, \epsilon)}$ we have $|f'(y)| < \rho < 1$. Then for all $y \in \overline{B(x, \epsilon)}$, since $f(x) = x$, by Mean Value Theorem there exists $\xi \in B(x, \epsilon)$ such that

$$|f(y) - x| = |f(y) - f(x)| = |f'(\xi)||y - x| \le \rho |y - x| \le \rho \epsilon.$$

This givens that $f(y) \in (x - \epsilon, x + \epsilon)$ for all $y \in \overline{B(x, \epsilon)}$. Thus $f(\overline{B(x, \epsilon)}) \subset B(x, \epsilon)$.

Let us prove by induction that

$$|f^n(y) - x| \le \rho^n \epsilon.$$

We already proved it for $n = 1$. Assume that it holds for $n \ge 1$. Then, applying mean value as before, for all $y \in B(x, \epsilon)$, there exists $\xi \in B(x, \epsilon)$ such that

$$|f^{n+1}(y) - f^{n+1}(x)| = |f(f^n(y)) - f(f^n(x))| = |f'(\xi)||f^n(y) - f^n(x)| \le \rho |f^n(y) - x|$$

[7] Here f is not necessarily invertible, so $f^{-1}(U)$ contains all preimages y such that $f^n(y) \in U$.

and by the induction assumption, since $f^{n+1}(x) = x$, this gives

$$|f^{n+1}(y) - x| = |f^{n+1}(y) - f^{n+1}(x)| \leq \rho |f^n(y) - x| \leq \rho^{n+1}\epsilon.$$

Since $\rho < 1$, $\lim_{n\to\infty} \rho^n = 0$. Thus, we get at the same time that

$$\lim_{n\to\infty} f^n(y) = x \quad \text{and} \quad \cap_{n\in\mathbb{N}} f^n(B(x,\epsilon)) = \{x\}.$$

The proof of part (2) is similar. □

Exercise 1.2.4. *Prove part (2) of Theorem 1.2.1.*

Exercise 1.2.5. *In our example $f(x) = 5/2x(1-x)$, one can check that $f'(0) = 5/2$ and $f'(3/5) = -1/2$ so that 0 is a repelling fixed point and $3/5$ is an attracting fixed point. Moreover, for each $\delta > 0$, $f([\delta, 1]) \subset (0,1)$ and all points points converge to $3/5$.*

The dynamics of this quadratic map is then very simple, it is an *attracting-repelling dynamics*. If one changes 5/2 with 4 and considers the map $f(x) = 4x(1-x)$, the behaviour is completely different and much more chaotic. See next section for more.

Exercise 1.2.6. *Program a computer program to plot some iterates of $f(x) = 4x(1-x)$ at some points. Is there any pattern? Compare with the case $f(x) = 3.9x(1-x)$.*

1.3 The quadratic family and attracting and repelling fixed points

The map f belongs to the quadratic family

$$f_\mu(x) = \mu x(1-x).$$

The behaviour of the maps for $0 \leq \mu < 4$ is also very simple and similar to the one for $\mu = 3$. There are only attracting and repelling fixed points and all the other points are attracted or repelled.

Exercise 1.3.1. *Consider the quadratic family f_μ for $\mu \in [0,4]$.*

(1) Check that for $\mu \in [0,4]$ the interval $I = [0,1]$ is mapped to itself, i.e. $f_\mu(I) \subset I$.

(2) Check that the fixed points of f_μ are 0 and $1 - 1/\mu$.

(3) Determine for which values of μ each of them is a repelling or attracting fixed point. What happens at $\mu = 4$?

The dynamics of f_μ for $\mu = 4$ is chaotic[8] (see Exercise 1.2.6)

For $\mu > 4$ the interval I is no longer invariant, i.e. there are points which are mapped outside I. It is still possible to consider the dynamics of f_μ, but one has to restrict the domain to an invariant subset of $[0,1]$, i.e. to the set C of the form

$$C = \cap_{n\in\mathbb{N}} f^{-n}(I). \tag{1.3}$$

If $x \in C \subset I$, for each $n \in \mathbb{N}$, $f^n(x) \in I$, so that $\mathcal{O}_f(x) \subset I$.

The dynamics of the quadratic family is very rich for $\mu > 4$ and display interesting chaotic phenomena as μ increasing, known as *period doubling*. We will not treat them in this course, we refer the interested reader to

[8]The dynamics of f_4 is very similar to the dynamics of the doubling map that we will see in §1.6. It is possible to show that these two maps are *conjugated* in the sense defined in §1.6 and hence they have similar dynamical properties, for example the same number of periodic points.

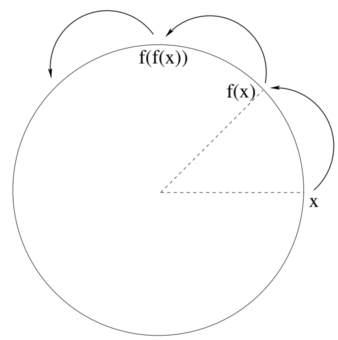


Figure 1.3: A rotation of S^1.

[4] R. Devaney *Chaotical Dyanamical Systems*, Springer

[5] K. Alligood, T. Sauer, j. Jorke, *Chaos: an Introduction to Dyanamical Systems*, Springer

Exercise 1.3.2. *Consider the map* $f : [0,1] \to [0,1]$.

$$f(x) = \begin{cases} 3x & \text{if } 0 \leq x \leq 1/2, \\ 3-3x & \text{if } 0 \leq x \geq 1/2 \end{cases}$$

Draw its graph. Prove that the invariant set given by (1.3) is a middle third Cantor set. *If you use the representation of numbers in* $[0,1]$ *by expansions in base* 3*, the middle third Cantor set is the set of all numbers in* $[0,1]$ *whose expansion in base* 3 *does not contain any digit* 1 *(if* $x = \sum_{i=1}^{\infty} x_i/3^i$*, where* $x_i \in \{0,1,2\}$*,* x *belongs to the middle third Cantor set iff all* $x_i \neq 1$*.)*

Exercise 1.3.3. *Consider the quadratic map with* $\mu = 9/2$*. Prove that the invariant set given by (1.3) is a* Cantor set*. A Cantor set is a set obtained as intersection* $C = \cap_n C_n$ *where at each* n *level the set* C_n *consist of a union of finitely many disjoint intervals and* C_{n+1} *is constructed by removing from each interval of* C_n *a finite number of intervals.*

1.4 Rotations of the circle: periodic points and dense orbits

Consider a circle of unit radius. More precisely, we will denote by S^1 the set

$$S^1 = \{(x,y) | \quad \sqrt{x^2+y^2} = 1\} \subset \mathbb{R}^2.$$

Identifying $\mathbb{R}^2$ with the complex plane $\mathbb{C}$, we can also write

$$S^1 = \{e^{2\pi i\theta}, \quad 0 \leq \theta < 1\} \subset \mathbb{C}.$$

Consider a rotation R_α of angle $2\pi\alpha$ on the circle (see Figure 1.3). It is given by

$$R_\alpha(e^{2\pi i\theta}) = e^{2\pi i(\theta+\alpha)} = e^{2\pi i\alpha}e^{2\pi i\theta}.$$

Since complex numbers in S^1 are multiplied by $e^{2\pi i\alpha}$, this is known as *multiplicative notation* for the rotation R_α.

There is a natural distance $d(z_1, z_2)$ between points on S^1, which is given by the *arc lenght* distance. We will renormalized it by dividing by 2π. For example, if $0 \le \theta_1 < \theta_2$ and $2\pi(\theta_2 - \theta_1) < \pi$ we have

$$d(e^{2\pi i\theta_1}, e^{2\pi i\theta_2}) = \frac{\text{arc lenght distance} = \ 2\pi(\theta_2 - \theta_1)}{2\pi} = \theta_2 - \theta_1.$$

This is clear by the geometric meaning that, since both points are rotated by the same angle $2\pi\alpha$, this distance is preserved i.e.

$$d(R_\alpha(z_1), R_\alpha(z_2)) = d(z_1, z_2), \qquad \text{for all } z_1, z_2 \in S^1.$$

Thus, the rotation of the circle is an example of an *isometry*, i.e. a map which preserves a distance.

There is another alternative way to describe a circle, that will be often more convenient. Imagine to *cut open* the circle to obtain an interval. Let $I/\sim$ denote the unit interval with the endpoints identified: the symbol $\sim$ recalls us that $0 \sim 1$ are glued together. Then $I/\sim$ is equivalent to a circle. More formally, consider $\mathbb{R}/\mathbb{Z}$, i.e. the space whose points are equivalence classes $x + \mathbb{Z}$ of real numbers x up to integers: two reals $x_1, x_2 \in \mathbb{R}$ are in the same equivalence class iff there exists $k \in \mathbb{Z}$ such that $x_1 = x_2 + k$. Then $\mathbb{R}/\mathbb{Z} = I/\sim$ since the unit interval $I = [0, 1]$ contains exactly one representative for each equivalence class with the only exception of 0 and 1, which belong to the same equivalence class, but are identifyed.

The map $\Psi : \mathbb{R}/\mathbb{Z} \to S^1$ given by

$$x \xrightarrow{\Psi} \Psi(x) = e^{2\pi i x} \tag{1.4}$$

establishes a one-to-one correspondence between $\mathbb{R}/\mathbb{Z}$ and S^1. The distance given by arc lenght divided by 2π, gives the following distance on $\mathbb{R}/\mathbb{Z}$:

$$d(x, y) = \min\{|x - y|, 1 - |x - y|\}. \tag{1.5}$$

Thus, we will use the same symbol d for both distances.

Exercise 1.4.1. *Check that the arc lenght distance divided by 2π becomes the distance in (1.5) on $\mathbb{R}/\mathbb{Z}$ under the identification given by Ψ, i.e.*

$$d(x, y) = \frac{\textit{arc lenght between } \Psi(x) \textit{ and } \Psi(y)}{2\pi}.$$

The rotation R_α, under this identification between S^1 and $\mathbb{R}/\mathbb{Z}$ becomes the map $R_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ given by

$$R_\alpha = x + \alpha \mod 1,$$

where mod 1 means that we subtract the integer part (for example $3.14 \mod 1 = 0.14$), hence taking the representative of the equivalence class $x + \alpha + \mathbb{Z}$ which lies in $[0, 1)$. We call α the rotation number of R_α (remark that the rotation angle is $2\pi\alpha$). More explicitely, if $\alpha \in [0, 1]$ we have

$$R_\alpha = \begin{cases} x + \alpha & \text{if } x + \alpha < 1 \\ x + \alpha - 1 & \text{if } x + \alpha \ge 1 \end{cases}$$

We call this *additive notation* (since here the rotation becomes addition mod 1).

Rotations of the circle display a very different behaviour according if the rotation number α is rational ($\alpha \in \mathbb{Q}$) or irrational ($\alpha \in \mathbb{R}\backslash\mathbb{Q}$). Recall that

Definition 1.4.1. *The orbit $\mathcal{O}_f(z_1)$ is* dense *if for all $z_2 \in S^1$ and for all $\epsilon > 0$ there exists $n > 0$ such that $R_\alpha^n(z_1) \in B(z_2, \epsilon)$ where $B(z_2, \epsilon)$ is the ball of radius ϵ and center z_2, i.e. $B(z_2, \epsilon) = \{z \in S^1 | \quad d(z, z_2) < \epsilon\}$.*

Theorem 1.4.1 (Dichotomy for Rotations)**.** *Let $R_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a rotation of the circle.*

(1) If $\alpha = p/q$ is rational, with $p, q \in \mathbb{Z}$, all orbits are periodic of period q;

(2) If α is irrational, for every point $z \in S^1$ the orbit $\mathcal{O}_{R_\alpha}(z)$ is dense.

In both cases the dynamics of the rotation is quite simple to describe: either all orbits are periodic, or all orbits are dense.

Proof. Let $\alpha = p/q$ with $p, q \in \mathbb{Z}$. Then for each $x \in \mathbb{R}/\mathbb{Z}$

$$R_\alpha^q(x) = x + q\frac{p}{q} \mod 1 = x + p \mod 1 = x.$$

Thus every point is periodic of period q. This proves (1).

Let us now prove (2). In this case it will be convenient to work on S^1 and use multiplicative notation. Assume that α is irrational. In particular, for each $z_1 = e^{2\pi i x_1} \in S^1$, for all $m \neq n$, $R_\alpha^n(e^{2\pi i x_1}) \neq R_\alpha^m(e^{2\pi i x_1})$. Indeed, if they were equal, $e^{2\pi i(x_1 + m\alpha)} = e^{2\pi i(x_1+n\alpha)}$ thus $2\pi(x_1 + m\alpha) = 2\pi(x_1 + n\alpha) + 2\pi k$ for some integer $k \in \mathbb{N}$. Thus, $m\alpha = n\alpha + k$. But this shows that $\alpha = k/(m - n)$, contradicting the assumption that α is irrational.

To show that the orbit of $z_1 \in S^1$ is dense, we have to show that for each $z_2 \in S^1$ and $\epsilon > 0$ there is a point of $\mathcal{O}_f(z_1)$ inside the ball $B(z_2, \epsilon)$. Let N be big enough so that $1/N < \epsilon$. Consider the points $z_1, R_\alpha(z_1), \dots, R_\alpha^{N-1}(z_1)$. Since as we proved before they are all distinct, by Pigeon Hole principle, there exists n, m such that $0 \leq n < m \leq N$ and

$$d(R_\alpha^n(z_1), R_\alpha^m(z_1)) \leq \frac{1}{N} < \epsilon.$$

This means that for some θ with $|\theta| < 1/N$ we have

$$R_\alpha^m(z_1) = e^{2\pi i\theta} R_\alpha^n(z_1) \;\Leftrightarrow\; e^{2\pi i m\alpha} z_1 = e^{2\pi i\theta} e^{2\pi i n\alpha} z_1 \;\Leftrightarrow\; \frac{e^{2\pi i m\alpha}}{e^{2\pi i n\alpha}} = e^{2\pi i\theta} \tag{1.6}$$

Consider now R_α^{m-n}. We claim that it is again a rotation by an agle smaller than ϵ. Indeed, from (1.6) we see that

$$R_\alpha^{m-n}(z_1) = e^{2\pi i m\alpha} e^{-2\pi i n\alpha} z_1 = \frac{e^{2\pi i m\alpha}}{e^{2\pi i n\alpha}} z_1 = e^{2\pi i\theta} z_1$$

R_α^{m-n} is a rotation by θ where $|\theta| < 1/N$, so

Thus, if we consider multiples $R_\alpha^{(m-n)}(z_1)$, $R_\alpha^{2(m-n)}(z_1)$, $R_\alpha^{3(m-n)}(z_1), \dots$ we obtain points

$$e^{2\pi i x_1}, e^{2\pi i(x_1+\theta)}, e^{2\pi i(x_1+2\theta)}, \dots, e^{2\pi i(x_1+k\theta)}, \dots,$$

whose spacing on S^1 is less than ϵ. Thus, there will be a $j > 0$ such that $R_\alpha^{j(m-n)}(z_1)$ enters the ball $B(z_2, \epsilon)$. □

Exercise 1.4.2. *Prove that if $\alpha = p/q$ and $(p, q) = 1$ i.e. p and q are coprime, than q is the minimal period, i.e. for each $x \in \mathbb{R}/\mathbb{Z}$ we have $R_\alpha^k(x) \neq x$ for each $1 \leq k < q$.*

Remark 1.4.1. *If $\alpha = p/q$ and $(p, q) = 1$ then $|p|$ gives the* winding number*, i.e. the number of "'turns"' that the orbit of any point does around the circle S^1 before closing up.*

1.4.1 Wely criterion and equidistribution modulo one

There are further properties which provide further information on dense orbits. A key one is *equidistribution*, which provides a more quantitative description of distribution of points in space and the way the orbits fills X.

Let $R_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a rotation and $\mathscr{O}^+_{R_\alpha}(x)$ an orbit. Recall from the introduction the following definition:

Definition 1.4.2 (Equidistribution)**.** *We say that the orbit $\mathscr{O}^+_{R_\alpha}(x)$ is* equidistributed *if for any $a < b$ real such that $[a,b] \subset [0,1]$ we have*

$$\lim_{n\to+\infty} \frac{1}{n} Card\{0 \leq k < n, \quad \textit{such that } R^k_\alpha(x) \in [a,b]\} = |b-a|.$$

The left hand side (LHS) for fixed n represents the proportion of time spent inside the interval $[a,b]$ by initial segment of the orbit up to time n, Thus, if the orbit is equidistributed, the time spent visiting a given interval $[a,b]$ is asymptotic to its *lenght*, and in particular it depends on the lenght only and *not* on the position of the inverval in space: intervals (arcs on the circle) with the same length are visited with the same frequency.

This definition is a special case of the more general definition for real sequences:

Definition 1.4.3 (Equidistribution modulo one)**.** *Let $(x_k)_{k\in\mathbb{N}}$ be a sequence of real numbers. We say that $(x_k)_{k\in\mathbb{N}}$ is* equidistributed modulo one *if for any $a < b$ real such that $[a,b] \subset [0,1]$ we have*

$$\lim_{n\to+\infty} \frac{1}{n} Card\{0 \leq k < n, \quad \textit{such that } x_k \mod 1 \in [a,b]\} = |b-a|.$$

Remark that equidistribution in particular implies density:

Exercise 1.4.3. *Show that if $(x_k)_{k\in\mathbb{N}}$ is* equidistributed modulo one *then in particular the sequence of fractional parts, namely $x_k = x_k \mod 1$ is dense in $[0,1]$.*

It is on the other hand a very effective and quantitative way of becoming dense. Notice that the converse is not true, i.e there are sequences which are dense modulo one but fail to be equdistributed modulo one.

A classical criterion to verify whether a sequence is equidistributed modulo one is the following, due to Weyl:

Theorem 1.4.2 (Weyl criterion for equidistribution)**.** *Let $(x_k)_{k\in\mathbb{N}}$ be a sequence of real numbers. The following are equivalent:*

(E1) The sequence $(x_k)_{k\in\mathbb{N}}$ is equidistributed modulo one*;*

(E2) For any continuous function $f : [0,1] \to \mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(x_k \mod 1) = \int_0^1 f(x)dx. \tag{1.7}$$

(E3) For any $\ell \in \mathbb{Z}\backslash\{0\}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i \ell x_k} = 0. \tag{1.8}$$

Weyl criterion is that $(E3)$ is equivalent to $(E1)$ so to verify equidistribution modulo one it is sufficient to verify $(E3)$. Remark that in case of the rotation, if we take $x_k = x + k\alpha, k \in \mathbb{N}$, $x_k \mod 1 = R^k_\alpha(x)$, i.e. the sequence of fractional parts $\{x_k\}, k \in \mathbb{N}$ is just the orbit of the rotation. In this case (or more in general if $x_k \mod 1$ are poins in the orbit of a dynamical system), $(E2)$ is one of the characterizations of a property known as *unique ergodicity.* We will encounter this property later on in the course. We added $(E2)$ in the formulation of the criterion since it is a useful intermediate step to prove the equivalence of $(E1)$ and $(E3)$.

Let us apply Weyl criterion to prove equidistribution for orbits of irrational rotations. We will comment on the proof of Weyl's criterion in the following section.

Theorem 1.4.3. *Assume that α is irrational. Then for every $x \in [0,1]$ the orbit $\mathscr{O}^+_{R_\alpha}(x)$ is equidistributed.*

Proof. Let us apply Weyl criterion. Fix $\ell \in \mathbb{Z}\backslash\{0\}$. Let us compute the sums which appear in $(E3)$, which, for $x_k \mod 1 = R^k_\alpha(x) = e^{2\pi(k\alpha+x)}$ become

$$\sum_{k=0}^{n-1} e^{(2\pi i \ell R^k_\alpha(x))} = \sum_{k=0}^{n-1} e^{2\pi\ell(k\alpha+x)} = e^{2\pi\ell x}\sum_{k=0}^{n-1}(e^{2\pi\ell\alpha})^k.$$

Thus, using the formula for the geometric progression with step $\lambda := e^{2\pi\ell\alpha}$, which gives $\sum_{k=0}^{n-1}\lambda^k = \frac{1-\lambda^n}{1-\lambda}$, we get

$$\left|\sum_{k=0}^{n-1} e^{(2\pi i\ell R^k_\alpha(x))}\right| = \left|e^{2\pi\ell x}\frac{1-e^{2\pi\ell n\alpha}}{1-e^{2\pi\ell\alpha}}\right| = \frac{|1-e^{2\pi\ell n\alpha}|}{|1-e^{2\pi\ell n\alpha}|} \leq \frac{2}{|1-e^{2\pi\ell n\alpha}|},$$

where the last is just the bound for the absolute value of the sumo of two numbers of absolute value one (triangle inequality). Thus, since ℓ is fixed,

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} e^{(2\pi i\ell R^k_\alpha(x))}\right| \leq \frac{1}{n}\frac{2}{|1-e^{2\pi\ell n\alpha}|}$$

tends to zero as n tends to infinity, which proves that $(E3)$ holds and thus, by Weyl criterion, that $(R^k_\alpha(x))_k$ is equidistributed. $\square$

Comments on the proof of Weyl criterion Let us add some comments on the proof of Theorem 1.4.2. Notice that the three properties $(E1), (E2), (E3)$ can all be written in the form

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f(x_k \mod 1) = \int_0^1 f(x)dx \tag{1.9}$$

for suitably chosen f. For $(E2)$ this is obvious, one just asks f to be continuous. For $(E1)$, as we already remaked in the first lecture, it is sufficient to take $f = \chi_{[a,b]}$. Then

$$\frac{1}{n}\sum_{k=0}^{n-1}\chi_{[a,b]}(x_k \mod 1) = \frac{1}{n}Card\{0 \leq k < n, \quad \text{such that } x_k \mod 1 \in [a,b]\}$$

and $\int_0^1 \chi_{[a,b]} = |b-a|$, so (1.9) for $f = \chi_{[a,b]}$ gives exactly the definition of equidistribution modulo one. Finally, to see how $(E3)$ can be expressed in the form (1.9), one needs to allow f to be complex valued, i.e. to consider $f : [0,1] \to \mathbb{C}$. Recall that if $\Re f$ and $\Im f$ denote respectively the real and imaginary part of f, the integral of a complex function is defined by:

$$\int_0^1 f(x)dx := \int_0^1 \Re f(x)dx + i\int_0^1 \Im f(x)dx.$$

Take now $f(x) = e^{2\pi i\ell x}$ in (1.9). Since $e^{2\pi i\ell x} = \cos(2\pi i\ell x) + i\sin(2\pi i\ell x)$, we then have that, for any integer $\ell \neq 0$

$$\int_0^1 e^{2\pi i\ell x}dx := \int_0^1 \cos(2\pi i\ell x)dx + i\int_0^1 \sin(2\pi i\ell x)dx = 0 + i0 = 0.$$

Thus, one can now see that $(E3)$ is simply the formulation of (1.9) for $f(x) = e^{2\pi i\ell x}$.

Extra: outline of the proof. To prove the equivalence of $(E1)$ and $(E2)$, one has to go from continuous to characteristic functions (by using that every characteristic function can be approximated from above and from below by continous functions and, conversely, every continous function can be approximated by *simple functions*, which are linear combinations of characteristic functions). By applying $(E2)$ to real and imaginary part of a continous function $f : [0,1] \to \mathbb{C}$, one can show that $(E2)$ is equivalent to $(E2)'$ where f is any continous complex valued function $f : [0,1] \to \mathbb{C}$. Then it is clear that $(E2)'$ implies $(E3)$ (since we are allowed to take $f(x) = e^{2\pi i\ell x}$). The hardest part if the proof that $(E3)$ implies $(E2)'$: this requires a result known as *Stone-Weierstrass Theorem*, which allows to show that every continous function can be approximated by a linear combination of the complex exponentials $e^{2\pi i\ell x}, \ell \in \mathbb{Z}$ (it implies in particular that trigonometric polynomials are dense in real valued continous functions).

1.5 Extra to Section ??

1.5.1 A conjugacy for the doubling map.

In order to obtain a conjugacy between the doubling map $f : X \to X$ and a shift is to consider the shift space $\Sigma/\!\sim$ with the equivalence relation $\underline{a} \sim \underline{b}$ if and only if there exists a $k \in \mathbb{N}$ such that $a_i = b_i$ for $0 \leq i < k$ and

$$a_k = 1, a_i = 1 \text{ for all } \ i > k, \qquad b_k = 0, b_i = 1 \text{ for all } \ i > k.$$

The sequences which are identified are sequences with tails of 0s and 1s which correspond to two possible choices of binary digits for a dyadic rational.

Theorem 1.5.1. *The map $\psi : \Sigma/\!\sim \to X$ is well defined and it is a conjugacy between the doubling map f and $\sigma : \Sigma/\!\sim \to \Sigma/\!\sim$.*

Exercise 1.5.1. *Prove Theorem 1.7.1.*

Exercise 1.5.2. *Let D be the set of dyadic rationals and let Σ' be the sequences which do not end with a tail of 0s or 1s. Verify that $\psi : \Sigma' \to X - D$ is a conjugacy.*

Full details of the proof might be added as an Extra later on, come back to this page if you are interested

Extra: Dirichlet's theorem

Let us show an application to number theory. The following result is known as Dirichlet's theorem:

Theorem 1.5.2 ((Dirichlet Theorem))**.** *If α is irrational, for each $\delta > 0$ there exists an integer q with $0 < q \leq 1/\delta$ and $p \in \mathbb{Z}$ such that*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{\delta}{q}$$

The theorem shows how well an irrational number can be approximated by rational numbers. It is a first important keystone in the theory of Diophantine Approximation.

Proof. Note that in additive notation the orbit of 0 contains the points

$$R_\alpha^n(0) = n\alpha \mod 1.$$

We saw in the proof of Theorem **??** that when α is irrational, the orbit $\mathcal{O}^+_{R_\alpha}(0)$ consists of distinct points. If we choose N large so that $1/N \leq \delta$, by Pigeon Hole principle there exists $0 \leq n < m \leq N$ such that

$$d(R^m(0), R^n(0)) < \frac{1}{N}.$$

Let us call $q = m - n$, so $0 < q \leq N$. Since R_α is an isometry,

$$d(0, R^q(0)) = d(0, R^{m-n}(0)) = d(R^n(0), R^{m-n}(R^n(0))) = d(R^m(0), R^n(0)) < \frac{1}{N}.$$

Recalling the definition of distance, this means exactly that there exists an integer $p \in \mathbb{Z}$ such that

$$|q\alpha - p| \leq \frac{1}{N} \leq \delta.$$

□

As a Corollary, one can show the following version of Dirichlet theorem:

Exercise 1.5.3. *Show that if α is irrational, there are infinitely many fractions p/q where $p \in \mathbb{Z}$, $q \in \mathbb{N}$ that solve the equation*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^2}.$$

1.6 The Doubling Map (coding and conjugacy)

Consider the following map on $[0,1]$, known as *doubling map*:

$$f(x) = 2x \mod 1 = \begin{cases} 2x & \text{if } 0 \leq x \leq 1/2, \\ 2x - 1 & \text{if } 0 \leq x \leq 1/2 \end{cases} \tag{1.10}$$

The map is well defined also on $[0,1]/\sim = \mathbb{R}/\mathbb{Z}$. To check that, we have to check that the points 0 and 1 which are indenfiyed have equivalent images. But this is true since $f(1) = 1 \sim 0 = f(0)$. So we can think of f as a map on $I/\sim = \mathbb{R}/\mathbb{Z}$. Since we saw that $\mathbb{R}/\mathbb{Z}$ is the same than S^1 (via the correspondence given by (1.4), we can see f in multiplicative coordinates as a map from $S^1 \to S^1$ given by

$$f(e^{2\pi i\theta}) = e^{2\pi i 2\theta} \tag{1.11}$$

Thus the angles are *doubled* and this explains the name *doubling map*. Moreover, one can see that the map f on S^1 is *continuous*.

Exercise 1.6.1. *Check that f given by (1.10) becomes (1.11) by the identification of $I/\sim$ with S^1.*

Remark that f is *not* ivertible: each point has two preimages:

$$f^{-1}(y) = \left\{\frac{y}{2}, \frac{y}{2} + \frac{1}{2}\right\}.$$

Remark also that f *expands* distances. If $d(x,y) < 1/4$, then

$$d(f(x), f(y)) = 2d(x,y).$$

Exercise 1.6.2. *Check the assertion above.*

Let us define the (Lebesgue) measure (think of lenght) of an interval $[a, b]$ to be $\lambda([a, b]) = |b - a|$. If $I = \cup_i I_i$ is a (finite or countable) union of *disjoint* intervals $I_i = [a_i, b_i]$, set

$$\lambda(I) = \sum_i |b_i - a_i|.$$

Then one can check that $\lambda(f([a, b])) = 2\lambda([a, b])$, but on the other side, since

$$f^{-1}[a, b] = \left[\frac{a}{2}, \frac{b}{2}\right] \cup \left[\frac{a+1}{2} +, \frac{b+1}{2} + 1\right],$$

we have

$$\lambda(f^{-1}[a, b]) = \frac{b-a}{2} + \frac{(b+1)-(a+1)}{2} = b - a = \lambda([a, b]).$$

This is an example of a map that preserves the measure λ (we will see the definition in Chapter 4).

Let us try to answer the following questions for f:

Q. 1 Are there periodic points?

Q. 2 Are there points with a dense orbit?

In order to answer these questions, we will show two powerful techniques in dynamical systems, *conjugacy* and *coding*.

1.6.1 Conjugacy and semi-conjugacy

Let X, Y be two spaces and $f : X \to X$ and $g : Y \to Y$ be two maps.

Definition 1.6.1. *A* conjugacy *between f and g is an* invertible *map $\psi : Y \to X$ such that $\psi g = f\psi$, i.e. for all $y \in Y$*

$$\psi(g(y)) = f(\psi(y)).$$

Since ψ is invertible, we can also write $g = \psi^{-1} \circ f \circ \psi$.

The relation $\psi g = f\psi$ is often expressed by saying that the diagram here below *commutes*, i.e. one can start from a point in $y \in Y$ in the left top corner and indifferently apply first the arrow g and then the arrow ψ on the right side of the diagram or first the arrow ψ on the left side and then the arrow g and the result is the same point in the bottom right corner.

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

Lemma 1.6.1. *If f and g are conjugated by ψ, then y is a periodic point of period n for g if and only if $\psi(x)$ is a periodic point of period n for f.*

Exercise 1.6.3. *Check by induction that if $\psi g = f\psi$, then $\psi g^n = f^n\psi$. If ψ is invertible, we also have $g^n = \psi^{-1} g^n \psi$.*

Proof. Assume that $g^n(y) = y$. Then, by the exercise above

$$f^n(\psi(y)) = \psi(g^n(x)) = \psi(x),$$

so $\psi(y)$ is a periodic point of period n for f.

Conversely, assume that $f^n(\psi(y)) = \psi(y)$. Then since ψ is invertible, by the previous exercise we also have $g^n = \psi^{-1} g^n \psi$. Thus

$$g^n(y) = \psi^{-1}(f^n(\psi(y))) = \psi^{-1}(\psi(y)) = y,$$

so y is periodic of period n for g. □

Thus, if the periodic points of the map g are easier to understand than the periodic points of the map f, through the conjugacy one can gain information about periodic points for f. We will see that this is exactly the case for the doubling map.

Definition 1.6.2. *A* semi-conjugacy *between f and g is an map $\psi : Y \to X$ such that $\psi g = f \psi$, i.e. for all $y \in Y$*

$$\psi(g(y)) = f(\psi(y)).$$

or, equivalently, such that the diagram below commutes:

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

We say that $g : Y \to Y$ is an extension *of $f : X \to X$ and that $f : X \to X$ is a* factor *of $g : Y \to Y$*

Exercise 1.6.4. *If f and g are semi-conjugated by ψ and y is a periodic point of period n for g, then $\psi(x)$ is a periodic point of period n for f.*

Remark 1.6.1. *There are examples of f and g which are semi-conjugated by ψ and such that $\psi(y)$ is a periodic point of period n for f, but y is not a periodic point for g. We will see such an example using the baker map in a few classes.*

1.6.2 Coding of the doubling map

We will define a conjucagy between the doubling map and an abstract space that will help us understand points with periodic and dense orbits.

Given $x \in [0, 1]$, we can express x in *binary expansion*, i.e. we can write

$$x = \sum_{i=1}^{\infty} \frac{x_i}{2^i}$$

where x_i are digits which are either 0 or 1. Binary expansions are useful to study the doubling map because if we apply the doubling map:

$$f(x) = 2x \mod 1 = \sum_{i=1}^{\infty} 2\frac{x_i}{2^i} \mod 1 = x_0 + \sum_{i=2}^{\infty} \frac{x_i}{2^{i-1}} \mod 1 = \sum_{i=2}^{\infty} \frac{x_i}{2^{i-1}} \qquad (1.12)$$

and if we know change the name of the index, setting $j = i-1$, we proved that if $x = \sum_{i=1}^{\infty} \frac{x_i}{2^i}$,

$$f(x) = \sum_{i=2}^{\infty} \frac{x_i}{2^{i-1}} = \sum_{j=0}^{\infty} \frac{x_{j+1}}{2^j}$$

Let us construct a map on the space of digits of binary expansion which mimic this behaviour.

Let $\Sigma^+ = \{0,1\}^{\mathbb{N}}$ be the set of all sequences of 0 and 1:

$$\Sigma^+ = \{(a_i)_{i=1}^{\infty}, \quad a_i \in \{0,1\}\}.$$

The *points* $(a_i)_{i=1}^{\infty} \in \Sigma^+$ are one-sided sequences of digits 0,1, for example a point is

$$0,0,1,1,0,1,0,0,0,1,1,0,1,1,1,1,0,0,1,0,1,0\dots$$

The *shift* map σ^+ is a map $\sigma^+ : \Sigma^+ \to \Sigma^+$ which maps a sequence to the *shifted* sequence:

$$\sigma((a_i)_{i=1}^{\infty}) = (b_i)_{i=1}^{\infty}, \qquad \text{where} \quad b_i = a_{i+1}.$$

The sequence $(b_i)_{i=1}^{\infty}$ is obtained from the sequence $(a_i)_{i=1}^{\infty}$ by dropping the first digit a_1 and by shifting all the other digits one place to the left. For example, if

$$\begin{aligned} (a_i)_{i=1}^{\infty} &= 0,0,1,1,0,1,0,0,0,1,1,0,1,1,1,1,0,0,1,0,1,0\dots, \\ (b_i)_{i=1}^{\infty} &= 0,1,1,0,1,0,0,0,1,1,0,1,1,1,1,0,0,1,0,1,0\dots, \end{aligned}$$

Remark that σ^+ is not invertible, because if we know $\sigma((a_i)_{i=1}^{\infty})$ we cannot recover $(a_i)_{i=1}^{\infty}$ since we lost the information about the first digit.

Define the following map $\psi : \Sigma^+ \to [0,1]$. For each $(a_i)_{i=1}^{\infty} \in \Sigma^+$ set

$$\psi((a_i)_{i=1}^{\infty}) = \sum_{i=1}^{\infty} \frac{a_i}{2^i} \in [0,1].$$

The map is well defined since the series $\sum_{i=1}^{\infty} \frac{a_i}{2^i} \le \sum_{i=1}^{\infty} \frac{1}{2^i}$ which is convergent. The map ψ associates to a sequence of 0 and 1 a number in $[0,1]$ which has a_i as digits of the binary expansion.

Clearly ψ is surjective: each real $x \in [0,1]$ has a binary expansion. It is not injective, since there are numbers which has to binary expansions. In the same way that in decimal expansion we can write $1.00000\dots = 0.999999\dots$, binary expansions which have an infinite tails of 1 yield the same number that an expansion with a tail of 0, for example

$$\frac{1}{2} = \frac{1}{2} + \sum_{i=2}^{\infty} \frac{0}{2^i} \quad \text{but also} \quad \frac{1}{2} = \sum_{i=2}^{\infty} \frac{1}{2^i}.$$

One can check that these ambiguty happens only for rational numbers of the form $p/2^n$, whose denominator is a power of 2. For all other numbers ψ is a bijection.

Proposition 1. *The map $\psi : \Sigma^+ \to [0,1]$ is a semi-conjucaty between the shift map $\sigma^+ : \Sigma^+ \to \Sigma^+$ and the doubling map $f : [0,1] \to [0,1]$.*

Proof. We have to prove that the following diagram commutes:

$$\begin{array}{ccc} \Sigma^+ & \xrightarrow{\sigma^+} & \Sigma^+ \\ \downarrow \psi & & \downarrow \psi \\ [0,1] & \xrightarrow{f} & [0,1] \end{array}$$

Take any $(a_i)_{i=1}^{\infty} \in \Sigma^+$. Let us first compute $\psi(\sigma^+((a_i)_{i=1}^{\infty}))$:

$$\psi(\sigma^+((a_i)_{i=1}^{\infty})) = \psi((b_i)_{i=1}^{\infty}) = \sum_{i=1}^{\infty} \frac{b_i}{2^i} = \sum_{i=1}^{\infty} \frac{a_{i+1}}{2^i},$$

since $b_i = a_{i+1}$. Let us now compare with $f(\psi((a_i)_{i=1}^{\infty}))$:

$$f(\psi((a_i)_{i=1}^{\infty})) = f\left(\sum_{i=1}^{\infty} \frac{a_{i+1}}{2^i}\right) = \sum_{i=1}^{\infty} \frac{2a_{i+1}}{2^i} \mod 1 = \sum_{j=0}^{\infty} \frac{a_{j+1}}{2^j}$$

in virtue of the computation done in (1.12). Thus the results are the same. This concludes the proof. □

We can now use the correspondence given by ψ to study the periodic points of the doubling map.

It is very easy to construct periodic points for σ^+: they are points $(a_i)_{i\in\mathbb{N}}$ whose digits are repeated periodically. For example, if we repeat the digits $0, 1, 1$ periodically, we get a periodic point of period 3:

$$\begin{aligned}
(a_i)_{i\in\mathbb{N}} &= 0,1,1,0,1,1,0,1,1,0,1,1,\dots \\
\sigma^+(a_i)_{i\in\mathbb{N}} &= 1,1,0,1,1,0,1,1,0,1,1,\dots \\
(\sigma^+)^2(a_i)_{i\in\mathbb{N}} &= 1,0,1,1,0,1,1,0,1,1,\dots \\
(\sigma^+)^3(a_i)_{i\in\mathbb{N}} &= 0,1,1,0,1,1,0,1,1,\dots \quad = (a_i)_{i\in\mathbb{N}}
\end{aligned}$$

More in general, if $a_{n+i} = a_i$ for all $i \in \mathbb{N}$, then

$$(\sigma^+)^n(a_i)_{i\in\mathbb{N}} = (a_{n+i})_{i\in\mathbb{N}} = (a_i)_{i\in\mathbb{N}},$$

so $(a_i)_{i\in\mathbb{N}}$ is periodic of period n.

Theorem 1.6.1. *The doubling map $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ has $2^n - 1$ periodic points of period n. Moreover, periodic points are dense.*

Proof. Since the shift map σ^+ and the doubling map f are semi-conjugated, periodic points of period n for $\sigma^+ : \Sigma^+ \to \Sigma^+$ are mapped to periodic points of period n for f. Periodic points of period n for σ^+ are all sequences whose digits repeat periodically with period n. Thus, there are 2^n such sequences, since we can choose freely the first n digits in $\{0, 1\}$ and then repeat them periodically. Remark that the sequence $0, 0, 0, \dots$ is mapped to 0 and the sequence $1, 1, 1, \dots$ is mapped to 1, which are the same point in $\mathbb{R}/\mathbb{Z} = I/\sim$. Thus, there are $2^n - 1$ periodic points of period n.

We leave the second part as an exercise. □

Example 1.6.1. *The periodic points of period 3 for σ^+ are the periodic sequences obtained repeating the blocks of digits:*

$$000 \quad 001 \quad 010 \quad 011 \quad 100\dots, \quad 101\dots, \quad 110\dots, \quad 111$$

Let us take the corresponding binary expansions. Since $a_{i+3} = a_i$ for all $i \in \mathbb{N}$ we have

$$\sum_{i=1}^{\infty} \frac{a_i}{2^i} = \left(\frac{a_0}{2} + \frac{a_1}{2} + \frac{a_3}{2^2}\right) + \left(\frac{a_0}{2^3} + \frac{a_1}{2^4} + \frac{a_3}{2^5}\right) + \dots = \sum_{j=0}^{\infty} \left(\frac{a_0}{2} + \frac{a_1}{2^2} + \frac{a_3}{2^2}\right) \frac{1}{(2^3)^j}$$

So, for example, starting from the sequence obtained repeating the block 101 we obtain

$$\sum_{j=0}^{\infty} \left(\frac{1}{2} + \frac{0}{4} + \frac{1}{8}\right) \frac{1}{8^j} = \frac{5}{8} \sum_{j=0}^{\infty} \frac{1}{8^j} = \frac{5}{8} \frac{1}{\left(1 - \frac{1}{8}\right)} = \frac{5}{7}.$$

Thus, we find the $7 = 2^3 - 1$ periodic points of period 3 for the doubling map are

$$0, \ \frac{1}{7}, \ \frac{2}{7}, \ \frac{3}{7}, \ \frac{4}{7}, \ \frac{5}{7}, \ \frac{6}{7}.$$

Exercise 1.6.5. *Periodic points for the doubling map can also be found directly solving the equation $f^n(x) = x$. List the fractions which correspond to periodic points of period n.*

Exercise 1.6.6. *Prove that the periodic points for the doubling map are dense. [Hint: use the previous exercise.]*

The map $\phi : \Sigma^+ \to [0,1]$ sends a sequence in 0 and 1 to the corresponding binary expansion. Conversely, given a point $x \in [0,1]$, we can construct its binary expansion as follows. Consider the two intervals

$$P_0 = \left[0, \frac{1}{2}\right), \qquad P_1 = \left[\frac{1}{2}, 1\right).$$

They give a *partition* $\{P_0, P_1\}$ of $[0,1]/\sim$, since $P_0 \cap P_1 = \emptyset$ and $P_0 \cup P_1 = [0,1]/\sim$. Let $\phi : I/\sim \to \Sigma^+$ be the map

$$x \to \phi(x) = (a_{k+1})_{k=1}^{\infty}, \qquad \text{where} \begin{cases} a_k = 0 & \text{if } T^{k-1}x \in P_0, \\ a_k = 0 & \text{if } T^{k-1}x \in P_0 \end{cases}$$

The sequence $a_1, a_2, \ldots, a_k, \ldots$ is called the *itineary* of $\mathcal{O}_f(x)$ with respect to the partition $\{P_0, P_1\}$: it is obtained by iterating $f^k(x)$ and recording which interval, whether P_0 or P_1, is visited at each k. In particular, if $a_1, a_2, \ldots, a_k, \ldots$ is called the *itineary* of $\mathcal{O}_f(x)$ we have

$$x \in P_{a_1},\ f(x) \in P_{a_2},\ f^2(x) \in P_{a_3}, \ldots\ f^{k-1}(x) \in P_{a_k}, \ldots$$

Remark 1.6.2. *The idea of* coding *an orbit by recording its* itineary *with respect to a partition is a very powerful technique in dynamical systems. It often allow to conjugate a dynamical system to a shift map on a space of symbols. These symbolic spaces will be studied in Chapter 2 and, even if at first they may seem more abstract, they are well studied and often easier to understand then the original system.*

Let us check that $\phi : I/\sim \to \Sigma^+$ is a right inverse for the map $\psi : \Sigma^+ \to I/\sim$ constructed before, i.e. $\phi \circ \psi : \Sigma^+ \to \Sigma^+$ is the identity map. Let us hence compute $\phi(\psi((a_i)_{i=1}^{\infty}))$. Since

$$\psi((a_i)_{i=1}^{\infty}) = \sum_{i=1}^{\infty} \frac{a_i}{2^i} \in \left[\frac{a_1}{2}, \frac{a_1+1}{2}\right) = \begin{cases} \left[0, \frac{1}{2}\right) & \text{if } a_1 = 0, \\ \left[\frac{1}{2}, 1\right) & \text{if } a_1 = 1, \end{cases},$$

we have that $\phi(\psi((a_i)_{i=1}^{\infty})) \in P_{a_1}$. Similarly, since for each $f \geq 1$

$$f^{k-1}\left(\sum_{i=1}^{\infty} \frac{a_i}{2^i}\right) = \sum_{i=1}^{\infty} 2^{k-1}\frac{a_i}{2^i} \mod 1 = \sum_{i=1}^{\infty} \frac{a_{i+k}}{2^i} \in \left[\frac{a_k}{2}, \frac{a_k+1}{2}\right),$$

we have that $f^{k-1}(\phi(\psi((a_i)_{i=1}^{\infty}))) \in P_{a_k}$, so, by definition of ϕ, this shows that

$$\phi(\psi((a_i)_{i=1}^{\infty}) = (a_i)_{i=1}^{\infty}.$$

Let us use conjucagy and coding to show costruct a dense orbit for the doubling map.

Theorem 1.6.2. *There exists a point whose orbit under the doubling map is* dense*.*

Proof. Step 1 We claim that to prove that an orbit $\mathcal{O}_f^+(x)$ is dense, it is enough to show that for each $n \geq 1$ it visits all intervals of the form $I(a_0, a_1, \ldots, a_n)$. Indeed, if this is the case, given $y \in I$ and $\epsilon > 0$, take N large enough so that $1/2^{N+1} \leq \epsilon$ and take the interval $I(a_0, a_1, \ldots, a_N)$ which contains y (one of them does since they partition $[0,1)$ by (P2) above). If we showed that there is a point $f^k(x)$ in the orbit $\mathcal{O}_f^+(x)$, which visits $I(a_0, a_1, \ldots, a_N)$, since both y and $f^k(x)$ belong to $I(a_1, \ldots, a_N)$ (which has size $1/2^{N+1}$ by (P1) above), we have $d(f^k(x), y) \leq 1/2^{N+1} < \epsilon$. This shows that $\mathcal{O}_f^+(x)$ is dense.

Step 2. To construct an orbit which visits all dyadic intervals, we are going to construct its itinerary as a sequence in Σ^+ first. Let us now list for each n all the possible sequences $a_0, a_1, \dots, a_n$ of length n (there are 2^{n+1} of them) and create a sequence $(\overline{a}_i)_{i=0}^{\infty}$ by just apposing all such sequences for $n = 0$, then $n = 1$, then $n = 2$ and so on:

$$0, 1, \quad 0, 0,\ 0, 1,\ 1, 0,\ 1, 0,\ 1, 1, \quad 0, 0, 0,\ 0, 0, 1,\ 0, 1, 0,\ 0, 1, 1,\ 1, 0, 0, \dots$$

Step 3. Let us consider now the point $\overline{x} := \psi((\overline{a}_i)_{i=0}^{\infty})$ which has the sequence $(\overline{a}_i)_{i=0}^{\infty}$ as digist of its binary expansion. Let us show that its orbit visits all intervals of the form $I(a_0, a_1, \dots, a_n)$ and thus that it is dense by Step 1. To see that, it is enough to find where the block $a_0, a_1, \dots, a_n$ appears inside $(\overline{a}_i)_{i=0}^{\infty}$, for example at $\overline{a}_k = a_0, \overline{a}_{k+1} = a_1, \dots, \overline{a}_{k+n} = a_n$. Then, since the itinerary of $f^k(\overline{x})$ by definition of itinerary is $\overline{a}_k, \overline{a}_{k+1}, \dots, \overline{a}_{k+n}, \dots$, this shows that

$$f^k(\overline{x}) \in I(\overline{a}_k, \overline{a}_{k+1}, \dots, \overline{a}_{k+n}) = I(a_0, a_1, \dots, a_n),$$

so $f^k(\overline{x})$ is the point in $\mathcal{O}_f^+(\overline{x})$ which visits $I(a_0, a_1, \dots, a_n)$. This concludes the proof that $\mathcal{O}_f^+(\overline{x})$ visits all dyadic intervals and hence that it is dense. □

Exercise 1.6.7. *Draw all intervals of the form $I(a_1, a_2, a_3)$ where $a_1, a_2, a_3 \in \{0, 1\}$.*

Linear expanding maps.

We remarked earlier that the doubling map doubles distances: if $x, y \in \mathbb{R}/\mathbb{Z}$ are any two points such that $d(x, y) < 1/4$, then $d(f(x), f(y)) = 2d(x, y)$, that is, the distance of their images is doubled. The doubling map is an example of an *expanding map*:

Definition 1.6.3. *A one-dimensional map $g : I \to I$ of an interval $I \subset \mathbb{R}$ is called an* expanding *if it is piecewise differentiable, that is we can decompose I into a finite union of intervals on each of which g is differentiable, and the derivative g' satisfies $|g'(x)| > 1$ for all $x \in I$.*

More precisely, the doubling belongs to the family of *linear expanding maps of the circle*: for each $m \in \mathbb{Z}$ with $|m| > 1$ the map $E_m : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is given by

$$E_m(x) = mx \mod 1 \qquad (\text{or equivalently} \quad E_m(z) = z^k \quad \text{on } S^1).$$

The doubling map is the same than E_2. These maps can be studied analogously, by considering expansion in base m instead than binary expansions. One can prove that they are semi-conjugated with the shift σ^+ on the space

$$\Sigma_m^+ = \{0, 1, \dots, m-1\}^{\mathbb{N}}$$

of one-sided sequences in the digits $0, \dots, m-1$.

1.7 Extra to Section ??

1.7.1 A conjugacy for the doubling map.

In order to obtain a conjugacy between the doubling map $f : X \to X$ and a shift is to consider the shift space $\Sigma/\sim$ with the equivalence relation $\underline{a} \sim \underline{b}$ if and only if there exists a $k \in \mathbb{N}$ such that $a_i = b_i$ for $0 \le i < k$ and

$$a_k = 1, a_i = 1 \text{ for all } \ i > k, \qquad b_k = 0, b_i = 1 \text{ for all } \ i > k.$$

The sequences which are identified are sequences with tails of 0s and 1s which correspond to two possible choices of binary digits for a dyadic rational.

Theorem 1.7.1. *The map $\psi : \Sigma/\!\sim \to X$ is well defined and it is a conjugacy between the doubling map f and $\sigma : \Sigma/\!\sim \to \Sigma/\!\sim$.*

Exercise 1.7.1. *Prove Theorem 1.7.1.*

Exercise 1.7.2. *Let D be the set of dyadic rationals and let Σ' be the sequences which do not end with a tail of 0s or 1s. Verify that $\psi : \Sigma' \to X - D$ is a conjugacy.*

i.e. the binary expression of $f(x)$ is such that the digits are *shifted by* 1.

1.8 Baker's map

Let $[0,1)^2 = [0,1) \times [0,1)$ be the unit square. Consider the following two dimensional map $F : [0,1)^2 \to [0,1)^2$

$$F(x,y) = \begin{cases} \left(2x, \frac{y}{2}\right) & \text{if } 0 \le x < \frac{1}{2}, \\ \left(2x-1, \frac{y+1}{2}\right) & \text{if } \frac{1}{2} \le x < 1. \end{cases}$$

Geometrically, F is obtained by cutting $[0,1)^2$ into two vertical rectangles $R_0 = [0,1/2) \times [0,1)$ and $R_1 = [1/2,1) \times [0,1)$, stretching and compressing each to obtain an interval of horizontal width 1 and vertical height $1/2$ and then putting them on top of each other. The name *baker's*

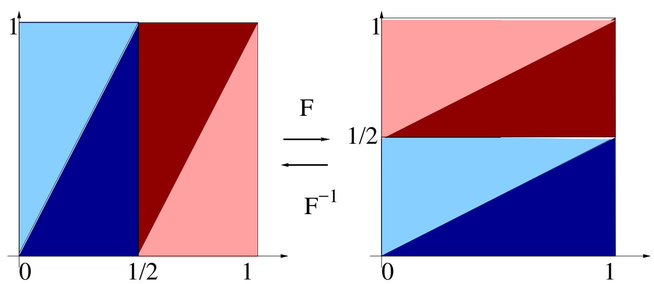


Figure 1.4: The action of the baker's map.

map comes becouse this mimic the movement made by a baker to prepare the bread dough[9]. Similar maps are often use in industrial processes since, as we will see formally later, they are very effective in quickly *mixing*.

Remark 1.8.1. *Notice while the horizontal direction is stretched by by a factor 2, the vertical direction is contracted by a factor $1/2$.*

The baker map is invertible. The inverse of the map F can be explicitely given by

$$F^{-1}(x,y) = \begin{cases} \left(\frac{x}{2}, 2y\right) & \text{if } 0 \le y < \frac{1}{2}, \\ \left(\frac{x+1}{2}, 2y-1\right) & \text{if } \frac{1}{2} \le y < 1. \end{cases}$$

Geometrically, F^{-1} cuts X into two horizontal squares and stretches each of them to double the height and divide by two the width and then places them one next to each other (in Figure 1.4, the right square now gives the departing rectangle decomposition and the left square shows the images of each rectangle under F^{-1}).

Unlike in the case of the doubling map, we now have to be more careful in identifying F as a map on $X = (\mathbb{R}/\mathbb{Z})^2$. The map $F : X \to X$

$$F(x,y) = \begin{cases} \left(\{2x\}, \frac{\{y\}}{2}\right) & \text{if } 0 \le \{x\} < \frac{1}{2}, \\ \left(\{2x\}, \frac{\{y\}+1}{2}\right) & \text{if } \frac{1}{2} \le \{x\} < 1. \end{cases}$$

[9]There are other versions of the baker's map where the dough is not cut, but folded over.

is well defined; here $\{y\} = y \mod 1$ denotes the fractional part of y.

If you compare the definition of the baker map F with the doubling map f in the previous section, you will notice that the horizontal coordinate is trasformed exactly as f. More precisely one can show that F is an extenstion of f (i.e. there is a semiconjugacy ψ such that $\psi \circ F = f \circ \psi$, see Exercise below). Extensions of non-invertible maps which are invertible are called *intervible extensions* or *natural extensions.*

Exercise 1.8.1. *Show that the doubling map $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ and the baker map $F : (\mathbb{R}/\mathbb{Z})^2 \to (\mathbb{R}/\mathbb{Z})^2$ are semi-conjugated and the semi-conjugacy is given by the projection $\pi : (\mathbb{R}/\mathbb{Z})^2 \to \mathbb{R}/\mathbb{Z}$ given by $\pi(x, y) = x$.*

To study the doubling map, we introduced the one-sided shift on two symbols ($\sigma^+ : \Sigma^+ \to \Sigma^+$). To study the baker map is natural to introduce the *bi-sided shift* on two symbols, that we now define.

Let $\Sigma = \{0, 1\}^{\mathbb{Z}}$ be the set of all bi-infinite sequences of 0 and 1:

$$\Sigma = \{(a_i)_{i=-\infty}^{\infty}, \quad a_i \in \{0, 1\}\}.$$

A *point* $\underline{a} \in \Sigma^+$ is a bi-sided sequence of digits 0,1, for example

$$\dots 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0 \dots$$

The (bi-sided) *shift* map σ is a map $\sigma : \Sigma \to \Sigma$ which maps a sequence to the *shifted* sequence:

$$\sigma((a_i)_{i=\infty}^{\infty}) = (b_i)_{i=-\infty}^{\infty}, \qquad \text{where} \quad b_i = a_{i+1}. \tag{1.13}$$

The sequence $(b_i)_{i=-\infty}^{\infty}$ is obtained from the sequence $(a_i)_{i=-\infty}^{\infty}$ by shifting all the digits one place to the left. For example, if

$$\begin{array}{lcccccccccccccccccc} (a_i)_{i=-\infty}^{\infty} & = & \dots & 0, & 0, & 1, & 1, & 0, & 1, & 0, & 0, & 0, & 1, & 1, & 0, & 1, & 1, & 1, & 1, & \dots \\ (b_i)_{i=-\infty}^{\infty} & = & 0, & 0, & 1, & 1, & 0, & 1, & 0, & 0, & 0, & 1, & 1, & 0, & 1, & 1, & 1, & 1, & \dots & \end{array}$$

Note that while the one-sided shift σ^+ was not invertible, because we were throwing away the first digit a_1 of the one-sided sequence $(a_i)_{i=1}^{\infty}$ before shifting to the left, the map σ is now invertible. The inverse σ^{-1} is simply the shift to the right.

One can show that the baker map F and the bi-sided shift σ are semi-conjugate if σ is restricted to a certain shift-invariant subspace (see Theorem 1.8.1 below).

In the case of the doubling map, the key was to use binary expansion. What to use now? We can get a hint of what is the semi-conjugacy using itineraries and trying to understand sets which share a common part of their itinerary.

Let R_0 and R_1 be the two basic rectangles

$$R_0 = \left[0, \frac{1}{2}\right) \times [0, 1)\,, \qquad R_1 = \left[\frac{1}{2}, 1\right) \times [0, 1)\,.$$

(See Figure 1.4, left square: R_0 is the left rectangle, R_1 the right one.)

The *(bi-infinite) itinerary* of (x, y) with respect to the partition $\{R_0, R_1\}$ is the sequence $(a_i)_{i=\infty}^{+\infty} \in \Sigma$ given by

$$\begin{cases} a_k = 0 & \text{if } F^k(x, y) \in R_0, \\ a_k = 1 & \text{if } F^k(x, y) \in R_1 \end{cases}$$

In particular, if $\dots, a_{-2}, a_{-1}, a_0, a_1, a_2, \dots$ is the *itineary* of $\mathcal{O}_F((x, y))$ we have

$$F^k(x, y) \in R_{a_k}, \quad \text{for all} \quad k \in \mathbb{Z}.$$

Note that here, since F is invertible, we can record not only the future but also the past.

Let us now define sets of points which share the same finite piece of itinerary. Given $n, m \in \mathbb{N}$ and $a_k \in \{0,1\}$ for $-m \le k \le n$, let

$$R_{-m,n}(a_{-m}, \dots, a_n) = \{(x,y) \in X | \quad F^k(x,y) \in R_{a_k} \text{ for } -m \le k \le n\}.$$

These are all points such that the block of the itineary from $-m$ to n is given by the digits

$$a_{-m}, a_{-m+1}, \dots, a_{-1}, a_0, a_1, \dots, a_{n-1}, a_n.$$

To construct such sets, let us rewrite them as

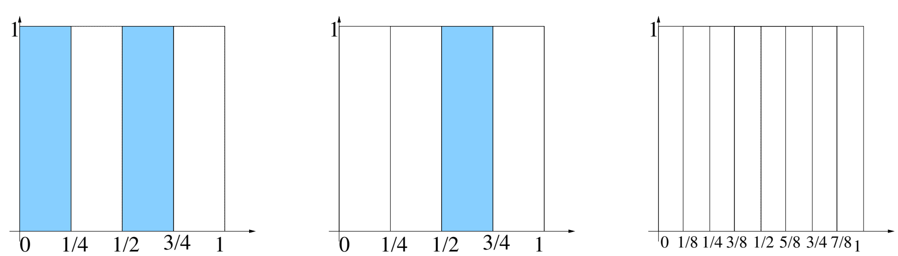
$$R_{-m,n}(a_{-m}, \dots, a_n) = \bigcap_{k=-m}^{n} F^{-k}(R_{a_k}).$$

Example 1.8.1. *Let us compute $F^{-1}(R_0)$. Either from the definition or from the geometric action of F^{-1}, one can see that (see Figure 1.5(a))*

$$F^{-1}(R_0) = \left[0, \frac{1}{4}\right) \times [0,1) \cup \left[\frac{1}{2}, \frac{3}{4}\right) \times [0,1).$$

Thus, (see Figure 1.5(b))

$$R_{0,1}(1,0) = R_1 \cap F^{-1}(R_0) = \left[\frac{1}{2}, \frac{3}{4}\right) \times [0,1).$$



(a) $F^{-1}(R_0)$ (b) $R_{0,1}(1,0)$ (c) $R_{0,2}(a_1, a_1, a_2)$

Figure 1.5: Examples of rectangles determined by future itineraries.

One can prove that all rectangles deterimed by forward itineraries, i.e. of the form $R_{0,n}(a_0, a_1, \dots, a_n)$, are thin vertical rectangles of width $1/2^{n+1}$ and full height, as in Figure 1.5(c), and as $a_0, \dots, a_n$ changes, they cover X. More precisely, recalling the intervals $I(a_0, \dots, a_n)$ defined for the doubling map[10], we have

$$R_{0,n}(a_0, a_1, \dots, a_n) = I(a_0, \dots, a_n) \times [0,1) = \left[\frac{k}{2^{n+1}}, \frac{k+1}{2^{n+1}}\right) \times [0,1) \quad \text{for some } 0 \le k < 2^{n+1}.$$

Let us now describe a set which share the same *past* itinerary.

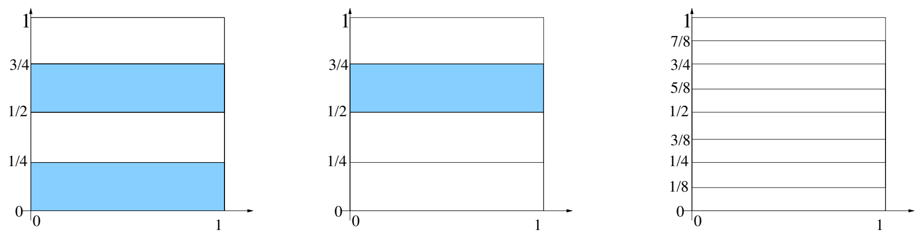
Example 1.8.2. *The image $F(R_0)$ is the bottom horizontal rectangle in the left square in Figure 1.4. The image $F^2(R_0)$ is shown in Figure 1.6(a) and is given by*

$$F^2(R_1) = [0,1) \times \left[0, \frac{1}{4}\right) \cup [0,1) \times \left[\frac{1}{2}, \frac{3}{4}\right).$$

[10]This is because the *future* history of F, i.e. whether $F^k(x,y)$ with $k \ge 0$ belongs to R_0 or R_1, is completely deterimed by the doubling map.

[Try to convince yourself by imagining the geometric action of F on these sets (or by writing an explicit formula)]. Hence, for example (see Figure 1.6(b))

$$R_{-2,-1}(0,1) = F^2(R_0) \cap F(R_1) = [0,1) \times \left[\frac{1}{2}, \frac{3}{4}\right).$$



(a) $F^2(R_0)$ (b) $R_{-2,-1}(0,1)$ (c) $R_{-3,-1}(a_{-3}, a_{-2}, a_{-1})$

Figure 1.6: Examples of rectangles determined by past itineraries.

In general, one can verify that each set of the form

$$R_{-n,-1}(a_{-n}, \dots, a_{-1})$$

(dependend only on the past itinerary) is a thin horizontal rectangle, of height $1/2^n$ and full width, as in Figure 1.6(c).

Exercise 1.8.2. *Draw the following sets:*

(a) $R_{-1,0}(0,1)$

(b) $R_{-1,1}(0,1,1)$

(c) $R_{-2,0}(1,0,1)$

In general $R_{-m,n}(a_{-m}, \dots, a_n)$ is a rectangle of horizontal width $1/2^{n+1}$ and height $1/2^m$.

The more we precise the backwards itinerary $a_{-1}, a_{-2}, \dots, a_{-n}$, the thinner the precision with which we determine the vertical component y. Moreover, from the geometric picture, you can guess that as $a_0, a_1, \dots, a_n, \dots$ give the digits of the binary expansion of x, $a_{-1}, a_{-2}, \dots, a_{-n}, \dots$ give the digits of the binary expansion of y. This is exactly the insight that we need to construct the semi-conjugacy with the full shift.

Now we are ready to construct a semi-conjugacty beteween the baker map and the full shift which is a conjugacy outside a measure zero set of points.

Denote by

$$T_1 = \{\underline{a} \in \Sigma : \exists i_0 \in \mathbb{Z} \text{ such that } a_i = 1 \forall i > i_0\}$$

the set of sequences with forward tails consisting only of 1s. We have $\sigma(T_1) = T_1$, i.e., T_1 is a shift-invariant subspace. Hence its complement $\tilde{\Sigma} = \Sigma \setminus T_1$ is also shift-invariant.

Theorem 1.8.1. *The baker map is semi-conjugated to the full shift $\sigma : \tilde{\Sigma} \to \tilde{\Sigma}$ via the map $\Psi : \Sigma \to X$ given by*

$$\Psi((a_i)_{i=-\infty}^{+\infty}) = (x,y) \quad \textit{where} \quad x = \sum_{i=1}^{\infty} \frac{a_{i-1}}{2^i} \mod 1, \quad y = \sum_{i=1}^{\infty} \frac{a_{-i}}{2^i} \mod 1.$$

As for the doubling map, binary expansions turns out to be crucial to build the map Ψ. While the *future* $(a_i)_{i=0}^{\infty}$ of the sequence $(a_i)_{i=-\infty}^{\infty}$ will be used to give the binary expansion of x, the *past* $(a_i)_{i=-\infty}^{-1}$ of the sequence $(a_i)_{i=-\infty}^{\infty}$ turns out to be related to the binary expansion of the vertical coordinate y.

Proof of Theorem 1.8.1. For every point (x, y), both x and y can be expressed in binary expansion. If x has a binary expansion of the form $a_0, \dots, a_{i_0}, 0, 1, 1, 1, 1, 1, \dots$ (i.e., with a forbidden tail), then x also has the binary expansion $a_0, \dots, a_{i_0}, 1, 0, 0, 0, 0, 0, \dots$ (exercise!). This shows that Ψ is surjective.

Thus it remains to check that $\Psi\sigma = F\Psi$. Let us first compute

$$\Psi\left(\sigma((a_i)_{i=-\infty}^{+\infty})\right) = \Psi\left((a_{i+1})_{i=-\infty}^{+\infty}\right) = \left(\sum_{i=1}^{\infty} \frac{a_i}{2^i} \mod 1, \sum_{i=1}^{\infty} \frac{a_{-i+1}}{2^i} \mod 1\right).$$

Let a_0 be the first digit of the binary expansion of x. We have that $a_0 = 0$ if $0 \le x < \frac{1}{2}$ and $a_0 = 1$ if $\frac{1}{2} \le x < 1$. For $x = \frac{1}{2}$ we have either $a_0 = 1$ (then $a_i = 0$ for $i \ge 1$) or $a_0 = 0$ (then $a_i = 1$ for $i \ge 1$). The latter has a forbidden tail and thus does not occur. Therefore $a_0 = 0$ if $0 \le x < \frac{1}{2}$ and $a_0 = 1$ if $\frac{1}{2} \le x < 1$, which allows us to write

$$F(x, y) = \left(2x \mod 1, \frac{y + a_0}{2}\right).$$

We compute

$$\begin{aligned} F(\Psi((a_i)_{i=-\infty}^{+\infty})) &= F\left(\sum_{i=1}^{\infty} \frac{a_{i-1}}{2^i}, \sum_{i=1}^{\infty} \frac{a_{-i}}{2^i}\right) \\ &= \left(2\sum_{i=1}^{\infty} \frac{a_{i-1}}{2^i} \mod 1, \frac{a_0}{2} + \frac{1}{2}\sum_{i=1}^{\infty} \frac{a_{-i}}{2^i}\right) = \left(\sum_{i=1}^{\infty} \frac{a_i}{2^i}, \sum_{i=1}^{\infty} \frac{a_{-i+1}}{2^i}\right), \end{aligned}$$

which is the same answer as for $\Psi\sigma$. □

Extra: Coding in dynamics

The idea of coding is very powerful in dynamical systems. Let $f : X \to X$ be a dynamical system and X a partition of X in finite (or sometimes countably) many pieces, that is a collection of sets

$$\{P_0, \dots, P_N\}, \qquad P_i \cap P_j = \emptyset \quad \forall\, i \neq j, \quad \text{and} \qquad \cup_{i=0}^{N} P_i = X.$$

Then we can defined the *itinerary* of $x \in X$ with respect to the partition $\{P_0, \dots, P_N\}$ as a sequence of digits a_k such that

$$f^k(x) \in P_{a_k}, \quad \text{for all } k.$$

Here the digits $a_k \in \{0, \dots, N\}$ (they take as many values as the number of elements in the partition). If f is not invertible, it makes sense to consider *forward itineraries*, so we get a sequence $(a_k)_{k=1}^{\infty} \in \{0, \dots, N\}^{\mathbb{N}}$. If f is invertible, we can also consider *backwards itineraries*, so we get a sequence $(a_k)_{k=-\infty}^{+\infty} \in \{0, \dots, N\}^{\mathbb{Z}}$.

If x has itinerary $(a_k)_{k=1}^{\infty} \in \{0, \dots, N\}^{\mathbb{N}}$ (or respectively $(a_k)_{k=-\infty}^{+\infty}$) then the itinerary of $f(x)$ will be shifted, so it will be given by $\sigma^+\left((a_k)_{k=1}^{\infty}\right)$ (or respectively $\sigma\left((a_k)_{k=-\infty}^{+\infty}\right)$). Thus, the coding maps an orbit of f into an orbit of the shift on a symbolic space. Two important questions to ask are:

- Does the itinerary completely determine the point x?
- Are all itineraries possible, i.e. does any sequence in Σ correspond to an actual itinerary?

In the case of the doubling map and of the baker map, both questions had positive answer. This allowed us to get a semi-conjugacy with the shift (one-sided or bi-sided). The itinerary is often enough to completely determine the point x: this is the case when the maps we are looking at are *expanding*. On the other hand, many times not all sequences arise as itineraries: we will describe in Chapter 3 more general symbolic spaces (subshifts of finite type) which will capture the behavior of many more maps.

1.9 Hyperbolic toral automorphisms

Consider the unit square $[0,1]\times[0,1]$. If you glue the two parallel vertical sides and the two parallel horizontal sides by using the identifications

$$(x,0)\sim(x,1),\quad x\in[0,1],\qquad (0,y)\sim(1,y),\quad y\in[0,1], \tag{1.14}$$

we get the surface of a doughnut, which is called a two-dimensional torus and denoted by $\mathbb{T}^2$ (when you glue the vertical sides first, you get a cylinder, then when you glue the horizontal sides you are gluing the two circles which bound the cylinder and you get the torus). More formally,

$$\mathbb{T}^2=\mathbb{R}^2/\mathbb{Z}^2=\mathbb{R}/\mathbb{Z}\times\mathbb{R}/\mathbb{Z}$$

is the set of equivalence classes of points $(x,y)\in\mathbb{R}^2$ modulo $\mathbb{Z}^2$: two points (x,y) and (x',y') in $\mathbb{R}^2$ are equivalent and we write

$$(x,y)\sim(x',y')\quad\text{if and only if}\quad (x-x',y-y')=(k,l),\quad \text{where } (k,l)\in\mathbb{Z}^2.$$

The unit square contains at least one representative for each equivalence class. Note though that since the points $(x,0)$ and $(x,1)$ for $x\in[0,1]$ differ by the vector $(0,1)$ they represent the same equivalence class. Similarly, the points $(0,y)$ and $(1,y)$ for $y\in[0,1]$, which differs by the vector $(1,0)$, also represent the same equivalence class, while all interior points represent distinct equivalence classes. Thus, the square with opposite sides glued by the identifications (1.14) contains exactly one representative for each equivalence class. In this sense, the space $\mathbb{R}^2/\mathbb{Z}^2$ of equivalence classes is represented by a square with opposite sides identified as in (1.14)

We will now study a class of maps of the torus. Let A be a 2×2 matrix with integer entries. Then A acts linearly on $\mathbb{R}^2$ (and in particular on the unit square):

$$\text{if}\quad A=\begin{pmatrix} a & b\\ c & d\end{pmatrix},\quad \text{where } a,b,c,d\in\mathbb{Z},\quad \text{then } A\begin{pmatrix} x\\ y\end{pmatrix}=\begin{pmatrix} ax+by\\ cx+dy\end{pmatrix}.$$

We claim that we can use this linear action to define a map $f_A:\mathbb{T}^2\to\mathbb{T}^2$ on the torus $\mathbb{T}^2=\mathbb{R}^2/\mathbb{Z}^2$, given by

$$f_A(x,y)=(ax+by \mod 1,\quad cx+dy \mod 1)\,. \tag{1.15}$$

We have to check that if $(x,y)\sim(x',y')$, then $f_A(x,y)\sim f_A(x',y')$. But $(x,y)\sim(x',y')$ means that $(x-x',y-y')=(k,l)$ where $(k,l)\in\mathbb{Z}^2$ and since A has integer entries, it maps integer vectors to integer vectors so

$$A\begin{pmatrix} x\\ y\end{pmatrix}-A\begin{pmatrix} x'\\ y'\end{pmatrix}=A\begin{pmatrix} k\\ l\end{pmatrix}\in\mathbb{Z}^2.$$

Thus, since f_A is defined taking entries $\mod 1$, we have $f_A(x,y) = f_A(x',y')$ as desired.

The map $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ is not invertible in general. However, if $det(A) = ad - bc = \pm 1$, then

$$A^{-1} = \frac{1}{det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \pm \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

is again a matrix with integer entries, so it induces a well-defined map $f_{A^{-1}} : \mathbb{T}^2 \to \mathbb{T}^2$. Thus, if $\det(A) = \pm 1$, then f_A is invertible and the inverse is given by

$$(f_A)^{-1} = f_{A^{-1}}.$$

Definition 1.9.1. *If A is an integer matrix, we say that $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ defined in (1.15) is a* toral endomorphism. *If furthermore $det(A) = \pm 1$, we say that $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ is a* toral automorphism.

Example 1.9.1 (Arnold CAT map)**.** *Let A be the matrix*

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Then the induced map $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ is

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

This map is known as Arnold's CAT map.[11] *To draw the action of the map, consider first the image of the unit square by the linear action of A: since the two bases vectors $\underline{e}_1, \underline{e}_2$ are mapped to*

$$A\,\underline{e}_1 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad A\,\underline{e}_2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

the image of the unit square is the parallelogram generated by $A\underline{e}_1, A\underline{e}_2$ (see Figure 1.7). To visualize f_A we need to consider the result modulo *1 in each coordinate (or, equivalently,* modulo $\mathbb{Z}^2$*), which means subtract to each point an integer vector (the vector whose components are the integer-parts of the x and y coordinate respectively) to get an equivalent point in unit square. Geometrically, this means* cutting and pasting *different triangles (each one consisting of all points of the parallelogram contained in a square of the unit square grid) back to the standard unit square, as shown in Figure 1.7.*

Definition 1.9.2. *The toral automorphism $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ given by a matrix A with integer entries and determinant ± 1 is called* hyperbolic *if A has no eigenvalues of modulus 1.*

Remark 1.9.1. *Note that if $det(A) = 1$, if $\lambda_1 > 1$ is one of the eigenvalues of A, the other eigenvalues is $\lambda_2 = 1/\lambda_1$ (since the product of the eigenvalues is the determinant). Thus, if one eigenvalue is greater than 1 in modulus, the other is automatically less than 1 in modulus.*

Example 1.9.2. *The matrix A has eigenvalues*

$$\lambda_1 = \frac{3+\sqrt{5}}{2} > 1 \quad \text{and} \quad \lambda_2 = \frac{2}{3+\sqrt{5}} = \frac{3-\sqrt{5}}{2} < 1.$$

Thus, f_A is a hyperbolic *toral automorphism.*

[11] In many books CAT is used as a shortening for Continuous Automorphism of the Torus. The first time the map was described in the book by Arnold, though, the action of the map is illustrated by drawing the face of a cat, as in Figure 1.7.

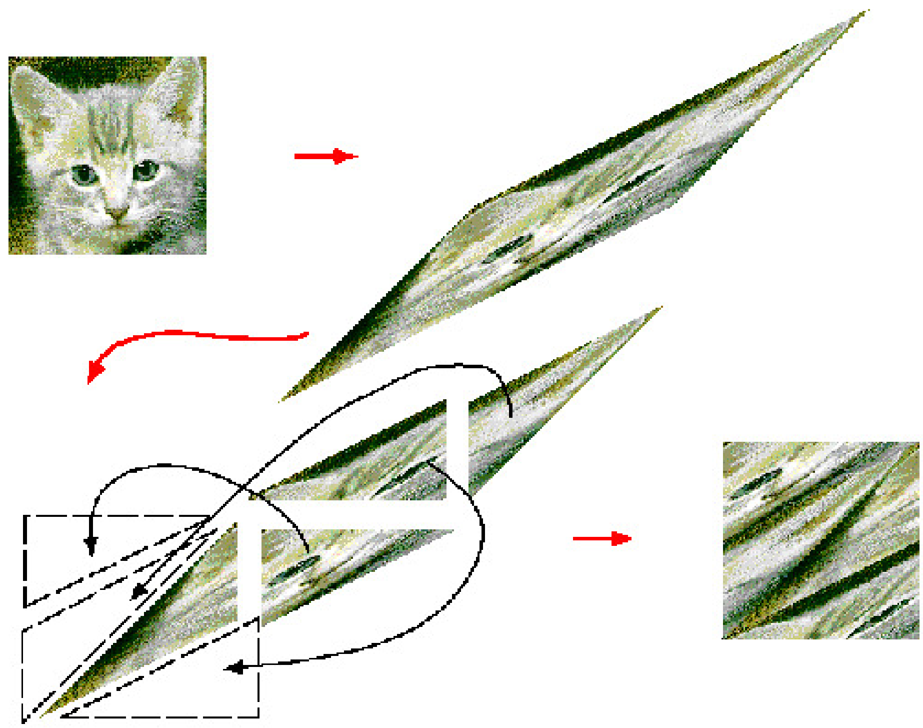


Figure 1.7: The Arnold CAT map.

The corresponding eigenvectors $\underline{v}_1$ *and* $\underline{v}_2$ *are*

$$\underline{v}_1 = \begin{pmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{pmatrix}, \qquad \underline{v}_2 = \begin{pmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{pmatrix}.$$

Thus, A expands all lines in the direction of v_1 *by* $\lambda_1 > 1$ *and contracts all lines in the direction of* v_2 *by multiplying them by* $\lambda_2 < 1$*. Since* f_A *is obtained by cutting and pasting the image of the unit square by A, the same is true for* f_A *(see Figure 1.7).*

Let us study periodic points of hyperbolic toral automorphisms.

Theorem 1.9.1. *Let* $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ *be a hyperbolic toral automorphism. The periodic points of* f_A *are exactly all points in* $[0,1) \times [0,1)$ *which have rational coordinates, that is all points*

$$\left(\frac{p_1}{q}, \frac{p_2}{q}\right), \qquad p_1, p_2 \in \mathbb{N}, \quad q \in \mathbb{N}\backslash\{0\}, \quad 0 \le p_1, p_2 < q.$$

Proof. If $(x, y) = \left(\frac{p_1}{q}, \frac{p_2}{q}\right)$ has rational coordinates, then

$$f_A^n(x, y) = \left(\frac{p_1^{(n)}}{q}, \frac{p_2^{(n)}}{q}\right),$$

where $p_1^{(n)}, p_2^{(n)}$ are integers $0 \le p_1^{(n)}, p_2^{(n)} < q$ which are given by

$$\begin{pmatrix} \frac{p_1^{(n)}}{q} \\ \frac{p_2^{(n)}}{q} \end{pmatrix} = A^n \begin{pmatrix} \frac{p_1}{q} \\ \frac{p_2}{q} \end{pmatrix} \mod \mathbb{Z}^2 = \begin{pmatrix} \frac{a_n p_1 + b_n p_2}{q} & \mod 1 \\ \frac{c_n p_1 + d_n p_2}{q} & \mod 1 \end{pmatrix}, \quad \text{if } A^n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix},$$

that is, which are given by

$$\begin{pmatrix} p_1^{(n)} \\ p_2^{(n)} \end{pmatrix} = \begin{pmatrix} a_n p_1 + b_n p_2 \mod q \\ c_n p_1 + d_n p_2 \mod q \end{pmatrix}, \quad \text{where } A^n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}.$$

Since there are at most q^2 choices for the pair $(p_1^{(n)}, p_2^{(n)})$, there exists $0 \le m \ne n \le q^2+1$ such that $f_A^m(x,y) = f_A^n(x,y)$. This means that (x,y) is eventually periodic and, since f_A is invertible, this implies that (x,y) is periodic (see Exercise 1.2(b)).

Conversely, if (x,y) is periodic, there is n such that $f_A^n(x,y) = (x,y)$ and by definition of f_A this means that there exists $k,l \in \mathbb{Z}$ such that

$$A^n \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} k \\ l \end{pmatrix} \quad \Leftrightarrow \quad (A^n - I)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} k \\ l \end{pmatrix}, \quad \text{where } I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Let us check that $(A^n - I)$ is invertible. Since f_A is hyperbolic, A has no eigenvalues of modulus 1, so A^n has no eigenvalues 1. Thus, there is no non-zero vector $\underline{v}$ that solves $(A^n - I)\underline{v} = 0$ and this shows that $(A^n - I)$ is invertible. Hence, we can solve:

$$\begin{pmatrix} x \\ y \end{pmatrix} = (A^n - I)^{-1} \begin{pmatrix} k \\ l \end{pmatrix}.$$

Since A has entries in $\mathbb{Z}$, $(A^n - I)$ has entries in $\mathbb{Z}$ and $(A^n - I)^{-1}$ has entries in $\mathbb{Q}$, so both x, y are rational numbers. □

See the Extra for an application of the previous thereom to explain why discretization of the CAT map (or in general, of a toral automorphism), for example by digital pixels on a screen, will eventually appear to be periodic.

One can precisely compute the number of periodic points of period n in a hyperbolic toral automorphism.

Theorem 1.9.2. *Let $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ be a hyperbolic toral automorphism associated to a matrix A with det$(A) = 1$ and eigenvalues λ_1 and λ_2. The number of periodic points of period n is $|\lambda_1^n + \lambda_2^n - 2|$.*

Proof. Fixed points of period n are solutions of $f_A^n(x,y) = (x,y)$. Equivalently, as we just proved,

$$(A^n - I)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} k \\ l \end{pmatrix}, \qquad \text{for some } k,l \in \mathbb{Z}. \tag{1.16}$$

The map $(A^n - I)$ maps the unit square $[0,1) \times [0,1)$ to a parallelogram P generated by the vectors $(A^n - I)\underline{e}_1$ and $(A^n - I)\underline{e}_2$. Thus, solutions to (1.16) correspond to integer points $(k,l) \in \mathbb{Z}^2$ which belong to the parallelogram P.

The number of such integer points is exactly the area of P in virtue of the following result known as Pick's theorem:

Theorem 1.9.3 (Pick's Theorem). *Consider a parallelogram $P \subset \mathbb{R}^2$ whose vertices are integer points in $\mathbb{Z}^2$ and let $P/\!\sim$ be obtained by identifying opposite parallel sides of P. Then the number of points in $P/\!\sim$ which integer coordinates is exactly equal to the area of P.*

A more general form of Pick's theorem and its proof are included as an Extra.

Thus, the number of integer points in P is exactly the area of P which is given by $|det(A^n - I)|$. Since the determinant is the product of the eigenvalues, it is enough to compute the eigenvalues of $(A^n - I)$. If $\underline{v}$ is an eigenvector of $(A^n - I)$ with eigenvalue μ, then

$$(A^n - I)\underline{v} = \mu \underline{v} \quad \Leftrightarrow \quad A^n \underline{v} = (\mu + 1)\underline{v},$$

so $\underline{v}$ is an eigenvector for A^n with eigenvector $\mu + 1$. Since A has eigenvectors λ_1, λ_2, A^n has eigenvectors λ_1^n, λ_2^n and then $(A^n - I)$ has eigenvectors $\lambda_1^n - 1, \lambda_2^n - 1$. So, putting everything together we proved that

$$\begin{aligned} Card\{(x, y) \text{ s.t. } f_A^n(x, y) = (x, y)\} = Area(P) = |det(A^n - I)| &= |(\lambda_1^n - 1)(\lambda_2^n - 1)| \\ &= |\lambda_1^n + \lambda_2^n - 2|, \end{aligned}$$

where in the last inequality we used that $\lambda_1^n \lambda_2^n = 1$, since λ_1^n and λ_2^n are the eigenvalues of A^n, which has determinant 1 since $det(A^n) = \det(A)^n$ and $det(A) = 1$ by assumption. □

Remark 1.9.2. *All definitions in this section generalize to higher dimensions. The k-dimensional torus $\mathbb{T}^k$ is $\mathbb{R}^k/\mathbb{Z}^k$. Given a $k \times k$ matrix A with integer entries, it determines a map $f_A : \mathbb{T}^k \to \mathbb{T}^k$. If $|det(A)| = 1$, f_A is called a toral automorphism. If A has no eigenvalues of modulus 1, f_A is an hyperbolic toral automorphism. Periodic points for f_A are exactly points with rational coordinates.*

Extra: A discrete toral automorphism

As an illustration of Theorem on periodic points of hyperbolic toral automorphisms, one can see Figure 1.8, in which iterates of a discretization of the Cat Map is plotted. If the image is made by a finite number of pixels, say q^2, each representing a point $(p_1/q, p_2/q)$ with rational coordinates, then an iterate of f_A is the identity by Theorem 1.9.1 proved in this section.

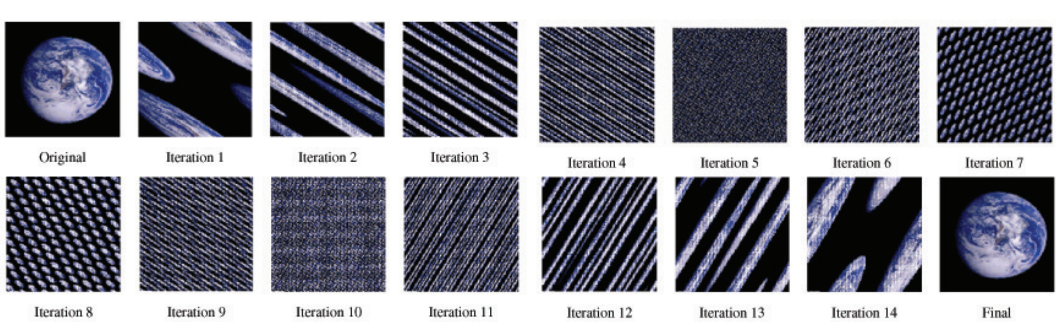


Figure 1.8: A discrete version of the Arnold CAT map.

Extra: Pick's theorem on the area of a parallelogram with integer vertices

Let us state and prove Pick's Theorem (which was used in the computation of the number of periodic points for a hyperbolic toral automorphism).

Theorem 1.9.4 (Pick's Theorem)**.** *The area of a parallelogram $P \subset \mathbb{R}^2$ whose vertices are integer points in $\mathbb{Z}^2$ is given by the number of points of $\mathbb{Z}^2$ which are contained, where points on the edges are counted as half and all vertices count as a single point. More precisely, let i be the number of points with integer coordinates in the interior of P and b be the number of points with integer coordinates on the perimeter of P. Then, $Area(P) = i + (b/2) + 1$.*

Let us remark that this form implies as a corollary the formulation used in the lecture notes. Indeed, consider $P/\!\sim$ obtained by glueing opposite parallel sides of P. Then the points with integer coordinates on the sides of P are glued together in pairs, and all the vertices of P are identifyed to produce the equivalence class of a unique point on $P/\!\sim$. Thus, the points with integer coordinates in $P/\!\sim$ are exactly $i + (b/2) + 1$. so the number of integer po $P/\!\sim$ Before the proof, let us present some lemmas.

Lemma 1.9.1. *If a polygon P as in the assumptions of Pick's theorem is divided into two smaller polygons, P_1 and P_2 by a path whose endpoints also belong to $\mathbb{Z}^2$, then, if Pick's formula holds for both P_1 and P_2, it also holds for P.*

Proof. The interior points with integer coordinates inside P, whose cardinality we denote by i, all fall either into the interior of P_1 (i_1 of them) or into the interior of P_2 (i_2 of them) or on the path that was drawn to divide P (i_3 of them), so $i = i_1 + i_2 + i_3$. Two of the boundary points of P (set $b_3 = 2$) are the endpoints of the dividing path that formed P_1 and P_2, while b_1 other boundary points are boundary points of P_1, and b_2 boundary points are boundary points of P_2, so $b_1 + b_2 + b_3 = b$.

By Pick's formula applied to each of the smaller polygons,

$$Area(P_1) = i_1 + \frac{b_1 + b_3 + i_3}{2} + 1, \qquad Area(P_2) = i_2 + \frac{b_2 + b_3 + i_3}{2} + 1,$$

and clearly $Area(P_1) + Area(P_2) = Area(P)$, so we get

$$\begin{aligned} Area(P) = Area(P_1) + Area(P_2) &= i_1 + (b_1 + b_3 + i_3)/2 + 1 + i_2 + (b_2 + b_3 + i_3)/2 + 1 \\ &= i_1 + i_2 + i_3 + (b_1 + b_2 + 2b_3)/2 + 2 \\ &= i + (b/2) + b_3/2 + 2, \end{aligned}$$

and since $b_3 = 2$, we have proved that $i + (b/2) + 1 = Area(P)$, that is, that Pick's formula holds for P. □

Lemma 1.9.2. *If P is a triangle with vertices with integer coordinates, Pick's formula holds.*

Sketch of Proof. The verification can be done by the following steps, using each time Lemma 1.9.1. Observe that the formula holds for any unit square (with vertices having integer coordinates). Deduce from this that the formula is correct for any rectangle with sides parallel to the axes and integer side lengths. Deduce now that it holds for right-angled triangles obtained by cutting such rectangles along a diagonal. Now any triangle with integer coordinates can be turned into a rectangle of this type by attaching (at most three) such right triangles; since the formula is correct for the right triangles and for the rectangle, it also follows for the original triangle. □

Proof of Pick's Theorem 1.9.4. Every n-sided polygon ($n > 3$) can be subdivided into two polygons each with fewer than n sides. By repeating this action, a polygon can be completely decomposed into triangles. By Lemma 1.9.2, Pick's formula is correct for each triangle and by Lemma 1.9.1 this proves Pick's formula for any polygon. □

1.10 Gauss map and continued fractions

In this lecture we will introduce the Gauss map, which is very important for its connection with continued fractions in number theory.

The *Gauss map* $G : [0,1] \to [0,1]$ is the following map:

$$G(x) = \begin{cases} 0 & \text{if } x = 0 \\ \left\{\frac{1}{x}\right\} = \frac{1}{x} \mod 1 & \text{if } 0 < x \leq 1 \end{cases}$$

Here $\{x\}$ denotes the *fractional part* of x. We can write $\{x\} = x - [x]$ where $[x]$ is the integer part. Equivalently, $\{x\} = x \mod 1$.

Note that

$$\left[\frac{1}{x}\right] = n \quad \Leftrightarrow \quad n \leq \frac{1}{x} < n+1 \quad \Leftrightarrow \quad \frac{1}{n+1} < x \leq \frac{1}{n}.$$

Thus, explicitely, one has the following expression (see the graph in Figure 1.9):

$$G(x) = \begin{cases} 0 & \text{if } x = 0 \\ \frac{1}{x} - n & \text{if } \frac{1}{n+1} < x \leq \frac{1}{n} \end{cases} \qquad \text{for } n \in \mathbb{N}.$$

The rescrition of G to an interval of the form $(1/n+1, 1/n]$ is called *branch*. Each *branch* $G : (1/n+1, 1/n] \to [0,1)$ is monotone, surjective (onto $[0,1)$) and invertible (see Figure 1.9).

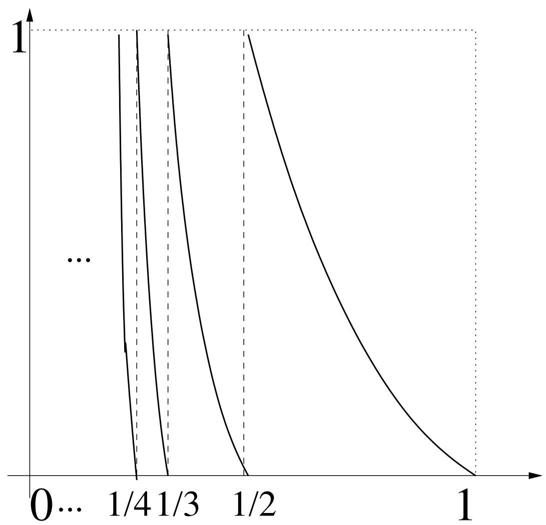


Figure 1.9: The first branches of the graph of the Gauss map.

The Gauss map is important for its connections with continued fractions.

A *finite continued fraction* (CF will be used as shortening for Continued Fraction) is an expression of the form

$$\cfrac{1}{a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots \frac{1}{a_n}}}}} \tag{1.17}$$

where $a_0, a_1, a_2, \dots, a_n \in \mathbb{N}\backslash\{0\}$ are called *entries* of the continued fraction expansion. We will denote the finite continued fraction expansion by $[a_0, a_1, a_2, \dots, a_n]$.

Every finite continued fraction expansion correspond to a rational number p/q (which can be obtained by clearing out denominators).

Example 1.10.1. *For example*

$$\frac{1}{2+\frac{1}{3}} = \frac{1}{\frac{2\cdot 3+1}{3}} = \frac{3}{7}.$$

Conversely, all rational numbers in $[0,1]$ admit a representation as a finite continued fraction[12].

Example 1.10.2. *For example*

$$\frac{3}{4} = \frac{1}{1+\frac{1}{3}}, \qquad \frac{49}{200} = \cfrac{1}{3+\cfrac{1}{4+\frac{1}{12+\frac{1}{4}}}}.$$

[12]This representatin is not unique: if the last digit a_n of a finite CF is 1, then $[a_0, \dots, a_{n-1}, 1] = [a_0, \dots, a_{n-1} + 1]$. If one requires that the last entry is different from one, though, then one can prove that the representation as finite continued fraction is unique.

Every *irrational* number $x \in (0,1)$ can be expressed through a (unique) *infinite* continued fraction[13], that we denote by

$$[a_0, a_1, a_2, a_3, \dots] = \cfrac{1}{a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}}}.$$

Example 1.10.3. *For example*

$$\pi = 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cfrac{1}{293 + \dots}}}},$$

$$\cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \dots}}}} = \frac{\sqrt{5} - 1}{2}.$$

The number $(\sqrt{5} - 1)/2$ *is known as* golden mean[14] *and it has the lowest possible continued fraction entries, all entries equal to one. Similarly, the number whose CF entries are all equal to* 2 *is known as* silver mean.

One can see that a number is rational if and only if the continued fraction expansion is finite.

If x is an irrational number whose infinite continued fraction expansion is $[a_0, a_1, a_2, \dots]$, one can *truncate* the continued fraction expansion at level n and obtain a rational number that we denote p_n/q_n

$$\frac{p_n}{q_n} = [a_0, a_1, a_2, \dots, a_n].$$

These numbers p_n/q_n are called *convergents* of the continued fraction.

Two of the important properties of convergents are the following:

1. One can prove that p_n/q_n converge to x exponentially fast, i.e.

$$\lim_{n\to\infty} \frac{p_n}{q_n} = x \quad \text{and} \quad \left| \frac{p_n}{q_n} - x \right| \leq \frac{1}{(\sqrt{2})^n}. \tag{1.18}$$

   Thus, the fractions p_n/q_n give *rational approximations* of x.

2. Convergents give *best approximations* among all rational approximations with denominator up to q_n, that is

$$\left| x - \frac{p_n}{q_n} \right| \leq \left| x - \frac{p}{q} \right|, \quad \forall \;\; p \in \mathbb{Z}, \quad 0 \leq q \leq q_n.$$

One can also see that the continued fraction expansion of an irrational number is unique.

To find the continued fraction expansion of a number, we will exploit the relation with the symbolic coding of the Gauss map, in the same way that binary expansions are related to the symbolic coding of the doubling map.

[13]To be precise, when we write such an infinite continued fraction expression, its value is the limit of the finite continued fraction expansion truncations $[a_0, a_1, a_2, a_3, \dots, a_n]$, each of which is a well defined rational number. One should first prove that this limit exist, see (1.18).

[14]The inverse of the golden mean is $\frac{\sqrt{5}+1}{2}$, known as *golden ratio*. It appears often in art and in nature since it is considered aesthetically pleasing: for example, the ratio of the width and height of the facade of the Partenon in Athens is exactly the golden ratio and a whole Renessaince treaty, Luca Pacioli's *De divina proportione*, written in 1509, is dedicated to the golden ratio in arts, science and architecture.

Let P_n be the subintervals of $[0,1)$ naturally determined by the domains of the branches of the Gauss map:

$$P_1 = \left(\frac{1}{2}, 1\right], \quad P_2 = \left(\frac{1}{3}, \frac{1}{2}\right], \quad P_3 = \left(\frac{1}{4}, \frac{1}{3}\right], \quad \dots, P_n = \left(\frac{1}{n+1}, \frac{1}{n}\right], \dots$$

Note that P_n accumulate towards 0 as n increases If we add $P_0 = \{0\}$, the collection $\{P_0, P_1, \dots, P_n, \dots\}$ is a (countable) partition[15] of $[0,1]$.

Theorem 1.10.1. *Let x be irrational. Let $a_0, a_1, \dots, a_n, \dots$ be the itinerary of $\mathcal{O}_G^+(x)$ with respect to the partition $\{P_0, P_1, P_2, \dots, P_n, \dots\}$, i.e.*

$$x \in P_{a_0}, G(x) \in P_{a_1}, \dots, G^2(x) \in P_{a_2}, \dots, G^k(x) \in P_{a_k}, \dots,$$

Then $x = [a_0, a_1, a_2, \dots, a_n, \dots]$. Thus, itineraries of the Guass map give the entries of the continued fraction expansions.

Remark 1.10.1. *If x is rational, then there exists n such that $G^n(x) = 0$ and hence $G^m(x) = 0$ for all $m \geq n$. In this case, $G^m(x) \in P_0$ for all $m \geq n$ so the itinerary is eventually zero. The theorem is still true if we consider the beginning of the itineary: the finite itinerary before the tail of 0 gives the entries of the finite continued fraction expansion of x.*

Proof. Let us first remark that

$$x \in P_n \quad \Leftrightarrow \quad \frac{1}{n+1} < x \leq \frac{1}{n} \quad \Leftrightarrow \quad n \leq \frac{1}{x} < n+1 \quad \Leftrightarrow \quad \left[\frac{1}{x}\right] = n. \tag{1.19}$$

In particular, $a_0 = [1/x]$ since $x \in P_{a_0}$. Thus,

$$G(x) = \left\{\frac{1}{x}\right\} = \frac{1}{x} - \left[\frac{1}{x}\right] = \frac{1}{x} - a_0 \quad \Leftrightarrow \quad x = \frac{1}{a_0 + G(x)}.$$

Let us prove by induction that

$$a_n = \left[\frac{1}{G^n(x)}\right] \quad \text{and} \quad x = \frac{1}{a_0 + \frac{1}{a_1 + \dots \frac{1}{a_n + G^{n+1}(x)}}} = [a_0, a_1, \dots, a_n + G^{n+1}(x)]. \tag{1.20}$$

We have already shown that this is true for $n = 0$. Assume that it is proved for n and consider $n+1$. Since $G^{n+1}(x) \in P_{a_{n+1}}$ by definition of itinerary, we have $a_{n+1} = \left[\frac{1}{G^{n+1}(x)}\right]$ by (1.19). This proves the first part of (1.20) for $n+1$. Then, recalling the definition of G we have

$$G^{n+2}(x) = \frac{1}{G^{n+1}(x)} - \left[\frac{1}{G^{n+1}(x)}\right] = \frac{1}{G^{n+1}(x)} - a_{n+1} \qquad \Leftrightarrow \qquad G^{n+1}(x) = \frac{1}{a_{n+1} + G^{n+2}(x)}$$

so that, plugging that in the second part of the inductive assumption (1.20) we get

$$x = \frac{1}{a_0 + \dots \frac{1}{a_n + G^{n+1}(x)}} = \frac{1}{a_0 + \dots \frac{1}{a_n + \frac{1}{a_{n+1} + G^{n+2}(x)}}},$$

which proves the second part of (1.20) for $n+1$. Thus, recursively, the itinerary is producing[16] the infinite continued fraction expansion of x. □

[15]Recall that a partition is a collection of disjoint sets whose union is the whole space.

[16]One should still prove that the finite continued fractions in (1.20) do converge, as n tends to infinity and that the limit is x. This can be done by the same method that one can use to show that convergents tend to x exponentially fast.

From the proof of the previous theorem, one can see the following.

Remark 1.10.2. *The Gauss map acts on the digits of the CF expansion as the one-sided shift, that is*

$$\text{if } x = [a_0, a_1, a_2, \dots, a_n, \dots]$$
$$\text{then } G(x) = [a_1, a_2, a_3, \dots, a_{n+1}, \dots].$$

One can characterize in terms of orbits of the Gauss map various class of numbers. For example:

1. *Rational numbers* are exactly the numbers x which have *finite* continued fraction expansion or equivalently such that there exists $n \in \mathbb{N}$ such that $G^n(x) = 0$ (*eventually mapped to zero by the Gauss map*).

2. *Quadratic irrationals*, that is numbers of the form $\frac{a+b\sqrt{c}}{d}$, where a, b, c, d are integers[17], are exactly numbers which have a *eventually periodic* continued fraction expansion or equivalently are *pre-periodic points for the Gauss map.*

In number theory (and in particular in Diophantine approximation) other class of numbers (for example Badly approximable numbers) can be characterized in terms of their continued fraction expansion[18].

Example 1.10.4. *We have already seen two examples of quadratic irrationals, the golden mean g and the silver mean s:*

$$g = \cfrac{1}{1 + \cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\dots}}}} = \frac{\sqrt{5}-1}{2}, \qquad s = \cfrac{1}{2 + \cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{2+\dots}}}} = \sqrt{2} - 1.$$

Both the golden mean and the silver mean are fixed points of the Gauss map: $G(g) = g$, $G(s) = s$. Similarly all other fixed points correspond to numbers whose continued fraction entries are all equal.

Example 1.10.5. *Let $\alpha = \frac{-3+\sqrt{15}}{2}$. Then one can check that $\alpha = [2, 3, 2, 3, 2, 3, \dots]$, so that the entries are periodic and the period is 2. Thus $G^2(\alpha) = \alpha$. Explicitely, since we know the itinerary of α, we can write down the equation satisfied by α. We know that*

$$G(\alpha) = \frac{1}{\alpha} - 2, \quad \textit{since} \left[\frac{1}{\alpha}\right] = 2, \quad \textit{and} \quad G(G(\alpha)) = \frac{1}{G(\alpha)} - 3 \quad \textit{since} \quad \left[\frac{1}{G(\alpha)}\right] = 3,$$

so that the equation $G^2(\alpha) = \alpha$ becomes

$$\frac{1}{\frac{1}{\alpha} - 2} - 3 = \alpha.$$

Using the ideas in the previous exercise, one can produce quadratic irrationals with any given periodic sequence of CF entries.

Exercise 1.10.1. *Prove that if $G^n(x) = 0$ then x has a representation as a finite continued fraction expansion and thus it is rational.*

Exercise 1.10.2. *Prove that if $G^n(x) = x$ then x satisfies an equation of degree two with integer entries. Conclude that x is a quadratic irrational.*

[17]Equivalently, one can define quadratic irrationals as solutions of equations of degree two with integer coefficients.

[18]One can defined Badly approximable numbers as the numbers for which there exists a number A such that all entries a_n of their continued fraction expansion are bounded by A. In particular, quadratic irrationals are badly approximable.