Chapter I
Introduction

The first chapter is devoted to simple and explicit examples of dynamical systems that illustrate some concepts and help to ask the appropriate questions. For simplicity, the systems under consideration are discrete and hence given by mappings acting on sets. The aim is to study the behavior of points under all iterates of a map (orbits of the points) and also to see what happens under perturbation of a map. A dynamical system consisting of a continuous map acting on a topological space is called transitive, if it possesses a dense orbit. The transitivity of a system will be guaranteed by the criterion of G. Birkhoff. An example of such a system is the rigid irrational rotation of the unit circle where every orbit of the system is dense on the circle. This example will lead us to the equidistribution (mod 1) theorem of H. Weyl. In sharp contrast to the stable systems of rigid rotations, a simple expansive map on the circle shows already a quite chaotic behavior described by the shift map in a sequence space. In this example, orbits of completely different behavior over a long-time interval (many iterates) coexist side by side. In the language of physics, the system shows a sensitive dependence on the initial conditions. It is a typical phenomenon that such an unstable behavior survives under a perturbation of the system, as will be demonstrated by a special case of the so-called structural stability theorem. Measure preserving mappings acting on a measure space will show strong recurrence properties and the question arises, how an orbit of such a system is distributed statistically in the space. As an answer we shall prove the individual ergodic theorem of G. Birkhoff following the strategy designed by A. M. Garsia.

The origin of the field of dynamical systems lies in the deep mathematical problems of celestial mechanics. That is why we shall first recall the N-body system whose dynamics is determined by the Newton equations.

I.1 N-body problem of celestial mechanics

In the N-body problem of celestial mechanics one studies N points $x_k \in \mathbb{R}^3$ in the 3-dimensional Euclidean space having masses $m_k > 0$. The *evolution in time* of these mass points,

$$x_k(t), \quad 1 \le k \le N,$$

is determined by the Newton equations

$$m_k \ddot{x}_k = \sum_{j \neq k} m_k m_j \frac{x_j - x_k}{|x_j - x_k|^3}, \quad 1 \le k \le N,$$

according to which every mass point x_k is attracted by every other mass point, which gives rise to an extremely complicated dynamics. The equation is not defined at *the collisions* $x_j = x_k$ for $j \neq k$. Introducing the collision sets $\Delta_{ij} = \{x \in \mathbb{R}^{3N} \mid x_i = x_j\}$ and

$$\Delta = \bigcup_{i<j} \Delta_{ij},$$

the Newton equations are defined on the *configuration space*

$$\mathbb{R}^{3N} \setminus \Delta \subset \mathbb{R}^{3N}$$

which is the set of points $x = (x_1, \ldots, x_N)$ without collisions. We reformulate the Newton equations as a system of first-order differential equations in the form of a *vector field* as

$$\begin{aligned} \dot{x}_k &= y_k \in \mathbb{R}^3, \\ \dot{y}_k &= \sum_{j \neq k} m_j \frac{x_j - x_k}{|x_j - x_k|^3} \in \mathbb{R}^3. \end{aligned}$$

The *phase space* of the N-body system is the set

$$\Omega = (\mathbb{R}^{3N} \setminus \Delta) \times \mathbb{R}^{3N} \subset \mathbb{R}^{6N}.$$

Denoting the points in the phase space by $z = (x, y) = (x_1, \ldots, x_N, y_1, \ldots, y_N)$, we write the equations in short form as

$$\dot{z} = V(z), \quad z \in \Omega \subset \mathbb{R}^{6N}.$$

The vector field $V\colon \Omega \subset \mathbb{R}^{6N} \to \mathbb{R}^{6N}$ is continuously differentiable and consequently, in particular, *locally Lipschitz-continuous*. The development of the system in time is described by a solution, which is a *continuously differentiable curve*

$$t \mapsto z(t) \in \Omega,$$

solving the equation

$$\dot{z}(t) = V(z(t))$$

for the time t in an open interval. Prescribing the point

$$z(0) = (x(0), y(0)) \in \Omega,$$

we are confronted with the Cauchy initial value problem. In view of the classical Cauchy–Lipschitz–Picard theorem in ordinary differential equations, there exists for every initial condition $z(0)$ exactly one solution $z(t)$ satisfying $z(t) = z(0)$ at the time $t = 0$ and this solution exists on an open interval

$$t \in I = I(z(0), V).$$

Locally the Cauchy initial value problem is not of great interest for us. We are interested in a long-term prediction of the future of the system as well as in a reconstruction of its past. Over long-time intervals solutions can behave very differently, even if their initial conditions are very close. Concerning the long-time behavior of the N-body-problem the following natural questions arise.

1. Are there solutions without collisions, which do exist for all times $t \in \mathbb{R}$, hence satisfying the estimate

$$0 < |x_j(t) - x_k(t)| < \infty, \quad j \neq k, \ t \in \mathbb{R}?$$

2. Are there singularities that are not collisions? Are there solutions that explode in finite time $t^* \in \mathbb{R}$ without collisions, so that

$$\mathrm{diam}\{x_1(t), \ldots, x_N(t)\} \to \infty, \quad t \to t^* \in \mathbb{R}?$$

   This question of P. Painlevé was answered only in 1988 by J. Xia who succeeded in finding such solutions in the $N = 5$-body-problem. His construction relies on the analysis of the 3-body-collision by R. McGehee. The history of the problem is described in [95] by D. Saari and J. Xia.

3. Does our solar system (in which one of the masses is much bigger than the others) keep its present nice shape for all future times or does, in the distant future, one of the planets, e.g. Jupiter, escape, or will a collision provoke a dramatic change of the situation?

The planetary system is an example of a classical dynamical system, defined by an ordinary differential equation on a manifold.

We now turn to discrete dynamical systems which are defined by a mapping acting on a set. The number of iterations of the map plays the role of the time.

I.2 Mappings as dynamical systems

Let X be a set, $x \in X$ a point in the set and let $\varphi\colon X \to X$ be a mapping of the set into itself. Iterating the map φ we obtain the following sequence of points:

$$\begin{aligned}
x_0 &= x = \varphi^0(x),\\
x_1 &= \varphi(x),\\
x_2 &= \varphi(x_1) = \varphi \circ \varphi(x) = \varphi^2(x),\\
&\ \vdots\\
x_j &= \varphi \circ \cdots \circ \varphi(x) = \varphi^j(x), \quad j \geq 1.
\end{aligned}$$

Definition. Given $x \in X$, the *sequence* $(\varphi^j(x))_{j\geq 0}$ is called the *parameterized orbit* of φ through the point x. The *set*

$$\mathcal{O}^+(x) := \bigcup_{j\geq 0}\{\varphi^j(x)\} \subset X$$

is called the *unparametrized orbit* (or simply *orbit*) of x under the mapping φ.

A point $x \in X$ is called a *periodic point* of φ, if there exists an integer $N \geq 1$ satisfying

$$\varphi^N(x) = x.$$

The corresponding (unparametrized) orbit consists of the *finitely many* points

$$x,\ \varphi(x),\ \varphi^2(x), \dots, \varphi^N(x) = x$$

and satisfies

$$\varphi^{N+1}(x) = \varphi(\varphi^N(x)) = \varphi(x), \quad \varphi^{N+2}(x) = \varphi^2(x), \quad \dots .$$

The sequence $(\varphi^j(x))_{j\geq 0}$ through a periodic point is periodic and the orbit is the finite set

$$\mathcal{O}^+(x) = \bigcup_{0\leq j\leq N-1}\{\varphi^j(x)\} \quad \text{and} \quad |\mathcal{O}^+(x)| \leq N.$$

The cardinality of the set is equal to $|\mathcal{O}^+(x)| = N$ if the period is *minimal*, i.e., if $\varphi^j(x) \neq x$ for $1 \leq j \leq N-1$ and $\varphi^N(x) = x$.

We are interested in the *asymptotic* behavior of the orbits. How do the sequences $(\varphi^j(x))_{j\geq 0}$ behave in the limit as $j \to \infty$?

Example (Contraction principle of Banach). We require additional conditions on the set and the mapping under consideration and assume (X, d) to be a complete metric space and the mapping $\varphi\colon X \to X$ to be a *contraction*. Hence, there exists a constant $0 \leq \lambda < 1$, satisfying

$$d(\varphi(x), \varphi(y)) \leq \lambda\, d(x, y) \quad \text{for all } x, y \in X.$$

It is well known that under these assumptions, there exists precisely one *fixed point* $x^* \in X$ satisfying $\varphi(x^*) = x^*$. In addition, for every point $x \in X$ the orbit $(\varphi^j(x))_{j\geq 0}$ converges to the fixed point,

$$\lim_{j\to\infty} \varphi^j(x) = x^*.$$

In this example the asymptotic behavior is easy to describe. *Every* orbit eventually comes to a standstill at the fixed point x^*. The set $\{x^*\}$ is a *global attractor* of the mapping φ. In addition, the system is *stable under perturbation*!

If the mapping $\varphi\colon X \to X$ is *bijective*, we can iterate also the inverse mapping $\varphi^{-1}\colon X \to X$ and study the (two-sided) *sequences*

$$(\varphi^j(x))_{j\in\mathbb{Z}}.$$

Accordingly, we distinguish between the positive orbit of a point, the negative orbit and the (full) orbit for which we introduce the following notation:

$$\mathcal{O}^+(x) = \bigcup_{j\geq 0} \varphi^j(x), \quad \mathcal{O}^-(x) = \bigcup_{j\geq 0} \varphi^{-j}(x), \quad \mathcal{O}(x) = \bigcup_{j\in\mathbb{Z}} \varphi^j(x).$$

For a given point $x \in X$ we now try to determine *the future as well as to reconstruct the past*!

Example (Rigid rotations). We look at the unit circle $S^1 \subset \mathbb{C}$, defined by

$$S^1 = \{z \in \mathbb{C} \mid |z| = 1\} = \{z = e^{2\pi i x} \mid x \in \mathbb{R}\}$$

and study the rotation φ of the circle around the angle $2\pi\alpha$ for a real number $\alpha \in \mathbb{R}$, in complex notation given by the map

$$\varphi\colon S^1 \to S^1, \quad z \mapsto \theta z, \quad \text{where } \theta = e^{2\pi i\alpha},$$

so that

$$\varphi(e^{2\pi i x}) = e^{2\pi i\alpha}e^{2\pi i x} = e^{2\pi i(x+\alpha)}.$$

In the covering space $\mathbb{R}$ of S^1 the map is the *translation* $\Phi\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x + \alpha$.

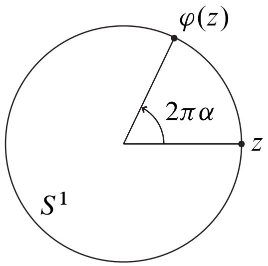

Figure I.1. Rigid rotation.

The relation between the mappings φ and Φ is described by the covering map

$$p\colon \mathbb{R} \to S^1, \quad x \mapsto p(x) := e^{2\pi i x} \in S^1$$

and one reads off that

$$p(x + j) = p(x), \quad j \in \mathbb{Z},$$
$$\Phi(x + j) = \Phi(x) + j, \quad j \in \mathbb{Z}$$

and

$$\varphi(p(x)) = p(\Phi(x)).$$

The last equation expresses the commutativity of the diagram

$$\begin{array}{ccc} S^1 & \xrightarrow{\varphi} & S^1 \\ {\scriptstyle p}\uparrow & & \uparrow{\scriptstyle p} \\ \mathbb{R} & \xrightarrow{\Phi} & \mathbb{R}. \end{array}$$

On the quotient space $\mathbb{R}/\mathbb{Z}$ the map Φ induces the map $\widehat{\Phi}$ on the equivalence classes defined by

$$\widehat{\Phi}\colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z} \cong S^1, \quad [x] \mapsto [x + \alpha],$$

or, in short notation, by $x \mapsto x + \alpha \mod 1$. Rigid rotations are *isometric*, if we choose as *metric* the smallest arc length between two points (or the smallest angle),

$$d(e^{2\pi i x}, e^{2\pi i y}) = \min_{j \in \mathbb{Z}} |x - y - j|, \quad x, y \in \mathbb{R}.$$

Then $0 \leq d \leq 1/2$. The metric on $\mathbb{R}/\mathbb{Z}$ is defined by the same formula,

$$d([x], [y]) = \min_{j \in \mathbb{Z}} |x - y - j|.$$

Proposition I.1. *If $\varphi\colon S^1 \to S^1$ is the rigid rotation $\varphi(z) = e^{2\pi i \alpha} z$, the following holds true.*

(i) *If $\alpha = p/q$ is* ***rational*** *and the integers $p, q \in \mathbb{Z}$ relatively prime, then every orbit is* ***periodic*** *with the same minimal period q,*

$$\varphi^q(z) = z \quad \textit{for all } z \in S^1.$$

(ii) *If $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is* ***irrational****, then every orbit is* ***dense*** *in S^1, i.e., the closure of an orbit is the circle,*

$$\overline{\mathcal{O}^+(z)} = \overline{\bigcup_{j \geq 0} \varphi^j(z)} = S^1.$$

For a fixed real number α *all* solutions behave in the same way, independent of their initial condition z. We have stability under the perturbation of the initial conditions, because φ is isometric.

In case (i) all solutions are periodic of the same period so that the orbits are finite sets. In case (ii) every orbit is dense and consists of infinitely many points.

Proof. (i) If $z \in S^1$ and $\alpha = p/q$ is a rational number, then $\varphi^j(z) = e^{2\pi i \frac{p}{q} j} z = z$ for $j = q$. The integer q is a minimal period, since p, q are relatively prime and

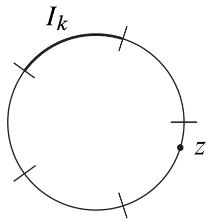

Figure I.2. Dirichlet's pigeon hole principle.

the orbit is the finite set $\mathcal{O}^+(z) = \{z, \varphi(z), \ldots, \varphi^{q-1}(z)\}$. In the covering space $\mathbb{R}$ we have $\Phi^j(x) = x + j\alpha$.

(ii) If $\alpha \in \mathbb{R}\backslash\mathbb{Q}$ is an irrational number and $z \in S^1$, then $\varphi^j(z) \neq \varphi^k(z)$ for $j \neq k$. Indeed, if $\varphi^j(z) = \varphi^k(z)$ and hence $e^{2\pi i(x+j\alpha)} = e^{2\pi i(x+k\alpha)}$, then it follows that $x + j\alpha = x + k\alpha + l$ for an integer $l \in \mathbb{Z}$, so that $\alpha = l/(j-k) \in \mathbb{Q}$ is a rational number, contradicting the assumption on the number.

Next, we apply the so-called *pigeon hole principle of Dirichlet*. We fix z and divide S^1 into N disjoint intervals $I_k = \{e^{2\pi i\beta} \mid \frac{k-1}{N} < \beta \leq \frac{k}{N}\}$ of length $1/N$ for $1 \leq k \leq N$. The $N+1$ points $z = \varphi^0(z), \ldots, \varphi^N(z)$ are different from each other on S^1. Therefore, there exists an interval that contains at least two of the points, let us say $\varphi^j(z)$ and $\varphi^k(z)$, where $0 \leq j < k \leq N$. By construction, $d(\varphi^j(z), \varphi^k(z)) < 1/N$ and since φ is isometric, it follows that $d(\varphi^{k-j}(z), z) < 1/N$, and

$$d(\varphi^{(k-j)(n+1)}(z), \varphi^{(k-j)n}(z)) = d(\varphi^{k-j}(z), z) < 1/N, \quad n \geq 0,$$

so that the points $\varphi^{(k-j)n}(z)$ for $n \geq 0$ traverse the circle in equidistant steps of length $< 1/N$. For every $w \in S^1$ there exists an integer $n \geq 0$ satisfying $d(\varphi^{(k-j)n}(z), w) < 1/N$. This is true for every integer $N > 0$, so that the set $\mathcal{O}^+(z)$ is dense in S^1. □

How is the orbit statistically distributed on the circle S^1? If $I \subset S^1$ is an interval we can ask, how often the orbit $(\varphi^j(z))_{j\geq 0}$ visits the interval I on the average. To be precise, we introduce the function

$$H(z, n, I) = \frac{1}{n}\#\{0 \leq j \leq n-1 \mid \varphi^j(z) \in I\}$$

and investigate the convergence as $n \to \infty$. We shall see for α irrational that

$$H(z, n, I) \to \frac{1}{2\pi}|I| = \frac{|I|}{|S^1|}$$

as $n \to \infty$, where the measure is defined by the arc length. In order to reformulate the problem, we introduce the characteristic function χ_I of the interval I,

$$\chi_I(z) = \begin{cases} 1, & z \in I, \\ 0, & z \notin I. \end{cases}$$

Then

$$H(z, n, I) = \frac{1}{n} \sum_{j=0}^{n-1} \chi_I(\varphi^j(z)).$$

More generally, instead of χ_I we can take any Riemann integrable (short: R-integrable) function $f : S^1 \to \mathbb{C}$ and prove the following classical result.

Theorem I.2 (Equidistribution (mod 1) by H. Weyl). *Let $\varphi \colon S^1 \to S^1$ be the rigid rotation $\varphi(e^{2\pi i x}) = e^{2\pi i(x+\alpha)}$ where $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is an irrational number. Then, for every Riemann integrable function $f : S^1 \to \mathbb{C}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\varphi^j(z)) = \frac{1}{2\pi} \int_{S^1} f := \frac{1}{2\pi} \int_0^{2\pi} f(e^{it})\, dt$$

*for **every point** $z \in S^1$.*

Remark. (i) We point out that the limit exists for *every* $z \in S^1$, if f is Riemann integrable. Later on, we shall show that the limit exists for *almost every* $z \in S^1$, if f is merely Lebesgue integrable.

(ii) If for any function $f : S^1 \to \mathbb{C}$ the limit on the left-hand side exists,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\varphi^j(z)) =: f^*(z) \in \mathbb{C},$$

then the limit is called the *mean value of the function f over the orbit* $\mathcal{O}^+(z)$, or mean value in time. The number $f^*(z)$ can depend on the orbit.

(iii) If $f : S^1 \to \mathbb{C}$ is a (Lebesgue) integrable function, the number

$$\bar{f} := \frac{1}{2\pi} \int_{S^1} f = \frac{1}{2\pi} \int_0^{2\pi} f(e^{it})\, dt = \int_0^1 f(e^{2\pi i t})\, dt$$

is called the *mean value of f over the space S^1*.

Considering the function $f : S^1 \to \mathbb{C}$ on the *covering space* $\mathbb{R}$ of S^1 we define the 1-periodic function $F : \mathbb{R} \to \mathbb{C}$ by $F(x) = f(e^{2\pi i x})$ and Theorem I.2 becomes

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} F(x + \alpha j) = \int_0^1 F(y)\, dy$$

for every $x \in \mathbb{R}$ and every 1-periodic, locally Riemann integrable function $F : \mathbb{R} \to \mathbb{C}$.

Corollary I.3. *If $I \subset S^1$ is an **interval**, then*

$$\frac{1}{n}\sum_{j=0}^{n-1} \chi_I(\varphi^j(z)) \longrightarrow \frac{1}{2\pi}\int_0^{2\pi} \chi_I(e^{it})\,dt = \frac{1}{2\pi}|I| = \frac{|I|}{|S^1|}$$

for every $z \in S^1$.

We see that statistically the points of an orbit $\mathcal{O}^+(z)$ are equidistributed on S^1. The mechanism to generate the orbit is *deterministic* and not stochastic!

Theorem I.2 implies Proposition I.1.

Corollary I.4. *If α is irrational, then $\overline{\mathcal{O}^+(z)} = S^1$ for every $z \in S^1$.*

Proof. If I is an interval, then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} \chi_I(\varphi^j(z)) = \frac{|I|}{|S^1|} \neq 0.$$

Hence, the interval I is *visited infinitely often*. This holds true for every interval, so that $\mathcal{O}^+(z)$ is indeed dense in S^1. □

Proof of Theorem I.2 [Definition of R-integrable, Weierstrass]. The proof is carried out in four steps.

(1) We first take the *trigonometric monomial* $f(z) = z^p$ where $p \in \mathbb{Z}$ and $z = e^{2\pi i x}$ for $x \in \mathbb{R}$. Abbreviating $\theta = e^{2\pi i \alpha}$, we have $f(\varphi^j(z)) = (\theta^p)^j z^p$ and therefore,

$$\frac{1}{n}\sum_{j=0}^{n-1} f(\varphi^j(z)) = \frac{1}{n} z^p \sum_{j=0}^{n-1}(\theta^p)^j = \begin{cases} 1, & p = 0,\\ \frac{1}{n} z^p \frac{\theta^{np}-1}{\theta^p-1}, & p \neq 0.\end{cases}$$

Because of $|\theta^{np} - 1| \leq 2$ and $\theta^p \neq 1$ for $p \neq 0$ (since α is irrational), it follows that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f(\varphi^j(z)) = \begin{cases} 1, & p = 0\\ 0, & p \neq 0\end{cases}$$
$$= \int_0^1 f(e^{2\pi i t})\,dt.$$

(2) Now, we take *linear combinations* and consider the trigonometric polynomial

$$P(z) = \sum_{k=-N}^{N} a_k z^k, \quad z \in S^1,\ a_k \in \mathbb{C}.$$

It follows from step (1) that $P^*(z) = a_0 = \bar{P}$.

(3) Next we *approximate* the R-integrable function $f: S^1 \to \mathbb{R}$ (in case of $\mathbb{C}$ we split the function into its real and its imaginary part). We claim that for $\varepsilon > 0$ there exist two trigonometric polynomials $P_\varepsilon^-, P_\varepsilon^+$, satisfying

$$(*) \quad \begin{cases} P_\varepsilon^-(z) \le f(z) \le P_\varepsilon^+(z), & z \in S^1, \\ \int_{S^1} P_\varepsilon^+ - \int_{S^1} P_\varepsilon^- < \varepsilon. \end{cases}$$

This can be seen as follows. Since the function f is *R-integrable*, there exist according to a classical theorem by Darboux two-step functions (belonging to lower and upper sums of f), for which $(*)$ holds true with $\varepsilon/4$. Moving these step functions down, respectively up, we approximate them by continuous functions, satisfying $(*)$ with $\varepsilon/2$. Since every continuous, periodic function can be *uniformly* approximated by trigonometric polynomials (K. Weierstrass), the claim follows with ε.

(4) Finally, *integrating* $(*)$, we obtain the estimates

$$-\varepsilon + \int f \le \int P_\varepsilon^- \le \int f \le \int P_\varepsilon^+ \le \int f + \varepsilon.$$

For a function $g: S^1 \to \mathbb{R}$ we abbreviate

$$S_n(g, z) = \frac{1}{n} \sum_{j=0}^{n-1} g(\varphi^j(z)).$$

With this abbreviation we can estimate

$$-\varepsilon + \int f \le \int P_\varepsilon^- \overset{(2)}{=} \lim_{n\to\infty} S_n(P_\varepsilon^-, z) = \underline{\lim}\, S_n(P_\varepsilon^-, z) \overset{(*)}{\le} \underline{\lim}\, S_n(f, z)$$

and

$$\overline{\lim}\, S_n(f, z) \overset{(*)}{\le} \overline{\lim}\, S_n(P_\varepsilon^+, z) \overset{(2)}{=} \lim_{n\to\infty} S_n(P_\varepsilon^+, z) \overset{(2)}{=} \int P_\varepsilon^+ \le \int f + \varepsilon,$$

so that altogether

$$-\varepsilon + \int f \le \underline{\lim}\, S_n(f, z) \le \overline{\lim}\, S_n(f, z) \le \int f + \varepsilon.$$

This holds true for every $\varepsilon > 0$ and therefore,

$$\underline{\lim}\, S_n(f, z) = \overline{\lim}\, S_n(f, z) = \lim_{n\to\infty} S_n(f, z) = \int f.$$

This is true for every $z \in S^1$ and the equidistribution theorem is proved. □

The equidistribution theorem is not only valid for S^1, but also for the n-torus

$$T^n := S^1 \times S^1 \times \cdots \times S^1 \subset \mathbb{C}^n$$

where $z \in T^n \iff z = (z_1, \dots, z_n), \quad |z_j| = 1$, and we can write z_j as $z_j = e^{2\pi i x_j}$ with a real number $x_j \in \mathbb{R}$. The *covering map* p is defined by

$$p\colon \mathbb{R}^n \to T^n, \quad x = (x_1, \dots, x_n) \mapsto (e^{2\pi i x_1}, \dots, e^{2\pi i x_n}),$$

so that $p(x+j) = p(x)$ for every integer vector $j \in \mathbb{Z}^n$; introducing the frequency vector $\omega = (\omega_1, \dots, \omega_n) \in \mathbb{R}^n$, we define the mapping $\varphi\colon T^n \to T^n$ of the torus by

$$(z_1, \dots, z_n) \mapsto (e^{2\pi i \omega_1} z_1, \dots, e^{2\pi i \omega_n} z_n).$$

On the covering space $\mathbb{R}^n$ of the torus the *translation* $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$, defined by $\Phi(x) = x + \omega$, satisfies $\Phi(x + j) = \Phi(x) + j$ for $j \in \mathbb{Z}^n$. The induced map on the quotient is denoted by

$$\widehat{\Phi}\colon \mathbb{R}^n/\mathbb{Z}^n \to \mathbb{R}^n/\mathbb{Z}^n \cong T^n, \quad [x] \mapsto \widehat{\Phi}([x]) = [x + \omega].$$

With the projection $\hat{p}\colon \mathbb{R}^n/\mathbb{Z}^n \to T^n$, defined by $\hat{p}([x]) = p(x)$, the diagram

$$\begin{array}{ccc} T^n & \xrightarrow{\ \varphi\ } & T^n \\ \uparrow{\scriptstyle \hat{p}} & & \uparrow{\scriptstyle \hat{p}} \\ \mathbb{R}^n/\mathbb{Z}^n & \xrightarrow{\ \widehat{\Phi}\ } & \mathbb{R}^n/\mathbb{Z}^n \end{array}$$

is commutative, so that $\varphi \circ \hat{p}([x]) = \hat{p} \circ \widehat{\Phi}([x])$.

In order *to visualize* the mapping we choose the representative x of the equivalence class $[x]$ in the fundamental domain $[0,1]^n = [0,1] \times \cdots \times [0,1] \subset \mathbb{R}^n$(with identified sides), then, the map $\widehat{\Phi}$ is a translation in $[0,1]^n$. If a point x is pushed out on one side, it enters again as illustrated in Figure I.3.

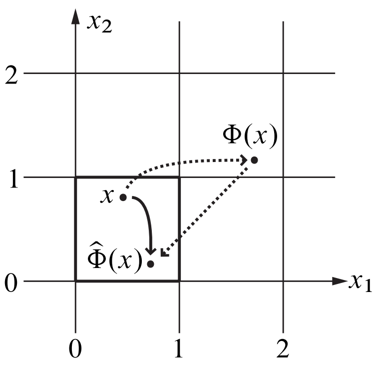

Figure I.3. The mapping $\widehat{\Phi}$ in the fundamental domain $[0,1] \times [0,1]$ of T^2.

We can embed the torus T^2 into $\mathbb{R}^3$ by means of the mapping $\psi\colon \mathbb{R}^2 \to \mathbb{R}^3$, $\psi(x_1, x_2) = (\xi_1, \xi_2, \xi_3)$, defined by

$$\begin{aligned}\xi_1 &= (a + b\cos 2\pi x_1)\cos 2\pi x_2,\\ \xi_2 &= (a + b\cos 2\pi x_1)\sin 2\pi x_2,\\ \xi_3 &= b\sin 2\pi x_1,\end{aligned}$$

where $a > b > 0$, see Figure I.4. The image $\psi(\mathbb{R}^2)$ is an embedded torus and the induced mapping $\hat{\psi}\colon \mathbb{R}^2/\mathbb{Z}^2 \to \mathbb{R}^3$ is bijective onto the torus $\psi(\mathbb{R}^2)$. Introducing the frequencies $\omega = (\omega_1, \omega_2)$ we define the translation φ on the embedded torus by

$$\varphi(\psi(x)) = \psi(x + \omega) = \psi(\Phi(x)).$$

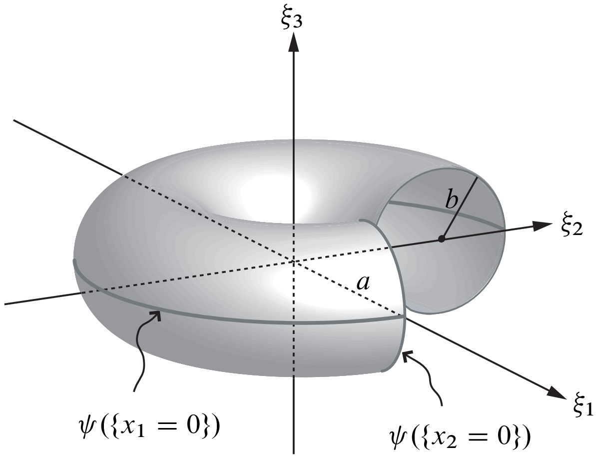

Figure I.4. The embedded torus.

The image of the line $x + t\omega$ on $\mathbb{R}^2$ is the curve $t \mapsto \psi(x + t\omega)$ on the embedded torus. Requiring

$$\langle \omega, j\rangle := \sum_{k=1}^{2} \omega_k j_k \notin \mathbb{Z} \quad \text{for all } j = (j_1, j_2) \in \mathbb{Z}^2 \setminus \{0\},$$

the curve spirals around on the torus without self intersections. Indeed, arguing by contradiction and assuming that $\psi(x + t_1\omega) = \psi(x + t_2\omega)$ for $t_1 \neq t_2$, we obtain $x + t_1\omega = x + t_2\omega \mod \mathbb{Z}^2$, and hence

$$x + t_1\omega = x + t_2\omega + r$$

for an integer vector $0 \neq r = (r_1, r_2) \in \mathbb{Z}^2$. Consequently, $\omega = \tau^{-1} r$, where $\tau := t_1 - t_2$, so that $\langle \omega, j \rangle = 0$ for the integer vector $j := (r_2, -r_1)$ which contradicts our assumption on the frequencies. The next theorem shows that not only the curve $\psi(x + t\omega)$, but already the set of points

$$\varphi^s(\psi(x)) = \psi(x + s\omega), \qquad \text{for all integers } s \in \mathbb{N}_0$$

are dense on the torus $\psi(\mathbb{R}^2)$.

Theorem I.5 (Equidistribution mod $\mathbb{Z}^n$). *If $\varphi\colon T^n \to T^n$ is the translation $\varphi(z_1, \ldots, z_n) = (e^{2\pi i \omega_1} z_1, \ldots, e^{2\pi i \omega_n} z_n)$ and if*

$$\langle \omega, j \rangle = \sum_{k=1}^{n} \omega_k j_k \notin \mathbb{Z} \quad \textit{for all } j = (j_1, \ldots, j_n) \in \mathbb{Z}^n \setminus \{0\},$$

then for every R-integrable function $f : T^n \to \mathbb{C}$ the equality

$$\lim_{n \to \infty} \frac{1}{n} \sum_{s=0}^{n-1} f(\varphi^s(z)) = \frac{1}{m(T^n)} \int_{T^n} f$$

holds for every point $z \in T^n$, where

$$\begin{aligned} \frac{1}{m(T^n)} \int_{T^n} f &= \int_0^1 \ldots \int_0^1 f(e^{2\pi i x_1}, \ldots, e^{2\pi i x_n})\, dx_1 \ldots dx_n \\ &= \frac{1}{(2\pi)^n} \int_0^{2\pi} \ldots \int_0^{2\pi} f(e^{i x_1}, \ldots, e^{i x_n})\, dx_1 \ldots dx_n. \end{aligned}$$

Proof. Exercise. Hint: the proof is analogous to the proof of Theorem I.2. As for step (1) one takes $f(z_1, \ldots, z_n) = z_1^{p_1} \ldots z_n^{p_n}$ with the integer vector $p = (p_1, \ldots, p_n) \in \mathbb{Z}^n$, so that $f(\varphi^s(z)) = (e^{2\pi i \langle \omega, p \rangle})^s f(z)$. By assumption, $e^{2\pi i \langle \omega, p \rangle} \neq 1$ if $p \neq 0$, while $|e^{2\pi i \langle \omega, p \rangle}| = 1$. The steps (2), (3) and (4) are as before. □

Corollary I.6 (Equidistribution). *If $\langle \omega, j \rangle \notin \mathbb{Z}$ for all $j \in \mathbb{Z}^n \setminus \{0\}$ and if $I \subset T^n$ is an interval satisfying $m(I) > 0$, then*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_I(\varphi^j(z)) = \frac{m(I)}{m(T^n)}$$

for every $z \in T^n$.

Corollary I.7 (Kronecker, density). *If $\langle \omega, j \rangle \notin \mathbb{Z}$ for all $j \in \mathbb{Z}^n \setminus \{0\}$, then every orbit φ is dense on the torus:*

$$\overline{\bigcup_{j \geq 0} \varphi^j(z)} = \overline{\mathcal{O}^+(z)} = T^n$$

for every point $z \in T^n$.

Proof. If $I \subset T^n$ is an interval satisfying $m(I) > 0$ we consider the step function $f = \chi_I$ and conclude from Corollary I.6 that every orbit visits the interval I infinitely often. This holds true for every open interval and the proof is complete. □

Corollary I.8 (Kronecker, rational approximation). *Assume that* $\langle \omega, j \rangle \notin \mathbb{Z}$ *for all* $j \in \mathbb{Z}^n \setminus \{0\}$ *and let* $\varepsilon > 0$ *and* $N \geq 1$ *be given. Then there exist an integer* $s \geq N$ *and an integer vector* $j \in \mathbb{Z}^n$ *satisfying*

$$|s\omega - j| < \varepsilon$$

or equivalently

$$|\omega - j/s| < \varepsilon/s.$$

Proof. Exercise. Hint: consider orbits of the translation map $\hat{\Phi}$ on the quotient $\mathbb{R}^n/\mathbb{Z}^n$ and use the metric $d([x],[y]) = \min_{j \in \mathbb{Z}^n} |x - y - j|$. Then,

$$\hat{\Phi}^s([x]) = [\Phi^s(x)] = [x + s\omega].$$

The orbit through the point $p(0) \in T^n$ is dense. If $\varepsilon > 0$, there exists an integer $s \geq N$ satisfying $d(\hat{\Phi}^s([0]),[0]) = \min_{j \in \mathbb{Z}^n} |s\omega - j| < \varepsilon$. □

I.3 Transitive dynamical systems

In the following (X, d) is a metric space and $\varphi\colon X \to X$ a continuous map.

Definition. The dynamical system (X, φ) is called *transitive*, if φ possesses *a dense orbit* i.e., if there exists a point $x \in X$ whose orbit is dense, so that its closure satisfies

$$\overline{\mathcal{O}^+(x)} = \overline{\bigcup_{j \geq 0} \varphi^j(x)} = X.$$

The system (X, φ) is called *minimal*, if *every* orbit of φ is dense.

Example. The irrational rotations of S^1 are *transitive and minimal.*

If the system (X, φ) is transitive, then there exists for every two non-empty open sets $\emptyset \neq U, V \subset X$ an integer $n \geq 0$ satisfying

$$\varphi^n(U) \cap V \neq \emptyset,$$

as illustrated in Figure I.5. Indeed, according to the assumption, there exists a dense orbit $(\varphi^j(x))_{j \geq 0}$ and therefore there are two integers j, k satisfying $\varphi^j(x) \in U$ and $\varphi^k(x) \in V$. Assuming $k \geq j$, we define $y = \varphi^j(x) \in U$ and set $n = k - j$, then $\varphi^n(y) = \varphi^n(\varphi^j(x)) = \varphi^k(x) \in V$, proving the claim.

Under *additional assumptions* the converse also holds true, as the next result will show.

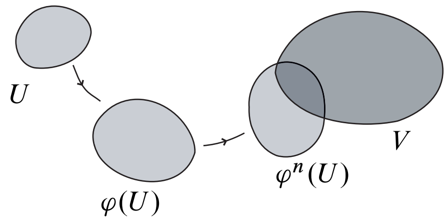

Figure I.5. Necessary condition for transitivity.

Theorem I.9 (Transitivity theorem by G. Birkhoff). *We assume that the metric space (X, d) is complete and that X possesses a countable basis of open sets. Let $\varphi \colon X \to X$ be a* ***continuous*** *map.*

If for every pair $\emptyset \neq U, V \subset X$ of open sets there exists an integer $n = n(U, V) \geq 0$ satisfying

$$\varphi^n(U) \cap V \neq \emptyset,$$

then there exists a ***dense*** *set $R^+ \subset X$, so that for every point $p \in R^+$,*

$$\overline{\mathcal{O}^+(p)} = X.$$

In addition, R^+ is of ***second Baire*** *category.*

Remark. The assumptions on the metric space (X, d) are fulfilled, in particular in the following cases.

- *(X, d) is a complete, separable, metric space.* Indeed, in a metric space, separability (i.e., the existence of a countable, dense subset) is equivalent to the existence of a countable basis of open sets. To see this, it is sufficient to find a countable system of open sets $(B_i)_{i\geq 1}$, so that for every $x \in X$ and every neighborhood U_x of x there exists an index i satisfying $B_i \subset U_x$.
  This is easy to accomplish. Indeed, if $(y_n)_{n\geq 1}$ is a dense sequence in X, then there exists, for every $x \in X$ and every $\varepsilon > 0$, a point y_n satisfying $d(x, y_n) < \varepsilon$. Therefore, the system of open balls $\{B(y_n, 1/m) \mid m, n \geq 1\}$ has the desired properties.
- *(X, d) is a compact metric space.* Due to the compactness every sequence has a convergent subsequence. This applies, in particular, to every Cauchy sequence. A Cauchy sequence, however, possessing a convergent subsequence is convergent. This proves the completeness of the metric space.
  It remains to show that X is *separable*. The open balls $\{B(x^n, 1/n) \mid x^n \in X\}$ are an open covering of X for every $n \in \mathbb{N}$. If we choose the finite subcovers $B(x_1^n, 1/n), \dots, B(x_{m_n}^n, 1/n)$, the set of points $\{x_i^n \mid n \geq 1, 1 \leq i \leq m_n\}$ is countable and dense in X.

- The space X is a closed subset of $\mathbb{R}^n$.

Proof of Theorem I.9 [Completeness, Baire category theorem]. For two open sets $\emptyset \neq U, V \subset X$ there exists by assumption an integer $n \geq 0$ satisfying $\varphi^n(U) \cap V \neq \emptyset$. Hence $U \cap \varphi^{-n}(V) \neq \emptyset$, where $\varphi^{-n}(V) := \{x \in X \mid \varphi^n(x) \in V\}$ is the preimage of V under the iterated map φ^n. Consequently,

$$U \cap \bigcup_{j \geq 0} \varphi^{-j}(V) \neq \emptyset,$$

and in view of the assumption this holds true for *every* open set $U \neq \emptyset$. Therefore, the open set

$$\mathcal{O}^-(V) = \bigcup_{j \geq 0} \varphi^{-j}(V)$$

is dense in X. This holds true for every open set $V \neq \emptyset$. If $(V_j)_{j \geq 1}$ is a *countable basis of open sets* in X, then the sets $\mathcal{O}^-(V_j)$ are open and dense. The countable intersection

$$R^+ := \bigcap_{j \geq 1} \mathcal{O}^-(V_j)$$

is still a dense subset of X in view of the following result.

Lemma I.10 (Baire category theorem). *If (X, d) is a complete metric space and if $(V_j)_{j \geq 1}$ is a* countable *family of* open and dense *subsets of X, then the countable intersection $\bigcap_{j \geq 1} V_j$ is dense in X.*

Postponing the proof of the Baire category theorem, we first complete the proof of Theorem I.9 and choose a point $p \in R^+$. Then

$$p \in \mathcal{O}^-(V_j) = \bigcup_{s \geq 0} \varphi^{-s}(V_j)$$

for every $j \geq 0$. Hence for every integer j there exists an integer $s \geq 0$ satisfying $\varphi^s(p) \in V_j$, so that

$$\mathcal{O}^+(p) \cap V_j \neq \emptyset$$

for every j. Because the family $(V_j)_{j \geq 1}$ is a basis of open sets, every open subset $U \neq \emptyset$ of X contains some subset $V_j \subset U$, so that

$$\mathcal{O}^+(p) \cap U \neq \emptyset.$$

Consequently, the orbit $\mathcal{O}^+(p)$ is dense in X. This holds true for every point $p \in R^+$ and completes the proof of Theorem I.9. □

It remains to prove the lemma.

Proof of Lemma I.10 [Completeness]. We have to show that for a given point $x \in X$ and a given real number $\varepsilon > 0$ there exists a point x^* satisfying

$$x^* \in B(x,\varepsilon) \cap \bigcap_{j \geq 1} V_j$$

where $B(x,\varepsilon)$ is an open ball of radius ε centered at x. Since $B(x,\varepsilon)$ is open and V_1 open and dense, the intersection $B(x,\varepsilon) \cap V_1$ is open and not empty. Therefore, there exists a point $x_1 \in B(x,\varepsilon)$ satisfying

$$K_1 := \overline{B(x_1, r_1)} \subset B(x,\varepsilon) \cap V_1$$

for a radius $0 < r_1 < 1/2$. Since $B(x_1, r_1)$ is an open ball and since V_2 is open and dense, we find a point $x_2 \in B(x_1, r_1)$ satisfying

$$K_2 := \overline{B(x_2, r_2)} \subset B(x_1, r_1) \cap V_2$$

for a radius $0 < r_2 < 1/2^2$. Proceeding inductively, we find for every $j \geq 2$ a point x_j satisfying

$$K_j := \overline{B(x_j, r_j)} \subset B(x_{j-1}, r_{j-1}) \cap V_j$$

for a radius $0 < r_j < 1/2^j$. Then

$$B(x,\varepsilon) \supset K_1 \supset K_2 \supset \cdots$$

and $\operatorname{diam}(K_j) \to 0$. Since K_j is a nested sequence of closed sets, it follows from the completeness of the metric space X that the countable intersection

$$\bigcap_{j \geq 1} K_j = \{x^*\}$$

consists of a single point $x^* \in X$. In view of our construction, $x^* \in K_j \subset B(x,\varepsilon) \cap V_j$ for every $j \geq 1$ and therefore,

$$x^* \in B(x,\varepsilon) \cap \bigcap_{j \geq 1} V_j.$$

This completes the proof of the lemma. □

Definition. A set $A \subset X$ is called *invariant* under the map $\varphi\colon X \to X$ if

$$\varphi^{-1}(A) := \{x \in X \mid \varphi(x) \in A\} = A.$$

We note that the set A is invariant precisely if

$$\varphi(A) \subset A \quad \text{and} \quad \varphi(A^c) \subset A^c$$

where A^c denotes the complement of the set A in X.

Proposition I.11. *If the dynamical system (X, φ) is minimal, then every* ***closed invariant*** *set $A \neq \emptyset$ is already the whole space, $A = X$.*

Proof [Definitions]. Since A is invariant under φ, $\varphi(A) \subset A$. Hence if $x \in A$,then the positive orbit $\mathcal{O}^+(x)$ is contained in A. Since (X, φ) is minimal, it is dense in X and taking its closure, we conclude

$$X = \overline{\mathcal{O}^+(x)} \subset \bar{A} = A \subset X,$$

so that $A = X$, as claimed in the proposition. □

I.4 Structural stability

In order to illustrate the concept of structural stability, we study the special example of an expanding map on the circle defined by

$$\varphi\colon S^1 \to S^1, \quad z \mapsto z^2, \quad z = e^{2\pi i x},\ x \in \mathbb{R},$$

or $\varphi(e^{2\pi i x}) = e^{2\pi i(2x)}$ for $x \in \mathbb{R}$. This map is *not* bijective, but two-to-one. The covering map on $\mathbb{R}$ is the function

$$\Phi\colon \mathbb{R} \to \mathbb{R}, \quad \Phi(x) = 2x$$

satisfying $\Phi(x + j) = \Phi(x) + 2j$ for $j \in \mathbb{Z}$. The projection map $p\colon \mathbb{R} \to S^1$, defined by $p(x) = e^{2\pi i x}$, satisfies $\varphi(p(x)) = p(\Phi(x))$, so that the diagram

$$\begin{array}{ccc} S^1 & \xrightarrow{\varphi} & S^1 \\ {\scriptstyle p}\uparrow & & \uparrow{\scriptstyle p} \\ \mathbb{R} & \xrightarrow{\Phi} & \mathbb{R} \end{array}$$

is commutative. The map Φ induces the map $\widehat{\Phi}\colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$, $\widehat{\Phi}([x]) = [\Phi(x)]$, on the quotient space, in short notation, $\widehat{\Phi}(x) = 2x \mod \mathbb{Z}$. With the *homeomorphism* $\hat{p}\colon \mathbb{R}/\mathbb{Z} \to S^1$, defined by $\hat{p}([x]) := p(x) = e^{2\pi i x}$, we obtain $\varphi \circ \hat{p}([x]) = \hat{p} \circ \widehat{\Phi}([x])$, so that the diagram

$$\begin{array}{ccc} S^1 & \xrightarrow{\varphi} & S^1 \\ {\scriptstyle \hat{p}}\uparrow & & \uparrow{\scriptstyle \hat{p}} \\ \mathbb{R}/\mathbb{Z} & \xrightarrow{\widehat{\Phi}} & \mathbb{R}/\mathbb{Z} \end{array}$$

commutes. The (restricted) projection $p\colon (0, 1] \to S^1$ is *bijective* and, identifying $x \in (0, 1]$ with its equivalence class $[x]$, the map $\widehat{\Phi}$ can be represented in the *fundamental domain* $(0, 1]$, by means of the formula

$$\widehat{\Phi}\colon (0,1] \to (0,1], \quad \widehat{\Phi}(x) = \begin{cases} 2x, & 0 < x \leq 1/2, \\ 2x - 1, & 1/2 < x \leq 1, \end{cases}$$

illustrated in Figure I.6. We recall that every real number in $0 < x \leq 1$ can be represented by a unique dyadic expansion containing infinitely many nonzero digits,

$$x = \frac{x_1}{2} + \frac{x_2}{2^2} + \cdots = \sum_{j \geq 1} \frac{x_j}{2^j}$$

where $x_j \in \{0, 1\}$. The standard notation for the dyadic expansion is the following

$$x = 0.x_1 x_2 x_3 \ldots .$$

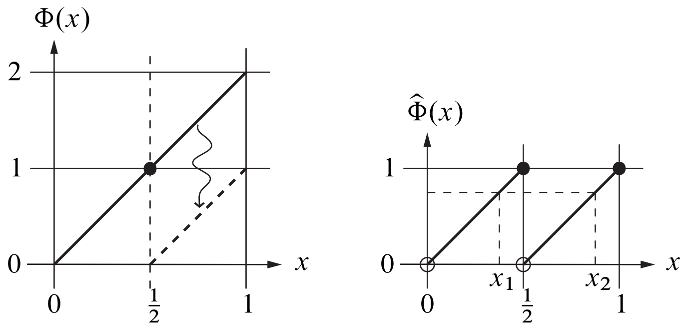

Figure I.6. The maps Φ and $\widehat{\Phi}$ in the fundamental domain $(0, 1]$.

In this notation, the multiplication of x by 2 corresponds to the shift of the point, namely

$$\begin{aligned} \Phi(x) &= 2x \\ &= x_1 + \frac{x_2}{2} + \frac{x_3}{2^2} + \cdots \\ &= x_1.x_2 x_3 \ldots \\ &= 0.x_2 x_3 x_4 \ldots \quad \bmod 1. \end{aligned}$$

Consequently,

$$\widehat{\Phi}(x) = \widehat{\Phi}(0.x_1 x_2 x_3 \ldots) = 0.x_2 x_3 x_4 \ldots,$$

and we see that the mapping $\widehat{\Phi}$ is a *shift map*. This shift map simply forgets the first entry in a sequence and shifts all other entries one place to the left. Of course, $\widehat{\Phi}$ is not bijective, since every image point $y = \widehat{\Phi}(x)$ has two preimages, namely $0.1y_1 y_2 \ldots$ and $0.0y_1 y_2 \ldots$.

We now introduce the space

$$\mathcal{F} = \{(x_j)_{j \geq 1} \mid x_j \in \{0, 1\}\}$$

of *one-sided sequences of the two symbols* 0 and 1, and consider the subset

$$\mathcal{F}_0 = \{(x_j)_{j \geq 1} \mid x_j \neq 0 \text{ for infinitely many } j\}.$$

There is a *bijective* mapping

$$\psi \colon (0, 1] \to \mathcal{F}_0,$$

defined by the dyadic expansion containing infinitely many nonzero digits as follows:

$$\psi(x) = (x_j)_{j \geq 1} \text{ if } x = 0.x_1 x_2 \cdots \in (0, 1].$$

Denoting by $\sigma \colon \mathcal{F} \to \mathcal{F}$ the shift $(x_1, x_2, x_3, \dots) \mapsto (x_2, x_3, x_4, \dots)$, in formulas defined as $\sigma((x_j)_{j \geq 1}) = (x_{j+1})_{j \geq 1}$, we obtain the commutative diagram

$$\begin{array}{ccc} (0,1] & \xrightarrow{\hat{\Phi}} & (0,1] \\ \psi \downarrow & & \downarrow \psi \\ \mathcal{F}_0 & \xrightarrow{\sigma} & \mathcal{F}_0 \end{array}$$

illustrating the equation

$$\psi \circ \hat{\Phi}(x) = \sigma \circ \psi(x).$$

This way, we have represented the *analytical* mapping $\varphi \colon z \mapsto z^2$ on the circle S^1 by the *shift map* σ in the *sequence space* $\mathcal{F}_0$. We will show that there coexist orbits of completely different long-time behavior.

Proposition I.12. *If $\varphi \colon S^1 \to S^1$ is the mapping $z \mapsto z^2$ and m the Lebesgue measure on S^1, the following holds true.*

(i) *The set of periodic points of φ is countable and dense.*

(ii) *The set $R^+ := \{z \in S^1 \mid \overline{\mathcal{O}^+(z)} = S^1\}$ of initial points whose orbits are dense, is a **dense set** in S^1.*

(iii) *$m(R^+) = m(S^1)$ and R^+ is of second Baire category.*

In particular, the mapping φ is transitive, but not minimal.

In contrast to the rigid rotations of S^1 this dynamical system possesses, in every open set, points whose orbits behave asymptotically *completely differently*. The distant future is *no longer predictable*, if not every digit in the dyadic expansion of the initial condition z for the orbit $\mathcal{O}^+(z)$ is known. For example, if $x = 0.x_1 x_2 \dots x_N * * * \dots$, then $\hat{\Phi}^N(x) = 0. * * * \dots$, where the stars stand for the unknown digits.

In physics one talks about the *sensitive dependence* of the orbits on the initial conditions. It is *hopeless* to gain an insight into the orbit structure of all solutions over infinitely long times by *solving the Cauchy initial value problem*!

Proof of Proposition I.12. In order to prove the statement (i), we assume $n \geq 1$. From $\varphi^n(z) = z \iff z^{2^n} = z \iff z^{2^n-1} = 1$ it follows that every complex *root of unity* of order $2^n - 1$ is a periodic point of the period n, and vice versa. There exist exactly $2^n - 1$ such roots of unity, and they are *equidistantly* distributed on S^1, so that the periodic points are countable and dense.

(ii) It is sufficient to verify the assumptions of the transitivity theorem (Theorem I.9). Since S^1 is compact, we have to show that, for every pair $\emptyset \neq U, V \subset S^1$ of open sets, there exists an integer n satisfying $\varphi^n(U) \cap V \neq \emptyset$. An open set $U \subset S^1$ contains the image of a *binary interval* I_n^k under the projection map $p|_{(0,1]}$, where

$$I_n^k := \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] \subset (0,1], \quad k = 0, 1, 2, \ldots, 2^n - 1,$$

the integer n being sufficiently large. Applying the map $\Phi(x) = 2x$ we obtain

$$\begin{aligned}\Phi^n(I_n^k) &= 2^n\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] \\ &= (k, k+1] \\ &= (0,1] \quad \text{mod } 1.\end{aligned}$$

In order to see this in terms of dyadic expansions, we take the real number $x \in I_n^k$, represented as

$$\begin{aligned}x &= 0.x_1x_2x_3\ldots x_nx_{n+1}\ldots \\ &= \frac{x_1}{2} + \frac{x_2}{2^2} + \cdots + \frac{x_n}{2^n} + \frac{x_{n+1}}{2^{n+1}} + \cdots \\ &= \frac{k}{2^n} + \frac{x_{n+1}}{2^{n+1}} + \cdots.\end{aligned}$$

Then

$$\begin{aligned}\Phi^n(x) &= 2^n x \\ &= k + \frac{x_{n+1}}{2} + \frac{x_{n+2}}{2^2} + \cdots \\ &= k.x_{n+1}x_{n+2}\ldots \\ &= 0.x_{n+1}x_{n+2}\ldots \quad \text{mod } 1\end{aligned}$$

and therefore $\widehat{\Phi}^n(I_n^k) = (0,1]$. In view of $p(I_n^k) \subset U$, we find

$$S^1 = p((k, k+1]) = p \circ \Phi^n(I_n^k) = \varphi^n \circ p(I_n^k) \subset \varphi^n(U).$$

Having verified that the assumptions of Theorem I.9 are met, the statement (ii) follows.

(iii) The statement (iii) will be proved later in Section I.5. □

Considering the mapping $\varphi_0\colon S^1 \to S^1$ defined by $\varphi_0(z) = z^2$, we shall study what happens to its complex orbit structure under a perturbation. It turns out that the complex structure is stable under perturbations, as will be proved in the following statement.

- *For every continuously differentiable mapping $\varphi\colon S^1 \to S^1$ in a sufficiently small C^1-neighborhood of φ_0, there exists a unique homeomorphism $h\colon S^1 \to S^1$, so that the diagram*

$$\begin{array}{ccc} S^1 & \xrightarrow{\varphi_0} & S^1 \\ \big\downarrow h & & \big\downarrow h \\ S^1 & \xrightarrow{\varphi} & S^1 \end{array}$$

  commutes, i.e.,

$$\varphi \circ h = h \circ \varphi_0.$$

For the iterates, we then have $\varphi^j(z) = h \circ \varphi_0^j \circ h^{-1}(z)$. The mappings φ and φ_0 are called *topologically conjugated*, and the mapping φ_0 is called structurally stable.

Definition. A C^1-mapping φ_0 is called (C^1-)*structurally stable*, if every C^1-mapping φ in a sufficiently small C^1-neighborhood of φ_0 is topologically conjugated to φ_0.

It is useful to describe the mappings in the *covering space* $\mathbb{R}$ of S^1. The unperturbed mapping is equal to $\varphi_0(e^{2\pi i x}) = e^{2\pi i \Phi(x)}$ where $\Phi(x) = 2x$, and the perturbed mapping is represented by

$$\varphi(e^{2\pi i x}) = e^{2\pi i \psi(x)}$$

where the mapping $\psi\colon \mathbb{R} \to \mathbb{R}$ satisfies

$$\psi(x) = \Phi(x) + \hat{\psi}(x) \quad \text{and} \quad \hat{\psi}(x+1) = \hat{\psi}(x), \quad x \in \mathbb{R}.$$

The homeomorphism $h\colon S^1 \to S^1$ we are looking for is represented as

$$h(e^{2\pi i x}) = e^{2\pi i u(x)},$$

with a homeomorphism $u\colon \mathbb{R} \to \mathbb{R}$. In particular, u is a continuous, strictly increasing function satisfying $u(x+1) = u(x) + 1$ and so is of the form

$$u(x) = x + \hat{u}(x), \quad \text{where } \hat{u}(x+1) = \hat{u}(x),\ x \in \mathbb{R}.$$

The inverse mapping $u^{-1}\colon \mathbb{R} \to \mathbb{R}$ is also continuous, strictly increasing and satisfies $u^{-1}(x+1) = u^{-1}(x) + 1$, so that

$$h^{-1}(e^{2\pi i x}) = e^{2\pi i u^{-1}(x)}, \quad x \in \mathbb{R}.$$

Indeed,

$$\begin{aligned} h^{-1}(h(e^{2\pi i x})) &= h^{-1}(e^{2\pi i u(x)}) \\ &= e^{2\pi i u^{-1}(u(x))} \\ &= e^{2\pi i x} \end{aligned}$$

for all $x \in \mathbb{R}$. The *functional equation* to be solved becomes $\varphi \circ h(z) = h \circ \varphi_0(z)$, for all $z \in S^1$, or equivalently,

$$e^{2\pi i \psi(u(x))} = e^{2\pi i u(\Phi(x))}$$

and we shall *solve* the nonlinear equation

$$\psi(u(x)) = u(\Phi(x))$$

for the unknown mapping u. The following theorem is a special case of a general phenomenon encountered in expanding mappings for which we refer to the monograph [113], Chapter 4.11] by W. Szlenk.

Theorem I.13 (Structural stability of the map $\varphi(z) = z^2$). *We consider on $\mathbb{R}$ the mapping $\psi(x) = 2x + \hat{\psi}(x)$ satisfying $\hat{\psi}(x+1) = \hat{\psi}(x)$ and assume $\hat{\psi}$ to be Lipschitz small in the sense that $|\hat{\psi}(x) - \hat{\psi}(y)| \le L|x-y|$ for all $x, y \in \mathbb{R}$ with a Lipschitz constant satisfying $0 \le L < 1$. Then, there exists a unique, strictly increasing homeomorphism $u\colon \mathbb{R} \to \mathbb{R}$ satisfying $u(x+1) = u(x) + 1$ for every $x \in \mathbb{R}$ and solving the equation*

$$\psi(u(x)) = u(\Phi(x)) = u(2x)$$

for all $x \in \mathbb{R}$. For the 1-periodic function $\hat{u}$, defined by $u(x) = x + \hat{u}(x)$, the estimate

$$|\hat{u}|_\infty \le |\hat{\psi}|_\infty$$

holds in the supremum norm.

Postponing the proof we observe that the homeomorphism u is unique, so that one can ask whether u is a C^1-diffeomorphism if the map ψ is continuously differentiable. In general, this is *not* the case, as we will convince ourselves next. We assume that $u\colon \mathbb{R} \to \mathbb{R}$ is a diffeomorphism *and* that the function $\hat{\psi}$ is of class C^1 and satisfies $\hat{\psi}(0) = 0$ and $\hat{\psi}'(0) \ne 0$. Using the equation $\psi(u(x)) = u(2x)$ we obtain

$$u(u^{-1}(0)) = 0 = \hat{\psi}(0) = \psi(0) = \psi(u(u^{-1}(0))) = u(2u^{-1}(0))$$

and conclude that $2u^{-1}(0) = u^{-1}(0)$. Therefore, $u(0) = 0$ and differentiating the equation $\psi(u(x)) = u(2x)$ in x at $x = 0$, results in

$$\psi'(0)u'(0) = 2u'(0).$$

Since u is a diffeomorphism, $u'(0) \neq 0$ and therefore $\psi'(0) = 2$ contradicting $\psi'(0) = 2 + \hat{\psi}'(0) \neq 2$. We see that the *eigenvalue* of ψ' at the fixed point $\psi(0) = 0$ is an invariant under a *differentiable* conjugation.

Proof of Theorem I.13 [Expansion $\rightsquigarrow$ Contraction]. (1) In the first step we shall show that the map $\psi\colon \mathbb{R} \to \mathbb{R}$ *is bijective and Lipschitz-continuous*, and that its inverse mapping ψ^{-1} *is a contraction.*

Due to $\psi(x) - \psi(y) = 2(x - y) + \hat{\psi}(x) - \hat{\psi}(y)$ we obtain for all $x \geq y$ the estimate

$$(*) \quad \underbrace{(2 - L)}_{=:r_1}(x - y) \leq \psi(x) - \psi(y) \leq \underbrace{(2 + L)}_{=:r_2}(x - y).$$

From $0 \leq L < 1$ we deduce $1 < r_1 < r_2 < 3$ and hence ψ is Lipschitz-continuous and strictly increasing, and therefore bijective in view of the intermediate value theorem. Consequently, there exists the inverse map $\psi^{-1}\colon \mathbb{R} \to \mathbb{R}$. By inserting $\psi^{-1}(x') = x$ and $\psi^{-1}(y') = y$ into the estimate $(*)$, we find

$$r_1\big(\psi^{-1}(x') - \psi^{-1}(y')\big) \leq x' - y' \leq r_2\big(\psi^{-1}(x') - \psi^{-1}(y')\big)$$

for all $x' \geq y'$, so that

$$\frac{1}{r_2}(x' - y') \leq \psi^{-1}(x') - \psi^{-1}(y') \leq \frac{1}{r_1}(x' - y').$$

It follows that $|\psi^{-1}(x') - \psi^{-1}(y')| \leq K|x' - y'|$ where the constant K is defined as

$$K := \max\{r_1^{-1}, r_2^{-1}\} = r_1^{-1} < 1.$$

We have proved that the mapping ψ^{-1} is a contraction. (The inverse ψ^{-1} is a map in $\mathbb{R}$, it is *not* the covering of a map of S^1.)

(2) In order to *solve the functional equation* $\psi(u(x)) = u(2x)$ *for the mapping* u we shall solve the equivalent equation

$$u(x) = \psi^{-1}(u(2x))$$

by the contraction principle and introduce the metric space

$$X = \{u \in C^0(\mathbb{R}, \mathbb{R}) \mid u(x + 1) = u(x) + 1 \text{ and increasing}\}$$

equipped with the metric

$$d_X(u, v) = \sup_{x \in \mathbb{R}} |u(x) - v(x)| = \max_{0 \leq x \leq 1} |u(x) - v(x)| =: |u - v|_\infty.$$

We have used that the difference $u - v$ is 1-periodic. As one can easily verify, the metric space(X, d_X) is *complete.* We note that the elements of X are chosen to be increasing functions (and not strictly increasing), since otherwise the space X would not be complete. We claim that the mapping T, defined by the formula

$$T\colon X \to C^0(\mathbb{R}, \mathbb{R}), \quad (Tu)(x) = \psi^{-1}(u(2x)), \quad x \in \mathbb{R},$$

maps the space X into itself and satisfies $d_X(Tu, Tv) \le K d_X(u, v)$, for all u and v in X, so that it is a contraction with the contraction constant $K < 1$ introduced above. Indeed, since u and ψ^{-1} are increasing, the function Tu is also increasing. From $\psi(x+1) = 2(x+1) + \hat{\psi}(x) = \psi(x) + 2$ one concludes that the inverse map satisfies $\psi^{-1}(y+2) = \psi^{-1}(y) + 1$. Using this, we compute,

$$\begin{aligned}(Tu)(x+1) &= \psi^{-1}(u(2x+2) \\ &= \psi^{-1}(u(2x) + 2) \\ &= \psi^{-1}(u(2x)) + 1 \\ &= (Tu)(x) + 1.\end{aligned}$$

Consequently, $Tu \in X$ and T maps our metric space into itself. It remains to show that T is a contraction. This follows immediately from the contraction property of ψ^{-1},

$$\begin{aligned}d_X(Tu, Tv) &= \max_{0 \le x \le 1} |\psi^{-1}(u(2x)) - \psi^{-1}(v(2x))| \\ &\le K \max_{0 \le x \le 1} |u(2x) - v(2x)| \\ &= K d_X(u, v).\end{aligned}$$

By the contraction principle of Banach there exists a *unique fixed point* $u \in X$ satisfying $Tu = u$, so that $u(x) = \psi^{-1}(u(2x))$ for all x.

(3) In order to show that u *is strictly increasing* we argue by contradiction and assume that there exist real numbers $\alpha < \beta$ in the interval $[0, 1]$ satisfying $u(\alpha) = u(\beta)$. Since u is increasing,

$$u(x) = \text{const.}, \quad \alpha \le x \le \beta.$$

The interval contains a binary interval and so we find integers $n \ge 1$ and $k \in \{0, 1, \ldots, 2^n - 1\}$, for which

$$u(x) = \text{const.}, \quad \frac{k}{2^n} \le x \le \frac{k+1}{2^n}.$$

Using the equation $\psi(u(x)) = u(2x)$ we deduce from $u(\frac{k}{2^n}) = u(\frac{k+1}{2^n})$, that also the equation

$$u\left(\frac{k}{2^{n-1}}\right) = u\left(\frac{k+1}{2^{n-1}}\right)$$

holds true. Indeed, $u(\frac{k}{2^{n-1}}) = u(2\frac{k}{2^n}) = \psi(u(\frac{k}{2^n})) = \psi(u(\frac{k+1}{2^n})) = u(2\frac{k+1}{2^n}) = u(\frac{k+1}{2^{n-1}})$. We repeat this procedure, until in the denominator $2^{n-n} = 1$ shows up. At this point we have $u(k) = u(k+1)$ in *contradiction* to $u(k+1) = u(k) + 1$ and hence proving that u must be strictly increasing.

(4) From $u(x+1) = u(x) + 1$ it follows that $u(x+n) = u(x) + n$ for every $n \in \mathbb{Z}$ from which we conclude that $\lim_{x\to\pm\infty} u(x) = \pm\infty$. Since u is continuous by construction, the surjectivity of u follows from the intermediate value theorem. The injectivity of u is a consequence of the strict monotonicity. Hence u is a strictly increasing bijection of $\mathbb{R}$ onto itself and therefore a homeomorphism of $\mathbb{R}$.

(5) In order to verify the announced *estimate* we deduce from the equation $\psi(u(x)) = u(2x)$ and the definition $\psi(x) = 2x + \hat{\psi}(x)$ that $2u(x) + \hat{\psi}(u(x)) = u(2x)$. Recalling $u(x) = x + \hat{u}(x)$ we obtain the equation

$$2x + 2\hat{u}(x) + \hat{\psi}(u(x)) = 2x + \hat{u}(2x)$$

and estimate

$$\begin{aligned} |\hat{u}(x)| &= \left|\frac{1}{2}\hat{u}(2x) - \frac{1}{2}\hat{\psi}(u(x))\right| \\ &\leq \frac{1}{2}|\hat{u}|_\infty + \frac{1}{2}|\hat{\psi}|_\infty \end{aligned}$$

for every $x \in \mathbb{R}$. Taking the supremum on the left-hand side, the desired estimate $|\hat{u}|_\infty \leq |\hat{\psi}|_\infty$ follows and Theorem I.13 is proved. □

I.5 Measure preserving maps and the ergodic theorem

The previous examples (with the exception of the contractions) are *measure preserving* with respect to the Lebesgue measure. This section deals with the *part played by the measures* in dynamics.

A measure space is a triple $(X, \mathcal{A}, m)$, in which X is a set, $\mathcal{A}$ a σ-algebra of subsets of X (called *measurable* sets), and $m\colon \mathcal{A} \to [0,\infty]$ a measure. In the following we assume the measure space to be *finite*, assuming that $m(X) < \infty$. We denote by $L = L(X, \mathcal{A}, m)$ the vector space of integrable functions $f\colon X \to \mathbb{R} \cup \{\pm\infty\}$. These are the measurable functions, for which the (Lebesgue-)integral is defined and finite. To facilitate the notation, we sometimes omit the measures in the integrals and suppress the integration domain, if it is the whole space. We also suppress the variable over which it is integrated and write $\int f := \int_X f(x)\, dm(x)$ for the integral. To avoid an accumulation of brackets, we simply write, e.g. Tx instead of $T(x)$ or $T^j x$ instead of $T^j(x)$.

Definition. A mapping $T\colon X \to X$ is called *measurable*, if $T^{-1}(A) \in \mathcal{A}$ for every $A \in \mathcal{A}$, where $T^{-1}(A) = \{x \in X \mid T(x) \in A\}$ is the preimage of A. A measurable mapping $T: X \to X$ is called *measure preserving*, if

$$m(T^{-1}(A)) = m(A)$$

for every $A \in \mathcal{A}$. We note that T does not need to be bijective.

It is useful to observe that a mapping T is measure preserving precisely if

$$(*) \qquad \int f \circ T = \int f \quad \text{for } \textit{all } f \in L(X, \mathcal{A}, m).$$

Indeed, if $A \in \mathcal{A}$, then

$$\int \chi_A(Tx)\, dm = \int \chi_{T^{-1}(A)}(x)\, dm = m(T^{-1}(A)),$$
$$\int \chi_A(x)\, dm = m(A).$$

The equation $(*)$ holds true, in particular, for the characteristic function $f = \chi_A$ of the set A and hence $m(T^{-1}(A)) = m(A)$. If, conversely, the map T is measure preserving, then the equation $(*)$ follows for the characteristic functions $f = \chi_A$ of measurable sets $A \in \mathcal{A}$. But then, it holds true for all step functions, and so for every integrable function $f \in L(X, \mathcal{A}, m)$.

Example. We recall the expanding map $\varphi: S^1 \to S^1$ of the circle defined by $z \mapsto z^2$ and consider the restriction T of its covering map to the fundamental domain (0,1] which is equipped with the Lebesgue measure. The map $T: (0,1] \to (0,1]$ is defined by

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

For the open interval $(a,b) \subset (0,1]$ we have $T^{-1}((a,b)) = \left(\frac{a}{2}, \frac{b}{2}\right) \cup \left(\frac{1}{2}(a+1), \frac{1}{2}(b+1)\right)$, where the union of the sets is disjoint, so that

$$m(T^{-1}(a,b)) = \frac{1}{2}(b-a) + \frac{1}{2}(b-a) = b - a = m((a,b)),$$

as illustrated in Figure I.7. This holds true for every interval in $(0,1]$ and hence also for every open subset. It follows that $m(T^{-1}A) = m(A)$ for every Lebesgue measurable set $A \subset (0,1]$, because a measurable set is the countable intersection of open sets up to a null set. Therefore, the map T is *measure preserving*. Alternatively we can also check the criterion $(*)$ above. If $f: (0,1] \to \mathbb{R}$ if integrable, then $\int_0^1 f(Tx)\, dx = \int_0^{1/2} f(2x)\, dx + \int_{1/2}^1 f(2x-1)\, dx = \frac{1}{2}\int_0^1 f(y)\, dy + \frac{1}{2}\int_0^1 f(y)\, dy = \int_0^1 f(y)\, dy$, where we have used the substitutions $y = 2x$ respectively $y = 2x - 1$, proving once more that T is measure preserving.

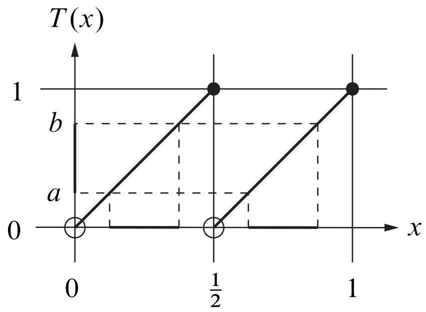

Figure I.7. Preimages of (a, b) under the mapping T in the example.

If the measure is finite the measure preserving maps have strong *recurrence properties* as was already known to H. Poincaré.

Theorem I.14 (Recurrence theorem of Poincaré). *Let $(X, \mathcal{A}, m)$ be a **finite** measure space and $T: X \to X$ a **measure preserving** map. For a measurable set $A \in \mathcal{A}$ we define the subset $A_0 \subset A$ by*

$$A_0 := \{x \in A \mid T^j x \in A \text{ for infinitely many integers } j \geq 0\}.$$

Then the set A_0 is measurable and its measure is equal to $m(A_0) = m(A)$.

The theorem shows that almost every point in A returns to A infinitely often. The theorem is only valid for finite measure spaces, as the translation $x \mapsto x + 1$ on $\mathbb{R}$ shows.

Proof. For the integers $n \geq 1$ we introduce the sets $C_n := \{x \in A \mid T^j x \notin A \text{ for all } j \geq n\}$, so that

$$A_0 = A \setminus \bigcup_{n \geq 1} C_n.$$

In order to prove the theorem it suffices to show that $C_n \in \mathcal{A}$ is measurable and $m(C_n) = 0$ for every $n \geq 1$. Since the set A belongs to $\mathcal{A}$ and since T is measurable, we conclude that $C_n = A \setminus \bigcup_{j \geq n} T^{-j}(A) \in \mathcal{A}$ is, indeed, a measurable set. Moreover, using the notation $T^{-0}(A) := A$,

$$C_n \subset \bigcup_{j \geq 0} T^{-j}(A) \setminus \bigcup_{j \geq n} T^{-j}(A)$$

from which we conclude, because the measure $m(X)$ is finite, that

$$m(C_n) \leq m\Big(\bigcup_{j \geq 0} T^{-j}(A)\Big) - m\Big(\bigcup_{j \geq n} T^{-j}(A)\Big).$$

Since T is measure preserving, both unions have the same measure, in view of

$$\bigcup_{j\geq n} T^{-j}(A) = T^{-n}\Big(\bigcup_{j\geq 0} T^{-j}(A)\Big).$$

Thus $m(C_n) = 0$ and the theorem is proved. □

We shall estimate when the orbit returns for the first time. The time of return is determined by the measure of the set. We assume that $(X, \mathcal{A}, m)$ is a *finite* measure space and $T\colon X \to X$ a *measure preserving* map and let $A \in \mathcal{A}$ be a measurable set satisfying $m(A) > 0$. We claim that if $N := [m(X)/m(A)]$ then there exists an integer j in $1 \leq j \leq N$ for which

$$m(T^{-j}(A) \cap A) > 0.$$

Arguing indirectly, we assume that $m(T^{-j}(A) \cap A) = 0$ for $1 \leq j \leq N$. Then, the sets $T^{-n}(A)$ and $T^{-m}(A)$ are *almost disjoint* for $0 \leq m < n \leq N$, in view of

$$\begin{aligned} m\big(T^{-n}(A) \cap T^{-m}(A)\big) &= m\big(T^{-m}(T^{-(n-m)}(A) \cap A)\big) \\ &= m\big(T^{-(n-m)}(A) \cap A)\big) \\ &= 0. \end{aligned}$$

Since T is measure preserving, it follows that

$$m(X) \geq \sum_{j=0}^{N} m(T^{-j}(A)) = m(A)(1+N) > m(A)(1 + \frac{m(X)}{m(A)} - 1) = m(X),$$

leading to the contradiction $m(X) > m(X)$ and hence proving the claim.

Theorem I.15. *Assume the triple $(X, \mathcal{A}, m)$ to be a finite measure space and the map $T\colon X \to X$ to be measure preserving. Assume in addition that (X, d) is a metric space possessing a countable basis of open sets and assume that all open sets are measurable and of positive measure. Then, there exists for almost every point $x \in X$ a sequence $j_k \to \infty$ of integers so that*

$$T^{j_k}(x) \to x.$$

In this sense, almost every point is recurrent.

Proof. In view of the postulated countable basis of open sets we find a dense sequence $(x_k)_{k\geq 1}$ in X. For every $n \geq 1$ the open balls $B(x_k, 1/n)$ cover the set X. By Theorem I.14 we find a null set $N = N(k,n)$ having the property that every point $x \in B(x_k, 1/n) \setminus N$ returns infinitely often to the ball $B(x_k, 1/n)$. We denote the countable union of these null sets over all k and n by the same letter N. Thus a point $x \in X \setminus N$ returns infinitely often into every ball $B(x_k, 1/n)$ to which it belongs. Since every neighborhood of x contains such a ball, the theorem is proved. □

In the following, we investigate how the points of an orbit $\mathcal{O}^+(x)$ under a measure preserving map T are *statistically* distributed in the space X. Given $A \in \mathcal{A}$ and $x \in X$, we ask, how often does the orbit $(T^j x)_{j \geq 0}$ visit the set A *on the average*? Note that

$$\frac{1}{n} \#\{0 \leq j \leq n-1 \mid T^j x \in A\} = \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(T^j x).$$

We are interested in the convergence of the sum as $n \to \infty$.

The theorem by G. Birkhoff (1932) provides an answer to the question. According to this theorem the pointwise limit exists (in $\mathbb{R}$) for *almost all* $x \in X$, and, in addition, we also have L^1-convergence towards the limit function. We recall that in the equidistribution theorem of H. Weyl (Theorem I.2), we have convergence for every point, assuming the functions to be Riemann integrable instead of Lebesgue integrable.

Theorem I.16 (Ergodic theorem of G. Birkhoff). *We consider the* ***finite*** *measure space* $(X, \mathcal{A}, m)$ *and assume the map* $T\colon X \to X$ *to be* ***measure preserving***. *For every integrable function* $f \in L(X, \mathcal{A}, m)$ *there exist a function* $f^* \in L(X, \mathcal{A}, m)$ *and a null set* $N \subset X$ *(*N ***depending*** *on* f*) satisfying* $T^{-1}(N) = N$ *and*

(i) $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = f^*(x)$ *for all* $x \in X \setminus N$,

(ii) $f^*(Tx) = f^*(x)$ *for all* $x \in X$,

(iii) $\int_X f^* = \int_X f$,

(iv) $\int_X \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) - f^*(x) \right| dm \to 0$, $n \to \infty$.

Postponing the *proof* to the end of this section, we first introduce the concept of ergodicity and draw some consequences from the ergodic theorem. We recall that the subset $A \subset X$ is called T-invariant, if $T^{-1}(A) = A$.

Definition. Assume $(X, \mathcal{A}, m)$ to be a finite measure space. A measure preserving map $T\colon X \to X$ is called *ergodic*, if, for *every* T-invariant set $A \in \mathcal{A}$, the following holds true:

$$m(A) = m(X) \quad \text{or} \quad m(A) = 0.$$

In an ergodic system it is *not* possible to split X into two *invariant subsets* of *positive* measure.

The following proposition characterizes the ergodicity in a different way; the map T is ergodic, precisely if the T-invariant (measurable) functions are constant almost everywhere.

Proposition I.17 (Criterion for ergodicity). *We consider the finite measure space* $(X, \mathcal{A}, m)$ *and the measure preserving map* $T\colon X \to X$. *The following two statements are equivalent.*

(i) *T is ergodic.*

(ii) *For every* $f \in L(X, \mathcal{A}, m)$ *we conclude from* $f(T(x)) = f(x)$ *for all* $x \in X$ *that* $f(x) =$ *constant almost everywhere.*

Proof. (ii) $\Longrightarrow$ (i): We assume that the set $A \in \mathcal{A}$ satisfies $T^{-1}(A) = A$. Its characteristic function satisfies $\chi_A(Tx) = \chi_{T^{-1}(A)}(x) = \chi_A(x)$. In view of the assumption, χ_A is constant almost everywhere and we conclude that either $m(A) = 0$ or $m(A) = m(X)$.

(i) $\Longrightarrow$ (ii): We assume that the map T is ergodic and that the function $f \in L(X, \mathcal{A}, m)$ satisfies $f(Tx) = f(x)$ for all $x \in X$. If f is *not constant almost everywhere,* there exists a real number $c \in \mathbb{R}$ such that the set $A := \{x \in X \mid f(x) \geq c\}$ has the measure $0 < m(A) < m(X)$. Since f is invariant, also the set A is invariant, in contradiction to the ergodicity of T. □

The proof shows that in the statement (ii) of Proposition I.17 in place of $f \in L(X, \mathcal{A}, m)$ we can also take $f \in L^p(X, \mathcal{A}, m)$ for any $1 \leq p \leq \infty$ (remember that $m(X) < \infty$) or the measurable functions $f : X \to \overline{\mathbb{R}}$. Another characterization of ergodicity follows from the ergodic theorem.

Theorem I.18 (Ergodicity criterion). *Let* $(X, \mathcal{A}, m)$ *be a finite measure space and let the map* $T : X \to X$ *be measure preserving. Then the following two statements are equivalent.*

(i) T *is ergodic.*

(ii) *For every* $f \in L(X, \mathcal{A}, m)$ *there exists a null set* $N = N(f)$ *such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \frac{1}{m(X)} \int_X f$$

for every $x \in X \setminus N$, *i.e., for almost every orbit the mean value of the function* $f \in L$ ***over the orbit*** *equals the mean value of* f ***over the space*** X.

Proof [Theorem I.16, Proposition I.17]. (i) $\Longrightarrow$ (ii): If the map T is ergodic and $f \in L$, then, also $f^* \in L$ and $f^*(T(x)) = f^*(x)$ for every $x \in X$, in view of Theorem I.16. Hence, it follows from Proposition I.17 that $f^*(x) = c \in \mathbb{R}$ almost everywhere and consequently,

$$\int_X f = \int_X f^* = \int_X c = c\, m(X),$$

hence $c = \frac{1}{m(X)} \int f$ and the statement (ii) follows from Theorem I.16.

(ii) $\Longrightarrow$ (i): If $f \in L$ satisfies $f(Tx) = f(x)$ for all $x \in X$, we show that $f(x)$ is constant almost everywhere. Then, the statement (i) follows from Proposition I.17. In view of Theorem I.16, there exists a null set $N \subset X$, such that for $x \in X \setminus N$

we have

$$\begin{aligned} f(x) &\overset{\text{Invar.}}{=} \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(x)) \quad \text{for } all\ n \geq 1 \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(x)) \\ &\overset{\text{Vor.}}{=} \frac{1}{m(X)}\int f \end{aligned}$$

showing that $f(x)$ is constant for almost all x. □

Corollary I.19 (Equidistribution). *Let $(X, \mathcal{A}, m)$ be a finite measure space and let the map $T: X \to X$ be **ergodic**. If $A \in \mathcal{A}$ satisfies $m(A) > 0$, then there exists a null set $N = N(A) \subset X$ such that for $x \in X \setminus N$ the following holds true:*

$$\frac{1}{n}\#\{0 \leq j \leq n-1 \mid T^j x \in A\} \to \frac{m(A)}{m(X)}.$$

In general, $N \neq \emptyset$ and N depends on the set A.

Proof. We take the characteristic function $f = \chi_A$. Then $f \in L$ and according to Theorem I.18 there exists a null set $N = N(A)$ such that for $x \in X \setminus N$,

$$\frac{1}{n}\sum_{j=0}^{n-1} \chi_A(T^j x) \to \frac{1}{m(X)}\int \chi_A = \frac{m(A)}{m(X)}$$

as $n \to \infty$. □

Concerning the next consequence from the ergodic theorem, we note that for every *finite* measure space $(X, \mathcal{A}, m)$ satisfying $m(X) > 0$ a new standardized measure μ can be defined by $\mu(A) = m(A)/m(X)$ for every set $A \in \mathcal{A}$. We then have $\mu(X) = 1$ so that $(X, \mathcal{A}, \mu)$ is a *probability space*.

Corollary I.20. *Let $(X, \mathcal{A}, m)$ be a measure space satisfying $m(X) = 1$ and let $T: X \to X$ be a measure preserving map. Then, the following two statements are equivalent:*

(i) *T is ergodic.*

(ii) *For **all** $A, B \in \mathcal{A}$ we have*

$$\frac{1}{n}\sum_{j=0}^{n-1} m(T^{-j}(A) \cap B) \to m(A)m(B).$$

Proof [Ergodic theorem, convergence theorem of Lebesgue]. (i) $\Longrightarrow$ (ii): Taking the characteristic function $f = \chi_A$, we conclude from the ergodic theorem (Theorem I.16) that

$$\frac{1}{n}\sum_{j=0}^{n-1} \chi_A(T^j x) \to m(A) \quad \text{for almost every } x.$$

Therefore,

$$(*) \quad \frac{1}{n}\sum_{j=0}^{n-1} \chi_A(T^j x)\chi_B(x) \to m(A)\chi_B(x) \quad \text{for almost every } x.$$

In view of $\chi_A(T^j x)\chi_B(x) = \chi_{T^{-j}(A)}(x)\chi_B(x) = \chi_{T^{-j}(A)\cap B}(x)$ the left side function in the formula $(*)$ is majorized by the characteristic function $\chi_X \equiv 1$ of the whole space. By integrating $(*)$ and using the convergence theorem of Lebesgue we obtain

$$\frac{1}{n}\sum_{j=0}^{n-1} m(T^{-j}(A) \cap B) \to m(A)m(B).$$

(ii) $\Longrightarrow$ (i): If $E \in \mathcal{A}$ is a T-invariant set, we take $A = B = E$ in the statement (ii) and find

$$m(E) = \frac{1}{n}\sum_{j=0}^{n-1} m(E) \to m(E)^2$$

from which we conclude that $m(E) = 0$ or $m(E) = 1$. Therefore, the map T is ergodic. □

Ergodicity is a property of a measure preserving map on a measure space, while transitivity is a property of a continuous map on a topological space. Next, we equip a measure space with a topology and show that under additional assumptions an ergodic system is also transitive.

Proposition I.21 (Transitivity of ergodic systems). *Let $(X, \mathcal{A}, m)$ be a finite measure space and let $T : X \to X$ be an* ***ergodic*** *map. Moreover, assume that (X, d) is a metric space possessing a countable basis of open sets. Furthermore, assume that every open set $U \neq \emptyset$ is measurable and of positive measure $m(U) > 0$. Under these assumptions, there exists a null set $N \subset X$ such that*

$$\overline{\mathcal{O}^+(x)} = X \text{ for } x \in X \setminus N,$$

i.e., almost all orbits are dense.

If, in addition, the map $T : X \to X$ is continuous and if the metric space (X, d) is complete, the set

$$R^+ := \{x \in X \mid \overline{\mathcal{O}^+(x)} = X\}$$

is dense in X and of second Baire category.

The set R^+ of initial points of dense orbits is large in the sense of the measure ($m(R^+) = m(X)$) and, under the additional assumptions, it is also large in the functional analytic sense of the Baire category.

Example. The assumptions on the space X are fulfilled for the circle S^1 and the torus T^n equipped with the Lebesgue measure and we shall later on look at these examples again.

Proof of Proposition I.21 [Ergodic theorem, category theorem of Baire]. (1) We take the countable base $(V_k)_{k\geq 1}$ of open sets of X satisfying $V_k \neq \emptyset$. According to Corollary I.19 (*equidistribution*) there exists for V_k a null set $N_k \subset X$ such that for all $x \in X \setminus N_k$ we have $T^j x \in V_k$ for infinitely many integers j, using that $m(V_k) > 0$. In particular, $\mathcal{O}^+(x) \cap V_k \neq \emptyset$ for $x \in X \setminus N_k$. This holds true for every $k \geq 1$. The countable union $N := \bigcup_{k\geq 1} N_k$ is a null set, so that $m(N) = 0$ and, of course, $\mathcal{O}^+(x) \cap V_k \neq \emptyset$ for $x \in X \setminus N$ and for every $k \geq 1$. If $\emptyset \neq V \subset X$ is any open set, then, according to the definition of the base, there exists an index k such that $V_k \subset V$ and therefore,

$$\mathcal{O}^+(x) \cap V \neq \emptyset, \quad x \in X \setminus N.$$

This holds true for every open set $\emptyset \neq V$ so that

$$\overline{\mathcal{O}^+(x)} = X, \quad x \in X \setminus N.$$

Moreover, $m(R^+) = m(X)$, since N is a null set.

(2) We assume now that the map T is in addition *continuous*. Then for every $k \geq 1$ the set

$$\mathcal{O}^-(V_k) := \bigcup_{j\geq 1} T^{-j}(V_k)$$

is open. If $\emptyset \neq U$ is open, then $m(U) > 0$, and due to $m(V_k) > 0$, we can argue as above to find a point $x \in U$ satisfying $T^j x \in V_k$ for infinitely many j, in particular,

$$U \cap \mathcal{O}^-(V_k) \neq \emptyset.$$

Therefore, the open set $\mathcal{O}^-(V_k)$ is dense, and this holds true for every $k \geq 1$. If X is *complete*, then, according to the category theorem of Baire (Lemma I.10) the countable intersection

$$R := \bigcap_{j\geq 1} \mathcal{O}^-(V_j)$$

of open and dense sets is dense in X and, by definition, of second Baire category. As in the proof of Theorem I.9 one shows for $p \in R$ that

$$\overline{\mathcal{O}^+(p)} = X.$$

Therefore, $R \subset R^+$ and the theorem is proved. □

Before we prove the ergodicity of our dynamical systems on the circle and on the torus described above, it is useful to recall in a short interlude some results about Fourier series. The proofs for most of the statements can be found in standard textbooks on analysis as e.g. in the textbook of K. Stromberg [111], or in the classic [31] by H. Dym and H. P. McKean. The theorem by L. Carleson, as well as the example by A. N. Kolmogorov below, can be found in the book [46] by L. Grafakos.

Facts about Fourier series. In the following we consider integrable functions $f \in L^1(S^1)$ which are measurable functions $x \mapsto f(e^{2\pi i x})$, periodic of period 1 and, abbreviating $f(e^{2\pi i x}) \equiv f(x)$, satisfy

$$\int_0^1 |f(x)|\, dx < \infty.$$

Fourier coefficients. With $f \in L^1(S^1)$ one can associate the *sequence* $(\hat{f}(n))_{n\in\mathbb{Z}}$ of *numbers* (called *Fourier coefficients*) defined by

$$\hat{f}(n) := \int_0^1 f(x)e^{-2\pi i n x}\, dx, \quad n \in \mathbb{Z}.$$

Two classical statements about Fourier coefficients are the following.

- *Riemann–Lebesgue lemma:* If $f \in L^1(S^1)$, then

$$\hat{f}(n) \to 0, \quad |n| \to \infty.$$

- If $f, g \in L^1(S^1)$, the following statement holds true:

$$\hat{f}(n) = \hat{g}(n) \text{ for all } n \in \mathbb{Z} \implies f = g \text{ almost everywhere.}$$

  In other words an element of L^1 is uniquely determined by its Fourier coefficients so that the map $f \mapsto \hat{f}$ is *injective*.

The Hilbert space $L^2(S^1)$. The *Hilbert space* $L^2(S^1)$ is equipped with the scalar product

$$(f, g) = \int_0^1 f(x)\overline{g(x)}\, dx$$

and possesses the orthonormal system $(e_n)_{n\in\mathbb{Z}}$ defined by

$$e_n := e^{2\pi i n x}, \quad n \in \mathbb{Z},$$

and satisfying $(e_n, e_k) = \delta_{ij}$. Thus, for $f \in L^2(S^1)$ the Fourier coefficients can be written as

$$\hat{f}(n) = (f, e_n), \quad n \in \mathbb{Z}.$$

Fourier series. If $f \in L^1(S^1)$, we abbreviate $S_n(f) := \sum_{|k|\le n} \hat{f}(k)e_k$. The trigonometrical polynomial $S_n(f)$ belongs to $C^\infty(S^1)$ and is equal to

$$S_n(f)(x) = \sum_{|k|\le n} \hat{f}(k)e^{2\pi i k x}.$$

A classical result in Hilbert space is as follows.

- *Theorem of Riesz–Fischer*: If $f \in L^2(S^1)$, then its Fourier series

$$f = \sum_{k\in\mathbb{Z}} \hat{f}(k)e_k$$

  converges in $L^2(S^1)$,

$$\|f - S_n(f)\|_{L^2(S^1)} \to 0 \quad \text{for } n \to \infty,$$

  where $\|f\|_{L^2(S^1)} := (f, f)^{1/2}$ is the norm in a Hilbert space.

In this sense the functions $(e_n)_{n\in\mathbb{Z}}$ constitute a *Hilbert basis* of $L^2(S^1)$, i.e., $\overline{\text{span}}((e_n)_{n\in\mathbb{Z}}) = L^2(S^1)$. A Hilbert space is separable precisely if it possesses a countable Hilbert base. Such a space is *isometrically isomorphic* to the sequence space

$$\ell^2 := \{c = (c_n)_{n\in\mathbb{Z}} \mid \sum_{k\in\mathbb{Z}} |c_k|^2 =: \|c\|_2^2 < \infty\}.$$

The isomorphism is the linear map $f \mapsto \hat{f} = (\hat{f}(n))_{n\in\mathbb{Z}}$ which satisfies

$$\|f\|^2_{L^2(S^1)} = \sum_{k\in\mathbb{Z}} |\hat{f}(k)|^2 = \|(\hat{f}(n))_{n\in\mathbb{Z}}\|_2^2,$$

called *Plancherel identity*.

Pointwise convergence. The question of *pointwise* convergence is subtle, as the following results show.

- Given a null set $N \subset [0,1]$, there exists a continuous function $f \in C(S^1)$ such that $S_n(f)(x)$ *diverges* for *all* $x \in N$.
- *A. N. Kolmogorov* (1926): There exists an element $f \in L^1(S^1)$ such that $S_n(f)(x)$ *diverges* for *every* $x \in S^1$!
- *L. Carleson* (1966): If $f \in L^2(S^1)$, then

$$S_n(f)(x) \to f(x) \quad \text{almost everywhere.}$$

- *R.A. Hunt* (1968): If $p > 1$ and $f \in L^p(S^1)$, then

$$S_n(f)(x) \to f(x) \quad \text{almost everywhere.}$$

- Assume that $f \in C^p(S^1)$ and $p \geq 1$. Then

$$\sup_{x \in [0,1]} |f(x) - S_n(f)(x)| \leq c(f) n^{-p+\frac{1}{2}},$$

  where $c(f)$ is a constant. The convergence is *uniform* and the faster, the smoother the function is.

After these recollections of Fourier series, we return to our simple examples of dynamical systems and investigate their ergodicity.

Examples. (1) *Rigid rotations.* We return to the mapping

$$\varphi \colon S^1 \to S^1, \quad \varphi(z) = \vartheta z,$$

where $\vartheta = e^{2\pi i \alpha}$ and $z = e^{2\pi i x}$ for a real number $x \in \mathbb{R}$.

*Claim***.** We first claim that the rigid rotation φ is *measure preserving* with respect to the Lebesgue measure.

Proof. We assume $f \colon S^1 \to \mathbb{C}$ to be integrable and define

$$F(x) := f(e^{2\pi i x}).$$

Then $F(x+1) = F(x)$ for $x \in \mathbb{R}$ and F is locally integrable. Recalling the translation $\Phi(x) = x + \alpha$ we calculate

$$\begin{aligned}
\int_0^1 F(\Phi(x))\,dx &= \int_0^1 F(x+\alpha)\,dx \qquad (y = x+\alpha)\\
&= \int_\alpha^{\alpha+1} F(y)\,dy\\
&= \int_\alpha^1 F(y)\,dy + \int_1^{\alpha+1} F(y-1)\,dy\\
&= \int_\alpha^1 F(y)\,dy + \int_0^\alpha F(y)\,dy\\
&= \int_0^1 F(y)\,dy.
\end{aligned}$$

One concludes that $\int_{S^1} f \circ \varphi = \int_{S^1} f$ for every $f \in L(S^1)$, so that the map is indeed measure preserving as claimed. □

- The case $\alpha = p/q \in \mathbb{Q}$.
  In this case there exist invariant functions $f\colon S^1 \to \mathbb{C}$ which are not constant, as the example $f(z) = z^q = e^{2\pi i q x}$ shows. Indeed, due to $\vartheta^q = 1$ we find
  $$f(\varphi(z)) = (\vartheta z)^q = z^q = f(z), \quad z \in S^1.$$
  Having found an invariant function which is not constant we conclude from Proposition I.17 that the map φ is *not ergodic*.
- The case $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.
  In this case φ is *ergodic*. To prove this, it suffices to show that all invariant functions $f \in L^2(S^1)$ are constant almost everywhere, cf. the remark following Proposition I.17. If $f \in L^2(S^1)$ we look at its *Fourier series*
  $$f(z) = \sum_{k\in\mathbb{Z}} f_k z^k, \quad z = e^{2\pi i x}$$
  which converges in L^2. Then,
  $$f(\varphi(z)) = f(\vartheta z) = \sum_{k\in\mathbb{Z}} (f_k \vartheta^k) z^k.$$
  If f is invariant, $f(\varphi(z)) = f(z)$, we conclude from the *uniqueness of the Fourier coefficients* that
  $$f_k = \vartheta^k f_k, \quad k \in \mathbb{Z}.$$
  Due to $\vartheta^k \neq 1$ if $k \neq 0$ (since α irrational) it follows that $f_k = 0$ for all $k \neq 0$, so that $f(z) = f_0$ is constant almost everywhere. This proves that the irrational rotation of the circle is ergodic.

Exercise. Consider the map $\varphi\colon T^n \to T^n$ on the torus which in the covering space $\mathbb{R}^n$ is the translation
$$\Phi(x) = x + \omega, \quad x \in \mathbb{R}^n,$$
and assume that
$$\langle \omega, j \rangle \notin \mathbb{Z}, \quad j \in \mathbb{Z}^n \setminus \{0\}.$$
Prove that φ is *ergodic* with respect to the Lebesgue measure on T^n.

(2) *Expansion*. Next, we return to our expanding map $\varphi\colon S^1 \to S^1$ defined by $\varphi(z) = z^2$. The map φ is measure preserving with respect to the Lebesgue measure, as we have already seen, and we prove the ergodicity of the map φ by demonstrating that the integrable invariant functions are constant almost everywhere. Let $f \in L(S^1)$ and introduce the function $g(z) := f(\varphi(z)) = f(z^2)$. If $F(x) :=$

$f(e^{2\pi ix})$, then $F(x+1) = F(x)$ for all $x \in \mathbb{R}$, and $G(x) := g(e^{2\pi ix}) = F(2x)$. Computing the *Fourier coefficients* g_{2k} of order $2k$ of the function G, we obtain

$$\begin{aligned} g_{2k} &:= \int_0^1 G(x)e^{-2\pi i2kx}\,dx \\ &= \int_0^1 F(2x)e^{-2\pi i2kx}\,dx \\ &= \int_0^{1/2} F(2x)e^{-2\pi i2kx}\,dx + \int_{1/2}^1 F(2x-1)e^{-2\pi i2kx}\,dx \\ &= \frac{1}{2}\int_0^1 F(y)e^{-2\pi iky}\,dy + \frac{1}{2}\int_0^1 F(y)e^{-2\pi iky}\,dy \\ &= \int_0^1 F(y)e^{-2\pi iky}\,dy \\ &=: f_k. \end{aligned}$$

If f is invariant, then $G(x) = F(2x) = F(x)$ and hence $g_k = f_k$. In view of the above calculation, $f_{2k} = g_{2k} = f_k$ for all $k \in \mathbb{Z}$ and we see that

$$f_k = f_{2k} = f_{2^2k} = \cdots = f_{2^nk} = \cdots.$$

Since $\lim_{|k|\to\infty} f_k = 0$ in view of the Riemann–Lebesgue lemma, we conclude that

$$f_k = 0, \quad \text{for all } k \neq 0.$$

Therefore, $F(x) = f_0$ is constant almost everywhere and according to Proposition I.17, our map is ergodic.

Remark. We have verified that the map $\varphi\colon z \mapsto z^2$ on the circle S^1 satisfies all the assumptions of Proposition I.21, from which *the statement* (iii) *in Proposition* I.12 *now follows.* Namely, the set $R^+ := \{z \in S^1 \mid \overline{\mathcal{O}^+(z)} = S^1\}$ of initial points of dense orbits has full measure,

$$m(R^+) = m(S^1),$$

and, in addition, the set R^+ is *dense* in S^1 and of second Baire category.

Exercise. Every real number $0 < x \leq 1$ can be represented by a unique decimal expansion

$$x = 0.x_1x_2x_3\cdots = \sum_{j\geq 1}\frac{x_j}{10^j}$$

containing infinitely many non zero digits $x_j \in \{0,1,2,\ldots,9\}$. Demonstrate that on the average, the number of zeros in the decimal expansion is equal to $1/10$ for *almost all* $0 < x \leq 1$. Hint: Consider the mapping $\varphi\colon S^1 \to S^1$ defined by $\varphi(z) = z^{10}$ in the fundamental domain of the covering space. It is the map $T\colon (0,1] \to (0,1]$ defined by $T(x) = 10x \mod 1$. The map T is measure preserving and ergodic. Consider now the real number $x = 0.x_1x_2x_3\ldots$ and the interval $A := (0, 1/10)$, then

$$T^j x \in A \iff x_{j+1} = 0.$$

Apply the equidistribution theorem.

Proof of the ergodic theorem. We conclude this chapter with the proof of the ergodic theorem following the arguments of A. M. Garsia (1965) and using the maximal ergodic lemma. We recall the statement of the theorem.

Theorem I.16. *We assume that $(X, \mathcal{A}, m)$ is a **finite** measure space and $T\colon X \to X$ a **measure preserving** map. For every integrable function $f \in L(X, \mathcal{A}, m)$ there exist a function $f^* \in L(X, \mathcal{A}, m)$ and a null set $N \subset X$ (N **depending** on f) satisfying $T^{-1}(N) = N$ and*

(i) $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = f^*(x)$ *for all* $x \in X \setminus N$,

(ii) $f^*(Tx) = f^*(x)$ *for all* $x \in X$,

(iii) $\int_X f^* = \int_X f$,

(iv) $\int_X \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) - f^*(x) \right| dm \to 0$, $n \to \infty$.

Proof [Convergence theorems from integration theory]. If $f \in L(X, \mathcal{A}, m)$ we abbreviate in the following

$$S_n(x) = S_n(f, x) = \sum_{j=0}^{n-1} f(T^j x), \quad n \geq 1.$$

Our first aim is to prove that the limit $\lim_{n\to\infty} \frac{1}{n} S_n(x)$ does exist almost everywhere, by showing that

$$m\Big(\Big\{x \in X \mid \underline{\lim} \frac{1}{n} S_n(x) < \overline{\lim} \frac{1}{n} S_n(x)\Big\}\Big) = 0.$$

(1) *Maximal ergodic lemma.* We introduce the sequence of integrable functions

$$S_0(x) \equiv 0,$$
$$S_n^+(x) = \max_{0 \leq k \leq n} S_k(x)$$

and observe that $S_n^+(x) \geq 0$ for all $x \in X$.

Lemma I.22. *Let $f \in L(X, \mathcal{A}, m)$ and define the subsets*

$$A_n = \{x \in X \mid S_n^+(x) > 0\}.$$

Then

$$\int_{A_n} f\,dm \geq 0, \quad \textit{for all } n \geq 1.$$

Proof. From the definition of S_n^+ it follows that

$$f(x) + S_n^+(Tx) \geq f(x) + S_k(Tx) = S_{k+1}(x), \quad 0 \leq k \leq n.$$

If $x \in A_n$, then there exists an integer $1 \leq k \leq n$ for which $S_k(x) > 0$. Consequently,

$$f(x) + S_n^+(Tx) \geq \max_{1\leq k\leq n} S_k(x) = \max_{0\leq k\leq n} S_k(x) = S_n^+(x)$$

and hence $f(x) \geq S_n^+(x) - S_n^+(Tx)$. By using that S_n^+ is equal to zero outside of A_n and $S_n^+(Tx) \geq 0$ everywhere, we conclude from this inequality and from the measure preservation of the map T that

$$\begin{aligned}\int_{A_n} f &\geq \int_{A_n} S_n^+(x) - \int_{A_n} S_n^+(Tx)\\ &= \int_X S_n^+(x) - \int_{A_n} S_n^+(Tx)\\ &\geq \int_X S_n^+(x) - \int_X S_n^+(Tx)\\ &= \int_X S_n^+(x) - \int_X S_n^+(x)\\ &= 0\end{aligned}$$

and the lemma is proved. □

(2) For $a < b$ we define the subset Y of X by

$$Y = Y(a,b) = \left\{x \in X \mid \underline{\lim}\, \frac{1}{n} S_n(x) < a < b < \overline{\lim}\, \frac{1}{n} S_n(x)\right\}.$$

Lemma I.23. *The set Y is measurable and invariant under the map T, i.e., $T^{-1}(Y) = Y$.*

Proof. The statement $Y \in \mathcal{A}$ follows from elementary measure theory, since the functions S_n are all measurable. To prove the invariance, we have to show that $\underline{\lim}\, \frac{1}{n} S_n(x) = \underline{\lim}\, \frac{1}{n} S_n(Tx)$ and $\overline{\lim}\, \frac{1}{n} S_n(x) = \overline{\lim}\, \frac{1}{n} S_n(Tx)$. However, this follows immediately from the identity

$$\frac{1}{n} S_n(Tx) = \frac{1}{n} S_{n+1}(x) - \frac{1}{n} f(x) = \frac{1}{n+1} S_{n+1}(x) \underbrace{\frac{n+1}{n}}_{\to 1} - \underbrace{\frac{1}{n} f(x)}_{\to 0}$$

by taking the $\underline{\lim}$ respectively $\overline{\lim}$, and the lemma is proved. □

Lemma I.24. *From $m(X) < \infty$ it follows that $m(Y) = 0$.*

Proof. We apply Lemma I.22 to the T-invariant set Y (instead of X) and to the function g (instead of f), defined by

$$g(x) := f(x) - b.$$

Since $m(X) < \infty$ the function $g \in L(X, \mathcal{A}, m)$ is integrable. Hence, setting

$$A_n = \{x \in Y \mid S_n^+(g, x) > 0\},$$

we have

$$\int_{A_n} (f - b) = \int_Y \chi_{A_n}(f - b) \geq 0.$$

According to the definition of Y there exists for every point $x \in Y$ an integer j for which

$$\frac{1}{j} S_j(g, x) = \frac{1}{j} S_j(f, x) - b > 0$$

and hence $S_j^+(g, x) > 0$. Consequently every $x \in Y$ is contained in some set A_j, so that $Y = \bigcup_{n \geq 1} A_n$. The monotonicity of the sequence of sets $A_n \subset A_{n+1} \subset \cdots$ implies $\lim \chi_{A_n}(x) = \chi_Y(x)$. Using the convergence theorem of Lebesgue we obtain in the limit as $n \to \infty$,

$$\int_Y (f - b) \geq 0.$$

In exactly the same way one proves $\int_Y (a - f) \geq 0$. Addition of the inequalities results in the inequality

$$(a - b) \int_Y dm = (a - b) m(Y) \geq 0.$$

Since $a < b$, one concludes $m(Y) = 0$ and the lemma is proved. $\square$

(3) *Pointwise convergence.* In view of Lemma I.24 the set

$$N_0 := \left\{x \in X \mid \underline{\lim} \frac{1}{n} S_n(x) < \overline{\lim} \frac{1}{n} S_n(x)\right\},$$

is the countable union of null sets

$$N_0 = \bigcup_{\substack{a<b \\ a,b \in \mathbb{Q}}} Y(a, b),$$

and hence a null set, so that $m(N_0) = 0$. For $x \notin N_0$, hence for *almost all* $x \in X$, we abbreviate the limit

$$\underline{\lim} \frac{1}{n} S_n(x) = \overline{\lim} \frac{1}{n} S_n(x) = \lim_{n \to \infty} \frac{1}{n} S_n(x) =: \varphi(x),$$

where $-\infty \leq \varphi(x) \leq +\infty$. Defining $\varphi(x) = 0$ on the null set N_0 we show next that $\varphi(x) \in \mathbb{R}$ is a real number almost everywhere. This will follow once we have proved that $\varphi \in L(X, \mathcal{A}, m)$. Since T is measure preserving, we have $\int_X |f(T^j x)|\, dm = \int_X |f|\, dm$ for every $j \geq 0$ and therefore,

$$\int_X \left| \frac{1}{n} S_n(x) \right| dm \leq \int_X |f|\, dm.$$

Using the Lemma of Fatou we can estimate

$$\int_X |\varphi| = \int_X \underline{\lim} \left| \frac{1}{n} S_n(x) \right| \leq \underline{\lim} \int_X \left| \frac{1}{n} S_n(x) \right| \leq \int_X |f| < \infty.$$

Hence $\varphi \in L(X, \mathcal{A}, m)$ and therefore, $\varphi(x) \in \mathbb{R}$ almost everywhere. The set

$$N = N_0 \cup \{x \in X \mid |\varphi(x)| = \infty\},$$

on which $\lim_{n\to\infty} \frac{1}{n} S_n(x)$ does not exist in $\mathbb{R}$, is therefore a null set. Since the limits $\underline{\lim}$, $\overline{\lim}$ are invariant under T (due to Lemma I.23), the set N is also invariant under T. We now *define* the function $f^*\colon X \to \mathbb{R}$ showing up in the statement of the theorem as follows:

$$f^*(x) := \begin{cases} \lim_{n\to\infty} \frac{1}{n} S_n(f, x), & x \notin N, \\ 0, & x \in N. \end{cases}$$

Then $f^* \in L(X, \mathcal{A}, m)$ and one sees, as in Lemma I.23, that $f^*(Tx) = f^*(x)$ for all $x \in X$. So far, we have proved the statements (i) and (ii) of the ergodic theorem.

(4) It remains to show that

(a) $\frac{1}{n} S_n(f, x) \to f^*$ in L^1,
(b) $\int_X f^* = \int_X f$.

We begin by proving a *special case*. We assume that f is a bounded function assuming that $|f| \leq K$ for a constant $K > 0$. Then,

$$\left| \frac{1}{n} S_n(x) \right| \leq K,$$

and the statement (a) follows by means of the dominated convergence theorem of Lebesgue. Since the map T is measure preserving,

$$\int_X \frac{1}{n} S_n(x) = \int_X f, \quad n \geq 1,$$

and the statement (b) follows, again using the convergence theorem of Lebesgue,

$$\int_X f = \lim_{n\to\infty} \int_X \frac{1}{n} S_n(x) = \int_X \lim_{n\to\infty} \frac{1}{n} S_n(x) = \int_X f^*.$$

(5) We next prove the *general case by approximation* with the truncated functions of $f \in L(X, \mathcal{A}, m)$, defined by

$$f_N(x) = \begin{cases} f(x), & |f(x)| \le N, \\ 0, & |f(x)| > N. \end{cases}$$

Then, $|f_N| \le N$, and in view of the special case we have the pointwise convergence

$$\frac{1}{n} S_n(f_N, x) \to f_N^*(x), \quad n \to \infty$$

almost everywhere and also the convergence in L^1. Let $\varepsilon > 0$. By the triangle inequality one estimates

$$\int \left| \frac{1}{n} S_n(f) - f^* \right| dm \le \int \left| \frac{1}{n} S_n(f) - \frac{1}{n} S_n(f_N) \right| dm + \int \left| \frac{1}{n} S_n(f_N) - f_N^* \right| dm + \int |f_N^* - f^*| \, dm,$$

and we are going to estimate each term on the right-hand side by $\varepsilon/3$ if N and n are sufficiently large.

Since $f \in L$ we have $|f| < \infty$ almost everywhere, and since $f_N(x) = f(x)\chi_{|f| \le N}(x) \to f(x)$ we conclude by the convergence theorem of Lebesgue that

$$f_N \to f \text{ in } L^1.$$

Therefore, choosing N large enough,

$$\int |f_N - f| \le \varepsilon/3.$$

Using that the map T is measure preserving we can now estimate the first term as follows.

$$\begin{aligned} \int \left| \frac{1}{n} S_n(f_N) - \frac{1}{n} S_n(f) \right| dm &= \int \frac{1}{n} \left| \sum_{j=0}^{n-1} (f_N - f)(T^j x) \right| dm \\ &\le \frac{1}{n} \sum_{j=0}^{n-1} \int |(f_N - f)(T^j x)| \, dm \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \int |(f_N - f)(x)| \, dm \\ &= \int |f_N - f| \, dm \\ &\le \varepsilon/3. \end{aligned}$$

This holds true for *all* n and the *first term* is taken care of. We now apply the *Lemma of Fatou* to the first term to obtain as $n \to \infty$,

$$\int |f_N^* - f^*| \, dm \leq \varepsilon/3,$$

which is the desired estimate of the *third term*. For the *second term* we conclude

$$\int \left| \frac{1}{n} S_n(f_N) - f_N^* \right| dm \leq \varepsilon/3$$

for n large enough, from the convergence statement in the special case. Altogether,

$$\int \left| \frac{1}{n} S_n(f) - f^* \right| \leq \varepsilon$$

for n large enough. This is true for every $\varepsilon > 0$. Therefore,

$$\frac{1}{n} S_n(f) \to f^* \text{ in } L^1$$

and the proof of the statement (a) in the general case is completed. As for the statement (b) we recall that, because of the measure preservation of the map T,

$$\int \frac{1}{n} S_n(f) \, dm = \int f$$

for all $n \geq 1$, and using the L^1-convergence we finally obtain

$$\int f = \lim_{n \to \infty} \int \frac{1}{n} S_n(f) = \int f^*.$$

Herewith, the ergodic theorem of Birkhoff is proved. □

Literature. There are several monographs on dynamical systems which cover most of the topics treated in the first four chapters. Among them [49] by B. Hasselblatt and A. Katok, [113] by W. Szlenk, [63] by R. Mañé and [91] by C. Robinson. The special topic of ergodic theory is treated, for example, in the monographs [23] by I. Cornfield, S. V. Fomin and Ya. G. Sinai, [108] and [109] by Ya. G. Sinai, [88] by M. Pollicott, [60] by U. Krengel, [37] H. Fuerstenberg and [115] by P. Walters. For surveys about specific problems in dynamical systems including historical information and many references we recommend the Handbooks of Dynamical Systems [50] and [35].