Ergodic Ramsey Theory - SNSB Lecture

Laurenţiu Leuştean

January 31, 2011

Contents

Part I

Topological Dynamics and Partition Ramsey Theory

Chapter 1

Topological Dynamical Systems

In the sequel, we shall use the following notations:

$$\begin{aligned}\mathcal{P}(D) &= \text{the power set of } D,\\ \mathbb{N} &= \{0,1,2,\ldots\}, \quad \mathbb{Z}_+ = \{1,2,\ldots\},\\ [m,n] &= \{m, m+1, \ldots, n\} \quad for\ m \le n \in \mathbb{Z}.\end{aligned}$$

Definition 1.0.1. *A* **topological dynamical system** *(TDS for short) is a pair (X,T), where X is a compact Hausdorff nonempty topological space and $T : X \to X$ is a continuous mapping. The TDS (X,T) is called* **invertible** *if T is a homeomorphism.*

An invertible TDS (X,T) defines two "one-sided" TDSs, namely the **forward system** (X,T) and the **backward system** (X,T^{-1}).

Topological dynamics is about what happens when the map T is applied repeatedly. If one takes a point $x \in X$, then we are primarily interested in the behaviour of $T^n x$ as n tends to infinity. Some basic questions one might ask are:

(i) Will two points that are close to each other initially, stay close even after a long time?

(ii) Will a point return to its original position (at least very near to it)?

(iii) Will a certain point x never leave a certain region or will it come arbitrarily close to any other given point ot X?

Let (X,T) be a TDS and $x \in X$. The **forward orbit** of x is given by

$$\mathrm{orb}_+(x) = \{T^n x \mid n \in \mathbb{N}\} = \{x, Tx, T^2x, \ldots\}. \tag{1.1}$$

If (X,T) is invertible, the **(total) orbit** of x is

$$\mathrm{orb}(x) = \{T^n x \mid n \in \mathbb{Z}\}. \tag{1.2}$$

We shall write

$$\overline{\mathrm{orb}_+}(x) = \overline{\{T^n x \mid n \in \mathbb{N}\}} \quad \text{and} \quad \overline{\mathrm{orb}}(x) = \overline{\{T^n x \mid n \in \mathbb{Z}\}} \tag{1.3}$$

for the closure of the forward and the total orbit, respectively.

Furthermore, we shall use the notation

$$\mathrm{orb}_{>0}(x) = \{T^n x \mid n \in \mathbb{Z}_+\} = \mathrm{orb}_+(x) \setminus \{x\} = \mathrm{orb}_+(Tx) = \{Tx, T^2x, T^3x, \dots\}. \tag{1.4}$$

It is obvious that many notions, like the forward orbit of a point x, do make sense in the more general setting of a continuous self-map of a topological space. However, we restrict ourselves to compact Hausdorff spaces and reserve the term TDS for this special situation.

Lemma 1.0.2. *Let (X,T) be a TDS and $U \subseteq X$.*

(i) $T(orb_+(x)) = orb_{>0}(x)$.

(ii) For all $x \in X$, $orb_+(x) \cap U \neq \emptyset$ iff $x \in \bigcup_{n\geq 0} T^{-n}(U)$.

(iii) If (X,T) is invertible, then for all $x \in X$, $orb(x) \cap U \neq \emptyset$ iff $x \in \bigcup_{n\in\mathbb{Z}} T^{n}(U)$.

Definition 1.0.3. *Let (X,T) be a TDS. A point $x \in X$ is called* **periodic** *if there is $n \geq 1$ such that $T^n x = x$.*

Thus, x is periodic if and only if $x \in \mathrm{orb}_{>0}(x)$.

1.1 Examples

Let us give some examples of topological dynamical systems.

1.1.1 Finite state spaces

Let X be a finite set with the discrete metric. Then X is a compact metric space and every map $T : X \to X$ is continuous. The TDS (X,T) is invertible if and only if T is injective (or surjective).

1.1.2 Finite-dimensional linear nonexpansive mappings

Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be linear and nonexpansive with respect to the chosen norm, i.e.:

$$\|Tx - Ty\| \leq \|x - y\| \text{ for all } x, y \in \mathbb{R}^n. \tag{1.5}$$

Lemma 1.1.1. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be linear. The following are equivalent* *(i) T is nonexpansive*

(ii) $\|Tx\| \leq \|x\|$ for all $x \in \mathbb{R}^n$.

Proof. $(i) \Rightarrow (ii)$ Take $y = 0$ in (1.5) and use the fact that $T0 = 0$.
$(ii) \Rightarrow (i)$ Since T is linear, $\|Tx - Ty\| = \|T(x - y)\| \leq \|x - y\|$. □

Then the unit ball $K := \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ is compact and $T|_K$ is a continuous self-map of K.

Hence, $(K, T|_K)$ is a TDS.

1.1.3 Translations on compact groups

Let G be a compact group.

For every $a \in G$, let

$$L_a : G \to G, \quad L_a(g) = a \cdot g.$$

be the left translation. By D.0.4, L_a is a homeomorphism for all $a \in G$.

Hence, (G, L_a) is an invertible TDS.

1.1.4 Rotations on the circle group

The unit circle $\mathbb{S}^1 = \{z \in \mathbb{C} \mid |z| = 1\}$ with the group operation being multiplication is an abelian compact group, called the **circle group**.

Since the group is abelian, left and right translations coincide, we call them **rotations** and denote them R_a for $a \in \mathbb{S}^1$.

Hence, $(\mathbb{S}^1, R_a)$ is an invertible TDS.

1.1.5 Rotations on the n-torus $\mathbb{T}^n$

The **n-dimensional torus**, often called the **n-torus for short** is the topological space

$$\mathbb{T}^n := \mathbb{S}^1 \times \ldots \times \mathbb{S}^1.$$

with the product topology. The 2-dimensional torus is simply called the **torus**.

If we define the multiplication on $\mathbb{T}^n$ pointwise, the n-torus $\mathbb{T}^n$ becomes another example of an abelian compact group. For any $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{T}^n$, the **rotation** by $\mathbf{a}$ is given by

$$R_{\mathbf{a}} : \mathbb{T}^n \to \mathbb{T}^n, \quad R_{\mathbf{a}}(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} = (a_1 x_1, \ldots, a_n x_n) \quad \text{for all } \mathbf{x} = (x_1, x_2 \ldots, x_n) \in \mathbb{T}^n. \tag{1.6}$$

Then $(\mathbb{T}^n, R_{\mathbf{a}})$ is a TDS.

1.1.6 The tent map

Let $[0,1]$ be the unit interval and define the **tent map** by

$$T:[0,1]\to[0,1],\quad T(x)=1-|2x-1|=\begin{cases}2x & \text{if } x<\frac{1}{2}\\ 2(1-x) & \text{if } x\geq\frac{1}{2}.\end{cases}\tag{1.7}$$

It is easy to see that T is well-defined and continuous. Since $[0,1]$ is a compact subset of $\mathbb{R}$, we get that (X,T) is a TDS.

1.2 The shift

Let W be a finite nonempty set of **symbols** which we will call the **alphabet**. **We assume** $|W|\geq 2$. Elements of W are also called **letters**, and they will typically be denoted by $a,b,c,\ldots$ or by digits $0,1,2,\ldots$.

Although in real life sequences of symbols are finite, it is often extremely useful to treat long sequences as infinite in both directions (or **bi-infinite**).

Definition 1.2.1. *The* **full W-shift** *is the set $W^{\mathbb{Z}}$ of all bi-infinite sequences of symbols from W, i.e. sequences taking values in W indexed by $\mathbb{Z}$. The* **full r-shift** *(or simply* **r-shift***) is the full shift over the alphabet* $\{0,1,\ldots,r-1\}$.

We shall denote with boldface letters $\mathbf{x},\mathbf{y},\mathbf{z},\ldots$ the elements of $W^{\mathbb{Z}}$ and call them also **points** of $W^{\mathbb{Z}}$. Points from the full 2-shift are also called **binary sequences**. If W has size $|W|=r$, then there is a natural correspondence between the full W-shift and the full r-shift, and sometimes the distinction between them is blurred. For example, it can be convenient to refer to the full shift on $\{+1,-1\}$ as the full 2-shift.

Bi-infinite sequences are denoted by $\mathbf{x}=(x_n)_{n\in\mathbb{Z}}$, or by

$$\mathbf{x}=\ldots x_{-2}x_{-1}x_0x_1x_2\ldots.\tag{1.8}$$

The symbol x_i is the ith **coordinate** of $\mathbf{x}$. When writing a specific sequence, you need to specify which is the 0th coordinate. This is conveniently done with a "decimal point" to separate the x_i's with $i\geq 0$ from those with $i<0$. For example,

$$\mathbf{x}=\ldots 010.1101\ldots$$

means that $x_{-3}=0, x_{-2}=1, x_{-1}=0, x_0=1, x_1=1, x_2=0, x_3=1$, and so on.

A **block** or **word** over W is a finite sequence of symbols from W. We will write blocks without separating their symbols by commas or other punctuation, so that a typical block over $W=\{a,b\}$ looks like *aababbabbb*. It is convenient to include the sequence of **no** symbols, called the **empty block** (or **empty word**) and denoted by ε.

The **length** of a block u is the number of symbols it contains, and is denoted by $|u|$. Thus if $u=a_1a_2\ldots a_k$ is a nonempty block, then $|u|=k$, while $|\varepsilon|=0$. A **k-block** is simply a block of length k. The set of all k-blocks over W is denoted W^k. A **subblock** or **subword** of $u = a_1a_2\ldots a_k$ is a block of the form $a_ia_{i+1}\ldots a_j$, where $1 \le i \le j \le k$. By convenience, the empty block ε is a subblock of every block. Denote

$$W^+ = \bigcup_{n\ge 1} W^n, \quad W^* = W^+ \cup \{\varepsilon\} = \bigcup_{n\ge 0} W^n. \tag{1.9}$$

If $u = a_1\ldots a_n$, $v = b_1\ldots b_m \in A^\star$, define uv to be $a_1\ldots a_nb_1\ldots b_m$ (an element of W^{m+n}). By convention, $\varepsilon u = u\varepsilon = u$ for all blocks u. This gives a binary operation on $W^\star$ called **concatenation** or **juxtaposition**. If $u, v \in W^+$ then $uv \in W^+$ too. Note that uv is in general not the same as vu, although they have the same length. If $n \ge 1$, then u^n denotes the concatenation of n copies of u, and we put $u^0 = \varepsilon$. The law of exponents $u^mu^n = u^{m+n}$ then holds for all integers $m, n \ge 0$. The point $\ldots uuu.uuu\ldots$ is denoted by u^∞.

If $\mathbf{x} \in W^{\mathbb{Z}}$ and $i \le j$, then we will denote the block of coordinates in $\mathbf{x}$ from position i to position j by

$$\mathbf{x}_{[i,j]} = x_ix_{i+1}\ldots x_{j-1}x_j. \tag{1.10}$$

If $i > j$, define $\mathbf{x}_{[i,j]}$ to be ε. It is also convenient to define

$$\mathbf{x}_{[i,j)} = x_ix_{i+1}\ldots x_{j-1}. \tag{1.11}$$

The **central** $(2k+1)$**-block** of $\mathbf{x}$ is $\mathbf{x}_{[-k,k]} = x_{-k}x_{-k+1}\ldots x_{k-1}x_k$.

If $\mathbf{x} \in W^{\mathbb{Z}}$ and u is a block over W, we will say that u **occurs in** $\mathbf{x}$ (or that $\mathbf{x}$ **contains** u) if there are indices i and j so that $u = \mathbf{x}_{[i,j]}$. Note that the empty block ε occurs in every $\mathbf{x}$, since $\varepsilon = \mathbf{x}_{[1,0]}$.

The index n in a point $\mathbf{x} = (x_n)_{n\in\mathbb{Z}}$ can be thought of as indicating time, so that, for example, the time-0 coordinate of $\mathbf{x}$ is x_0. The passage of time corresponds to shifting the sequence one place to the left, and this gives a map or transformation from $W^{\mathbb{Z}}$ to itself.

Definition 1.2.2. *The* **(left) shift map** T *on* $W^{\mathbb{Z}}$ *is defined by*

$$T : W^{\mathbb{Z}} \to W^{\mathbb{Z}}, \quad (T\mathbf{x})_n = x_{n+1} \text{ for all } n \in \mathbb{Z}. \tag{1.12}$$

In the sequel, we shall give a metric on $W^{\mathbb{Z}}$. The metric should capture the idea that points are close when large central blocks of their coordinates agree.

If $\mathbf{x} = (x_n)_{n\in\mathbb{Z}}, \mathbf{y} = (y_n)_{n\in\mathbb{Z}}$ are two sequences in $W^{\mathbb{Z}}$ such that $\mathbf{x} \ne \mathbf{y}$, then there exists $N \ge 0$ such that $x_N \ne y_N$ or $x_{-N} \ne y_{-N}$, so the set $\{n \ge 0 \mid x_n \ne y_n \text{ or } x_{-n} \ne y_{-n}\}$ is nonempty. Then $N(\mathbf{x}, \mathbf{y}) = \min\{n \ge 0 \mid x_n \ne y_n \text{ or } x_{-n} \ne y_{-n}\}$ is well-defined. Thus,

$$N(\mathbf{x}, \mathbf{y}) = 0 \quad \text{if } x_0 \ne y_0, \text{ and} \tag{1.13}$$

$$N(\mathbf{x}, \mathbf{y}) = 1 + \max\{k \ge 0 \mid \mathbf{x}_{[-k,k]} = \mathbf{y}_{[-k,k]}\} \quad \text{if } x_0 = y_0. \tag{1.14}$$

Let us define $d : W^{\mathbb{Z}} \times W^{\mathbb{Z}} \to [0, +\infty)$ by

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \begin{cases} 2^{-N(\mathbf{x},\mathbf{y})+1} & \text{if } \mathbf{x} \neq \mathbf{y} \\ 0 & \text{if } \mathbf{x} = \mathbf{y} \end{cases} \\ &= \begin{cases} 2 & \text{if } \mathbf{x} \neq \mathbf{y} \text{ and } x_0 \neq y_0 \\ 2^{-k} & \text{if } \mathbf{x} \neq \mathbf{y},\ x_0 = y_0 \text{ and } k \geq 0 \text{ is maximal with } \mathbf{x}_{[-k,k]} = \mathbf{y}_{[-k,k]} \\ 0 & \text{if } \mathbf{x} = \mathbf{y}. \end{cases} \end{aligned} \tag{1.15}$$

In other words, to measure the distance between $\mathbf{x}$ and $\mathbf{y}$, we find the largest k for which the central $(2k+1)$-blocks of $\mathbf{x}$ and $\mathbf{y}$ agree, and use 2^{-k} as the distance (with the conventions that if $\mathbf{x} = \mathbf{y}$ then $k = \infty$ and $2^{-\infty} = 0$, while if $x_0 \neq y_0$, then $k = -1$).

For every $k \geq 0$ and $\mathbf{x} \in W^{\mathbb{Z}}$, let $B_{2^{-k}}(\mathbf{x})$ be the open ball with center $\mathbf{x}$ and radius 2^{-k} and $\overline{B}_{2^{-k}}(\mathbf{x})$ be the closed ball with center $\mathbf{x}$ and radius 2^{-k}.

Proposition 1.2.3. *(i) If $\mathbf{x}, \mathbf{y} \in W^{\mathbb{Z}}$, then for all $k \geq 0$,*

$$d(\mathbf{x}, \mathbf{y}) \leq 2^{-k} \text{ iff } d(\mathbf{x}, \mathbf{y}) < 2^{-k+1} \text{ iff } \mathbf{x}_{[-k,k]} = \mathbf{y}_{[-k,k]}.$$

(ii) d is a metric on $W^{\mathbb{Z}}$.

(iii) For all $\mathbf{x} \in W^{\mathbb{Z}}$, $\overline{B}_2(\mathbf{x}) = W^{\mathbb{Z}}$, and, for all $k \geq 0$,

$$B_{2^{-k+1}}(\mathbf{x}) = \overline{B}_{2^{-k}}(\mathbf{x}) = \{\mathbf{y} \in W^{\mathbb{Z}} \mid \mathbf{y}_{[-k,k]} = \mathbf{x}_{[-k,k]}\}.$$

(iv) Let $(\mathbf{x}^{(n)})$ be a sequence in $W^{\mathbb{Z}}$ and $\mathbf{x} \in W^{\mathbb{Z}}$. Then $\lim_{n\to\infty} \mathbf{x}^{(n)} = \mathbf{x}$ exactly when, for each $k \geq 0$, there is n_k such that

$$\mathbf{x}^{(n)}_{[-k,k]} = \mathbf{x}_{[-k,k]}$$

for all $n \geq n_k$.

Proof. (i) If $\mathbf{x} = \mathbf{y}$ or $\mathbf{x} \neq \mathbf{y}$ and $x_0 \neq y_0$, the conclusion is trivial. We can assume that $\mathbf{x} \neq \mathbf{y}$ and $x_0 = y_0$. Then $d(\mathbf{x}, \mathbf{y}) \leq 2^{-k}$ iff $2^{-N(\mathbf{x},\mathbf{y})+1} \leq 2^{-k}$ iff $-N(\mathbf{x}, \mathbf{y}) + 1 \leq -k$ iff $k \leq N(\mathbf{x}, \mathbf{y}) - 1$ iff $\mathbf{x}_{[-k,k]} = \mathbf{y}_{[-k,k]}$, by (1.14)

(ii) It remains to verify the triangle inequality. Let $\mathbf{x}, \mathbf{y}, \mathbf{z}$ be pairwise distinct points of $W^{\mathbb{Z}}$. If $d(\mathbf{x}, \mathbf{y}) = 2$ or $d(\mathbf{y}, \mathbf{z}) = 2$, this is obvious, so we can assume that $d(\mathbf{x}, \mathbf{y}) = 2^{-k}$ and $d(\mathbf{y}, \mathbf{z}) = 2^{-l}$ with $k, l \geq 0$. By (i), we get that $\mathbf{x}_{[-k,k]} = \mathbf{y}_{[-k,k]}$ and $\mathbf{y}_{[-l,l]} = \mathbf{z}_{[-l,l]}$. If we put $m := \min\{k, l\} \geq 0$, it follows that $\mathbf{x}_{[-m,m]} = \mathbf{z}_{[-m,m]}$, hence

$$d(\mathbf{x}, \mathbf{z}) \leq 2^{-m} \leq 2^{-k} + 2^{-l} = d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}).$$

(iii) By (i).

(iv) We have that

$$\lim_{n\to\infty} \mathbf{x}^{(n)} = \mathbf{x} \quad \text{iff} \quad \text{for all } k \geq 0 \text{ there exists } n_k \text{ such that } d(\mathbf{x}^{(n)}, \mathbf{x}) \leq 2^{-k} \text{ for all } n \geq n_k$$
$$\text{iff} \quad \text{for all } k \geq 0 \text{ there exists } n_k \text{ such that } \mathbf{x}^{(n)}_{[-k,k]} = \mathbf{x}_{[-k,k]} \text{ for all } n \geq n_k.$$

□

Thus, a sequence of points in a full shift converges exactly when, for each $k \geq 0$, the central $(2k+1)$-blocks stabilize starting at some element of the sequence. For example, if

$$\mathbf{x}^{(n)} = (10^n)^\infty = \ldots 10^n 10^n . 10^n 10^n \ldots,$$

then $\lim_{n\to\infty} \mathbf{x}^{(n)} = \ldots 0000.10000 \ldots$.

Proposition 1.2.4. *(i) T is invertible, its inverse being the right shift*

$$T^{-1} : W^{\mathbb{Z}} \to W^{\mathbb{Z}}, \quad (T^{-1}\mathbf{x})_n = x_{n-1} \text{ for all } n \in \mathbb{Z}. \tag{1.16}$$

(ii) For all $\mathbf{x}, \mathbf{y} \in W^{\mathbb{Z}}$,

$$d(T\mathbf{x}, T\mathbf{y}) \leq 2d(\mathbf{x}, \mathbf{y}) \text{ and } d(T^{-1}\mathbf{x}, T^{-1}\mathbf{y}) \leq 2d(\mathbf{x}, \mathbf{y}).$$

Hence, both T and T^{-1} are Lipschitz continuous.

Proof. (i) It is easy to see.

(ii) The cases $d(\mathbf{x}, \mathbf{y}) = 0$ and $d(\mathbf{x}, \mathbf{y}) = 2$ are obvious, so we can assume $d(\mathbf{x}, \mathbf{y}) = 2^{-k}$ with $k \geq 0$, so that $\mathbf{x}_{[-k,k]} = \mathbf{y}_{[-k,k]}$. It follows that

$$(T\mathbf{x})_i = \mathbf{x}_{i+1} = \mathbf{y}_{i+1} = (T\mathbf{y})_i \quad \text{for all } i = -(k+1), -k, -(k-1), \ldots, k-1, \text{and}$$
$$(T^{-1}\mathbf{x})_i = \mathbf{x}_{i-1} = \mathbf{y}_{i-1} = (T^{-1}\mathbf{y})_i \quad \text{for all } i = -(k-1), \ldots, k-1, k, k+1,$$

so that $(T\mathbf{x})_{[-(k-1),k-1]} = (T\mathbf{y})_{[-(k-1),k-1]}$ and $(T^{-1}\mathbf{x})_{[-(k-1),k-1]} = (T^{-1}\mathbf{y})_{[-(k-1),k-1]}$. By Proposition 1.2.3.(i), we get that

$$d(T\mathbf{x}, T\mathbf{y}), d(T^{-1}\mathbf{x}, T^{-1}\mathbf{y}) \leq 2^{-(k-1)} = 2d(\mathbf{x}, \mathbf{y}).$$

□

Theorem 1.2.5. *$(W^{\mathbb{Z}}, T)$ is an invertible TDS.*

Proof. By Proposition 1.2.4, T is a homeomorphism. Furthermore, $W^{\mathbb{Z}}$ is Hasudorff, since it is a metric space. It remains to prove that $W^{\mathbb{Z}}$ is compact. We shall actually show that $W^{\mathbb{Z}}$ is sequentially compact. Given a sequence $(\mathbf{x}^{(n)})_{n\geq 1}$ in $W^{\mathbb{Z}}$, we construct a convergent subsequence using Cantor diagonalization as follows.

First consider the 0th coordinates $\mathbf{x}_0^{(n)}$ for $n \geq 1$. Since there are only finitely many symbols, there is an infinite set $S_0 \subseteq \mathbb{Z}_+$ for which $\mathbf{x}_0^{(n)}$ is the same for all $n \in S_0$.

Next, the central 3-blocks $\mathbf{x}_{[-1,1]}^{(n)}$ for $n \in S_0$ all belong to the finite set of possible 3-blocks, so there is an infinite subset $S_1 \subseteq S_0$ so that $\mathbf{x}_{[-1,1]}^{(n)}$ is the same for all $n \in S_1$. Continuing this way, we find for each $k \geq 1$ an infinite set $S_k \subseteq S_{k-1}$ so that all blocks $\mathbf{x}_{[-k,k]}^{(n)}$ are equal for $n \in S_k$.

Define $\mathbf{x} \in W^{\mathbb{Z}}$ as follows: for any $k \geq 0$, take $n \in S_k$ arbitrary and define $x_k = x_k^{(n)}$, $x_{-k} = x_{-k}^{(n)}$. By our construction, $x_k^{(n)}$, resp. $x_{-k}^{(n)}$, have the same values for all $n \in S_k$, so $\mathbf{x}$ is well-defined. Furthermore, since $(S_k)_{k\geq 0}$ is decreasing, we have that $\mathbf{x}_{[-k,k]} = \mathbf{x}_{[-k,k]}^{(n)}$ for all $n \in S_k$.

Define inductively a strictly increasing sequence of natural numbers $(n_k)_{k\geq 0}$ by: n_0 is any element in S_0, and, for $k \geq 0$, n_{k+1} is the smallest element in S_{k+1} strictly greater than n_k.

Then $(\mathbf{x}^{(n_k)})_{k\geq 0}$ is a subsequence of $\mathbf{x}^{(n)}$ such that $\lim_{k\to\infty} \mathbf{x}^{(n_k)} = \mathbf{x}$, by Proposition 1.2.3.(iv). □

1.2.1 Cylinder sets and product topology

For every $n \in \mathbb{Z}$, let

$$\pi_n : W^{\mathbb{Z}} \to W, \quad \pi_n(\mathbf{x}) = x_n. \tag{1.17}$$

be the nth-projection.

An **elementary cylinder** is a set of the form

$$C_n^w \;=\; \pi_n^{-1}(\{w\}) = \{\mathbf{x} \in W^{\mathbb{Z}} \mid x_n = w\}, \qquad \text{where } n \in \mathbb{Z},\, w \in W.$$

A **cylinder** in $W^{\mathbb{Z}}$ is a set of the form

$$\begin{aligned} C_{n_1,\ldots,n_t}^{w_1,\ldots,w_t} &= \{\mathbf{x} \in W^{\mathbb{Z}} \mid x_{n_i} = w_i \text{ for all } i = 1,\ldots,t\} \\ &= \bigcap_{i=1}^{t} C_{n_i}^{w_i} \end{aligned}$$

where $t \geq 1$, $n_1, \ldots, n_t \in \mathbb{Z}$ are pairwise distinct and $w_1, \ldots, w_t \in W$. A particular case of cylinder is the following: if u is a block over X and $n \in \mathbb{Z}$, define $C_n(u)$ as the set of points in which the block u occurs starting at position n. Thus,

$$C_n(u) \;=\; \{\mathbf{x} \in W^{\mathbb{Z}} \mid \mathbf{x}_{[n,n+|u|-1]} = u\} = C_{n,n+1,\ldots,n+|u|-1}^{u_1,u_2,\ldots,u_{|u|}}.$$

Notation 1.2.6. *We shall use the notations $\mathcal{C}$ for the set of all cylinders and $\mathcal{C}_e$ for the set of elementary cylinders.*

The following lemma collects some obvious properties of cylinders.

Lemma 1.2.7. *(i) For all* $n \in \mathbb{Z}$, $W^{\mathbb{Z}} = \bigcup_{w\in W} C_n^w$.

(ii) For all $m, n \in \mathbb{Z}$, $u, w \in W$,

$$C_n^w \cap C_m^u = \begin{cases} \emptyset & \textit{if } m = n \textit{ and } w \neq u, \\ C_n^w & \textit{if } m = n \textit{ and } w = u, \\ C_{n,m}^{w,u} & \textit{if } m \neq n. \end{cases}$$

$$W^{\mathbb{Z}} \setminus C_n^w = \bigcup_{z\in W,\, z\neq w} C_n^z, \qquad C_n^w \setminus C_m^u = \bigcup_{z\in W,\, z\neq u} C_n^w \cap C_m^z.$$

(iii) For all $k \geq 0$ *and* $\mathbf{x} \in W^{\mathbb{Z}}$,

$$B_{2^{-k}}(\mathbf{x}) = C_{-k-1}(\mathbf{x}_{[-k-1,k+1]}).$$

(iv) For all $n \in \mathbb{Z}$, $w \in W$,

$$T(C_n^w) = C_{n-1}^w \textit{ and } T^{-1}(C_n^w) = C_{n+1}^w.$$

(v) For all $t \geq 1$, $n_1 < n_2 < \ldots < n_t \in \mathbb{Z}$, *and* $w_1, \ldots, w_t \in W$,

$$T(C_{n_1,\ldots,n_t}^{w_1,\ldots,w_t}) = C_{n_1-1,\ldots,n_t-1}^{w_1,\ldots,w_t} \textit{ and } T^{-1}(C_{n_1,\ldots,n_t}^{w_1,\ldots,w_t}) = C_{n_1+1,\ldots,n_t+1}^{w_1,\ldots,w_t}$$

Let us consider the discrete metric on W:

$$d(x, y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases}$$

Since W is finite, we have that (W, d) is a compact metric space. Furthermore, a subbasis for the metric topology is given by

$$\mathcal{S}_W := \{\{w\} \mid w \in W\}. \tag{1.18}$$

Let us consider the product topology on $W^{\mathbb{Z}}$.

Proposition 1.2.8. *(i) The set* $\mathcal{C}_e$ *of elementary cylinders is a subbasis for the product topology on* $W^{\mathbb{Z}}$.

(ii) The set $\mathcal{C}$ *of cylinders is a basis for the product topology on* $W^{\mathbb{Z}}$.

(iii) Cylinders are clopen sets in the product topology.

Proof. (i) By the fact that $\mathcal{S}_W$ is a subbasis on W and apply B.7.2.(ii).

(ii) Any cylinder is a finite intersection of elementary cylinders.

(iii) Since $C_n^w = \pi_n^{-1}(\{w\})$ and $\{w\}$ is closed in W, we have that elementary cylinders are closed. As cylinders are finite intersections of elementary cylinders, they are closed too.

□

Proposition 1.2.9. *The metric d given by (1.15) induces the product topology on $W^{\mathbb{Z}}$.*

Proof. By Lemma 1.2.7.(iii), any ball $B_{2^{-k}}(\mathbf{x})\,(k \geq 0)$ is a cylinder, hence is open in the product topology. Let us prove now that every elementary cylinder $C_n^w\,(n \in \mathbb{Z}, w \in W)$ is open in the metric topology. Let $\mathbf{y} \in C_n^w$ and take $k \geq 0$ such that $k \geq |n| - 1$, so $n \in [-k-1, k+1]$. Then $B_{2^{-k}}(\mathbf{y}) \subseteq C_n^w$, since $\mathbf{z} \in B_{2^{-k}}(\mathbf{y}) = C_{-k-1}(\mathbf{y}_{[-k-1,k+1]})$, implies that $z_n = y_n = w$. □

1.3 Basic constructions

1.3.1 Homomorphisms, factors, extensions

Definition 1.3.1. *Let (X,T) and (Y,S) be two TDSs. A* **homomorphism** *from (X,T) to (Y,S) is a continuous map $\varphi : X \to Y$ such that the following diagram commutes:*

$$\begin{array}{ccc} X & \xrightarrow{\ \varphi\ } & Y \\ {\scriptstyle T}\downarrow & & \downarrow{\scriptstyle S} \\ X & \xrightarrow{\ \varphi\ } & Y \end{array}$$

which means $\varphi \circ T = S \circ \varphi$. We use the notation $\varphi : (X,T) \to (Y,S)$.

A homomorphism $\varphi : (X,T) \to (Y,S)$ is an **isomorphism** *if $\varphi : X \to Y$ is a homeomorphism; in this case the TDSs are called* **isomorphic**.

If $\varphi : (X,T) \to (Y,S)$ is a homomorphism (resp. isomorphism), it is easy to see by induction on n that $\varphi \circ T^n = S^n \circ \varphi$ for all $n \geq 1$ (resp. for all $n \in \mathbb{Z}$).

An **automorphism** of a TDS (X,T) is a self-isomorphism $\varphi : (X,T) \to (X,T)$. Hence, $\varphi : (X,T) \to (X,T)$ is an automorphism of (X,T) if and only if $\varphi : X \to X$ is a homeomorphism that commutes with T.

Definition 1.3.2. *Let (X,T) and (Y,S) be two TDSs. We say that (Y,S) is a* **factor** *of (X,T) or that (X,T) is an* **extension** *of (Y,S) if there exists a surjective homomorphism $\varphi : (X,T) \to (Y,S)$.*

1.3.2 Invariant and strongly invariant sets

In the following, (X,T) is a TDS.

Definition 1.3.3. *A nonempty subset $A \subseteq X$ is called*

(i) **invariant under** T *or* T**-invariant** *if* $T(A) \subseteq A$.

(ii) **strongly invariant under** T *or* **strongly** T**-invariant** *if* $T^{-1}(A) = A$.

Trivial strongly T-invariant subsets of X are $\emptyset$ and X.

Lemma 1.3.4. *Let (X, T) be a TDS.*

(i) Any strongly T-invariant set is also T-invariant.

(ii) The complement of a strongly T-invariant set is strongly T-invariant.

(iii) The closure of a T-invariant set is also T-invariant.

(iv) The union of any family of (strongly) T-invariant sets is (strongly) T-invariant.

(v) The intersection of any family of (strongly) T-invariant sets is (strongly) T-invariant.

(vi) If A is T-invariant, then $T^n(A) \subseteq A$ and $T^n(A)$ is T-invariant for all $n \geq 0$.

(vii) If A is strongly T-invariant, then $T^n(A) \subseteq A$ and $T^{-n}(A) = A$ for all $n \geq 0$; in particular, $T^{-n}(A)$ is strongly T-invariant for all $n \geq 0$.

(viii) For any $x \in X$, the forward orbit $orb_+(x)$ of x is the smallest T-invariant set containing x and $\overline{orb_+}(x)$ is the smallest T-invariant closed set containing x.

Proof. (i) By A.0.7.(v).

(ii) If $T^{-1}(A) = A$, then $T^{-1}(X \setminus A) = X \setminus T^{-1}(A) = X \setminus A$.

(iii) If $T(A) \subseteq A$, then $T(\overline{A}) \subseteq \overline{T(A)} \subseteq \overline{A}$, by B.4.2.

(iv) Let $(A_i)_{i \in I}$ be a family of subsets of X. If $T(A_i) \subseteq A_i$ for all $i \in I$, then

$$T\Big(\bigcup_{i \in I} A_i\Big) = \bigcup_{i \in I} T(A_i) \subseteq \bigcup_{i \in I} A_i.$$

If $T^{-1}(A_i) = A_i$ for all $i \in I$, then

$$T^{-1}\Big(\bigcup_{i \in I} A_i\Big) = \bigcup_{i \in I} T^{-1}(A_i) = \bigcup_{i \in I} A_i.$$

(v) Let $(A_i)_{i \in I}$ be a family of subsets of X. If $T(A_i) \subseteq A_i$ for all $i \in I$, then

$$T\Big(\bigcap_{i \in I} A_i\Big) \subseteq \bigcap_{i \in I} T(A_i) \subseteq \bigcap_{i \in I} A_i.$$

If $T^{-1}(A_i) = A_i$ for all $i \in I$, then

$$T^{-1}\Big(\bigcap_{i \in I} A_i\Big) = \bigcap_{i \in I} T^{-1}(A_i) = \bigcap_{i \in I} A_i.$$

(vi) By A.0.7.(i).

(vii) By (i), A is T-invariant, hence we can apply (vi) to conclude that $T^n(A) \subseteq A$ for all $n \geq 0$. Apply A.0.7.(vi) to obtain that $T^{-n}(A) = A$ for all $n \geq 0$.

(viii) By Lemma 1.0.2, We have that $T(\mathrm{orb}_+(x)) = \mathrm{orb}_{>0}(x) \subseteq \mathrm{orb}_+(x)$, hence $\mathrm{orb}_+(x)$ is T-invariant . If B is a T-invariant set containing x, then $Tx \in T(B) \subseteq B$ and, by induction, $T^n x \in B$ for all $n \geq 1$. Thus, $\mathrm{orb}_+(x) \subseteq B$.

By (iii), $\overline{\mathrm{orb}_+}(x)$ is also T-invariant. Furthermore, if B is a closed T-invariant set containing x, then $\mathrm{orb}_+(x) \subseteq B$ and, since B is closed, $\overline{\mathrm{orb}_+}(x) \subseteq B$.

□

Lemma 1.3.5. *Let (X, T) be an invertible TDS.*

(i) $A \subseteq X$ is strongly T-invariant if and only if $T(A) = A$ if and only if A is strongly T^{-1}-invariant.

(ii) The closure of a strongly T-invariant set is also strongly T-invariant .

(iii) If $A \subseteq X$ is strongly T-invariant , then $T^n(A) = A$ for all $n \in \mathbb{Z}$; in particular, $T^n(A)$ is strongly T-invariant for all $n \in \mathbb{Z}$.

(iv) For any $x \in X$, the orbit orb(x) of x is the smallest strongly T-invariant set containing x and $\overline{orb}(x)$ is the smallest strongly T-invariant closed set containing x.

(v) For any nonempty open set U of X, $\bigcup_{n\in\mathbb{Z}} T^n(U)$ is a nonempty open strongly T-invariant set, and $X \setminus \bigcup_{n\in\mathbb{Z}} T^n(U)$ is a proper closed strongly T-invariant subset of X.

Proof. (i) Using the fact that T is a homeomorphism, we get that $A \subseteq X$ is strongly T-invariant if and only if $T^{-1}(A) = A$ if and only if $T(T^{-1}(A)) = T(A)$ if and only if $A = T(A)$.

(ii) Let A be strongly T-invariant. By (i) and B.4.6, we get that $T(\overline{A}) = \overline{T(A)} = \overline{A}$.

(iii) Apply (i) and A.0.8.(ii).

(iv)

$$T(\mathrm{orb}(x)) = T\left(\bigcup_{n\in\mathbb{Z}} T^n x\right) = \bigcup_{n\in\mathbb{Z}} T^{n+1} x = \mathrm{orb}(x),$$

so $\mathrm{orb}(x)$ is strongly T-invariant. If B is a strongly T-invariant set containing x, then for all $n \in \mathbb{Z}$, $T^n x \in T^n(B) = B$, by(iii). Thus, $\mathrm{orb}(x) \subseteq B$.

By (ii), $\overline{\mathrm{orb}}(x)$ is also strongly T-invariant. Furthermore, if B is a closed strongly T-invariant set containing x, then $\mathrm{orb}(x) \subseteq B$ and, since B is closed, $\overline{\mathrm{orb}}(x) \subseteq B$.

(v) Let $A := \bigcup_{n\in\mathbb{Z}} T^n(U)$. Then A is open, since T^n is an open mapping for all $n \in \mathbb{Z}$, and A is nonempty, since $\emptyset \neq U = T^0(U) \subseteq A$. Furthermore,

$$T(A) = T\left(\bigcup_{n\in\mathbb{Z}} T^n(U)\right) = \bigcup_{n\in\mathbb{Z}} T^{n+1}(U) = A.$$

Finally, $X \setminus A$ is proper, closed and strongly T-invariant, as a complement of an open strongly T-invariant set).

□

1.3.3 Subsystems

Let (X,T) be a TDS, $A \subseteq X$ be a nonempty closed T-invariant set and

$$j_A : A \to X, \quad j_A(x) = x$$

be the inclusion.

Notation 1.3.6. *We shall use the notation T_A for the mapping obtained from T by restricting both the domain and the codomain to A.*

$$T_A : A \to A, \quad T_A x = Tx \quad \textit{for all } x \in A. \tag{1.19}$$

Obviously, T_A is continuous.

Then A is compact Hausdorff and $T_A : A \to A$ is continuous, hence (A, T_A) is a TDS.

Definition 1.3.7. *A* **subsystem** *of the TDS (X,T) is any TDS of the form (A,T_A), where A is a nonempty closed T-invariant set.*

For simplicity, we shall say that A **is a subsystem of** (X,T). Obviously, X is a trivial subsystem of itself. A **proper** subsystem is one different from (X,T).

Lemma 1.3.8. *Let (X,T) be a TDS.*

(i) For any subsystem A of (X,T), $j_A : (A,T_A) \to (X,T)$ is an injective homomorphism.

(ii) Any subsystem of a subsystem of (X,T) is also a subsystem of (X,T).

(iii) For any $x \in X$, $\overline{orb_+}(x)$ is a subsystem of (X,T).

(iv) If (X,T) is invertible, and $A \subseteq X$ is a nonempty closed strongly T-invariant set, then the subsystem (A,T_A) is invertible.

(v) If (X,T) is invertible, then $\overline{orb}(x)$ is an invertible subsystem of (X,T).

Proof. (i),(ii),(iv) are easy to see.
(iii),(v) follow by Lemma 1.3.4.(viii) and Lemma 1.3.5.(iv). □

The next proposition shows that every TDS contains a surjective subsystem.

Proposition 1.3.9. *Let A be a subsystem of a TDS (X,T). Then there exists a nonempty closed set $B \subseteq A$ such that $T(B) = B$.*

Proof. Using the fact that X is compact Hausdorff, A is closed (hence compact) and T^n is continuous, we get that $T^n(A)$ is compact (hence closed) in X for all $n \geq 0$. Furthermore, by A.0.7.(i), $(T^n(A))_{n\geq 0}$ is a decreasing sequence. Applying B.10.5, it follows that

$$B := \bigcap_{n\geq 0} T^n(A)$$

is nonempty. Furthermore, $B \subseteq A$ and B is closed, as intersection of closed sets.

Claim $T(B) = B$.

Proof of Claim " $\subseteq$ " B is T-invariant as the intersection of a family of T-invariant sets, by Lemma 1.3.4.(vi),(v.

" $\supseteq$ " Let $x \in B$ and set $B_{n+1} := T^{-1}(\{x\}) \cap T^n(A)$ for all $n \geq 0$. Since $\{x\}$ is closed in the compact Hausdorff space X and T is continuous, we get that $T^{-1}(\{x\})$ is also closed, hence, B_{n+1} is closed. Furthermore, $(B_{n+1})_{n\geq 0}$ is a decreasing sequence.

Let us prove that B_{n+1} is nonempty for all $n \geq 0$. Since $x \in B$, we get that $x \in T^{n+1}(A)$, so $x = Ty$ for some $y \in T^n(A)$. Thus, $y \in B_{n+1}$.

We can apply again B.10.5 to conclude that

$$\emptyset \neq \bigcap_{n\geq 0} B_{n+1} = T^{-1}(\{x\}) \cap \bigcap_{n\geq 0} T^n(A) = T^{-1}(\{x\}) \cap B.$$

Thus, there exists $y \in B$ such that $Ty = x$, i.e. $x \in T(B)$.

□

Applying the above proposition for $A := X$, we get the following useful results.

Corollary 1.3.10. *If (X,T) is a TDS, then there exists a nonempty closed set $B \subseteq X$ such that $T(B) = B$.*

Corollary 1.3.11. *In an invertible TDS (X,T), any nonempty closed T-invariant subset contains a nonempty closed strongly T-invariant set.*

Proof. Apply Proposition 1.3.9 and Proposition 1.3.5.(i). □

1.3.4 Products

Let $(X_1,T_1),\dots,(X_n,T_n)$ be TDSs, where $n \geq 2$. The **product** TDS is defined by:

$$X := \prod_{i=1}^{n} X_i = X_1 \times \dots \times X_n$$

$$T := \prod_{i=1}^{n} T_i = T_1 \times \dots \times T_n : X \to X, \quad \text{that is } T(x_1,\dots,x_n) = (T_1x_1,\dots,T_nx_n).$$

For any $i = 1, \ldots, n$, let us consider the natural projections

$$\pi_i : \prod_{i=1}^{n} X_i \to X_i, \quad \pi_i(x_1, \ldots, x_n) = x_i.$$

Proposition 1.3.12. *(i) (X, T) is a TDS.*

(ii) (X_i, T_i) is a factor of (X, T) for all $i = 1, \ldots, n$.

(iii) (X, T) is invertible whenever (X_i, T_i) $(i = 1, \ldots, n)$ are invertible TDSs.

Proof. (i) X is compact Hausdorff as a product of compact Hausdorff spaces. Furthermore, T is continuos as a product of continuous functions, by B.7.4.

(ii) It is easy to see that $\pi_i : (X, T) \to (X_i, T_i)$ is a surjective homomorphism: π_i is surjective, continuous, and for all $x = (x_1, \ldots, x_n) \in X$, we have that

$$(\pi_i \circ T)(x) = \pi_i(Tx) = T_i x_i \quad \text{and} \quad (T_i \circ \pi_i)(x) = T_i x_i.$$

(iii) T is a homeomorphism as a product of homeomorphisms, by B.7.4.

□

Example 1.3.13. The TDS $(\mathbb{T}^n, R_a)$ (see Example 1.1.5) is the n-fold product of the TDSs $(\mathbb{S}^1, R_{a_i})$, $i = 1, \ldots, n$ (see Example 1.1.4).

1.3.5 Disjoint unions

Let (X_1, T_1) and (X_2, T_2) be TDSs and consider the disjoint union $X := X_1 \sqcup X_2$ of the topological spaces X_1, X_2.

Let us define

$$T : X \to X, \quad Tx = \begin{cases} T_1 x & \text{if } x \in X_1, \\ T_2 x & \text{if } x \in X_2. \end{cases}$$

Proposition 1.3.14. *(X, T) is a TDS, called the* **disjoint union** *of the TDSs (X_1, T_1) and (X_2, T_2).*

Proof. Apply B.6.2 and B.10.6.(vdisj-union-compact). □

Lemma 1.3.15. *Let (X, T) be a disjoint union of (X_1, T_1) and (X_2, T_2).*

(i) both (X_1, T_1) and (X_2, T_2) are subsystems of (X, T).

(ii) If (X_1, T_1) and (X_2, T_2) are both invertible, then (X, T) is invertible too.

Proof. (i) X_1 is nonempty closed and T-invariant, since $T(X_1) = T_1(X_1) \subseteq X_1$. Furthermore, $T_1 = T_{X_1}$. Similarly for X_2.

(ii) The inverse $T^{-1} : X \to X$ of T is given by

$$T^{-1}x = \begin{cases} T_1^{-1}x & \text{if } x \in X_1, \\ T_2^{-1}x & \text{if } x \in X_2. \end{cases}$$

and is continuous, by B.6.2.(ii).

□

1.4 Transitivity

Definition 1.4.1. *Let (X, T) be a TDS. A point $x \in X$ is called* **forward transitive** *if its forward orbit $orb_+(x)$ is dense in X. If there is at least one forward transitive point, the TDS is called* **(topologically) forward transitive**.

The property of a TDS being forward transitive expresses the fact that if we start at the point x we can reach, at least approximately, any other point in X after some time.

Definition 1.4.2. *Let (X, T) be an invertible TDS. A point $x \in X$ is called* **transitive** *if its orbit $orb(x)$ is dense in X. The TDS is called* **(topologically) transitive** *if there is at least one transitive point.*

The following is obvious.

Lemma 1.4.3. *Let (X, T) be a TDS.*

(i) For every $x \in X$, $(\overline{orb_+}(x), T_{\overline{orb_+}(x)})$ is a forward transitive subsystem of (X, T).

(ii) If (X, T) is invertible, then $(\overline{orb}(x), T_{\overline{orb}(x)})$ is a transitive subsystem of (X, T) for all $x \in X$.

Lemma 1.4.4. *Let (X, T) be a TDS and $x \in X$.*

(i) x is a forward transitive point if and only if $x \in \bigcup_{n\geq 0} T^{-n}(U)$ for every nonempty open subset U of X.

(ii) Assume that (X, T) is invertible. Then x is a transitive point if and only if $x \in \bigcup_{n\in\mathbb{Z}} T^n(U)$ for every nonempty open subset U of X.

Proof. (i) Applying B.1.5.(ii) and Lemma 1.0.2.(ii), we get that x is forward transitive if and only if $\text{orb}_+(x) \cap U \neq \emptyset$ for any nonempty open set U iff $x \in \bigcup_{n\geq 0} T^{-n}(U)$ for any nonempty open set U.

(ii) Similarly, using Lemma 1.0.2.(iii).

□

Lemma 1.4.5. *Let (X, T) be a TDS with X metrizable and $(U_n)_{n\geq 1}$ be a countable basis of X (which exists, by B.10.11).*

(i) $\{x \in X \mid \overline{orb_+}(x) = X\} = \bigcap_{n \geq 1} \bigcup_{k \geq 0} T^{-k}(U_n)$.

(ii) If (X, T) is invertible, then $\{x \in X \mid \overline{orb}(x) = X\} = \bigcap_{n \geq 1} \bigcup_{k \in \mathbb{Z}} T^{k}(U_n)$.

Proof. As the proof of the above lemma, using B.1.5.(iii). □

Theorem 1.4.6. *Let (X, T) be an invertible TDS and assume that X is metrizable. The following are equivalent:*

(i) *(X, T) is transitive.*

(ii) *If U is a nonempty open subset of X such that $T(U) = U$, then U is dense in X.*

(iii) *If $E \neq X$ is a proper closed subset of X such that $T(E) = E$, then E is nowhere dense in X.*

(iv) *for any nonempty open subset U of X, $\bigcup_{n \in \mathbb{Z}} T^n(U)$ is dense in X.*

(v) *for any nonempty open subsets U, V of X, there exists $n \in \mathbb{Z}$ such that $T^n(U) \cap V \neq \emptyset$.*

(vi) *The set of transitive points is residual.*

Proof. (i) ⇒ (ii) Let x be a transitive point, so that $\mathrm{orb}(x)$ is dense. Let U be a nonempty open set satisfying $T(U) = U$. Since $\mathrm{orb}(x) \cap U \neq \emptyset$, we have that $T^k x \in U$ for some $k \in \mathbb{Z}$. It follows that for all $n \in \mathbb{Z}$, $T^n x = T^{n-k}(T^k x) \in T^{n-k}(U) = U$, by A.0.7.(vi). Hence, $\mathrm{orb}(x) \subseteq U$ and, since $\overline{\mathrm{orb}}(x) = X$, we must have $\overline{U} = X$.
(ii) ⇔ (iii) By B.1.5.(iv).
(iv) ⇔ (v) follows immediately from B.1.5.
(ii) ⇒ (iv) Apply Proposition 1.3.5.(v).
(iv) ⇒ (vi) Let $(U_n)_{n \geq 1}$ be a countable basis of X. By Lemma 1.4.5, the set of transitive points is $\bigcap_{n \geq 1} \bigcup_{k \in \mathbb{Z}} T^k(U_n)$, which is an intersection of countably many open dense sets, by (iv). Hence, the set of transitive points is residual, by B.11.3.
(vi) ⇒ (i) Since X is compact Hausdorff, we get that X is a Baire space, by Baire Category Theorem B.11.7. Apply now B.11.6 to conclude that there exist transitive points.
□

1.5 Minimality

Definition 1.5.1. *A TDS (X, T) is called* **minimal** *if there are no non-trivial closed T-invariant sets in X.*

This means that if $A \subseteq X$ is closed and $T(A) \subseteq A$, then $A = \emptyset$ or $A = X$. Equivalently, (X,T) is minimal if and only if it does not have proper subsystems. Hence, "irreducible" appears to be the adequate term. However, the term "minimal" is generally used in topological dynamics.

Proposition 1.5.2. *(i) $(X, 1_X)$ is minimal if and only if $|X| = 1$.*

(ii) If (X,T) is minimal, then T is surjective.

(iii) A factor of a minimal TDS is also minimal.

(iv) If a product TDS is minimal, then so are each of its components.

(v) If (X_1, T_{X_1}), (X_2, T_{X_2}) are two minimal subsystems of a TDS (X,T), then either $X_1 \cap X_2 = \emptyset$ or $X_1 = X_2$.

Proof. Exercise. □

As a consequence of the above proposition, minimality is an isomorphism invariant, i.e. if two TDSs are isomorphic and one of them is minimal, so is the other.

Proposition 1.5.3. *Let (X,T) be a TDS. The following are equivalent:*

(i) (X,T) is minimal.

(ii) Every $x \in X$ is forward transitive.

(iii) $X = \bigcup_{n\geq 0} T^{-n}(U)$ for every nonempty open subset U of X.

(iv) For every nonempty open subset U of X, there are $n_1, \ldots, n_k \geq 0$ such that $X = \bigcup_{i=1}^{k} T^{-n_i}(U)$.

Proof. (i) ⇒ (ii) By Lemma 1.3.8.(iii).
(ii) ⇒ (i) Assume that $A \neq \emptyset$ is a closed T-invariant set and let $x \in A$ be arbitrary. Then $X = \overline{\text{orb}_+(x)} \subseteq A$, by Proposition 1.3.4.(viii). Hence, $X = A$.
(ii) ⇔ (iii) Apply Lemma 1.4.4.(i).
(iv) ⇒ (iii) Obviously.
(iii) ⇒ (iv) By the compactness of X, since $T^{-n}(U)$ is open for all $n \geq 0$. □

Corollary 1.5.4. *Every minimal TDS is forward transitive.*

Theorem 1.5.5. *Any TDS (X,T) has a minimal subsystem.*

Proof. Let $\mathcal{M}$ be the family of all nonempty closed T-invariant subsets of X with the partial ordering by inclusion. Then, of course, $X \in \mathcal{M}$, so $\mathcal{M}$ is non-empty. Let $(A_i)_{i\in I}$ be a chain in $\mathcal{M}$ and take $A := \bigcap_{i\in I} A_i$. Then $A \in \mathcal{M}$, since A is nonempty (by B.10.4), A is closed, and A is T-invariant (by Proposition 1.3.4.(v)). Thus, by Zorn's Lemma A.0.6 there exists a minimal element $F \in \mathcal{M}$. Then (F, T_F) is a minimal subsystem of (X,T). □

1.6 Topological recurrence

We now turn to the question whether a state returns (at least approximately) to itself from time to time.

Let $A \subseteq X$ be arbitrary and consider the successive sites $x, Tx, T^2x, \ldots, T^nx, \ldots$ of an arbitrary point $x \in A$ as time runs through $0, 1, 2, \ldots, n, \ldots$. The set of all points which return (= are back) to A at time $n \geq 1$ is

$$\{x \in A \mid T^n x \in A\} = A \cap T^{-n}(A).$$

Notation 1.6.1. *We shall use the following notations:*

(i) A_{ret} *is the set of those points of* A *which return to* A **at least once**.

(ii) A_{inf} *is the set of those points of* A *which return to* A **infinitely often**.

(iii) For every $x \in A$, $rt(x, A)$ *is the set of return times of* x *in* A.

Thus,

$$A_{ret} = A \cap \bigcup_{n \geq 1} T^{-n}(A), \quad A_{inf} = A \cap \bigcap_{n \geq 1} \bigcup_{m \geq n} T^{-m}(A),$$
$$rt(x, A) = \{n \geq 1 \mid T^n x \in A\} = \{n \geq 1 \mid x \in T^{-n}(A)\}.$$

Furthermore, for every $x \in A$ we have that $x \in A_{ret}$ if and only if $rt(x, A)$ is nonempty, and $x \in A_{inf}$ if and only if $rt(x, A)$ is infinite.

Definition 1.6.2. *Let* (X, T) *be a TDS. A point* $x \in X$ *is called*

(i) **recurrent** *if* $x \in U_{ret}$ *for every open neighborhood* U *of* x.

(ii) **infinitely recurrent** *if* $x \in U_{inf}$ *for every open neighborhood* U *of* x.

Thus, x is recurrent if and only if x returns at least once to U for every open neighborhood U if and only if $x \in \overline{\mathrm{orb}_{>0}(x)}$.

Definition 1.6.3. *A set* $S \subseteq \mathbb{Z}_+$ *is called* **syndetic** *if there exists an integer* $N \geq 1$ *such that* $[k, k+N] \bigcap S \neq \emptyset$ *for any* $k \in \mathbb{Z}_+$.

Thus syndetic sets have "bounded gaps". Any syndetic set is obviously infinite.

Definition 1.6.4. *Let* (X, T) *be a TDS. A point* $x \in X$ *is called* **almost periodic** *or* **uniformly recurrent** *if for every open neighborhood* U *of* x *the set of return times* $rt(x, U)$ *is syndetic.*

Lemma 1.6.5. *(i) Any periodic point is almost periodic.*

(ii) Any almost periodic point is recurrent.

Proof. (i) Let x be a periodic point. Let $N \geq 1$ be the smallest positive integer such that $T^N x = x$. Then for every $k \geq 1$, there exists $n \in [k, k+N]$ such that $T^n x = x$, in particular $n \in rt(x, U)$ for every open neighborhood U of x.

(ii) Obviously.

□

Lemma 1.6.6. *(i) If $\varphi : (X, T) \to (Y, S)$ is a homomorphism of TDSs and $x \in X$ is recurrent (almost periodic) in (X, T), then $\varphi(x)$ is recurrent (almost periodic) in (Y, S).*

(ii) If (A, T_A) is a subsystem of (X, T) and $x \in A$, then x is recurrent (almost periodic) in (X, T) if and only if x is recurrent (almost periodic) in (A, T_A).

Proof. Exercise. □

As a consequence, isomorphisms map recurrent (almost periodic) points in recurrent (almost periodic) points.

Proposition 1.6.7. *Let (X, T) be a TDS and $x \in X$. The following are equivalent:*

(i) x is recurrent.

(ii) x is infinitely recurrent.

Proof. Exercise. □

Lemma 1.6.8. *Let (X, T) be a TDS and assume that X is metrizable. For any $x \in X$, the following are equivalent:*

(i) x is recurrent.

(ii) $\lim\limits_{k\to\infty} T^{n_k} x = x$ for some sequence (n_k) in $\mathbb{Z}_+$.

(iii) $\lim\limits_{k\to\infty} T^{n_k} x = x$ for some sequence (n_k) in $\mathbb{Z}_+$ such that $\lim\limits_{k\to\infty} n_k = \infty$.

Proof. Exercise. □

Proposition 1.6.9. *[G. D. Birkhoff]*
Every point in a minimal TDS (X, T) is almost periodic.

Proof. Assume that (X, T) is minimal and let $x \in X$, and U be a an open neighborhood of x. Applying Proposition 1.5.3.(iv), there are $n_1, \ldots, n_k \geq 0$ such that $X = \bigcup_{i=1}^k T^{-n_i}(U)$. Let $N := \max\{n_1, \ldots, n_k\}$. For each $n \geq 1$, there exists $i = 1, \ldots, k$ such that $T^n x \in T^{-n_i}(U)$, that is $T^{n+n_i} x \in U$. It follows that $n + n_i \in [n, n+N] \cap rt(x, U)$. □

Combining Theorem 1.5.5 with Proposition 1.6.9, we immediately obtain the

Theorem 1.6.10 (Birkhoff Recurrence Theorem). *Every TDS contains at least one point x which is almost periodic (and hence recurrent).*

Corollary 1.6.11. *Let (X,T) be a TDS and assume that X is metrizable. Then there exists $x \in X$ satisfying $\lim_{k\to\infty} T^{n_k}x = x$ for some sequence (n_k) in $\mathbb{Z}_+$ such that $\lim_{k\to\infty} n_k = \infty$.*

Proof. Apply Theorem 1.6.10 and Lemma 1.6.8. □

Proposition 1.6.12. *Let (X,T) be a TDS and $x \in X$. The following are equivalent:*

(i) x is almost periodic.

(ii) For any open neighborhood U of x there exists $N \geq 1$ such that

$$orb_+(x) \subseteq \bigcup_{k=0}^{N} T^{-k}(U).$$

(iii) $(\overline{orb_+}(x), T_{\overline{orb_+(x)}})$ is a minimal subsystem.

Proof. Exercise. □

1.6.1 An application to a result of Hilbert

The following result, due to Hilbert [53], is presumably the first result of Ramsey theory. Hilbert used this lemma to prove his irreducibility theorem: If the polynomial $P(X,Y) \in \mathbb{Z}[X,Y]$ is irreducible, then there exists some $a \in \mathbb{N}$ with $P(a,Y) \in \mathbb{Z}[Y]$.

The **finite sums** of a set D of natural numbers are all those numbers that can be obtained by adding up the elements of some finite nonempty subset of D. The set of all finite sums over D will be denoted by $FS(D)$. Thus,

$$FS(D) = \left\{ \sum_{m\in F} m \mid F \text{ is a finite nonempty subset of } D \right\}. \tag{1.20}$$

If $D = \{n_1, n_2, \dots, n_l\}$, we shall denote $FS(D)$ by $FS(n_1, \dots, n_l)$.

Theorem 1.6.13 (Hilbert (1892). *Let $r \in \mathbb{Z}_+$ and $\mathbb{N} = \bigcup_{i=1}^{r} C_i$. Then for any $l \geq 1$ there exist $n_1 \leq n_2 \leq \ldots \leq n_l \in \mathbb{N}$ such that infinitely many translates of $FS(n_1, \dots, n_l)$ belong to the same C_i. That is,*

$$\bigcup_{a\in B} \left(a + FS(n_1, \dots, n_l)\right) \subseteq C_i$$

for some finite sequence $n_1 \leq n_2 \leq \ldots \leq n_l$ in $\mathbb{N}$ and some infinite set $B \subseteq \mathbb{N}$.

Proof. Let $W = \{1, 2, \ldots, r\}$ and consider the full shift $(W^{\mathbb{Z}}, T)$. Let $\mathbf{x} \in W^Z$ be defined by:

$$x_n = \begin{cases} i & \text{if } n \geq 0 \text{ and } n \in C_i \\ \text{arbitrarily} & \text{if } n < 0. \end{cases}$$

Step 1 Assume that $\mathbf{x}$ is recurrent.

We construct a finite sequence (W_k), $k = 0, 1, \ldots, l$ of blocks of $\mathbf{x}$ inductively as follows:

(i) Let $N := x_0$ and define $W_0 := N$.

(ii) Assume that $W_0, \ldots, W_k$ were defined. Since $\mathbf{x}$ is recurrent, the block W_k occurs in $\mathbf{x}$ a second time (see H2.5). Hence, there exists a (possibly empty) block Y_{k+1} such that $W_k Y_{k+1} W_k$ occurs in $\mathbf{x}$. Define $W_{k+1} := W_k Y_{k+1} W_k$.

For every $k = 1, \ldots, l$, let n_k be the length of $W_k Y_{k+1}$, so that $1 \leq n_1 \leq \ldots \leq n_l$. Let us remark that

$$W_k = \mathbf{x}_{[0, |W_k|-1]}, \quad |W_{k+1}| = |W_k| + n_k,$$

and that if some symbol occurs at position p in W_k, then it occurs also at position $p + n_k$ in W_{k+1}.

Let $1 \leq i_1 < i_2 < \ldots < i_p \leq l$, where $1 \leq p \leq l$. Then N occurs at position 0 in $\mathbf{x}$, at position n_{i_1} in W_{i_1}, at position $n_{i_1} + n_{i_1+1}$ in W_{i_1+1}, at position $n_{i_1} + n_{i_1+2}$ in W_{i_1+2}, and so on, at position $n_{i_1} + n_{i_2}$ in W_{i_2}. Applying the above argument repeatedly, we get that N occurs at position $n_{i_1} + n_{i_2} + \ldots + n_{i_p}$ in W_{i_p}, hence in $\mathbf{x}$. It follows that N occurs in $\mathbf{x}$ at any position in $FS\big(n_1, \ldots, n_l\big)$.

Applying again the fact that $\mathbf{x}$ is recurrent, we get that the block W_l occurs in $\mathbf{x}$ at an infinite number of positions, say $0 = p_1 < p_2 < \ldots < p_k < \ldots$. Take $B = \{p_k \mid k \geq 1\}$ to get that N occurs at any position in $\bigcup_{a \in B} \big(a + FS\big(n_1, \ldots, n_l\big)\big)$. That is,

$$\bigcup_{a \in B} \big(a + FS\big(n_1, \ldots, n_l\big)\big) \subseteq C_N.$$

Step 2 Let us consider the general case, when $\mathbf{x}$ is not necessarily recurrent. Consider the subsystem $(\overline{\text{orb}_+}(x), T_{\overline{\text{orb}_+(x)}})$, and apply Birkhoff recurrence theorem 1.6.10 to get a recurrent point $\mathbf{y}$ of this TDS. We have two cases:

Case 1: $\mathbf{y} = T^m \mathbf{x}$ for some $m \geq 0$. Applying Step 1 for $\mathbf{y}$, we get that $N := y_0 = x_m$ occurs in $\mathbf{y}$ at any position in $\bigcup_{a \in B} \big(a + FS\big(n_1, \ldots, n_l\big)\big)$. Letting $C := m + B$, we get that C is infinite and

$$\bigcup_{a \in C} \big(a + FS\big(n_1, \ldots, n_l\big)\big) \subseteq C_N$$

Case 2: $\mathbf{y} \notin \mathrm{orb}_+(x)$. Then $\lim_{k\to\infty} T^{m_k}\mathbf{x} = \mathbf{y}$ for some strictly increasing sequence (m_k) of natural numbers. Applying Step 1 for the recurrent point $\mathbf{y}$, we get that $N := y_0$ occurs at any position $p \in FS\big(n_1, \ldots, n_l\big)$ for some finite sequence $n_1 \le n_2 \le \ldots \le n_l$ in $\mathbb{N}$.

Take $n := n_1 + n_2 + \ldots + n_l$. It follows that there exists $K \ge 0$ such that $(T^{m_k}\mathbf{x})_{[-n,n]} = \mathbf{y}_{[-n,n]}$ for all $k \ge K$. Let $B = \{m_k \mid k \ge K\}$. Then B is infinite, and

$$x_{m_k+p} = (T^{m_k}\mathbf{x})_p = y_p = N \text{ for all } p \in FS\big(n_1, \ldots, n_l\big), \text{ and all } m_k \in B.$$

Thus

$$\bigcup_{a\in B} \big(a + FS\big(n_1, \ldots, n_l\big)\big) \subseteq C_N.$$

$\square$

1.7 Multiple recurrence

Let X be a compact metric space, $l \geq 1$, and $T_1, \ldots, T_l : X \to X$ be continuous mappings.

Definition 1.7.1. *We say that a point $x \in X$ is* **multiply recurrent** *(for $T_1, \ldots, T_l$) if there exists a sequence (n_k) in $\mathbb{N}$ with $\lim_{k\to\infty} n_k = \infty$ such that*

$$\lim_{k\to\infty} T_1^{n_k} x = \lim_{k\to\infty} T_2^{n_k} x = \ldots = \lim_{k\to\infty} T_l^{n_k} x = x. \tag{1.21}$$

Furthermore, the mappings $T_1, \ldots, T_l : X \to X$ are said to be **commuting** if $T_i \circ T_j = T_j \circ T_i$ for all $i, j = 1, \ldots, l$. This implies $T_i^n \circ T_j^m = T_j^m \circ T_i^n$ for all $m, n \in \mathbb{Z}_+$; if the T_i's are homeomorphisms, then $T_i^n \circ T_j^m = T_j^m \circ T_i^n$ holds for all $m, n \in \mathbb{Z}$.

In this section, we extend Birkhoff's Recurrence Theorem. We shall prove the following result.

Theorem 1.7.2 (Multiple Recurrence Theorem (MRT))**.**
Let $l \geq 1$ and $T_1, \ldots, T_l : X \to X$ be commuting homeomorphisms of a compact metric space (X, d). Then there exists a multiply recurrent point for $T_1, \ldots, T_l$.

Corollary 1.7.3.
Let (X, d) be a compact metric space and $T : X \to X$ be a homeomorphism. For all $l \geq 1$, there exists a multiply recurrent point for $T, T^2, \ldots, T^l$.

Proof. Let $T_i := T^i$ for all $1 \leq i \leq l$. Then $T_1, \ldots, T_l$ are commuting homeomorphisms of the compact metric space (X, d), so we can apply MRT to conclude that there exists a multiply recurrent point $x \in X$. □

Corollary 1.7.4.
Let (X, d) be a compact metric space and $T : X \to X$ be a continuous mapping. For all $l \geq 1$, there exists a multiply recurrent point for $T, T^2, \ldots, T^l$.

Proof. Exercise. □

1.7.1 Some useful lemmas

In the sequel, (X, d) is a compact metric space, $l \geq 1$, and $T_1, \ldots, T_l : X \to X$ are continuous mappings.

Consider the product TDS $(X^l, \tilde{T})$:

$$X^l = \underbrace{X \times X \times \ldots \times X}_{l}, \qquad \tilde{T} := \prod_{i=1}^{l} T_i.$$

Then the metric $d_l(\mathbf{x}, \mathbf{y}) = \max_{i=1,\ldots,l} d(x_i, y_i)$ induces the product topology on X^l, by B.7.5.

For every $\emptyset \neq Y \subseteq X$, let

$$Y^l_\Delta := \{\mathbf{y} = (y, y, \dots, y) \mid y \in Y\}$$

be the diagonal of Y. For every $i = 1, \dots, l$, let

$$\tilde{T}_i : X^l \to X^l, \quad \tilde{T}_i = \underbrace{T_i \times \dots \times T_i}_{l}.$$

Lemma 1.7.5. *(i) $d_l(\mathbf{x}, \mathbf{y}) = d(x, y)$ for all $\mathbf{x}, \mathbf{y} \in X^l_\Delta$.*

(ii) For all $x \in X$, $(B_\varepsilon(x))^l_\Delta = \{\mathbf{y} \in X^l_\Delta \mid d_l(\mathbf{x}, \mathbf{y}) < \varepsilon\} = B_\varepsilon(\mathbf{x}) \cap X^l_\Delta$.

(iii) V is open in X^l_Δ if and only if $V = U^l_\Delta$ for some open subset U of X.

(iv) Let $Y \subseteq X$ be a nonempty closed set. Then

(a) Y^l_Δ is a compact metric space.

(b) For all $i = 1, \dots, l$, $\tilde{T}_i(Y^l_\Delta) = (T_i(Y))^l_\Delta$.

We have the following characterization of multiply recurrent points.

Lemma 1.7.6. *Let $x \in X$ and $\mathbf{x} = (x, \dots, x) \in X^l_\Delta$. The following are equivalent:*

(i) x is multiply recurrent for $T_1, \dots, T_l$.

(ii) $\mathbf{x}$ is a recurrent point in $(X^l, \tilde{T})$.

(iii) For all $\varepsilon > 0$ there exists $N \geq 1$ such that $d_l(\mathbf{x}, \tilde{T}^N \mathbf{x}) < \varepsilon$.

(iv) For all $\varepsilon > 0$ there exists $N \geq 1$ such that $d(x, T_i^N x) < \varepsilon$ for all $i = 1, \dots, l$.

Proof. Exercise. □

Lemma 1.7.7. *Assume that $T_1, \dots, T_l : X \to X$ are commuting homeomorphisms. Then*

(i) X contains a subset X_0 which is minimal with the property that it is nonempty closed and strongly T_i-invariant for all $i = 1, \dots, l$.

(ii) For every nonempty open subset U of X_0, there are $M \geq 1$ and $n_{ij} \in \mathbb{Z}, i = 1, \dots, l,\ j = 1, \dots, M$ such that $X_0 = \bigcup_{j=1}^{M} \big(T_1^{n_{1j}} \circ \dots \circ T_l^{n_{lj}}\big)(U)$.

(iii) $(X_0)^l_\Delta$ is strongly $\tilde{T}_i$-invariant for all $i = 1, \dots, l$.

Proof. Exercise. □

The following lemma is one of the most important steps in proving Theorem 1.7.2. According to Furstenberg, its proof is due to Rufus Bowen.

Lemma 1.7.8. *Let (X,T) be a TDS with (X,d) metric space. Let $A \subseteq X$ be a subset with the property that*

$$\text{for every } \varepsilon > 0 \text{ and for all } x \in A \text{ there exist } y \in A \text{ and } n \geq 1 \text{ with } d(T^n y, x) < \varepsilon. \quad (1.22)$$

Then for every $\varepsilon > 0$ there exist a point $z \in A$ and $N \geq 1$ satisfying $d(T^N z, z) < \varepsilon$.

Proof. Let $\varepsilon > 0$ be given. We define inductively sequences $\varepsilon_1 > \varepsilon_2 > \dots$ of positive parameters, $z_0, z_1, \dots,$ of points in A, and $p_1, p_2, \dots, p_n, \dots$ of positive integers satisfying the following for all $k \geq 1$:

(i) $\varepsilon_k < \frac{\varepsilon}{2^{k+1}}$,

(ii) $d(z_k, T^{p_{k+1}} z_{k+1}) < \varepsilon_{k+1}$, and

(iii) for all $u, v \in X$, $d(u,v) < \varepsilon_{k+1}$ implies

$$d(T^{p_k}u, T^{p_k}v) < \varepsilon_k,\ d(T^{p_{k-1}+p_k}u, T^{p_{k-1}+p_k}v) < \varepsilon_k, \dots, d(T^{p_1+\dots+p_k}u, T^{p_1+\dots+p_k}v) < \varepsilon_k.$$

Let $z_0 \in A$ be arbitrarily. Let $\varepsilon_1 < \varepsilon/4$ and apply (1.22) to get $z_1 \in A$ and $p_1 \geq 1$ such that

$$d(T^{p_1} z_1, z_0) < \varepsilon_1.$$

Since $T^{p_1} : X \to X$ is uniformly continuous, there exists $\delta > 0$ such that for all $u, v \in X$,

$$d(u,v) < \delta \quad \text{implies} \quad d(T^{p_1}u, T^{p_1}v) < \varepsilon_1.$$

Let $\varepsilon_2 < \min\{\delta, \varepsilon_1/2\}$ and apply again (1.22) to get $z_2 \in A$ and $p_2 \geq 1$ such that

$$d(z_1, T^{p_2} z_2) < \varepsilon_2.$$

Since $T^{p_2}, T^{p_1+p_2} : X \to X$ are uniformly continuous, there exists $\delta > 0$ such that for all $u, v \in X$,

$$d(u,v) < \delta \quad \text{implies} \quad d(T^{p_1}u, T^{p_1}v) < \varepsilon_2,\ d(T^{p_1+p_2}u, T^{p_1+p_2}v) < \varepsilon_2.$$

Let $\varepsilon_3 < \min\{\delta, \varepsilon_2/2\}$ and apply again (1.22) to get $z_3 \in A$ and $p_3 \geq 1$ such that

$$d(z_2, T^{p_3} z_3) < \varepsilon_3.$$

Assume $\varepsilon_1, \dots, \varepsilon_k$, $z_0, z_1, \dots, z_k$, and $p_1, \dots, p_k$ were defined. Since T^{p_k},$T^{p_{k-1}+p_k}$, $T^{p_1+\dots+p_k}$: $X \to X$ are uniformly continuous, there exist $\delta_1, \dots, \delta_k > 0$ such that for all $u, v \in X$,

$$d(u,v) < \delta_k \quad \text{implies} \quad d(T^{p_k}u, T^{p_k}v) < \varepsilon_k, \text{ and for all } i = 1, \dots, k-1,$$
$$d(u,v) < \delta_i \quad \text{implies} \quad d(T^{p_i+\dots+p_k}u, T^{p_i+\dots+p_k}v) < \varepsilon_k.$$

Let $\varepsilon_{k+1} < \min\{\delta_1, \ldots, \delta_k, \varepsilon_k/2\}$ and apply again (1.22) to get $z_{k+1} \in A$ and $p_{k+1} \geq 1$ such that

$$d(z_k, T^{p_{k+1}} z_{k+1}) < \varepsilon_{k+1}.$$

By sequential compactness, the sequence (z_n) has a convergent subsequence. In particular, there exist $1 \leq i < j$ such that $d(z_i, z_j) < \varepsilon/2$. It follows that

$$\begin{array}{rcll} d(z_i, T^{p_{i+1}} z_{i+1}) & < & \varepsilon_{i+1}, & \text{by (ii) for } k = i \\ d(T^{p_{i+1}} z_{i+1}, T^{p_{i+1}+p_{i+2}} z_{i+2}) & < & \varepsilon_{i+1}, & \text{by (ii), (iii) for } k = i+1, \\ d(T^{p_{i+1}+p_{i+2}} z_{i+2}, T^{p_{i+1}+p_{i+2}+p_{i+3}} z_{i+3}) & < & \varepsilon_{i+2}, & \text{by (ii), (iii) for } k = i+2, \\ d(T^{p_{i+1}+p_{i+2}+\ldots p_{j-1}} z_{j-1}, T^{p_{i+1}+p_{i+2}+\ldots p_j} z_j) & < & \varepsilon_{j-1}, & \text{by (ii), (iii) for } k = j-1. \end{array}$$

Hence,

$$\begin{array}{rcl} d(z_i, T^{p_{i+1}+p_{i+2}+\ldots+p_j} z_j,) & \leq & \varepsilon_{i+1} + \varepsilon_{i+1} + \ldots + \varepsilon_{j-1} < \frac{\varepsilon}{2^{i+2}} + \frac{\varepsilon}{2^{i+2}} + \frac{\varepsilon}{2^{i+3}} + \ldots \frac{\varepsilon}{2^j} \\ & < & \varepsilon/8 + \varepsilon/8 \sum_{k=0}^{\infty} 1/2^k = \varepsilon/8 + \varepsilon/4 < \varepsilon/2. \end{array}$$

By the triangle inequality we then have

$$d(z_j, T^{p_{i+1}+p_{i+2}+\ldots+p_j} z_j) \leq d(z_j, z_i) + d(z_i, T^{p_{i+1}+p_{i+2}+\ldots p_j} z_j) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

The conclusion of the lemma follows on taking $x := z_j$ and $N := p_{i+1} + p_{i+2} + \ldots p_j$. □

1.7.2 Proof of the Multiple Recurrence Theorem

In the sequel, we give a proof of Theorem 1.7.2.
Let us denote with $MRT(l)$ the statement of the theorem. We prove it by induction on $l \geq 1$.

M$RT(1)$ follows from Birkhoff Recurrence Theorem (see Corollary 1.6.11).

M$RT(l-1) \Rightarrow MRT(l)$ Let $l \geq 2$ and $T_1, \ldots, T_l : X \to X$ be l commuting homeomorphisms of X. By Lemma 1.7.7.(i), we can assume that X does not contain a proper nonempty closed subset Y such that $T_i(Y) = Y$ for all $i = 1, \ldots, l$.

Claim 1: For all $\varepsilon > 0$ there exist $\mathbf{x}, \mathbf{y} \in X^l_\Delta$ and $N \geq 1$ such that $d_l(\mathbf{x}, \tilde{T}^N \mathbf{y}) < \varepsilon$.

Proof: For every $i = 1, \ldots, l-1$, let $S_i := T_i \circ T_l^{-1}$. Then $S_1, \ldots, S_{l-1}$ are commuting homeomorphisms, so we can apply $MRT(l-1)$ to get the existence of $x \in X$ such that, for all $\varepsilon > 0$, there exists $N \geq 1$ satisfying $d(x, S_i^N x) < \varepsilon$ for all $i = 1, \ldots, l-1$. By letting $y := T_l^{-N} x$, and $\mathbf{x}, \mathbf{y} \in X^l_\Delta, \mathbf{x} = (x, x, \ldots, x), \mathbf{y} = (y, y, \ldots, y)$, we get that

$$d_l(\mathbf{x}, \tilde{T}^N \mathbf{y}) = \max\{d(x, S_1^N x), \ldots, d(x, S_{l-1}^N x), d(x, x)\} < \varepsilon. \quad \square$$

Claim 2: For all $\varepsilon > 0$ and for all $\mathbf{x} \in X^l_\Delta$ there exist $\mathbf{y} \in X^l_\Delta$ and $N \geq 1$ such that $d_l(\mathbf{x}, \tilde{T}^N \mathbf{y}) < \varepsilon$.

Proof: Let $U := B_{\varepsilon/2}(x) \subseteq X$. Applying Lemma 1.7.7.(ii), we get the existence of $M \geq 1$ and $n_{ij} \in \mathbb{Z}, i = 1, \ldots, l,\ j = 1, \ldots, M$ such that $X = \bigcup_{j=1}^{M} \left(T_1^{n_{1j}} \circ \ldots \circ T_l^{n_{lj}}\right)(U)$. As an immediate consequence,

$$X^l_\Delta = \left(\bigcup_{j=1}^{M} \left(T_1^{n_{1j}} \circ \ldots \circ T_l^{n_{lj}}\right)(U)\right)^l_\Delta = \bigcup_{j=1}^{M} \left(\tilde{T}_1^{\,n_{1j}} \circ \ldots \circ \tilde{T}_l^{\,n_{lj}}\right)(U^l_\Delta). \quad (1.23)$$

Let us denote, for all $j = 1, \ldots, M$,

$$S_j := \left(\tilde{T}_1^{\,n_{1j}} \circ \ldots \circ \tilde{T}_l^{\,n_{lj}}\right)^{-1} = \tilde{T}_1^{\,-n_{1j}} \circ \ldots \circ \tilde{T}_l^{\,-n_{lj}}, \quad \text{since } \tilde{T}_i\text{'s commute.} \quad (1.24)$$

X^l_Δ is compact and strongly S_j-invariant, by Lemma 1.7.7.(iii), so $S_j : X^l_\Delta \to X^l_\Delta$ is uniformly continuous. We get then for all $j = 1, \ldots, M$ the existence of $\delta_j > 0$ such that for all $\mathbf{z}, \mathbf{u} \in X^l_\Delta$,

$$d_l(\mathbf{z}, \mathbf{u}) < \delta_j \quad \text{implies} \quad d_l(S_j \mathbf{z}, S_j \mathbf{u}) < \varepsilon/2. \quad (1.25)$$

Take $\delta := \min\{\delta_1, \ldots, \delta_j\} > 0$ and apply Claim 1 to get $\mathbf{z}_0, \mathbf{u}_0 \in X^l_\Delta$ and $N \geq 1$ such that

$$d_l(\mathbf{u}_0, \tilde{T}^N \mathbf{z}_0) < \delta. \quad (1.26)$$

Since $\mathbf{u}_0 \in X^l_\Delta$, by (1.23) there exists $j_0 = 1, \ldots, M$ such that $S_{j_0} \mathbf{u}_0 \in U^l_\Delta$, hence

$$d_l(\mathbf{x}, S_{j_0} \mathbf{u}_0) < \varepsilon/2. \quad (1.27)$$

Let $\mathbf{y} := S_{j_0} \mathbf{z}_0$. Applying (1.25), (1.26), and the fact that $\tilde{T}^N$ and S_{j_0} commute, we get that

$$d_l(\tilde{T}^N \mathbf{y}, S_{j_0} \mathbf{u}_0) = d_l(S_{j_0}(\tilde{T}^N \mathbf{z}_0), S_{j_0} \mathbf{u}_0) < \varepsilon/2. \quad (1.28)$$

Finally, it follows that

$$\begin{aligned} d_l(\tilde{T}^N \mathbf{y}, \mathbf{x}) &\leq d_l(\tilde{T}^N \mathbf{y}, S_{j_0} \mathbf{u}_0) + d_l(S_{j_0} \mathbf{u}_0, \mathbf{x}) \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \quad \square \end{aligned}$$

Claim 3: For all $\varepsilon > 0$ there exist $\mathbf{x} \in X^l_\Delta$ and $N \geq 1$ such that $d_l(\mathbf{x}, \tilde{T}^N \mathbf{x}) < \varepsilon$.

Proof: follows from Claim 2, after applying Lemma 1.7.8 with $A = X^l_\Delta$. □

Claim 4: For all $\varepsilon > 0$ the set

$$Y_\varepsilon = \{\mathbf{x} \in X^l_\Delta \mid \text{ there exists } N \geq 1 \text{ such that } d_l(\mathbf{x}, \tilde{T}^N \mathbf{x}) < \varepsilon\} \quad (1.29)$$

is dense in X^l_Δ.

Proof: Let $\varepsilon > 0$. We shall prove that $Y_\varepsilon \cap U_\Delta^l \neq \emptyset$ for any open subset U of X. As in the proof of Claim 2, we get

$$M \geq 1,\ n_{ij} \in \mathbb{Z}, i = 1, \ldots, l,\ j = 1, \ldots, M,\ \ S_j = \tilde{T}_1^{-n_{1j}} \circ \ldots \circ \tilde{T}_l^{-n_{lj}}$$

satisfying

(i) $X_\Delta^l = \bigcup_{j=1}^{M} S_j^{-1}(U_\Delta^l)$, and

(ii) there exists $\delta > 0$ such that for all $j = 1, \ldots, M$, and for all $\mathbf{z}, \mathbf{u} \in X_\Delta^l$,

$$d_l(\mathbf{z}, \mathbf{u}) < \delta \quad \text{implies} \quad d_l(S_j\mathbf{z}, S_j\mathbf{u}) < \varepsilon.$$

By Claim 3, Y_δ is nonempty. Let $\mathbf{x} \in Y_\delta$ and $N \geq 1$ be such that $d_l(\mathbf{x}, \tilde{T}^N\mathbf{x}) < \delta$. Since $\mathbf{x} \in X_\Delta^l$, there exists $j_0 = 1, \ldots, M$ such that $\mathbf{y} := S_{j_0}\mathbf{x} \in U_\Delta^l$. Since $\tilde{T}^N$ and S_{j_0} commute, it follows that

$$d_l(\mathbf{y}, \tilde{T}^N\mathbf{y}) = d_l(S_{j_0}\mathbf{x}, S_{j_0}(\tilde{T}^N\mathbf{x})) < \varepsilon,$$

hence $\mathbf{y} \in U_\Delta^l \cap Y_\varepsilon$.

□

Claim 5: $MRT(l)$ is true, that is there exists $\mathbf{x} \in X_\Delta^l$ such that, for all $\varepsilon > 0$, there exists $N \geq 1$ such that

$$d_l(\tilde{T}^N\mathbf{x}, \mathbf{x}) < \varepsilon.$$

Proof: For every $n \geq 1$, by Claim 5, By Claim 5, $Y_{1/n}$ is dense in X_Δ^l. Furthermore, $Y_{1/n} = U_\Delta^l$, where

$$U = \bigcup_{N \geq 1} \bigcap_{i=1}^{l} \{x \in X \mid d(x, T_i^N x) < 1/n\}.$$

It is easy to see that U is open in X, hence $Y_{1/n}$ is is open in X_Δ^l. Thus, $Y := \bigcap_{n \geq 1} Y_{1/n}$ is a residual set and we can apply B.11.6 to conclude that Y is nonempty. Then any $\mathbf{x} \in Y$ satisfies the claim. □

Chapter 2

Ramsey Theory

Ramsey theory is that branch of combinatorics which deals with structure which is preserved under partitions. The theme of Ramsey theory:

"Complete disorder is impossible." (T.S. Motzkin)

Thus, inside any large structure, no matter how chaotic, will lie a smaller substructure with great regularity. One looks typically at the following kind of question: *If a particular structure (e.g. algebraic, combinatorial or geometric) is arbitrarily partitioned into finitely many classes, what kind of substructure must always remain intact in at least one class?*

Ramsey theorems are natural, and they can be very powerful, as they assume very little information; they are usually very easy to state, but can have very complicated combinatorial proofs.

Ramsey theory owes its name to a very general theorem of Ramsey from 1930 [89], popularized by Erdös in the 30's.

A number of results in Ramsey theory have the following general form:

(*) *Let X be a set. For any $r \in \mathbb{Z}_+$, and any r-partition $X = \bigcup_{i=1}^{r} C_i$ of X, at least one of the classes possesses some property P.*

X could be $\mathbb{N}, \mathbb{Z}, \mathbb{N}^d, \mathbb{Z}^d \, (d \geq 1), \ldots$. The statement can be expressed also in terms of finite colourings of X. For any $r \geq 1$, an r-**colouring** of X is a mapping $c : X \to \{1, 2, \ldots, r\}$. Then (*) becomes:

For any finite colouring of a set X, there exists a monochromatic subset of X having some property P.

An **affine image** of a set $F \subseteq \mathbb{N}$ (resp. $F \subseteq \mathbb{Z}$) is a set of the form

$$a + bF = \{a + bf \mid f \in F\} \quad \text{where } a \in \mathbb{N}, b \in \mathbb{Z}_+ (\text{ resp. } a \in \mathbb{Z}, b \in \mathbb{Z} \setminus \{0\}). \quad (2.1)$$

2.1 van der Waerden theorem

One of the most fundamental results of Ramsey theory is the celebrated van der Waerden theorem:

Theorem 2.1.1 (van der Waerden)**.**
Let $r \geq 1$ *and* $\mathbb{N} = \bigcup_{i=1}^{r} C_i$. *For any* $k \geq 1$*, there exists* $i \in [1, r]$ *such that* C_i *contains an arithmetic progression of length* k.

This result was conjectured by Baudet and proved by van der Waerden in 1927 [111]. The theorem gained a wider audience when it was included in Khintchine's famous book *Three pearls in number theory* [60].

Let us denote with **(vdW1)** the above formulation of van der Waerden theorem and consider the following statements:

(vdW2) Let $r \geq 1$ and $\mathbb{N} = \bigcup_{i=1}^{r} C_i$. There exists $i \in [1, r]$ such that C_i contains arithmetic progression of arbitrary finite length.

(vdW3) Let $r \geq 1$ and $\mathbb{N} = \bigcup_{i=1}^{r} C_i$. For any finite set $F \subseteq \mathbb{N}$ there exists $i \in [1, r]$ such that C_i contains affine images of F.

(vdW4) Let $r \geq 1$ and $\mathbb{N} = \bigcup_{i=1}^{r} C_i$. There exists $i \in [1, r]$ such that C_i contains affine images of every finite set $F \subseteq \mathbb{N}$.

Let (**vdWi**$*$), $i = 1, 2, 3, 4$ be the statements obtained from (**vdWi**), $i = 1, 2, 3, 4$ by changing $\mathbb{N}$ to $\mathbb{Z}$ in their formulations.

Proposition 2.1.2. (**vdWi**), (**vdWi**$*$), $i = 1, 2, 3, 4$ *are all equivalent.*

Proof. Exercise. □

(vdW2) states that for any finite partition of $\mathbb{N}$, one of the cells contains arithmetic progressions of arbitrary finite length. Equivalently, any finite colouring of $\mathbb{N}$ contains monochromatic arithmetic progressions of arbitrary finite length.

We remark that one cannot, in general, expect to get from any finite colouring of $\mathbb{N}$ a monochromatic infinite arithmetic progression (why?).

2.1.1 Topological dynamics proof of van der Waerden Theorem

The topological dynamics proof we give here is due to Furstenberg and Weiss [39].

Proposition 2.1.3.
Let $l \geq 1$ *and* $\varepsilon > 0$. *For any compact metric space* (X, d) *and homeomorphism* $T : X \to X$ *there exist* $x \in X$ *and* $N \geq 1$ *such that*

$$d(x, T^{iN}x) < \varepsilon \text{ for all } 1 \leq i \leq l. \tag{2.2}$$

Proof. Apply Corollary 1.7.3 and Lemma 1.7.6.(iv) □

Let us denote with **(vdW-dynamic)** the statement of the above proposition.

Theorem 2.1.4. (vdW-dynamic) *implies* **(vdW1*)**.

Proof. Let $r, k \geq 1$ and let $\mathbb{Z} = \bigcup_{i=1}^{r} C_i$. Set $W = \{1, 2, \ldots, r\}$ and consider the full shift $(W^{\mathbb{Z}}, T)$. Let $\gamma \in W^Z$ be defined by:

$$\gamma_n = i \quad \text{if and only if } n \in C_i.$$

Let $X := \overline{\{T^n\gamma \mid n \in \mathbb{Z}\}}$ be the orbit closure of γ and consider the subsystem (X, T_X).

Applying **(vdW-dynamic)** with $\varepsilon := 2$ and $l := k-1$, we get $\mathbf{x} \in X$ and $N \geq 1$ such that

$$d(\mathbf{x}, T^{jN}\mathbf{x}) < 2 \text{ for all } 1 \leq j \leq k-1.$$

Thus, by Proposition 1.2.3.(i),

$$x_0 = \left(T^N\mathbf{x}\right)_0 = \ldots = \left(T^{(k-1)N}\mathbf{x}\right)_0, \quad \text{i.e. } x_0 = x_N = \ldots = x_{(k-1)N}.$$

Since $\mathbf{x} \in X$, by letting $p = (k-1)N - 1$, we get the existence of $M \in \mathbb{Z}$ such that

$$d(\mathbf{x}, T^M\gamma) < 2^{-p}, \quad \text{hence, } \mathbf{x}_{[-(k-1)N,(k-1)N]} = (T^M\gamma)_{[-(k-1)N,(k-1)N]}.$$

Let $i := x_0$. It follows that $i = x_0 = x_N = \ldots x_{(k-1)N}$, hence

$$i = (T^M\gamma)_0 = (T^M\gamma)_N = \ldots = (T^M\gamma)_{(k-1)N}, \quad \text{i.e. } i = \gamma_M = \gamma_{M+N} = \ldots = \gamma_{M+(k-1)N}.$$

By the definition of γ, it follows that the k-term arithmetic progression

$$\{M, M+N, M+2N\ldots, M+(k-1)N\} \tag{2.3}$$

is contained in C_i.

□

Theorem 2.1.5. (vdW1) *implies* **(vdW-dynamic)**.

Proof. Let $l \geq 1$, $\varepsilon > 0$, (X, d) be a compact metric space, and $T : X \to X$ be a homeomorphism. Since X is compact, it is totally bounded (see B.10.15). Thus, there exists a finite cover of X by $\varepsilon/2$-balls. From this we can construct a finite cover of X by pairwise disjoint sets $U_1, \ldots, U_r$ of less than ε diameter (see A.1.3).

Let $y \in X$ and define for all $i = 1, \ldots, r$,

$$C_i := \{n \in \mathbb{N} \mid T^n y \in U_i\}.$$

Then $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, and the C_i's are pairwise disjoint, so by taking the nonempty ones of them we get a finite partition of $\mathbb{N}$.

Applying **(vdW1)**, one of the cells C_i contains an arithmetic progression $\{a, a+N, \dots, a+lN\}$ of length $l+1$, where $a \in \mathbb{N}$, and $N \geq 1$, since $l \geq 1$. This means that

$$T^a y \in U_i,\ T^{a+N} y \in U_i, \dots, T^{a+lN} y \in U_i.$$

By letting $x := T^a y$, it follows that $\{x, T^N x, \dots, T^{lN} x\} \subseteq U_i$. Since U_i is of diameter less than ε, the conclusion follows. □

2.1.2 The compactness principle

The compactness principle, in very general terms, is a way of going from the infinite to the finite. It gives us a ”finite” (or finitary) Ramsey-type statement providing the corresponding ”infinite” Ramsey-type statement is true.

Theorem 2.1.6 (The Compactness Principle)**.**
Let $r \geq 1$ and let $\mathcal{F}$ be a family of finite subsets of $\mathbb{Z}_+$. Assume that for every r-colouring of $\mathbb{Z}_+$ there is a monochromatic member of $\mathcal{F}$. Then there exists a least positive integer $N = N(\mathcal{F}, r)$ such that, for every r-colouring of $[1, N]$, there is a monochromatic member of $\mathcal{F}$.

Proof. The proof we give is essentially what is known as Cantor's diagonal argument. Let $r \geq 1$ be fixed and assume that every r-colouring of $\mathbb{Z}_+$ admits a monochromatic member of $\mathcal{F}$. Assume by contradiction that for each $n \geq 1$ there exists an r-colouring

$$\chi_n : [1, n] \to [1, r]$$

with no monochromatic member of $\mathcal{F}$. We proceed by constructing a specific r-colouring χ of $\mathbb{Z}_+$. Since there are only finitely many colours, among $\chi_1(1), \chi_2(1), \dots$, there must be some colour that appears an infinite number of times. Call this colour c_1, and let $\mathcal{C}_1$ be the infinite set of all colourings χ_j with $\chi_j(1) = c_1$. Within the set of colours $\{\chi_j(2) \mid \chi_j \in \mathcal{C}_1\}$ there must be some colour c_2 that occurs an infinite number of times. Let $\mathcal{C}_2 \subseteq \mathcal{C}_1$ be the infinite set of all colourings $\chi_j \in \mathcal{C}_1$ with $\chi_j(2) = c_2$. Continuing in this way, we find for each $k \geq 2$ a colour c_k such that the family of colourings

$$\mathcal{C}_k = \{\chi_j \in \mathcal{C}_{k-1} \mid \chi_j(k) = c_k\}$$

is infinite. We define the r-colouring

$$\chi : \mathbb{Z}_+ \to [1, r], \quad \chi(k) = c_k.$$

Then χ has the property that for every $k \geq 1$, $\mathcal{C}_k$ is the collection of colourings χ_j with $\chi(i) = \chi_j(i)$ for all $i = 1, \dots, k$.

By assumption, χ admits a monochromatic member of $\mathcal{F}$, say S. Let $M := \max S$ and take some arbitrary colouring $\chi_j \in \mathcal{C}_M$. Then $\chi_j|_S = \chi|_S$, hence $S \in \mathcal{F}$ is monochromatic under χ_j. This contradicts our assumption that all of the χ_n's avoid monochromatic members of $\mathcal{F}$. □

Remark 2.1.7. *The compactness principle does not give us any bound for $N(\mathcal{F}, r)$; it only gives us its existence.*

Corollary 2.1.8. *Let $r \geq 1$ and let $\mathcal{F}$ be a family of finite subsets of $\mathbb{Z}_+$. The following are equivalent:*

(i) For every r-colouring of $\mathbb{Z}_+$ there is a monochromatic member of $\mathcal{F}$.

(ii) There exists a least positive integer $N = N(\mathcal{F}, r)$ such that, for every r-colouring of $[1, N]$, there is a monochromatic member of $\mathcal{F}$.

(iii) There exists a least positive integer $N = N(\mathcal{F}, r)$ such that, for all $m \geq N$ and for every r-colouring of $[1, m]$, there is a monochromatic member of $\mathcal{F}$.

Proof. $(i) \Rightarrow (ii)$ By the Compactness Principle.
$(ii) \Rightarrow (iii)$ If $m \geq N(\mathcal{F}, r)$, and χ is an r-colouring of $[1, m]$, then we can apply (ii) for its restriction to $[1, N(\mathcal{F}, r)]$ to get a monochromatic member of $\mathcal{F}$.
$(iii) \Rightarrow (i)$ is obvious. □

2.1.3 Finitary version of van der Waerden theorem

As a consequence of the Compactness Principle, we get the following

Theorem 2.1.9 (Finitary van der Waerden theorem)**.**
Let $r, k \geq 1$. There exists a least positive integer $W = W(k, r)$ such that for any $n \geq W$ and for any partition $[1, n] = \bigcup_{i=1}^{r} C_i$ of $[1, n]$, some C_i contains an arithmetic progression of length k.

In terms of colourings, there exists a least positive integer $W = W(k, r)$ such that for all $n \geq W$, and for any r-colouring of $[1, n]$ there is a monochromatic arithmetic progression of length k. In fact, by Corollary 2.1.8, van der Waerden theorem and its finitary version are equivalent.

Definition 2.1.10. *The numbers $W(r, k)$ are called the* **van der Waerden numbers***.*

We have that $W(1, k) = k$ for any $k \geq 1$, since one colour produces only trivial colourings. $W(r, 2) = r + 1$, since we may construct a colouring that avoids arithmetic progressions of length 2 by using each color at most once, but once we use a color twice, a length 2 arithmetic progression is formed.

The combinatorial proof of van der Waerden theorem proceeds by a double induction on r and k and yields extremely large upper bounds for $W(k, r)$. Shelah [99] proved that van der Waerden numbers are primitive recursive. In 2001, Gowers [41] showed that van der Waerden numbers with $r \geq 2$ are bounded by

$$W(r, k) \leq 2^{2^{r^{2^{2^{k+9}}}}}. \tag{2.4}$$

There are only a few known nontrivial van der Waerden numbers. We refer to

http://www.st.ewi.tudelft.nl/sat/waerden.php

for known values and lower bounds for van der Waerden numbers.

2.1.4 Multidimensional van der Waerden Theorem

An **affine image** of a set $F \subseteq \mathbb{N}^d$ (resp. $F \subseteq \mathbb{Z}^d$) is a set of the form

$$a + bF = \{a + bf \mid f \in F\} \quad \text{where } a \in \mathbb{N}^d,\ b \in \mathbb{Z}_+ \ (\text{ resp. } a \in \mathbb{Z}^d,\ b \in \mathbb{Z} \setminus \{0\}). \tag{2.5}$$

Here is the formulation of the multidimensional analogue of van der Waerden's theorem. It was first proved by Grünwald (also referred to in the literature by the name of Gallai), who apparently never published his proof (Grünwald's authorship is acknowledged in [87, p.123]).

Theorem 2.1.11 (Multidimensional van der Waerden)**.**
Let $d \geq 1, r \geq 1$, and $\mathbb{N}^d = \bigcup_{i=1}^{r} C_i$ be an r-partition of $\mathbb{N}^d$. There exists $i \in [1, r]$ such that C_i contains affine images of every finite set $F \subseteq \mathbb{N}^d$.

Proof. Exercise. □

2.1.5 Polynomial van der Waerden's theorem

The following generalization of van der Waerden theorem is due to Bergelson and Leibman [13], who proved it using topological dynamics methods. A combinatorial proof was obtained in 2000 by Walters [112].

Theorem 2.1.12 (Polynomial van der Waerden theorem)**.** *[13]*
Let $k \geq 1$, and $p_1, \ldots, p_k : \mathbb{Z} \to \mathbb{Z}$ be polynomials of one variable with integer coefficients, which vanish at the origin (i.e. $p_i(0) = 0$ for all $i = 1, \ldots, k$). For any finite colouring of $\mathbb{Z}$, there exists a monochromatic configuration of the form

$$\{a + p_1(d), \ldots, a + p_k(d)\}, \quad a, d \in \mathbb{Z},\ d \neq 0.$$

The case with a single polynomial was proved by Furstenberg [34] and Sarkozy [96] independently.

Remark that by specializing to the linear case $p_i(n) := in$, $i = 1, \ldots, k$ one recovers the ordinary van der Waerden theorem.

2.2 The ultrafilter approach to Ramsey theory

We present now a different approach to Ramsey theory, based on *ultrafilters* via the *Stone-Čech compactification*. We refer to [55] or to the surveys [11, 7, 8] for details.

Definition 2.2.1. *Let D be any set. A* **filter** *on D is a nonempty set $\mathcal{F}$ of subsets of D with the following properties:*

(i) $\emptyset \notin \mathcal{F}$.

(ii) If $A, B \in \mathcal{F}$, *then* $A \cap B \in \mathcal{F}$.

(iii) If $A \in \mathcal{F}$ *and* $A \subseteq B \subseteq D$, *then* $B \in \mathcal{F}$.

We remark that $D \in \mathcal{F}$ for any filter $\mathcal{F}$ on D. A classic example of a filter is the set of neighborhoods of a point in a topological space. If D is an infinite set, an example of a filter on D is the family of **cofinite** subsets of D, defined to be those subsets of D whose complement is finite.

Definition 2.2.2. *An* **ultrafilter** *on D is a filter on D which is not properly contained in any other filter on D.*

Proposition 2.2.3. *Let* $\mathcal{U} \subseteq \mathcal{P}(D)$. *The following are equivalent.*

(i) $\mathcal{U}$ *is an ultrafilter on* D.

(ii) $\mathcal{U}$ *has the finite intersection property and for each* $A \in \mathcal{P}(D) \setminus \mathcal{U}$ *there is some* $B \in \mathcal{U}$ *such that* $A \cap B = \emptyset$.

(iii) $\mathcal{U}$ *is maximal with respect to the finite intersection property. (That is,* $\mathcal{U}$ *is a maximal member of* $\{\mathcal{V} \subseteq \mathcal{P}(D) \mid \mathcal{V}$ *has the finite intersection property*$\}$.*)*

(iv) $\mathcal{U}$ *is a filter on D and for any collection* $C_1, \dots, C_n$ *of subsets of D, if* $\bigcup_{i=1}^{n} C_i \in \mathcal{U}$, *then* $C_j \in \mathcal{U}$ *for some* $j = 1, \dots n$.

(v) $\mathcal{U}$ *is a filter on D and for all* $A \subseteq D$ *either* $A \in \mathcal{U}$ *or* $D \setminus A \in \mathcal{U}$.

Proof. Exercise. See [55, Theorem 3.6, p.49]. □

If $a \in D$, then $e(a) := \{A \in \mathcal{P}(D) \mid a \in A\}$ is easily seen to be an ultrafilter on D, called the **principal ultrafilter** defined by a. It is immediate the fact that $e(a) = e(b)$ if and only if $a = b$, so e is an embedding of D into the set of ultrafilters of D.

Proposition 2.2.4. *Let $\mathcal{U}$ be an ultrafilter on D. The following are equivalent:*

(i) $\mathcal{U}$ *is a principal ultrafilter.*

(ii) There is some $F \in \mathcal{P}_f(D)$ *such that* $F \in \mathcal{U}$.

(iii) *The set* $\{A \subseteq D \mid D \setminus A$ *is finite* $\}$ *is not contained in* $\mathcal{U}$.

(iv) $\bigcap_{A\in\mathcal{U}} A \neq \emptyset$.

(v) *There is some* $x \in D$ *such that* $\bigcap_{A\in\mathcal{U}} A = \{x\}$.

Proof. Exercise. See [55, Theorem 3.7, p.50]. □

Proposition 2.2.5. *Let* D *be set and let* $\mathcal{A}$ *be a subset of* $\mathcal{P}(D)$ *which has the finite intersection property. Then there is an ultrafilter* $\mathcal{U}$ *on* D *such that* $\mathcal{A} \subseteq \mathcal{U}$.

Proof. Exercise. □

Corollary 2.2.6. *Let* D *be set, let* $\mathcal{F}$ *be a filter on* D, *and let* $A \subseteq D$. *Then* $A \notin \mathcal{F}$ *if and only if there is some ultrafilter* $\mathcal{U}$ *with* $\mathcal{F} \cup \{D \setminus A\} \subseteq \mathcal{U}$.

Proof. Exercise. □

To see that non-principal ultrafilters exist, take, for example,

$$\mathcal{A} = \{A \subseteq \mathbb{Z}_+ \mid \mathbb{Z}_+ \setminus A \text{ is finite}\}.$$

Clearly $\mathcal{A}$ has the finite intersection property, so there is an ultrafilter $\mathcal{U}$ on $\mathbb{Z}_+$ such that $\mathcal{A} \subseteq \mathcal{U}$. It is easy to see that such $\mathcal{U}$ cannot be principal.

The following result shows that questions in Ramsey theory are questions about ultrafilters.

Proposition 2.2.7. *Let* D *be a set and let* $\mathcal{G} \subseteq \mathcal{P}(D)$. *The following are equivalent.*

(i) *Whenever* $r \geq 1$ *and* $D = \bigcup_{i=1}^{r} C_i$, *there exists* $i \in [1, r]$ *and* $G \in \mathcal{G}$ *such that* $G \subseteq C_i$.

(ii) *There is an ultrafilter* $\mathcal{U}$ *on* D *such that for every member* A *of* $\mathcal{U}$, *there exists* $G \in \mathcal{G}$ *with* $G \subseteq A$.

Proof. Exercise. □

Those more familiar with measures may find it helpful to view an ultrafilter on D as a $\{0, 1\}$-valued finitely additive measure on $\mathcal{P}(D)$. Given an ultrafilter p on D, define a mapping $\mu_p : \mathcal{P}(D) \to \{0, 1\}$ by $\mu_p(A) = 1 \Leftrightarrow A \in p$. It is easy to see that $\mu_p(\emptyset) = 0$, $\mu_p(D) = 1$, and the fact that for any finite collection of pairwise disjoint sets $C_1, \ldots, C_n$, one has $\mu_p\left(\bigcup_{i=1}^{n} C_i\right) = \sum_{i=1}^{n} \mu_p(C_i)$. The members of the ultrafilters are the "big" sets.

2.2.1 The Stone-Čech compactification

Let D be a discrete topological space. We shall denote with p, q ultrafilters on $\mathcal{D}$ and we shall use the following notations

$$\beta D = \{p \mid p \text{ ultrafilter on } D\}, \tag{2.6}$$

$$\widehat{A} = \{p \in \beta D \mid A \in p\} \quad \text{for any } A \subseteq D, \tag{2.7}$$

$$\mathcal{B} = \{\widehat{A} \mid A \subseteq D\}. \tag{2.8}$$

Lemma 2.2.8. *Let $A, B \subseteq D$.*

(i) $\widehat{A \cap B} = \widehat{A} \cap \widehat{B}$ *and* $\widehat{A \cup B} = \widehat{A} \cup \widehat{B}$.

(ii) $\widehat{D \setminus A} = \beta D \setminus \widehat{A}$.

(iii) $\widehat{A} = \emptyset$ *if and only if* $A = \emptyset$.

(iv) $\widehat{A} = \beta D$ *if and only if* $A = D$.

(v) $\widehat{A} = \widehat{B}$ *if and only if* $A = B$.

Proof. Exercise. See [55, Lemma 3.17, p.53]. □

It follows that the family $\mathcal{B}$ forms a basis for a topology on βD. We define the topology of βD to be the topology which has these sets as a basis.

We consider any $a \in D$ as an element of βD by identifying it with the principal filter $e(a)$ defined by a.

Theorem 2.2.9. *βD is the Stone-Čech compactification of D.*

Proof. See [55, Theorem 3.27, p.56]. □

Being a nice compact Hausdorff space, βD is, for infinite discrete spaces D, quite a strange object.

Proposition 2.2.10. *Let D be an infinite discrete topological space.*

(i) $|\beta D| = 2^{2^{|D|}}$. *In particular,* $|\beta \mathbb{Z}_+| = 2^c$, *where c is the cardinality of the continuum,* $c = |\mathbb{R}| = 2^{\aleph_0}$.

(ii) βD is not metrizable.

(iii) Any infinite closed subset of βD contains a homeomorphic copy of all $\beta \mathbb{Z}_+$.

Proof. (i) See [55, Section 3.6, p.66].

(ii) Otherwise, being a compact and hence separable metric space, it would have cardinality not exceeding c.

(iii) See [55, Theorem 3.59, p.66].

□

2.2.2 Topological semigroups

In the sequel, $(S,+)$ is a semigroup. For every $A, B \subseteq S$, $A + B = \{a + b \mid a \in A, b \in B\}$.

An element $x \in S$ is an **idempotent** if and only if $x + x = x$. We shall denote with $E(S)$ the set of all idempotents of S.

Definition 2.2.11. *Let $\emptyset \neq L, R, I \subseteq S$.*

(i) L is a **left ideal** *of S if and only if $S + L \subseteq L$.*

(ii) R is a **right ideal** *of S if and only if $R + S \subseteq R$.*

(iii) I is an **ideal** *of S if and only if I is both a left and a right ideal of S.*

Of special importance is the notion of **minimal** left and right ideals. By this we mean simply left or right ideals which are minimal with respect to set inclusion.

Let $(S,+)$ be a semigroup with S a topological space and define for each $x \in S$, the functions

$$\rho_x,\ \lambda_x : S \to S, \quad \rho_x(y) = y + x,\ \ \lambda_x(y) = x + y. \tag{2.9}$$

Definition 2.2.12. *(i) $(S,+)$ is a* **right topological semigroup** *if ρ_x is continuous for all $x \in S$.*

(ii) $(S,+)$ is a **left topological semigroup** *if λ_x is continuous for all $x \in S$.*

(iii) $(S,+)$ is a **semitopological semigroup** *if it is both a left and a right topological semigroup.*

(iv) $(S,+)$ is a **topological semigroup** *if $+ : S \times S \to S$ is continuous.*

We shall be concerned with compact Hausdorff right topological semigroups. Of fundamental importance is the following result.

Theorem 2.2.13. *Any compact Hausdorff right topological semigroup has an idempotent.*

Proof. See [55, Theorem 2.5, p.33]. □

Proposition 2.2.14. *Let $(S,+)$ be a compact Hausdorff right topological semigroup. Then every left ideal of S contains a minimal left ideal. Minimal left ideals are closed, and each minimal left ideal has an idempotent.*

Proof. See [55, Corollary 2.5, p.34]. □

Definition 2.2.15. *A* **minimal idempotent** *of $(S,+)$ is an idempotent which belongs to a minimal left ideal.*

Hence, any compact Hausdorff right topological semigroup has minimal idempotents.

2.2.3 The Stone-Čech compactification of $\mathbb{Z}_+$

Let us consider the discrete semigroup $(\mathbb{Z}_+, +)$ and its Stone-Čech compactification $\beta\mathbb{Z}_+$. It is natural to attempt to extend the addition $+$ from $\mathbb{Z}_+$ to $\beta\mathbb{Z}_+$. We recall that we consider $\mathbb{Z}_+ \subseteq \beta\mathbb{Z}_+$, by identifying $n \in \mathbb{Z}_+$ with the principal ultrafilter $e(n)$.

We define the following operation on $\beta\mathbb{Z}_+$: for all $p, q \in \beta\mathbb{Z}_+$,

$$p + q = \{A \subseteq \mathbb{Z}_+ \mid \{n \in \mathbb{Z}_+ \mid A - n \in q\} \in p\}. \tag{2.10}$$

Proposition 2.2.16. *(i) $+$ extends to $\beta\mathbb{Z}_+$ the addition $+$ on $\mathbb{Z}_+$.*

(ii) $(\beta\mathbb{Z}_+, +)$ is a right topological semigroup.

(iii) $(\beta\mathbb{Z}_+, +)$ is not commutative. In fact, for all non-principal ultrafilters $p, q \in \beta\mathbb{Z}_+$, we have that $p + q \neq q + p$.

Proof. (i), (ii) See [11, p. 43-44], or, for arbitrary discrete semigroups, [55, Chapter 4]. (iii) See [55, Theorem 6.9, p.109].

□

Proposition 2.2.17. *(i) Any idempotent ultrafilter is non-principal.*

(ii) There are minimal idempotents in $\beta\mathbb{Z}_+$.

Proof. (i) This follows from the fact that $(\mathbb{Z}_+, +)$ has no idempotents.

(ii) Apply the fact that $(\beta\mathbb{Z}_+, +)$ is a compact Hausdorff right topological semigroup.

□

Proposition 2.2.18. *Let p be an idempotent ultrafilter and define for all $A \subseteq \mathbb{Z}_+$,*

$$A^\star(p) := \{n \in A \mid A - n \in p\}. \tag{2.11}$$

Then

(i) For every $A \in p$, $A^\star(p) \in p$.

(ii) For each $n \in A^\star(p)$, $A^\star(p) - n \in p$.

Proof. (i) We have that $p + p = \{A \subseteq \mathbb{Z}_+ \mid \{n \in \mathbb{Z}_+ \mid (A - n) \in p\} \in p\}$. Hence, $A \in p = p + p$ implies $\{n \in \mathbb{Z}_+ \mid (A - n) \in p\} \in p$. In particular, $A^\star(p) = A \cap \{n \in \mathbb{Z}_+ \mid A - n \in p\} \in p$.

(ii) Let $n \in A^\star(p)$, and let $B := A - n$. Then $B \in p$ and, by (i), $B^\star(p) \in p$. We prove that $B^\star(p) \subseteq A^\star(p) - n$ and then apply (ii) from the definition of a filter to conclude that $A^\star(p) - n \in p$. Assume that $m \in B^\star(p)$. It follows that $m \in B$, hence $m + n \in A$. Furthermore, $B - m \in p$, that is $A - (n + m) \in p$. We get that $m + n \in A^\star(p)$, i.e. $m \in A^\star(p) - n$.

□

Property (i) from the above proposition is a shift-invariance property of idempotent ultrafilters.

2.2.4 Finite Sums Theorem

In this section, we shall give an ultrafilter proof of Hindman's classical Finite Sums theorem [54], which contains as very special cases two early classical results in Ramsey theory: Hilbert theorem 1.6.13 and Schur theorem. Hindman's original proof, elementary though difficult, was greatly simplified by Baumgartner [3]. A topological dynamics proof was given by Furstenberg and Weiss [39].

Given an infinite sequence $(x_n)_{n\geq 1}$ in $\mathbb{Z}_+$, the **IP-set** generated by (x_n) is the set $FS\big((x_n)_{n\geq 1}\big)$ of finite sums of elements of (x_n) with distinct indices:

$$FS\big((x_n)_{n\geq 1}\big) = \left\{\sum_{m\in F} x_m \mid F \text{ is a finite nonempty subset of } \mathbb{Z}_+\right\}. \tag{2.12}$$

The term "IP-set", coined by Furstenberg and Weiss [39], stands for *infinite-dimensional parallelepiped*, as IP-sets can be viewed as a natural generalization of the notion of a parallelepiped of dimension d.

Furthermore, for any finite sequence $(x_k)_{k=1}^n$, let

$$FS\big((x_k)_{k=1}^n\big) = \{\sum_{m\in F} x_m \mid F \text{ is a finite nonempty subset of } \{1,\dots,n\}\}. \tag{2.13}$$

Then $FS\big((x_n)_{n\geq 1}\big) = \bigcup_{n\geq 1} FS\big((x_k)_{k=1}^n\big)$.

Theorem 2.2.19. *Let $p \in \beta\mathbb{Z}_+$ be a minimal idempotent and let $A \in p$. There exists a sequence $(x_n)_{n\geq 1}$ in $\mathbb{Z}_+$ such that $FS\big((x_n)_{n\geq 1}\big) \subseteq A$.*

Proof. Let p be a minimal idempotent and $A \in p$. By Proposition 2.2.18.(i), we have that $A^\star(p) \in p$. We define $(x_n)_{n\geq 1}$ in $\mathbb{Z}_+$ such that $FS\big((x_k)_{k=1}^n\big) \subseteq A^\star(p)$ for all $n \geq 1$. Since $A^\star(p) \subseteq A$, the conclusion follows.
$n = 1$: Take $x_1 \in A^\star(p)$ arbitrary. Remark that $A^\star(p)$ is nonempty, since p is a filter, hence $\emptyset \notin A$.
$n \Rightarrow n+1$: Let $n \geq 1$ and assume that we have chosen $(x_k)_{k=1}^n$ satisfying $FS\big((x_k)_{k=1}^n\big) \subseteq A^\star(p)$. Let

$$E = FS\big((x_k)_{k=1}^n\big). \tag{2.14}$$

Then E is a finite subset of $\mathbb{Z}_+$ and for each $a \in E$ we have, by Proposition 2.2.18.(ii), that $A^\star(p) - a \in p$. Hence, $B := A^\star(p) \cap \bigcap_{a\in E} \big(A^\star(p) - a\big) \in p$, so we can pick $x_{n+1} \in B$. Then $x_{n+1} \in A^\star(p)$ and given $a \in E$, $x_{n+1} + a \in A^\star(p)$. Thus, $FS\big((x_k)_{k=1}^{n+1}\big) \subseteq A^\star(p)$. □

As an immediate corollary we obtain the Finite Sums theorem.

Corollary 2.2.20 (Finite Sums theorem)**.**
Let $r \geq 1$ and $\mathbb{Z}_+ = \bigcup_{i=1}^r C_i$. There exist $i \in [1, r]$ and a sequence $(x_n)_{n\geq 1}$ in $\mathbb{Z}_+$ such that such that $FS\big((x_n)_{n\geq 1}\big) \subseteq C_i$.

Proof. By Proposition 2.2.17.(ii), there exists a minimal idempotent $p \in \beta\mathbb{Z}_+$. Since $\mathbb{Z}_+ \in \beta\mathbb{Z}_+$, we can apply Proposition 2.2.3.(iv) to get $i \in [1, r]$ such that $C_i \in p$. The conclusion follows from Theorem 2.2.19. □

As an immediate corollary, we obtain Schur theorem, one of the earliest results in Ramsey theory.

Corollary 2.2.21 (Schur theorem). *[98]*
Let $r \geq 1$ and let $\mathbb{Z}_+ = \bigcup_{i=1}^{r} C_i$. There exist $i \in [1, r]$ and $x, y \in \mathbb{N}$ such that $\{x, y, x+y\} \subseteq C_i$.

Hilbert theorem 1.6.13, proved in Section 1.6.1 using topological dynamics, is also an immediate consequence of Finite Sums theorem.

Corollary 2.2.22 (see Hilbert theorem 1.6.13).
Let $r \in \mathbb{Z}_+$ and $\mathbb{N} = \bigcup_{i=1}^{r} C_i$. Then for any $l \geq 1$ there exist $n_1 \leq n_2 \leq \ldots \leq n_l \in \mathbb{N}$ such that infinitely many translates of $FS\big(n_1, \ldots, n_l\big)$ belong to the same C_i. That is,

$$\bigcup_{a \in B} \big(a + FS\big(n_1, \ldots, n_l\big)\big) \subseteq C_i$$

for some finite sequence $n_1 \leq n_2 \leq \ldots \leq n_l$ in $\mathbb{N}$ and some infinite set $B \subseteq \mathbb{N}$.

Proof. Exercise. □

2.2.5 Ultrafilter proof of van der Waerden

Theorem 2.2.23. *Let $p \in \beta\mathbb{Z}_+$ be a minimal idempotent and let $A \in p$. Then A contains arbitrarily long arithmetic progressions.*

Proof. See [11, Theorem 3.11, p. 50] or [,]. □

As an immediate corollary, we get van der Waerden theorem.

Corollary 2.2.24. *Let $r \geq 1$ and $\mathbb{Z}_+ = \bigcup_{i=1}^{r} C_i$. There exists $i \in [1, r]$ such that C_i contains arithmetic progression of arbitrary finite length.*

2.2.6 Ultralimits

Definition 2.2.25. *Let $p \in \beta\mathbb{Z}_+$, X be a Hausdorff topological space, $x \in X$, and $(x_n)_{n\geq 1}$ be a sequence in X. Then x is said to be a p-**limit** of (x_n) if*

$$\{n \in \mathbb{Z}_+ \mid x_n \in U\} \in p$$

for every open neighborhood U of x.

We write $p\text{−}\lim x_n = x$.

Proposition 2.2.26. *Let X be a Hausdorff topological space and $(x_n)_{n\geq 1}$ be a sequence in X.*

(i) For every $p \in \beta\mathbb{Z}_+$, the following are satisfied:

(a) The p-limit of (x_n), if exists, is unique.

(b) If X is compact, then $p\text{−}\lim x_n$ exists.

(c) If $f : X \to Y$ is continuous and $p\text{−}\lim x_n = x$, then $p\text{−}\lim f(x_n) = f(x)$.

(ii) $\lim_{n\to\infty} x_n = x$ implies $p\text{−}\lim x_n = x$ for every non-principal ultrafilter p.

Proof. Exercise. □

Proposition 2.2.27. *Let $(x_n)_{n\geq 1}, (y_n)_{n\geq 1}$ be bounded sequences in $\mathbb{R}$, and p be a non-principal ultrafilter on $\mathbb{Z}_+$.*

(i) (x_n) has a unique p-limit. If $a \leq x_n \leq b$, then $a \leq p\text{−}\lim x_n \leq b$.

(ii) For any $c \in \mathbb{R}$, $p\text{−}\lim cx_n = c \cdot p\text{−}\lim x_n$.

(iii) $p\text{−}\lim(x_n + y_n) = p\text{−}\lim x_n + p\text{−}\lim y_n$.

(iv) If $\lim_{n\to\infty}(x_n - y_n) = 0$, then $p\text{−}\lim x_n = p\text{−}\lim y_n$.

Proof. Exercise. □

2.3 Hales-Jewett Theorem

The Hales-Jewett theorem strips van der Waerden's theorem of its unessential elements and reveals the heart of Ramsey theory. It provides a focal point from which many results can be derived and acts as a cornerstone for much of the more advanced work.

To state the Hales-Jewett theorem, we need a little more terminology.

Let W be a finite set with $k := |W| \geq 2$, $n \geq 1$ and W^n be the *n-cube over* W. We refer to the elements of W^n as *strings* or *points*. We will typically use letters $\lambda, \mu, \ldots$ for strings.

Notation 2.3.1. *Let λ be a string. We introduce the notation $\lambda^{b\to a}$ for the string formed by changing all occurences of symbol b to symbol a in λ.*

Definition 2.3.2. *(i) A* line template over W *with wildcard symbol $\star \notin W$ is a string $\lambda \in (W \cup \{\star\})^n$.*

(ii) We say that the line template λ is degenerate *if λ contains no '$\star$' symbols.*

(iii) A combinatorial line *is a set*

$$L_c := \{\lambda^{\star \to a} \mid a \in W\} \subseteq W^n, \tag{2.15}$$

where λ is a nondegenerate line template over W^n. We also say that L_c is the combinatorial line associated with λ and λ is called a combinatorial line template

Example 2.3.3. Let $W := \{1, 2\}$ and $\lambda = 1 \star\star\star 221\star \in (W \cup \{\star\})^8$. The combinatorial line associated to λ is $\{1112211, 2222212\}$.

Remark 2.3.4. *(i) For any combinatorial line L_c, we have that $|L_c| = |W|$.*

(ii) If $W = [1, k]$, then a combinatorial line template $\lambda \in ([1, k] \cup \{\star\})^n$ is again a string over an alphabet of size $k+1$. If we use the symbol $'k+1'$ in place of $'\star'$, we see that a point in $[1, k+1]^n$ can be interpreted as a line in $[1, k]^n$, namely the combinatorial line

$$L_c = \{\lambda^{(k+1)\to i} \mid i \in [1, k]\}.$$

Theorem 2.3.5 (Hales-Jewett)**.**
Let $r, k \geq 1$. There exists $HJ = HJ(k, r) \geq 1$ such that for any $n \geq HJ$ and for any partition $[1, k]^n = \bigcup_{i=1}^{r} C_i$ of $[1, k]^n$, some C_i contains a combinatorial line.

Proof. See [46, Theorem 3, p.36] or [51]. For the Shelah proof see [99] or [46, Section 2.6, p. 54-60]. □

The original proof given by Hales-Jewett [51] proceeded by double induction (on the number r of colours and the size k of the monochromatic set). Shelah's celebrated proof of the theorem [99] uses simple induction (on k). It gives primitive recursive bounds for the Hales-Jewett theorem (and thus also for van der Waerden theorem).

Remark 2.3.6. *The following are obvious reformulations of the Hales-Jewett Theorem:*

(i) For all $r, k \geq 1$ there exists $HJ = HJ(k, r) \geq 1$ such that for any $n \geq HJ$, the following holds: if $[1, k]^n$ is r-coloured there exists a monochromatic combinatorial line.

(ii) For all $r, k \geq 1$ there exists $HJ = HJ(k, r) \geq 1$ such that for any $n \geq HJ$, the following holds: for any finite set W with $|W| = k$, if W^n is r-coloured there exists a monochromatic combinatorial line.

Proposition 2.3.7. *Hales-Jewett Theorem implies van der Waerden Theorem.*

Proof. We shall prove that Hales-Jewett theorem implies the finitary version of van der Waerden theorem. Let $k, r \geq 1$, and apply the Hales-Jewett theorem with $W = [0, k-1]$ instead of $W = [1, k]$. Then there exists $HJ = HJ(k, r) \geq 1$ such that for any $n \geq HJ$, if $[0, k-1]^n$ is r-coloured there exists a monochromatic combinatorial line.

The idea now is to identify the integers with their base k-representation. Thus, for all $n \geq 1$, we have a bijection

$$\varphi : [0, k-1]^n \to [0, k^n - 1], \quad \varphi(a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i k^{i-1}. \tag{2.16}$$

Let us take $N := k^{HJ}$. We shall prove that Theorem 2.1.9 holds with this N. Let $c : [1, N] \to [1, r]$ be an r-colouring of $[1, N]$. Then $c' : [0, N-1] \to [1, r]$, $c'(j) = c(j+1)$ is an r-colouring of $[0, N-1]$, and $C := c' \circ \varphi : [0, k-1]^{HJ} \to [1, r]$ is an r-colouring of the cube $[0, k-1]^{HJ}$.

By the Hales-Jewett theorem, there exists a monochromatic combinatorial line. Thus, there exists nondegenerate $\lambda \in ([0, k-1] \cup \{\star\})^{HJ}$ such that the associated combinatorial line

$$L_c = \{\lambda^{\star\to 0}, \lambda^{\star\to 1}, \ldots, \lambda^{\star\to k-1}\} \tag{2.17}$$

is monochromatic. Let $C_0 := C(\lambda^{\star\to i})$ for all $i \in [0, k-1]$, and denote $S := \{l \in [1, HJ] \mid \lambda_l = \star\}$.

For every $i \in [0, k-1]$, let $x_i := \varphi(\lambda^{\star\to i}) + 1 \in [1, N]$. Then

$$c(x_i) = c\big(\varphi(\lambda^{\star\to i}) + 1\big) = c'(\varphi(\lambda^{\star\to i})) = C(\lambda^{\star\to i}) = C_0, \text{ and} \tag{2.18}$$

$$x_i = \sum_{l\notin S} \lambda_l k^{l-1} + \sum_{l\in S} i k^{l-1} + 1 = a + id, \tag{2.19}$$

$$\text{where } a = \sum_{l\notin S} \lambda_l k^{l-1} + 1, \ d = \sum_{l\in S} k^{l-1}. \tag{2.20}$$

Thus, $\{x_0, \ldots, x_{k-1}\} \subseteq [1, N]$ is a monochromatic arithmetic progression of length k. $\square$

Part II

Ergodic Theory and Density Ramsey Theory

Chapter 3

Measure-preserving systems

In the following we shall consider only probability spaces.

Definition 3.0.8. *Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be probability spaces, and $T : X \to Y$ be a mapping.*

*(i) T is **measurable** if $T^{-1}(\mathcal{C}) \subseteq \mathcal{B}$.*

*(ii) T is **measure-preserving** if T is measurable and*

$$\mu(T^{-1}(A)) = \nu(A) \quad \text{for all } A \in \mathcal{C}.$$

*(iii) T is an **invertible measure-preserving** transformation if T is bijective and both T and T^{-1} are measure-preserving.*

We should write $T : (X, \mathcal{B}, \mu) \to (Y, \mathcal{C}, \nu)$ since the measure-preserving property depends on $\mathcal{B}, \mathcal{C}$ and μ, ν. Measure-preserving transformations are the structure preserving maps (morphisms) between probability spaces.

We shall be mainly interested in the case $(X, \mathcal{B}, \mu) = (Y, \mathcal{C}, \nu)$ since we wish to study the iterates T^n. When $T : X \to X$ is a measure-preserving transformation of $(X, \mathcal{B}, \mu)$ we also say that T **preserves** μ or that μ **is** T**-invariant**.

Definition 3.0.9. *(i) A **measure-preserving dynamical system** (MDS for short) is a quadruple $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B}, \mu)$ is a probability space and $T : X \to X$ is a measure-preserving transformation.*

*(ii) An **invertible measure-preserving dynamical system** is a quadruple $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B}, \mu)$ is a probability space and $T : X \to X$ is an invertible measure-preserving transformation.*

If $(X, \mathcal{B}, \mu, T)$ is a MDS, we also say that μ is T**-invariant**.

Lemma 3.0.10. *(i) $1_X : X \to X$, the identity on $(X, \mathcal{B}, \mu)$, is an invertible measure-preserving transformation.*

(ii) The composition of two measure-preserving transformations is a measure-preserving transformation.

Proof. Exercise. □

Lemma 3.0.11. *Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be probability spaces and $T : X \to Y$ be bijective such that both T and T^{-1} are measurable. The following are equivalent*

(i) T is measure-preserving.

(ii) $\mu(B) = \nu(T(B))$ for all $B \in \mathcal{B}$.

(iii) T^{-1} is measure-preserving.

Proof. (i)⇒(ii) Assume that T is measure-preserving. Then for all $B \in \mathcal{B}$, $\mu(B) = \mu(T^{-1}(T(B))) = \nu(T(B))$.
(ii)⇔(iii) Obviously, since for all $B \in \mathcal{B}$, $(T^{-1})^{-1}(B) = T(B)$.
(ii)⇒(i) Let $C \in \mathcal{C}$. Then $T^{-1}(C) \in \mathcal{B}$ since T is measurable. Hence, $\nu(C) = \nu(T(T^{-1}(C))) = \mu(T^{-1}(C))$. □

In practice it would be difficult to check, using Definition 3.0.8, whether a given transformation is measure-preserving or not, since one usually does not have explicit knowledge of all the members of $\mathcal{B}$. The following result is very useful.

Proposition 3.0.12. *Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be probability spaces and $T : X \to Y$ be a mapping. The following are equivalent*

(i) T is a measure-preserving transformation.

(ii) $T^{-1}(A) \in \mathcal{B}$ and $\mu(T^{-1}(A)) = \nu(A)$ for each $A \in \mathcal{S}$, where $\mathcal{S}$ is a semialgebra that generates $\mathcal{C}$.

Proof. $(ii) \Rightarrow (i)$ Let

$$\mathcal{F} = \{A \in \mathcal{C} \mid T^{-1}(A) \in \mathcal{B} \text{ and } \mu(T^{-1}(A)) = \nu(A)\}.$$

We want to show that $\mathcal{F} = \mathcal{C}$.

Claim 1: $\mathcal{F}$ is a monotone class.

Proof: If $(A_n)_{n\geq 1}$ is an increasing sequence in $\mathcal{F}$, then $\lim_{n\to\infty} A_n = \bigcup_{n\geq 1} A_n$. Furthermore, $(T^{-1}(A_n))$ is also increasing, hence $\lim_{n\to\infty} T^{-1}(A_n) = \bigcup_{n\geq 1} T^{-1}(A_n) = T^{-1}\left(\bigcup_{n\geq 1} A_n\right) = T^{-1}(\lim_{n\to\infty} A_n)$. We get that

(i) $T^{-1}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} T^{-1}(A_n) \in \mathcal{B}$, by C.2.2.(iii), and

(ii) $\nu(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \nu(A_n) = \lim_{n\to\infty} \mu(T^{-1}(A_n)) = \mu(\lim_{n\to\infty} T^{-1}(A_n)) = \mu(T^{-1}(\lim_{n\to\infty} A_n))$, by C.5.5.(ii).

Hence, $\lim_{n\to\infty} A_n \in \mathcal{F}$. The case when $(A_n)_{n\geq 1}$ is a decreasing sequence is similar. □

Claim 2: $\mathcal{A}(\mathcal{S}) \subseteq \mathcal{F}$.

Proof: By (ii), we have that $\mathcal{S} \subseteq \mathcal{F}$. If $A \in \mathcal{A}(\mathcal{S})$, by C.1.7, $A = \bigcup_{i=1}^m A_i$ for some pairwise disjoint sets $A_1, \ldots, A_m \in \mathcal{S}$. It follows that

(i) $T^{-1}(A) = \bigcup_{i=1}^m T^{-1}(A_m) \in \mathcal{B}$, by (ii), and, from the finite additivity of measures,

(ii) $\nu(A) = \sum_{i=1}^m \nu(A_i) = \sum_{i=1}^m \mu(T^{-1}(A_i)) = \mu\left(\bigcup_{i=1}^m T^{-1}(A_i)\right) = \mu(T^{-1}(A))$. □

Apply now Halmos' Monotone Class theorem C.2.6 to conclude that $\mathcal{C} = \sigma(\mathcal{S}) = \sigma(\mathcal{A}(\mathcal{S})) \subseteq \mathcal{F}$. Hence, $\mathcal{F} = \mathcal{C}$. □

3.1 The induced operator

For any measurable space $(X, \mathcal{B})$, we shall use the notations

(i) $\mathcal{M}_{\mathbb{C}}(X, \mathcal{B})$ is the set of all complex-valued measurable functions $f : X \to \mathbb{C}$.

(ii) $\mathcal{M}_{\mathbb{R}}(X, \mathcal{B})$ is the set of all real-valued measurable functions $f : X \to \mathbb{R}$.

Definition 3.1.1. *Let $(X, \mathcal{B}), (Y, \mathcal{C})$ be measurable spaces and $T : X \to Y$ be a measurable transformation. The operator*

$$U_T : \mathcal{M}_{\mathbb{C}}(Y, \mathcal{C}) \to \mathcal{M}_{\mathbb{C}}(X, \mathcal{B}), \quad U_T(f) = f \circ T. \tag{3.1}$$

is called the ***operator induced by*** *T.*

Definition 3.1.2. *A mapping $f \in \mathcal{M}_{\mathbb{C}}(Y, \mathcal{C})$ is said to be* ***T-invariant*** *if $U_T(f) = f$.*

The following lemmas collect some basic properties of the induced operator.

Lemma 3.1.3. *Let $(X, \mathcal{B}), (Y, \mathcal{C}), (Z, \mathcal{D})$ be measurable spaces, $T : X \to Y, S : Y \to Z$ be measurable transformations.*

(i) $U_{S\circ T} = U_T \circ U_S$.

(ii) U_T is linear and $U_T(f \cdot g) = (U_T f) \cdot (U_T g)$ for all $f, g \in \mathcal{M}_{\mathbb{C}}(Y, \mathcal{C})$.

(iii) If $f : Y \to \mathbb{C}$, $f(y) = c$ is a constant function, then $U_T(f)(x) = c$ for every $x \in X$.

(iv) $U_T(\mathcal{M}_{\mathbb{R}}(Y, \mathcal{C})) \subseteq \mathcal{M}_{\mathbb{R}}(X, \mathcal{B})$.

(v) If $f \in \mathcal{M}_{\mathbb{R}}(Y, \mathcal{C})$ is nonnegative, then $U_T f$ is nonnegative too, hence U_T is a positive operator.

(vi) *For all* $C \in \mathcal{C}$, $U_T(\chi_C) = \chi_{T^{-1}(C)}$.

(vii) *If* f *is a simple function in* $\mathcal{M}_{\mathbb{C}}(Y, \mathcal{C})$, $f = \sum_{i=1}^{n} c_i \chi_{C_i}$, $c_i \in \mathbb{C}, C_i \in \mathcal{C}$, *then* $U_T f$ *is a simple function in* $\mathcal{M}_{\mathbb{C}}(X, \mathcal{B})$, $U_T f = \sum_{i=1}^{n} c_i \chi_{T^{-1}(C_i)}$.

Proof. Exercise. □

Lemma 3.1.4. *Let* $(X, \mathcal{B})$ *be a measurable space and* $T : X \to X$ *be measurable.*

(i) $U_{1_X} = 1_{\mathcal{M}_{\mathbb{C}}(X,\mathcal{B})}$

(ii) $U_{T^n} = (U_T)^n$ *for all* $n \in \mathbb{N}$.

(iii) *If* $T : X \to X$ *is bijective and both* T *and* T^{-1} *are measurable, then* U_T *is invertible and its inverse is* $U_{T^{-1}}$. *Furthermore,* $U_{T^n} = (U_T)^n$ *for all* $n \in \mathbb{Z}$.

Proposition 3.1.5.
Let $(X, \mathcal{B}, \mu), (Y, \mathcal{C}, \nu)$ *be probability spaces and* $T : X \to Y$ *be a measurable transformation. The following are equivalent*

(i) T *is a measure-preserving transformation.*

(ii) *For all* $f \in \mathcal{M}_{\mathbb{C}}(Y, \mathcal{C})$,

$$\int_X U_T f \, d\mu = \int_Y f \, d\nu. \tag{3.2}$$

Proof. (i)⇒(ii) It suffices to prove the result when f is real-valued and, by considering positive and negative parts of f, it suffices to consider non-negative functions. So, suppose that $f \geq 0$. If f is a measurable simple function, $f = \sum_{i=1}^{n} c_i \chi_{C_i}$, then by Lemma 3.1.3.(vii), $U_T f = \sum_{i=1}^{n} c_i \chi_{T^{-1}(C_i)}$ is a measurable simple function, hence

$$\begin{aligned}
\int_X U_T f \, d\mu &= \sum_{i=1} c_i \mu(T^{-1}(C_i)) = \sum_{i=1} c_i \nu(C_i), \quad \text{as } T \text{ is measure-preserving} \\
&= \int_Y f \, d\nu.
\end{aligned}$$

Otherwise, by (C.7.10), there exists an increasing sequence of simple functions (s_n) such that $0 \leq s_n \leq f$ for all n, and $\lim_{n\to\infty} s_n(y) = f(y)$ for all $y \in Y$. Then $(U_T s_n)$ is an increasing sequence of simple functions such that $0 \leq U_T s_n \leq U_T f$, and for all $x \in X$,

$$\lim_{n\to\infty} (U_T s_n)(x) = \lim_{n\to\infty} s_n(Tx) = f(Tx) = U_T f(x).$$

Apply now (C.8.1) to get that

$$\int_X U_T f\, d\mu \quad = \quad \lim_{n\to\infty} \int_X U_T s_n\, d\mu = \lim_{n\to\infty} \int_Y s_n\, d\nu = \int_Y f\, d\nu.$$

(ii)⇒(i) Let $A \in \mathcal{C}$. Then $\chi_A \in \mathcal{M}_{\mathbb{C}}(Y, \mathcal{C})$ and $U_T(\chi_A) = \chi_{T^{-1}(A)}$ by Lemma 3.1.3.(vi). Applying (ii) with $f := \chi_A$, we get that

$$\nu(A) = \int_Y \chi_A\, d\nu = \int_X U_T(\chi_A)\, d\mu = \int_X \chi_{T^{-1}(A)}\, d\mu = \mu(T^{-1}(A)).$$

□

Theorem 3.1.6.
Let $(X, \mathcal{B}, \mu), (Y, \mathcal{C}, \nu)$ be probability spaces, and $T : X \to X$ be a measure-preserving transformation. For all $1 \le p < \infty$,

(i) $U_T(L^p(Y, \mathcal{C}, \nu)) \subseteq L^p(X, \mathcal{B}, \mu)$ and $U_T(L^p_{\mathbb{R}}(Y, \mathcal{C}, \nu)) \subseteq L^p_{\mathbb{R}}(X, \mathcal{B}, \mu)$,

(ii) the operator $U_T : L^p(Y, \mathcal{C}, \nu) \to L^p(X, \mathcal{B}, \mu)$ is a linear isometry, i.e.

$$\|U_T f\|_p = \|f\|_p \quad \text{for all } f \in L^p(Y, \mathcal{C}, \nu). \tag{3.3}$$

Proof. Let $f \in L^p(Y, \mathcal{C}, \nu)$ and let $g : Y \to \mathbb{C},\ g(y) := |f(y)|^p$. Then g is integrable, since $f \in L^p(Y, \mathcal{C}, \nu)$ and, furthermore, $U_T g(x) = g(Tx) = |f(Tx)|^p = |U_T f(x)|^p$. Applying Theorem 3.1.5 for g, it follows that

$$\int |U_T f|^p\, d\mu \quad = \quad \int U_T g\, d\mu = \int g\, d\nu = \int |f|^p\, d\nu.$$

Thus, $U_T f \in L^p(X, \mathcal{B}, \mu)$ and $\|U_T f\|_p = \|f\|_p$. □

Therefore a measure-preserving transformation $T : X \to Y$ induces a linear isometry of $L^p(Y, \mathcal{C}, \nu)$ and $L^p(X, \mathcal{B}, \mu)$ for all $1 \le p < \infty$.

Proposition 3.1.7. *If $(X, \mathcal{B}, \mu, T)$ is an invertible measure-preserving system, then U_T is an unitary operator on the Hilbert space $L^2(X, \mathcal{B}, \mu)$.*

Proof. U_T is invertible by Proposition 3.1.4. Furthermore, U_T is an isometry. □

The study of U_T is called the spectral study of T and this is useful in formulating concepts such as ergodicity and mixing.

3.2 Basic constructions

3.2.1 Invariant subsets

Let $(X, \mathcal{B}, \mu, T)$ be a MDS.

Definition 3.2.1. *A set $A \in \mathcal{B}$ is **invariant by** T, or T-**invariant** if $T^{-1}(A) = A$.*

The fundamental property of this concept is the following: if A is T-invariant, then so is $X \setminus A$. Thus, when A is T-invariant we obtain by restriction two well-defined transformations

$$T_A : A \to A, \quad T_{X \setminus A} : X \setminus A \to X \setminus A.$$

Hence, the existence of an invariant subset allows one to decompose the set X into two disjoint subsets and study the transformation T in each of these subsets.

Lemma 3.2.2. *(i) The set of all T-invariant subsets of X is a σ-algebra on X.*

(ii) If $A \in \mathcal{B}$ is T-invariant, then $(A, A \cap \mathcal{B}, \mu|_{A \cap \mathcal{B}}, T_A)$ is a MDS.

Proof. (i) Exercise.

(ii) We have that $(A, A \cap \mathcal{B}, \mu|_{A \cap \mathcal{B}})$ is a probability space. It remains to prove that T_A is measure-preserving. Let $B \in \mathcal{B}$. Then

$$\begin{aligned} T^{-1}(A \cap B) &= T^{-1}(A) \cap T^{-1}(B) = A \cap T^{-1}(B) \in A \cap \mathcal{B}, \quad \text{since } T \text{ is measurable} \\ \mu(T^{-1}(A \cap B)) &= \mu(A \cap B), \quad \text{since } A \cap B \in \mathcal{B}. \end{aligned}$$

□

We shall denote with $\mathcal{B}^T$ the σ-algebra of T-invariant subsets of X.

3.3 Bernoulli shift

Let $W = \{w_1, \ldots, w_k\}$ be a finite nonempty set with $k = |W| \geq 2$, $W^{\mathbb{Z}}$ be the full W-shift, and

$$T : W^{\mathbb{Z}} \to W^{\mathbb{Z}}, \quad (T\mathbf{x})_n = x_{n+1} \text{ for all } n \in \mathbb{Z} \tag{3.4}$$

be the shift map. We refer to Section 1.2 for details.

We consider the measurable space $(W, \mathcal{P}(W))$. Let $(p_1, \ldots, p_k)$ be a probability vector with non-zero entries, i.e. $p_i > 0$ for all $i = 1, \ldots, k$ and $\sum_{i=1}^{k} p_i = 1$. Define a probability measure $\nu : \mathcal{P}(W) \to [0, 1]$ by

$$\nu(\{w_i\}) = p_i, \ \nu(A) = \textstyle\sum_{w \in A} \nu(\{w\}) \text{ for any (finite) subset of } W.$$

The probability measure ν is called the $(p_1,\ldots,p_k)$-product measure. Thus, $(W,\mathcal{P}(W),\nu)$ is a probability space.

Consider the product probability space

$$\left(W^{\mathbb{Z}}, \mathcal{B}=\bigotimes_{i\in\mathbb{Z}}\mathcal{P}(W), \mu=\bigotimes_{i\in\mathbb{Z}}\nu\right)=\prod_{i\in\mathbb{Z}}(W,\mathcal{P}(W),\nu). \tag{3.5}$$

We refer to C.3 for details.

Let us recall the following notations:

$$\begin{aligned}
C_n^w &= \{\mathbf{x}\in W^{\mathbb{Z}} \mid x_n = w\}, \quad \text{where } n\in\mathbb{Z},\ w\in W, && (3.6)\\
C_{n_1,\ldots,n_t}^{w_{i_1},\ldots,w_{i_t}} &= \{\mathbf{x}\in W^{\mathbb{Z}} \mid x_{n_j}=w_{i_j} \text{ for all } j=1,\ldots,t\} = \bigcap_{j=1}^t C_{n_j}^{w_{i_j}}, && (3.7)\\
&\quad \text{where } t\geq 1, n_1<n_2<\ldots<n_t\in\mathbb{Z}, w_{i_1},\ldots,w_{i_t}\in W,\\
R_n^A &= \{\mathbf{x}\in W^{\mathbb{Z}} \mid x_n\in A\} = \bigcup_{w\in A} C_n^w, \quad \text{where } n\in\mathbb{Z},\ A\subseteq W, && (3.8)\\
R_{n_1,\ldots,n_t}^{A_1,\ldots,A_t} &= \{\mathbf{x}\in W^{\mathbb{Z}} \mid x_{n_i}\in A_i \text{ for all } i=1,\ldots,t\} = \bigcap_{i=1}^t R_{n_i}^{A_i} = \bigcap_{i=1}^t\bigcup_{w\in A_i} C_n^w && (3.9)\\
&\quad \text{where } t\geq 1, n_1<n_2<\ldots<n_t\in\mathbb{Z}, A_1,\ldots,A_t\subseteq W.
\end{aligned}$$

Let $\mathcal{R}$ be the set of all measurable rectangles. Then $\mathcal{B}$ is the σ-algebra generated by $\mathcal{R}$, and μ is the unique probability measure on $(X,\mathcal{B})$ such that

$$\mu(R_{n_1,\ldots,n_t}^{A_1,\ldots,A_t}) = \prod_{i=1}^t \nu(A_i) \quad \text{for every rectangle } R_{n_1,\ldots,n_t}^{A_1,\ldots,A_t}. \tag{3.10}$$

In particular,

$$\begin{aligned}
\mu(C_n^{w_i}) &= \mu(R_n^{\{w_i\}}) = \nu(\{w_i\}) = p_i,\\
\mu(C_{n_1,\ldots,n_t}^{w_{i_1},\ldots,w_{i_t}}) &= \prod_{j=1}^t p_{i_j}.
\end{aligned}$$

We recall that we use the notations $\mathcal{C}$ for the set of all cylinders $C_{n_1,\ldots,n_t}^{w_{i_1},\ldots,w_{i_t}}$ and $\mathcal{C}_e$ for the set of elementary cylinders C_n^w.

Proposition 3.3.1. *(i) $\mathcal{S}=\mathcal{C}\cup\{\emptyset\}$ is a semialgebra on $W^{\mathbb{Z}}$.*

(ii) $\mathcal{B}=\sigma(\mathcal{S})=\sigma(\mathcal{C}_e)$.

(iii) $\mathcal{B}$ coincides with the Borel σ-algebra on W.

Proof. (i) We have that $\emptyset \in \mathcal{S}$ and that $\mathcal{S}$ is closed under finite intersections as an immediate consequence of Lemma 1.2.7.(ii). Furthermore,

$$W^{\mathbb{Z}} \setminus C_{n_1,\ldots,n_t}^{w_{i_1},\ldots,w_{i_t}} = \bigcup_{u_1 \neq w_{i_1}} C_{n_1}^{u_1} \cup \bigcup_{u_2 \neq w_{i_2}} C_{n_1,n_2}^{w_{i_1},u_2} \cup \ldots \cup \bigcup_{u_t \neq w_{i_t}} C_{n_1,\ldots,n_{t-1},n_t}^{w_{i_1},\ldots,w_{i_{t-1}},u_t}$$

is a finite union of pairwise disjoint cylinders.

(ii) $\mathcal{B}$ is the σ-algebra generated by the set $\mathcal{R}$ of measurable rectangles. By (3.9), we have that $\mathcal{C}_e \subseteq \mathcal{R} \subseteq \mathcal{A}(\mathcal{C}_e)$, hence $\sigma(\mathcal{C}_e) \subseteq \mathcal{B} = \sigma(\mathcal{R}) \subseteq \sigma(\mathcal{A}(\mathcal{C}_e)) = \sigma(\mathcal{C}_e)$. Thus, $\mathcal{B} = \sigma(\mathcal{C}_e)$. Since $\mathcal{C}_e \subseteq \mathcal{S} \subseteq \mathcal{R}$, we also get that $\sigma(\mathcal{S}) = \mathcal{B}$.

(iii) Let $\mathcal{B}(W^{\mathbb{Z}})$ be the Borel σ-algebra on $W^{\mathbb{Z}}$. We have to prove that $\mathcal{B} = \mathcal{B}(W^{\mathbb{Z}})$.
"$\subseteq$" follows from the fact that the elementary cylinders are open sets in $W^{\mathbb{Z}}$.
$\supseteq$ The set $\mathcal{C}$ of cylinders is countable, since W is finite. Since $\mathcal{C}$ is a basis for the product topology on $W^{\mathbb{Z}}$, any open set U of $W^{\mathbb{Z}}$ is a union of sets in $\mathcal{C}$, hence U is an at most countable union of sets in $\mathcal{C}$. Thus, any open set is in $\sigma(\mathcal{C}) = \sigma(\mathcal{S}) = \mathcal{B}$. □

Proposition 3.3.2. *$(W^{\mathbb{Z}}, \mathcal{B}, \mu, T)$ is an invertible MDS.*

Proof. We know already that T is invertible, and by Proposition 3.0.11, it remains to prove that T is measure-preserving. We shall apply Proposition 3.0.12 for the semialgebra $\mathcal{S}$ that generates $\mathcal{B}$. Let $C_{n_1,\ldots,n_t}^{w_{i_1},\ldots,w_{i_t}} \in \mathcal{C}$. Using Lemma 1.2.7.(v), we get that

$$\begin{aligned} T^{-1}(C_{n_1,\ldots,n_t}^{w_{i_1},\ldots,w_{i_t}}) &= C_{n_1+1,\ldots,n_t+1}^{w_{i_1},\ldots,w_{i_t}} \in \mathcal{S} \subseteq \mathcal{B}, \\ \mu(C_{n_1,\ldots,n_t}^{w_{i_1},\ldots,w_{i_t}}) &= \prod_{j=1}^{t} p_{i_j} = \mu(C_{n_1+1,\ldots,n_t+1}^{w_{i_1},\ldots,w_{i_t}}) = \mu(T^{-1}(C_{n_1,\ldots,n_t}^{w_{i_1},\ldots,w_{i_t}})). \end{aligned}$$

□

The invertible MDS $(W^{\mathbb{Z}}, \mathcal{B}, \mu, T)$ is called the **Bernoulli shift** and is also denoted by $B(p_1, \ldots, p_k)$.

3.4 Recurrence

Let $(X, \mathcal{B}, \mu, T)$ be a MDS. In this section we discuss the problem of recurrence, one of the most basic questions to be asked about the natures of orbits of points and measurable sets.

Given a measurable set $A \in \mathcal{B}$, we recall the following notations:

(i) A_{ret} is the set of those points of A which return to A **at least once**.

(ii) A_{inf} is the set of those points of A which return to A **infinitely often**.

Using the notations

$$A^+ := \bigcup_{n\geq 1} T^{-n}(A), \quad A^\star := A \cup A^+ = \bigcup_{n\geq 0} T^{-n}(A),$$

we have that

$$\begin{aligned} A_{ret} &= A \cap \bigcup_{n\geq 1} T^{-n}(A) = A \cap A^+, \quad A \setminus A_{ret} = A \setminus A^+ = A^\star \setminus A^+, \\ A_{inf} &= A \cap \bigcap_{n\geq 1} \bigcup_{m\geq n} T^{-m}(A) = \bigcap_{n\geq 1} \bigcup_{m\geq n} (A \cap T^{-m}(A)) = A \cap \bigcap_{n\geq 1} T^{-n}(A^\star) \end{aligned}$$

A point $x \in A_{ret}$ is also said to be **recurrent with respect to** A, while a point $x \in A_{inf}$ is **infinitely recurrent with respect to** A.

Definition 3.4.1. *A measurable set $A \in \mathcal{B}$ is called* ***wandering*** *if the sets*

$$A, T^{-1}(A), \ldots, T^{-n}(A), \ldots$$

are pairwise disjoint.

Lemma 3.4.2. *Let $A \in \mathcal{B}$.*

(i) $A \setminus A_{ret}$ is wandering.

(ii) $A \setminus A_{inf} = A \cap \bigcup_{n\geq 0} T^{-n}(A \setminus A_{ret})$.

Proof. Exercise. □

Definition 3.4.3.

(i) T is ***recurrent*** *if for all $A \in \mathcal{B}$ almost all points of A return to A.*

(ii) T is ***infinitely recurrent*** *if for all $A \in \mathcal{B}$ almost all points of A return infinitely often to A.*

Thus, T is recurrent if and only if $\mu(A \setminus A_{ret}) = 0$ if and only if $\mu(A) = \mu(A_{ret})$. Furthermore, T is infinitely recurrent if and only if $\mu(A \setminus A_{inf}) = 0$ if and only if $\mu(A) = \mu(A_{inf})$.

Definition 3.4.4.

(i) T is ***conservative*** *if there are no wandering sets A with $\mu(A) > 0$.*

(ii) T is ***incompressible*** *if whenever $A \in \mathcal{B}$ and $T^{-1}(A) \subseteq A$, then $\mu(A \setminus T^{-1}(A)) = 0$.*

The following theorem and its proof are due to F. B. Wright [116]. The crucial point is the simple proof of (i)⇒(iv). The truth of this conclusion was already known before, but only by heavy technics: Halmos [49] and Taam [104].

Theorem 3.4.5. *Let $(X, \mathcal{B}, \mu)$ be a measure space and $T : X \to X$ be a measurable transformation. The following are equivalent*

(i) T is incompressible.

(ii) T is conservative.

(iii) T is recurrent.

(iv) T is infinitely recurrent.

(v) For all $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists $n \geq 1$ such that $\mu(A \cap T^{-n}(A)) > 0$.

(vi) For all $A \in \mathcal{B}$ with $\mu(A) > 0$, there exist infinitely many $n \geq 1$ such that $\mu(A \cap T^{-n}(A)) > 0$.

Proof. Let $A \in \mathcal{B}$.
$(i) \Rightarrow (iii)$ We have that $T^{-1}(A^\star) = A^+ \subseteq A^\star$ and $A \setminus A_{ret} = A^\star \setminus A^+ = A^\star \setminus T^{-1}(A^\star)$. Since T is incompressible, it follows that $\mu(A \setminus A_{ret}) = \mu(A^\star \setminus T^{-1}(A^\star)) = 0$. Hence, T is recurrent.

$(iii) \Rightarrow (i)$ Assume that $T^{-1}(A) \subseteq A$. Then $A^+ = T^{-1}(A)$, hence

$$\mu(A \setminus T^{-1}(A)) = \mu(A \setminus A^+) = \mu(A \setminus A_{ret}) = 0.$$

$(ii) \Rightarrow (iii)$ By Lemma 3.4.2.(i), $A \setminus A_{ret}$ is wandering, hence using the fact that T is conservative, $\mu(A \setminus A_{ret}) = 0$. Thus, T is recurrent.

$(iii) \Rightarrow (ii)$ Assume that A is wandering. Then the sets A and $T^{-n}(A)$ are disjoint for all $n \geq 1$, hence

$$A_{ret} = A \cap A^+ = \bigcup_{n \geq 1} (A \cap T^{-n}(A)) = \emptyset.$$

Since T is recurrent, we have that $\mu(A) = \mu(A_{ret}) = 0$.

$(iv) \Rightarrow (iii)$ Obvious.

$(i) \Rightarrow (iv)$ By Lemma 3.4.2, we have that

$$\begin{aligned} A \setminus A_{inf} &= A \cap \bigcup_{n \geq 0} T^{-n}(A \setminus A_{ret}) = A \cap \bigcup_{n \geq 0} T^{-n}(A^\star \setminus T^{-1}(A^\star)) \\ &= A \cap \bigcup_{n \geq 0} \big(T^{-n}(A^\star) \setminus T^{-n-1}(A^\star)\big). \end{aligned}$$

Since $T^{-1}(A^\star) \subseteq A^\star$, we get that $T^{-n-1}(A^\star) \subseteq T^{-n}(A^\star)$. Apply now the fact that T is incompressible to obtain $\mu\big(T^{-n}(A^\star) \setminus T^{-n-1}(A^\star)\big) = 0$ for all $n \geq 0$. Consequently, $\mu(A \setminus A_{inf}) = 0$, hence T is infinitely recurrent.

$(iii) \Rightarrow (v)$ Assume that $\mu(A \cap T^{-n}(A)) = 0$ for all $n \geq 1$. Then

$$\mu(A_{ret}) = \mu(A \cap A^+) = \mu\left(\bigcup_{n\geq 1}(A \cap T^{-n}(A))\right) \leq \sum_{n\geq 1}\mu(A \cap T^{-n}(A)) = 0,$$

hence $\mu(A_{ret}) = 0$. On the other hand, since T is recurrent, we have that $\mu(A_{ret}) = \mu(A) > 0$. We have got a contradiction.

$(v) \Rightarrow (ii)$ If A is a wandering set, then $A \cap T^{-n}(A) = \emptyset$, hence $\mu(A \cap T^{-n}(A)) = 0$ for all $n \geq 1$. By (v), we must have $\mu(A) = 0$.

$(vi) \Rightarrow (v)$ is obvious.

$(iv) \Rightarrow (vi)$ Assume that $\mu(A \cap T^{-n}(A)) > 0$ only for finitely many $n \geq 1$. Hence there exists $N \geq 1$ such that $\mu(A \cap T^{-n}(A)) = 0$ for all $n \geq N$. It follows that

$$\mu(A_{inf}) = \mu\left(\bigcap_{n\geq 1}\bigcup_{m\geq n} A \cap T^{-m}(A)\right) \leq \mu\left(\bigcup_{m\geq N} A \cap T^{-m}(A)\right) \leq \sum_{m\geq N}\mu(A \cap T^{-m}(A)) = 0.$$

On the other hand, T is infinitely recurrent, hence $\mu(A_{inf}) = \mu(A) > 0$. We have got a contradiction. □

3.4.1 Poincaré Recurrence Theorem

The mathematically simple Poincaré recurrence threorem has become famous because of its physical and philosophical implications, some of which are indicated in [83, p. 34-36]. It may be considered to be the most basic result in ergodic theory.

Theorem 3.4.6 (Poincaré Recurrence Theorem (1899))**.**
Let $(X, \mathcal{B}, \mu, T)$ be a MDS. Then for all $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists $n \geq 1$ such that $\mu(A \cap T^{-n}(A)) > 0$.

Proof. We prove that T is conservative and then apply Theorem 3.4.5 to get the conclusion. Let $A \in \mathcal{B}$ be a wandering set. Then $A, T^{-1}(A), \ldots$ is a sequence of mutually disjoint measurable sets having the same measure, since T is measure-preserving. If $\mu(A) > 0$, then

$$1 = \mu(X) \geq \mu\left(\bigcup_{n\geq 0} T^{-n}(A)\right) = \sum_{n=0}^{\infty}\mu\left(T^{-n}(A)\right) = \sum_{n=0}^{\infty}\mu(A) = \infty,$$

that is a contradictions. We must have then $\mu(A) = 0$, hence T is conservative. □

A quantitative version of Poincaré Recurrence Theorem is the following.

Proposition 3.4.7.
Let $(X, \mathcal{B}, \mu, T)$ be a MDS. If $A \in \mathcal{B}$ is such that $\mu(A) > 0$, then there exists $1 \leq N \leq \Phi$ such that

$$\mu(A \cap T^{-N}(A)) > 0,$$

where $\Phi = \left\lceil \dfrac{1}{\mu(A)} \right\rceil$.

Proof. Exercise. □

3.5 Ergodicity

Let $(X, \mathcal{B}, \mu, T)$ be a MDS. If $A \in \mathcal{B}$ is T-invariant (i.e. $T^{-1}(A) = A$), then also $X \setminus A$ is T-invariant and we could study T by studying two simpler transformations T_A and $T_{X\setminus A}$. If $\mu(A) \neq 0$ and $\mu(X \setminus A) \neq 0$ this has simplified the study of T. If $\mu(A) = 0$ (or $\mu(X \setminus A) = 0$) we can ignore A (or $X \setminus A$) and we have not significantly simplified T since neglecting a set of measure zero is allowed in measure theory. This raises the idea of studying transformations that cannot be decomposed as above and of trying to express every measure-preserving transformation in terms of the indecomposable ones. The indecomposable transformations are called ergodic.

Definition 3.5.1. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS. T is called **ergodic** if for all $A \in \mathcal{B}$,*

$$T^{-1}(A) = A \quad \text{implies} \quad \mu(A) = 0 \text{ or } \mu(X \setminus A) = 0. \tag{3.11}$$

*We also say that the MDS $(X, \mathcal{B}, \mu, T)$ is **ergodic** or that μ is T**-ergodic**.*

Since $\mu(X \setminus A) = \mu(X) - \mu(A) = 1 - \mu(A)$, we have that $\mu(X \setminus A) = 0$ is equivalent with $\mu(A) = 1$.

The next theorem characterizes ergodicity in terms of the induced operator U_T.

Proposition 3.5.2. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS. The following are equivalent*

(i) T is ergodic.

(ii) Whenever f is measurable and $f \circ T = f$ a.e., then f is constant a.e.

Proof. Exercise. □

Example 3.5.3. (i) The identity transformation 1_X is ergodic if and only if all members of $\mathcal{B}$ have measure 0 or 1.

(ii) The Bernoulli shift is ergodic.

(iii) Let $(\mathbb{S}^1, R_a)$ be the rotation on the circle group. Then it is ergodic if and only of a is not a root of unity.

Proof. (i) Obviously.

(ii) See Example 7.0.4 for the proof of a stronger fact.

(iii) See [113, Example (2), p.24].

□

Chapter 4

Invariant measures on compact metric spaces

Let (X, d) be a compact metric space and $\mathcal{B}(X)$ be the Borel σ-algebra on X. We shall denote with $\mathcal{M}(X)$ the set of all Borel probability measures on X. By a **Borel probability space** we mean a probability space $(X, \mathcal{B}(X), \mu)$, where μ is a Borel probability measure on X.

Definition 4.0.4. *A **Borel measure-preserving dynamical system** (Borel MDS for short) is a MDS $(X, \mathcal{B}(X), \mu, T)$, where (X, d) is a compact metric space, $\mathcal{B}(X)$ is the Borel σ-algebra on X.*

4.1 Ergodic decomposition

Definition 4.1.1. *Let $(X, \mathcal{B})$, $(Y, \mathcal{C})$ be measurable spaces. A **probability kernel** $y \mapsto \mu_y$ is an assignment of a probability measure μ_y on X to each $y \in Y$ in such a way that the map*

$$y \mapsto \int_X f \, d\mu_y$$

is measurable for every bounded measurable $f : X \to \mathbb{C}$.

Theorem 4.1.2. *[Ergodic decomposition]*
Let $(X, \mathcal{B}(X), \mu, T)$ be a Borel MDS and $\mathcal{B}^T$ be the σ-algebra of T-invariants sets of X. Then there exists a probability kernel $x \mapsto \mu_x$ between the measurable spaces $(X, \mathcal{B})$ and $(X, \mathcal{B}^T)$ satisfying

(i) μ_x is T-invariant and ergodic for almost all $x \in X$.

(ii) $\int_X f \, d\mu = \int_X \left(\int_X f \, d\mu_x \right) d\mu(x)$ for any bounded measurable $f : X \to \mathbb{C}$.

Proof. See [109, Theorem 2.9.21, p. 265] and [109, Theorem 2.9.22, p. 266]. See also [28, Theorem 6.3]. □

Let $A \in \mathcal{B}$. By taking $f := \chi_A$ in (ii) above we get that $\mu(A) = \int_X \mu_x(A) d\mu(x)$.

Chapter 5

Ergodic theorems

In the following, $(X, \mathcal{B}, \mu, T)$ is a MDS.

For every $f \in \mathcal{M}_{\mathbb{C}}(X, \mathcal{B})$, we consider the **ergodic average**

$$S_n f : X \to \mathbb{C}, \quad S_n f(x) := \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x). \tag{5.1}$$

We shall also use the following notations for $f \in \mathcal{M}_{\mathbb{R}}(X, \mathcal{B})$:

$$f^\star(x) := \sup_{n \geq 1} S_n f(x), \qquad f_\star(x) := \inf_{n \geq 1} S_n f(x), \tag{5.2}$$

$$\underline{f}(x) := \liminf_{n \to \infty} S_n f(x), \quad \overline{f}(x) := \limsup_{n \to \infty} S_n f(x). \tag{5.3}$$

Lemma 5.0.3. *Let $f \in \mathcal{M}_{\mathbb{C}}(X, \mathcal{B})$ and $n \geq 1$.*

(i) If f is T-invariant (a.e.), then $S_n f = f$ (a.e.).

(ii) $S_n f \in \mathcal{M}_{\mathbb{C}}(X, \mathcal{B})$.

(iii) $S_n f = \dfrac{1}{n} \displaystyle\sum_{k=0}^{n-1} U_{T^k} f$.

(iv) For any $p \geq 1$, $f \in L^p(X, \mathcal{B}, \mu)$ (resp. $L^p_{\mathbb{R}}(X, \mathcal{B}, \mu)$) implies $S_n f \in L^p(X, \mathcal{B}, \mu)$ (resp. $L^p_{\mathbb{R}}(X, \mathcal{B}, \mu)$).

(v) For all $x \in X$, $\dfrac{n+1}{n} S_{n+1}(x) - S_n f(Tx) = \dfrac{1}{n} f(x)$.

(vi) If $f \in \mathcal{M}_{\mathbb{R}}(X, \mathcal{B})$, then $\underline{f} \circ T = \underline{f}$ and $\overline{f} \circ T = \overline{f}$.

(vii) $\displaystyle\int_X S_n f \, d\mu = \int_X f \, d\mu$.

(viii) If $f \in L^1_{\mathbb{R}}(X,\mathcal{B},\mu)$ *is nonnegative, then* $S_n f \in L^1_{\mathbb{R}}(X,\mathcal{B},\mu)$ *is nonnegative and* $\|S_n f\|_1 = \|f\|_1$.

Proof. Exercise. □

Lemma 5.0.4. *Let* $A, B \in \mathcal{B}$ *and* $n \geq 1$.

(i) $S_n\chi_A = \frac{1}{n}\sum_{k=0}^{n-1}\chi_{T^{-k}(A)}$ *and* $\chi_B \cdot S_n\chi_A = \frac{1}{n}\sum_{k=0}^{n-1}\chi_{T^{-k}(A)\cap B}$.

(ii) $\int_X S_n\chi_A = \mu(A)$.

(iii) $\int_X \chi_B \cdot S_n\chi_A \, d\mu = \frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}(A)\cap B)$.

Proof. Exercise. □

5.1 Maximal Ergodic Theorems

A linear operator $U : L^1_{\mathbb{R}}(X,\mathcal{B},\mu) \to L^1_{\mathbb{R}}(X,\mathcal{B},\mu)$ is said to be **positive** if for all $f \in L^1_{\mathbb{R}}(X,\mathcal{B},\mu)$, $f \geq 0$ implies $Uf \geq 0$ a.e. We assume also that U is **nonexpansive**, i.e. $\|Uf\|_1 \leq \|f\|_1$ for all $f \in L^1_{\mathbb{R}}(X,\mathcal{B},\mu)$.

We recall the following notations for an arbitrary mapping $g : X \to \mathbb{R}$ and $\alpha \in \mathbb{R}$:

$$\{g > \alpha\} := g^{-1}((\alpha,\infty)), \quad \{g \geq \alpha\} := g^{-1}([\alpha,\infty)).$$

The following theorem was obtained by Hopf [56]; the proof we present here was given by Garsia [40].

Theorem 5.1.1 (Hopf Maximal Ergodic Theorem)**.**
Let $U : L^1_{\mathbb{R}}(X,\mathcal{B},\mu) \to L^1_{\mathbb{R}}(X,\mathcal{B},\mu)$ *be a nonexpansive positive linear operator. For all* $f \in L^1_{\mathbb{R}}(X,\mathcal{B},\mu)$*,*

$$\int_{\{f^\star > 0\}} f \, d\mu \geq 0. \tag{5.4}$$

where $f^\star = \sup_{n\geq 1}\frac{1}{n}\sum_{k=0}^{n-1}U^k f$.

Proof. First, let us remark that $f^\star$ is measurable, as a maximum of measurable functions. Hence, $\{f^\star > 0\}$ is a measurable set.

Define the sequence $(f_n)_{n\geq 0}$ in $L^1_{\mathbb{R}}(X,\mathcal{B},\mu)$ by:

$$f_0 := 0, \quad f_n := \sum_{k=0}^{n-1}U^k f \quad \text{for } n \geq 1$$

and the sequence $(F_n)_{n\geq 1}$ in $L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$ by:

$$F_n := \max_{0\leq k\leq n} f_k.$$

Let us remark that $F_n \geq f_0 \geq 0$ for all $n \geq 1$ and that $f_{n+1} = f + Uf_n$ for all $n \geq 0$.

Let $n \geq 1$ and $x \in \{F_n > 0\}$. It follows that

$$\begin{aligned} F_n(x) &= \max_{0\leq k\leq n} f_k(x) = \max_{1\leq k\leq n} f_k(x), \quad \text{since } F_n(x) > 0 \\ &\leq \max_{1\leq k\leq n+1} f_k(x) = \max_{0\leq k\leq n} f_{k+1}(x) = \max_{0\leq k\leq n} (f + Uf_k)(x) \leq (f + UF_n)(x), \\ &\quad \text{since } F_n \geq f_k, \text{ hence } UF_n \geq Uf_k \text{ by the positivity of } U. \end{aligned}$$

Claim For all $n \geq 1$, $\int_{\{F_n>0\}} f\, d\mu \geq 0$.

Proof:

$$\begin{aligned} \int_{\{F_n>0\}} f\, d\mu &\geq \int_{\{F_n>0\}} (F_n - UF_n)\, d\mu = \int_{\{F_n>0\}} F_n\, d\mu - \int_{\{F_n>0\}} UF_n\, d\mu \\ &= \int_X F_n\, d\mu - \int_{\{F_n>0\}} UF_n\, d\mu, \quad \text{since } F_n = 0 \text{ on } X \setminus \{F_n > 0\} \\ &\geq \int_X F_n\, d\mu - \int_X UF_n\, d\mu, \quad \text{since } F_n \geq 0, \text{ so } UF_n \geq 0, \\ &\quad \text{and hence } \int_{\{F_n>0\}} UF_n\, d\mu \leq \int_X UF_n\, d\mu \\ &= \int_X |F_n|\, d\mu - \int_X |UF_n|\, d\mu = \|F_n\|_1 - \|UF_n\|_1 \\ &\geq 0. \quad \square. \end{aligned}$$

Furthermore, $x \in \{f^\star > 0\}$ if and only if there exists $n \geq 1$ such that $f_n(x) > 0$ if and only if there exists $n \geq 1$ such that $F_n(x) > 0$. Thus,

$$\{f^\star > 0\} = \bigcup_{n\geq 1} \{F_n > 0\}.$$

Furthermore, since $(F_n)_{n\geq 1}$ is increasing, we get that $(\{F_n > 0\})_{n\geq 1}$ is an increasing sequence of measurable subsets of X. We can apply C.8.9.(v) to conclude that

$$\int_{\{f^\star>0\}} f\, d\mu = \lim_{n\to\infty} \int_{\{F_n>0\}} f\, d\mu \geq 0.$$

$\square$

As an immediate consequence, we get the Yosida-Kakutani maximal ergodic theorem [118].

Theorem 5.1.2 (Maximal Ergodic Theorem)**.**
Let $(X, \mathcal{B}, \mu, T)$ be a MDS. For all $f \in L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$,

$$\int_{\{f^\star > 0\}} f \, d\mu \geq 0, \tag{5.5}$$

where $f^\star = \sup_{n \geq 1} S_n f$ is defined in (5.2).

Proof. Apply Theorem 5.1.1 for the operator $U_T : L^1_{\mathbb{R}}(X, \mathcal{B}, \mu) \to L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$ induced by T, which is positive and nonexpansive, by see Lemma 3.1.3 and Theorem 3.1.6. □

The following *maximal ergodic inequality* was already known to Wiener [115].

Corollary 5.1.3.
Let $(X, \mathcal{B}, \mu, T)$ be a MDS. For all $f \in L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$ and all $\alpha \in \mathbb{R}$,

$$\int_{\{f^\star > \alpha\}} f \, d\mu \geq \alpha\mu(\{f^\star > \alpha\}). \tag{5.6}$$

Proof. Let $g := f - \alpha$. Then $g \in L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$ and

$$g^\star(x) = \sup_{n\geq 1} \frac{1}{n} \sum_{k=0}^{n-1} g(T^k x) = \sup_{n\geq 1} \frac{1}{n} \sum_{k=0}^{n-1} (f(T^k x) - \alpha) = f^\star(x) - \alpha.$$

Thus, $\{f^\star > \alpha\} = \{g^\star > 0\}$, so we can apply Theorem 5.1.2 for g to conclude that $\int_{\{g^\star > 0\}} g \, d\mu \geq 0$. On the other hand,

$$\int_{\{g^\star > 0\}} g \, d\mu = \int_{\{f^\star > \alpha\}} (f - \alpha) \, d\mu = \int_{\{f^\star > \alpha\}} f \, d\mu - \int_{\{f^\star > \alpha\}} \alpha \, d\mu = \int_{\{f^\star > \alpha\}} f \, d\mu - \alpha\mu(\{f^\star > \alpha\}).$$

□

Corollary 5.1.4.
Let $(X, \mathcal{B}, \mu, T)$ be a MDS and $A \subseteq X$ be T-invariant. For all $f \in L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$ and all $\alpha \in \mathbb{R}$,

$$\int_{A \cap \{f^\star > \alpha\}} f \, d\mu \geq \alpha\mu(A \cap \{f^\star > \alpha\}). \tag{5.7}$$

Proof. Apply Corollary 5.1.3 to the MDS $(A, \mathcal{B} \cap A, \mu|_{\mathcal{B}\cap A}, T_A)$. □

5.2 Birkhoff Ergodic Theorem

The following result of G. D. Birkhoff is the fundamental theorem of ergodic theory, known as the **Pointwise Ergodic Theorem** or as just the **Ergodic Theorem**.

Theorem 5.2.1.
Let $(X, \mathcal{B}, \mu, T)$ be a MDS and $f \in L^1(X, \mathcal{B}, \mu)$. Then

(i) $S_n f$ *converges a.e. to a function* f^+ *satisfying* $f^+ \circ T = f^+$ *a.e.*

(ii) $f^+ \in L^1(X, \mathcal{B}, \mu)$ *and, in fact,* $\|f^+\|_1 \leq \|f\|_1$.

(iii) If $A \in \mathcal{B}$ *is* T*-invariant, then* $\int_A f\, d\mu = \int_A f^+\, d\mu$.

(iv) If T *is ergodic, then* f^+ *is constant a.e., namely* $f^+ = \displaystyle\int_X f\, d\mu$ *a.e..*

Proof. By considering real and imaginary parts it suffices to consider $f \in L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$.

(i) Let $\underline{f} = \liminf_{n\to\infty} S_n f$, $\overline{f} = \limsup_{n\to\infty} S_n f$ be as in (5.3). Then, by Proposition 5.0.3.(vi), we have that

$$\underline{f} \circ T = \underline{f} \text{ and } \overline{f} \circ T = \overline{f}. \tag{5.8}$$

We have to show that $\overline{f} = \underline{f}$ a.e., i.e. that the set

$$A := \{x \in X \mid \underline{f}(x) < \overline{f}(x)\}. \tag{5.9}$$

has measure 0.

For each $\alpha, \beta \in \mathbb{R}$ with $\beta < \alpha$, let

$$E_{\alpha,\beta} := \{x \in X \mid \underline{f}(x) < \beta < \alpha < \overline{f}(x). \tag{5.10}$$

It is easy to see that $A = \bigcup\{E_{\alpha,\beta} \mid \beta < \alpha$ and α, β are both rational$\}$. Thus, in order to see that $\mu(A) = 0$ it is enough to show that $\mu(E_{\alpha,\beta}) = 0$ whenever $\beta < \alpha$.

Claim 1: $E_{\alpha,\beta} \subseteq \{f^\star > \alpha\}$ is T-invariant.

Proof: If $x \in E_{\alpha,\beta}$, then $\alpha < \overline{f}(x) \leq f^\star(x)$, hence $x \in \{f^\star > \alpha\}$. Furthermore, for all $x \in X$ we have that $x \in T^{-1}(E_{\alpha,\beta})$ iff $Tx \in E_{\alpha,\beta}$ iff $\underline{f}(Tx) < \beta < \alpha < \overline{f}(Tx)$ iff $\underline{f}(x) < \beta < \alpha < \overline{f}(x)$ iff $x \in E_{\alpha,\beta}$ by (5.8). □

We apply Corollary 5.1.4 to conclude that

$$\int_{E_{\alpha,\beta}} f\, d\mu = \int_{E_{\alpha,\beta}\cap\{f^\star>\alpha\}} f \geq \alpha\mu(E_{\alpha,\beta} \cap \{f^\star > \alpha\}) = \alpha\mu(E_{\alpha,\beta}). \tag{5.11}$$

Claim 2: $E_{\alpha,\beta} \subseteq \{(-f)^\star > -\beta\}$

Proof: If $x \in E_{\alpha,\beta}$, then $\underline{f}(x) = \liminf_{n\to\infty} S_n f(x) < \beta$, so there exists $n \geq 1$ such that $S_n f(x) < \beta$. We get that

$$-\beta < -S_n f(x) = S_n(-f)(x) \leq (-f)^\star(x). \quad \square$$

We apply again Corollary 5.1.4 for $-f$ and β to conclude that

$$\int_{E_{\alpha,\beta}} (-f)\, d\mu = \geq -\beta\mu(E_{\alpha,\beta}),$$

hence

$$\int_{E_{\alpha,\beta}} f \leq \beta\mu(E_{\alpha,\beta}).. \tag{5.12}$$

We conclude from (5.11) and (5.12) that $\alpha\mu(E_{\alpha,\beta}) \leq \beta\mu(E_{\alpha,\beta})$. Since $\beta < \alpha$, we must have $\mu(E_{\alpha,\beta})$. This gives $\underline{f} = \overline{f}$ a.e. as explained above.

Therefore $S_n f$ converges a.e. to $f^+ := \underline{f}$. Furthermore, $f^+ \circ T = f^+$ a.e., by (5.8). In general, if $f = g + ih : X \to \mathbb{C}$, then $S_n f$ converges a.e. to $f^+ := g^+ + ih^+ = \underline{g} + i\underline{h}$.

(ii) f^+ is measurable, as the limit of a sequence of measurable mappings. Let

$$g_n, h_n : X \to [0, +\infty), \quad g_n(x) := |S_n f(x)|, \quad h_n(x) := S_n(|f|)(x).$$

Then $\lim\limits_{n\to\infty} g_n(x) = \lim\limits_{n\to\infty} |S_n f(x)| = |f^+(x)|$ a.e.. Since $f \in L^1(X, \mathcal{B}, \mu)$, we have that $|f| \in L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$, so we can apply (i) for $|f|$ to conclude that $\lim\limits_{n\to\infty} h_n = |f|^+$ a.e.. Since obviously $0 \leq g_n \leq h_n$ for all $n \geq 1$, we get that

$$|f^+| \leq |f|^+ \text{ a.e.}. \tag{5.13}$$

It follows that

$$\begin{aligned}
\int_X |f^+|\, d\mu &\leq \int_X |f|^+\, d\mu = \int_X \lim_{n\to\infty} h_n\, d\mu = \int_X \liminf_n h_n\, d\mu && (5.14)\\
&\leq \liminf_n \int_X h_n\, d\mu \quad \text{by Fatou's Lemma} && (5.15)\\
&= \liminf_n \int_X |f|\, d\mu \quad \text{by Proposition 5.0.3.(vii)} && (5.16)\\
&= \int_X |f|\, d\mu = \|f\|_1 < \infty, \quad \text{since } f \in L^1(X, \mathcal{B}, \mu). && (5.17)
\end{aligned}$$

Thus, $f^+ \in L^1(X, \mathcal{B}, \mu)$ and $\|f^+\|_1 \leq \|f\|_1$.

(iii) Let A be T-invariant and define for each $m \geq 0$ and $k \in \mathbb{Z}$,

$$A_{m,k} = \left\{ x \in A \mid \frac{k}{2^m} \leq f^+(x) < \frac{k+1}{2^m} \right\}. \tag{5.18}$$

It is easy to see that each $A_{m,k}$ is T-invariant. Furthermore, for fixed $m \geq 0$, $(A_{m,k})_{k\in\mathbb{Z}}$ is a countable family of pairwise disjoint sets satisfying $A = \bigcup\limits_{k\in\mathbb{Z}} A_{m,k}$.

Let $\varepsilon > 0$, $m \geq 0, k \in \mathbb{Z}$. If $x \in A_{m,k}$ then $\dfrac{k}{2^m} - \varepsilon < \dfrac{k}{2^m} \leq f^+(x) = \underline{f}(x) = \liminf_{n\to\infty} S_n f(x)$, so $S_n f(x) > \frac{k}{2^m} - \varepsilon$ for all n from some N on. It follows that $f^\star(x) = \sup_{n\geq 1} S_n f(x) > \dfrac{k}{2^m} - \varepsilon$. Thus, we have proved that

$$A_{m,k} \subseteq \left\{ f^\star > \frac{k}{2^m} - \varepsilon \right\}. \tag{5.19}$$

We apply Corollary 5.1.4 to conclude that for all $\varepsilon > 0$,

$$\int_{A_{m,k}} f\, d\mu \geq \left(\frac{k}{2^m} - \varepsilon \right) \mu(A_{m,k}).$$

Hence, by letting $\varepsilon \to 0$, we have that

$$\int_{A_{m,k}} f\, d\mu \geq \frac{k}{2^m} \mu(A_{m,k}). \tag{5.20}$$

Then

$$\int_{A_{m,k}} f^+\, d\mu \quad \leq \quad \frac{k+1}{2^m} \mu(A_{m,k}) \leq \frac{1}{2^m} \mu(A_{m,k}) + \int_{A_{m,k}} f\, d\mu.$$

Summing over k, we get that for all $m \geq 0$,

$$\begin{aligned} \int_A f^+\, d\mu \quad &= \quad \int_{\cup_{k\in\mathbb{Z}} A_{m,k}} f\, d\mu = \sum_{k\in\mathbb{Z}} \int_{A_{m,k}} f^+\, d\mu \leq \sum_{k\in\mathbb{Z}} \left(\frac{1}{2^m} \mu(A_{m,k}) + \int_{A_{m,k}} f\, d\mu \right) \\ &= \quad \frac{1}{2^m} \sum_{k\in\mathbb{Z}} \mu(A_{m,k}) + \sum_{k\in\mathbb{Z}} \int_{A_{m,k}} f\, d\mu = \frac{\mu(A)}{2^m} + \int_A f\, d\mu. \end{aligned}$$

By letting $m \to \infty$, it follows that

$$\int_A f^+\, d\mu \quad \leq \textstyle\int_A f\, d\mu.$$

Applying the above reasoning to $-f$ instead of f gives

$$\int_A (-f)^+\, d\mu \leq \int_A -f\, d\mu,$$

hence

$$\begin{aligned} \int_A f\, d\mu \quad &\geq \quad -\int_A (-f)^+\, d\mu = -\int_A \underline{(-f)}\, d\mu = -\int_A -\overline{f}\, d\mu = \int_A \overline{f}\, d\mu = \int_A f^+\, d\mu, \\ &\text{since } f^+ = \overline{f} \text{ a.e..} \end{aligned}$$

(iv) Since $U_T(f^+) = f^+ \circ T = f^+$ a.e. and T is ergodic, we can use Theorem 3.5.2 to conclude that $f^+(x) = c$ a.e. for some constant $c \in \mathbb{C}$. By (iii), we get that

$$c = c\mu(X) = \int_X f^+ \, d\mu = \int_X f \, d\mu.$$

□

Let $(X, \mathcal{B}, \mu, T)$ be a MDS and $f \in L^1(X, \mathcal{B}, \mu)$. The **time mean** of f at $x \in X$ is defined to be

$$\lim_{n\to\infty} S_n f(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \quad \text{if the limit exists.} \tag{5.21}$$

The **space mean** or **phase mean** of f is defined to be

$$\int_X f \, d\mu. \tag{5.22}$$

The ergodic theorem implies that for ergodic transformations the space mean is equal almost everywhere with the time mean. This assertion, of great significance in the physical aspects of the theory, is sometimes (incorrectly) identified with the ergodic theorem.

5.3 Mean ergodic theorem

In this section we present the proof given in 1939 by Garrett Birkhoff [17] of the generalization of von Neumann mean ergodic theorem to uniformly convex Banach spaces.

5.3.1 Uniformly convex Banach spaces

Uniformly convex Banach spaces were introduced in 1936 by Clarkson in his seminal paper [20]. A Banach space X is called **uniformly convex** if for all $\varepsilon \in (0, 2]$ there exists $\delta \in (0, 1]$ such that for all $x, y \in X$,

$$\|x\| \leq 1, \quad \|y\| \leq 1 \text{ and } \|x - y\| \geq \varepsilon \text{ imply } \left\| \frac{1}{2}(x + y) \right\| \leq 1 - \delta. \tag{5.23}$$

A mapping $\eta : (0, 2] \to (0, 1]$ providing such a $\delta := \eta(\varepsilon)$ for given $\varepsilon \in (0, 2]$ is called a **modulus of uniform convexity**.

This is a geometric property of the unit ball of the space. If the midpoint of a line segment with endpoints on the surface of the unit ball approaches the surface, then the endpoints must come close together.

An example of a modulus of uniform convexity is Clarkson's **modulus of convexity** [20], defined for any Banach space X as the function $\delta_X : [0, 2] \to [0, 1]$ given by

$$\delta_X(\varepsilon) = \inf \left\{ 1 - \left\| \frac{1}{2}(x + y) \right\| \mid \|x\| \leq 1, \|y\| \leq 1, \|x - y\| \geq \varepsilon \right\}. \tag{5.24}$$

It is easy to see that $\delta_X(0) = 0$ and that δ_X is monotone increasing. Note that for uniformly convex Banach spaces X, δ_X is the largest modulus of uniform convexity. Furthermore, a Banach space X is uniformly convex if and only if $\delta_X(\varepsilon) > 0$ for $\varepsilon \in (0, 2]$.

Example 5.3.1. (i) Any Hilbert space H is uniformly convex with modulus of convexity $\delta_H(\varepsilon) = 1 - \sqrt{1 - \frac{\varepsilon^2}{4}}$. A modulus of uniform convexity for a Hilbert space H is given by $\eta(\varepsilon) := \varepsilon^2/8$.

(ii) L^p-spaces for $1 < p < \infty$ are uniformly convex.

We refer to [20], [24, Chapter 3] or [61, Chapter 5] for details on uniformly convex Banach spaces and moduli of convexity.

Remark 5.3.2. *Since there are no $x, y \in X$ satisfying condition (5.23) for $\varepsilon > 2$ we can simply extend any modulus of uniform convexity η to all strictly positive real numbers by stipulating $\eta'(\varepsilon) := \eta(\min(2, \varepsilon))$ if η is not already defined for $\varepsilon > 2$. We will make free use of this without further mentioning.*

Lemma 5.3.3. *Let X be a uniformly convex Banach space and η be a modulus of uniform convexity. Define*

$$u_\eta : (0, 2] \to (0, 1], \quad u_\eta(\varepsilon) = \frac{\varepsilon}{2} \cdot \eta(\varepsilon).$$

Then for all $\varepsilon > 0$ and for all $x, y \in X$

$$\|x\| \leq \|y\| \leq 1 \text{ and } \|x - y\| \geq \varepsilon \quad \text{imply} \quad \left\|\frac{1}{2}(x + y)\right\| \leq \|y\| - u_\eta(\varepsilon). \tag{5.25}$$

Proof. Exercise. □

We use the notation u_X for u_{δ_X}, where δ_X is the modulus of convexity.

5.3.2 Mean ergodic theorem in uniformly convex Banach spaces

If X is a Banach space and $T : X \to X$, the **Cesaro average** starting with $x \in X$ is the sequence $(x_n)_{n \geq 1}$ defined by

$$x_n := \frac{1}{n} \sum_{i=0}^{n-1} T^i x.$$

Lemma 5.3.4. *Let X be a Banach space, $T : X \to X$ be linear and (x_n) be the Cesaro average starting with x.*

(i) For all $n, k \geq 1$,

$$x_{n+k} = \frac{n}{n+k}x_n + \frac{1}{n+k}\sum_{i=0}^{k-1} T^{n+i}x, \tag{5.26}$$

$$x_{kn} = \frac{1}{k}\sum_{i=0}^{k-1} T^{in}x_n, \tag{5.27}$$

$$x_{2kn} = \frac{1}{k}\sum_{i=0}^{k-1} \frac{1}{2}T^{in}\left(x_n + T^{kn}x_n\right). \tag{5.28}$$

(ii) Assume moreover that T is nonexpansive. Then for all $n, k \geq 1$,

$$\|x_{n+k} - x_n\| \leq \frac{2k\|x\|}{n+k}, \tag{5.29}$$

$$\|x_{kn} - x_n\| \leq \max_{i=0,\ldots,k-1} \|T^{in}x_n - x_n\| \tag{5.30}$$

Proof. Exercise. □

Theorem 5.3.5 (Mean ergodic theorem for uniformly convex Banach spaces)**.** *Let X be a uniformly convex Banach space and $T : X \to X$ be a nonexpansive linear operator. Then for any $x \in X$, the Cesaro average (x_n) converges to a fixed point of T.*

Proof. Let η be a modulus of uniform convexity, $x \in X$ and $b > 0$ be such that $b \geq \|x\|$. We remark that $\|T^n x\| \leq \|x\|$ for all $n \geq 1$, hence

$$\|x_n\| \leq \frac{1}{n}\sum_{i=0}^{n-1} \|T^i x\| \leq \|x\| \leq b$$

for all $n \geq 1$. Let

$$\alpha := \inf_{n \geq 1} \|x_n\|$$

Given $\varepsilon > 0$, define

$$\gamma := \frac{\varepsilon}{16}\eta\left(\frac{\varepsilon}{8b}\right) > 0 \tag{5.31}$$

Then there exists $N \geq 1$ such that

$$\|x_N\| < \alpha + \gamma.$$

Denote for all $k \geq 1$,

$$y_k := \|T^{kN}x_N - x_N\|.$$

Claim: $y_k \leq \frac{\varepsilon}{8}$ **for all** $k \geq 1$**.**

Proof: If $y_k = 0$, then it is obvious, so we can assume in the sequel that $y_k \neq 0$. We get that for all $k \geq 1$

$$\left\|\frac{1}{b}T^{kN}x_N\right\| \leq \left\|\frac{1}{b}x_N\right\| \leq \frac{\|x\|}{b} \leq 1 \text{ and}$$
$$\frac{y_k}{b} = \left\|\frac{1}{b}(T^{kN}x_N - x_N)\right\| \leq \frac{1}{b}(\|T^{kN}x_N\| + \|x_N\|) \leq \frac{2\|x\|}{b} \leq 2.$$

Thus, applying Lemma 5.3.3 (with $x := \frac{1}{b}T^{kN}x_N, y := \frac{1}{b}x_N$, and $\varepsilon := \frac{y_k}{b} = \|x - y\|$), we get that

$$\left\|\frac{1}{2b}(T^{kN}x_N + x_N)\right\| \leq \frac{1}{b}\|x_N\| - u_X\left(\frac{y_k}{b}\right), \tag{5.32}$$

that is

$$\left\|\frac{1}{2}(T^{kN}x_N + x_N)\right\| \leq \|x_N\| - bu_X\left(\frac{y_k}{b}\right) \tag{5.33}$$

for all $k \geq 1$.

Using now (5.28) of Lemma 5.3.4, we obtain

$$\begin{aligned}
\|x_{2kN}\| &= \left\|\frac{1}{k}\sum_{i=0}^{k-1}\frac{1}{2}T^{iN}\left(x_N + T^{kN}x_N\right)\right\| \leq \frac{1}{k}\sum_{i=0}^{k-1}\left\|T^{iN}\left(\frac{1}{2}(x_N + T^{kN}x_N)\right)\right\| \\
&\leq \frac{1}{k}\sum_{i=0}^{k-1}\left\|\frac{1}{2}(x_N + T^{kN}x_N)\right\| = \left\|\frac{1}{2}(x_N + T^{kN}x_N)\right\| \\
&\leq \|x_N\| - bu_X\left(\frac{y_k}{b}\right) \leq \alpha + \gamma - bu_X\left(\frac{y_k}{b}\right).
\end{aligned}$$

On the other hand, $\|x_{2kN}\| \geq \alpha$. Thus we must have that

$$bu_X\left(\frac{y_k}{b}\right) \leq \gamma \quad \text{for all } \ k \geq 1. \tag{5.34}$$

Assume that $y_k > \frac{\varepsilon}{8}$. Then, since δ_X is increasing and $\delta_X \geq \eta$, we get that

$$bu_X\left(\frac{y_k}{b}\right) = b \cdot \frac{y_k}{2b} \cdot \delta_X\left(\frac{y_k}{b}\right) > \frac{\varepsilon}{16}\delta_X\left(\frac{\varepsilon}{8b}\right) \geq \frac{\varepsilon}{16}\eta\left(\frac{\varepsilon}{8b}\right) = \gamma, \tag{5.35}$$

that is a contradiction with (5.34). Hence, we must have $y_k \leq \frac{\varepsilon}{8}$ for all $k \geq 1$. This finishes the proof of the claim. □

Using the claim it follows that for all $m \geq 1$ and $0 \leq i < N$, we get that

$$\|x_{mN+i} - x_N\| \leq \frac{2b}{m} + \frac{\varepsilon}{8}, \tag{5.36}$$

since

$$\begin{aligned} \|x_{mN+i}-x_N\| &\le \|x_{mN+i}-x_{mN}\|+\|x_{mN}-x_N\| \\ &\le \frac{2ib}{mN+i}+\|x_{mN}-x_N\|, \quad \text{by (5.29) and the fact that } \|x\|\le b \\ &< \frac{2b}{m}+\|x_{mN}-x_N\|, \quad \text{since } 0\le i<N \text{ implies} \frac{2i}{mN+i}<\frac{2}{m} \\ &\le \frac{2b}{m}+\max_{j=0,\dots,m-1} y_j, \quad \text{by (5.30)} \\ &\le \frac{2b}{m}+\frac{\varepsilon}{8} \quad \text{by the above claim.} \end{aligned}$$

Let us define

$$M := \left\lceil \frac{16b}{\varepsilon} \right\rceil, \qquad P := MN.$$

For all $j \ge P$, there are $q \ge 0$ and $0 \le i < N$ such that $j - P = Nq + i$. It follows that

$$\begin{aligned} \|x_j - x_P\| &= \|x_{MN+Nq+i}-x_{MN}\| = \|x_{N(M+q)+i}-x_{MN}\| \\ &\le \|x_{N(M+q)+i}-x_N\|+\|x_{MN}-x_N\| \\ &< \frac{2b}{M+q}+\frac{\varepsilon}{8}+\frac{2b}{M}+\frac{\varepsilon}{8} \quad \text{by (5.36) with } m:=M \text{ and } m:=M+q \\ &\le \frac{\varepsilon}{4}+\frac{4b}{M}\le\frac{\varepsilon}{2}, \end{aligned}$$

It follows immediately that for all $j, l \ge P$, we have that

$$\|x_j - x_l\| \le \|x_j - x_P\| + \|x_l - x_P\| < \varepsilon.$$

Hence, (x_n) is a Cauchy sequence. As X is a Banach space, we get that (x_n) converges to some $z \in X$. Since

$$Tx_n = \frac{1}{n}\sum_{i=1}^{n} T^i x = \frac{n+1}{n}x_{n+1} - \frac{1}{n}x,$$

we get that $Tz = \lim_{n\to\infty} Tx_n = z$.

□

As an immediate corollary we get

Theorem 5.3.6 (von Neumann L^p ergodic theorem)**.** *[77, 78]*
Let $(X, \mathcal{B}, \mu, T)$ be a MDS and $f \in L^p(X, \mathcal{B}, \mu)$, where $1 < p < \infty$. Then $S_n f$ converges in $L^p(X, \mathcal{B}, \mu)$ to a function $f^+ \in L^p(X, \mathcal{B}, \mu)$ satisfying $f^+ \circ T = f^+$ a.e..

Proof. $L^p(X, \mathcal{B}, \mu)$ is a uniformly convex Banach space, and $U_T : L^p(X, \mathcal{B}, \mu) \to L^p(X, \mathcal{B}, \mu)$ is a linear nonexpansive mapping. Furthermore, the Cesaro average (x_n) is exactly the ergodic average $S_n f$. So, we can apply Theorem 5.3.5 to conclude that $S_n f$ converges to a function $f^+ \in L^p(X, \mathcal{B}, \mu)$ satisfying $U_T(f^+) = f^+$, i.e. $f^+ \circ T = f^+$ a.e.. □

We refer to [113, Corollaries 1.5(ii), p.36-37] for another proof of the von Neumann L^p ergodic theorem for $1 \leq p < \infty$.

Theorem 5.3.7 (von Neumann mean ergodic theorem)**.** *[77]*
Let $(X, \mathcal{B}, \mu, T)$ be a MDS. and $f \in L^2(X, \mathcal{B}, \mu)$. Then

(i) $S_n f$ converges in $L^2(X, \mathcal{B}, \mu)$ to a function $f^+ \in L^2(X, \mathcal{B}, \mu)$ satisfying $f^+ \circ T = f^+$ a.e..

(ii) If T is ergodic, then f^+ is constant a.e., namely $f^+ = \int_X f\, d\mu$ a.e..

Proof. (i) Take $p = 2$ in Theorem 5.3.6.

(ii) Since T is ergodic, we can use Theorem 3.5.2 to conclude that $f^+(x) = c$ a.e. for some constant $c \in \mathbb{C}$. We get that

$$\begin{aligned} c &= \int_X f^+\, d\mu = \left\langle f^+, \mathbf{1} \right\rangle = \lim_{n\to\infty} \langle S_n f, \mathbf{1}\rangle = \lim_{n\to\infty} \int_X S_n f\, d\mu = \lim_{n\to\infty} \int_X f\, d\mu \\ &\quad \text{by Proposition 5.0.3.(vii)} \\ &= \int_X f\, d\mu. \end{aligned}$$

□

5.3.3 A finitary version of the mean ergodic theorem

Let (a_n) be a sequence in a metric space (X, d). If $\lim_{n\to\infty} a_n = a$, then a function $\gamma : (0, \infty) \to \mathbb{N}$ is called a **rate of convergence** of (a_n) if

$$\forall \varepsilon > 0\, \forall n \geq \gamma(\varepsilon)\, (d(a_n, a) < \varepsilon)\,. \tag{5.37}$$

In [1], Avigad, Gerhardy and Towsner address the issue of finding an effective rate of convergence for the Cesaro average (x_n) in Hilbert spaces. They show that even for the separable Hilbert space L^2 there are simple computable such operators T and computable points $x \in L^2$ such that there is no computable rate of convergence of (x_n). In such a situation the best one can hope for is an effective bound on the following reformulation of the Cauchy property of (x_n) which in logic is called the **Herbrand normal form** of the latter:

$$\forall \varepsilon > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists N \in \mathbb{N}\, \forall i, j \in [N, N + g(N)]\ \ \|x_i - x_j\| < \varepsilon. \tag{5.38}$$

It is trivial to see that (5.38) is implied by the Cauchy property. However, ineffectively, also the converse implication holds.

The mathematical relevance of this reformulation of the Cauchy property was recently pointed out by Terence Tao [107, 108], who also uses the term **metastability**. The statement (5.38) looks easier to prove because (once one fixes the function g) one asks only for the sequence (x_n) to be **metastable** rather than **stable** - that is stable on a finite range $[N, N+g(N)]$ rather than an infinite range $[N, \infty)$.

In Tao's terminology, an effective uniform bound on '$\exists N$' in (5.38) is called a **rate of metastability**.

The following finitary version of Theorem 5.3.5 was obtained in [62] by a logical analysis of its proof.

Theorem 5.3.8. *Assume that X is a uniformly convex Banach space, η is a modulus of uniform convexity and $T : X \to X$ is a nonexpansive linear operator. Let $b > 0$. Then for all $x \in X$ with $\|x\| \le b$,*

$$\forall \varepsilon > 0 \, \forall g : \mathbb{Z}_+ \to \mathbb{Z}_+ \, \exists P \le \Phi(\varepsilon, g, b, \eta) \, \forall i, j \in [P, P+g(P)] \left(\|x_i - x_j\| < \varepsilon \right). \tag{5.39}$$

where (x_n) is the Cesaro average starting with x and

$$\Phi(\varepsilon, g, b, \eta) := M \cdot \tilde{h}^K(1), \tag{5.40}$$

with

$$\begin{array}{l} M := \left\lceil \dfrac{16b}{\varepsilon} \right\rceil, \quad \gamma := \dfrac{\varepsilon}{16}\eta\left(\dfrac{\varepsilon}{8b}\right), \quad K := \left\lceil \dfrac{b}{\gamma} \right\rceil, \\ h, \tilde{h} : \mathbb{Z}_+ \to \mathbb{Z}_+, \; h(n) := 2(Mn + g(Mn)), \quad \tilde{h}(n) := \max\limits_{i \le n} h(i). \end{array}$$

If $\eta(\varepsilon)$ can be written as $\varepsilon \cdot \tilde{\eta}(\varepsilon)$ with $0 < \varepsilon_1 \le \varepsilon_2 \to \tilde{\eta}(\varepsilon_1) \le \tilde{\eta}(\varepsilon_2)$, then we can replace η by $\tilde{\eta}$ and the constant '16' by '8' in the definition of γ in the bound above.

Note that our bound Φ is independent from T and depends on the space X and the starting point $x \in X$ only via the modulus of convexity η and the norm upper bound $b \ge \|x\|$. Moreover, it is easy to see that the bound depends on b and ε only via b/ε.

As an immediate consequence of the above theorem we get in the case of Hilbert spaces the following result.

Corollary 5.3.9. *Assume that X is a Hilbert space and $T : X \to X$ is a nonexpansive linear operator. Let $b > 0$. Then for all $x \in X$ with $\|x\| \le b$,*

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists P \le \Phi(\varepsilon, g, b) \, \forall i, j \in [P, P+g(P)] \left(\|x_i - x_j\| < \varepsilon \right). \tag{5.41}$$

where Φ is defined as above, but with $K := \left\lceil \dfrac{512b^2}{\varepsilon^2} \right\rceil$.

Proof. Use the fact that $\eta(\varepsilon) := \varepsilon^2/8$ is a modulus of convexity satisfying the requirements in the last claim of the theorem. □

By applying a logical analysis of the standard textbook proof of von Neumann mean ergodic theorem, Avigad, Gerhardy and Towsner [1] extracted for Hilbert spaces the following bound

$$\begin{array}{l} \Phi(\varepsilon, g, b) = h^K(1) \text{ where } h(n) = n + 2^{13}\rho^4 \tilde{g}((n+1)\tilde{g}(2n\rho)\rho^2) \text{ with} \\ \rho = \left\lceil \frac{b}{\varepsilon} \right\rceil, \; K = 512\rho^2, \text{ and } \tilde{g}(n) = \max\limits_{i \le n}(i + g(i)). \end{array}$$

Note that the number of iterations K in both this bound and in the bound in Corollary 5.3.9 coincide (disregading the different placement of '$\lceil\cdot\rceil$') whereas the function h being iterated in the bound in Corollary 5.3.9 is much simpler than that occurring in the above bound from [1].

5.4 Directions of research

5.4.1 More general averages

An idea of research is to obtain finitary versions (with effective bounds) of generalizations of the mean ergodic theorem obtained by replacing the Cesaro averages with **weighted averages**.

If X is a Banach space, $T : X \to X$, and $x \in X$, let us define

$$x_n = \sum_{j=1}^{\infty} a_{nj} T^j x, \tag{5.42}$$

where $A = (a_{nj})$ is a regular matrix.

By a **regular matrix** we understand an infinite matrix $A = (a_{nj})$ with the property that for every convergent sequence (z_j) of complex numbers, we have that

$$\lim_{n\to\infty} z_n = z \quad \text{implies} \quad \lim_{n\to\infty} \sum_{j=1}^{\infty} a_{nj} z_j = z.$$

We refer to [82] for details on regular matrices.

In the literature, there are more results stating convergence of the weighted averages for power bounded operators linear operators $T : X \to X$ (see [64, Chapter 8] for a nice exposution).

Direction of research: Try to get finitary versions of these result for nonexpansive linear operators in uniformly convex Banach spaces, by generalizing the proofs of Theorem 5.3.5 and of its finitary version, Theorem 5.3.8.

5.4.2 Multiple commuting transformations

In [108], Tao settles in full generality the norm convergence problem for several commuting transformations.

Theorem 5.4.1. *Let $l \geq 1$ be an integer. Assume that $T_1, \ldots, T_l : X \to X$ are commuting invertible measure-preserving transformations of a probability space $(X, \mathcal{B}, \mu)$. Then for any $f_1, \ldots, f_l \in L^\infty(X, \mathcal{B}, \mu)$, the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T_1^n x) \ldots f_l(T_l^n x)$$

are convergent in $L^2(X, \mathcal{B}, \mu)$.

The case $l = 1$ is essentially the Mean Ergodic Theorem. Tao deduces Theorem 5.4.1 from the following finitary version, proved by *finitary ergodic theory* techniques, inspired by those used by him an Green [48] to establish arbitrary arithmetic progressions in the primes.

Theorem 5.4.2 (Finitary norm convergence)**.** *[108, Theorem 1.6]*
Let $l \geq 1$ be an integer, let $F : \mathbb{Z}_+ \to \mathbb{Z}_+$ be a function, and let $\varepsilon > 0$. Then there exists an integer $M^* > 0$ with the following property: if $P \geq 1$ and $f_1, \dots, f_l : \mathbb{Z}_P^l \to [-1, 1]$ are arbitrary functions on $\mathbb{Z}_P^l$, then there exists an integer $1 \leq M \leq M^*$ such that we have the L^2 ***metastability***

$$\|S_N(f_1, \dots, f_l) - S_{N'}(f_1, \dots, f_l)\|_{L^2(\mathbb{Z}_P^l)} \leq \varepsilon \quad \textit{for all } M \leq N, N' \leq F(M),$$

The finitary version is obtained by replacing the general probability space $(X, \mathcal{B}, \mu)$ with the finite abelian group $\mathbb{Z}_P^l = (Z/p\mathbb{Z})^l$ for some large integer P, with the discrete σ-algebra, and the uniform probability measure $\mu(Y) := \dfrac{|Y|}{|\mathbb{Z}_P^l|}$ for all $Y \subseteq \mathbb{Z}_P^l$. If $e_1, \dots, e_l$ are the standard generators of $\mathbb{Z}_P^l$, we consider the standard l commuting shifts $T_i(x) := x + e_i$.

For any $N \geq 1$ and any functions $f_1, \dots, f_l : \mathbb{Z}_P^l \to \mathbb{R}$, we define the **multiple average**

$$S_N(f_1, \dots, f_l) : \mathbb{Z}_P^l \to \mathbb{R}, \quad S_N(f_1, \dots, f_l)(a) = \mathbf{E}_{n \in [0, N-1]} \prod_{i=1}^{l} f_i(a + ne_i),$$

with $[0, N-1] := \{0, 1, \dots, N-1\}$ and $\mathbf{E}_{n \in [0,N-1]} f(n) = \dfrac{1}{N} \sum_{n=0}^{N-1} f(n)$ for every $f : [0, N - 1] \to \mathbb{R}$.

Example 5.4.3. If $l = 2$ and $f_1, f_2 : \mathbb{Z}_P^2 \to \mathbb{R}$, then

$$S_N(f_1, f_2)(a_1, a_2) = \frac{1}{N} \sum_{n=0}^{N-1} f_1(a_1 + n, a_2) f_2(a_1, a_2 + n) \quad \text{for all } a_1, a_2 \in \mathbb{Z}_P.$$

Assume that $l = 1$. For all $f : \mathbb{Z}_P \to \mathbb{R}$,

$$S_N(f)(a) \;=\; \frac{1}{N} \sum_{n=0}^{N-1} f(a_1 + n) = \frac{1}{N} \sum_{n=0}^{N-1} f(T_1^n(a)),$$

where $T_1(a) = a + 1$ for all $a \in \mathbb{Z}_P$ is an invertible measure-preserving transformation. Thus, $S_N(f)$ is the ergodic average associated to f. Furthermore, the real Hilbert space $L^2(\mathbb{Z}_P)$ is endowed with the inner product

$$\langle f, g \rangle := \mathbb{E}_{a \in \mathbb{Z}_P} f(a) g(a) = \frac{1}{P} \sum_{a \in \mathbb{Z}_P} f(a) g(a). \tag{5.43}$$

Therefore, the $l = 1$ version of Theorem 5.4.2 follows from Corollary5.3.9.

Direction of research: Obtain an explicit uniform bound on M^*, whose ineffective existence is proved by Tao in Theorem 5.4.2.

5.5 Ergodicity again

Proposition 5.5.1. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS. The following are equivalent*

(i) T is ergodic.

(ii) For all $A \in \mathcal{B}$, if $\mu(T^{-1}(A)\Delta A) = 0$ implies $\mu(A) = 0$ or $\mu(A) = 1$.

(iii) For all $A, B \in \mathcal{B}$ such that $\mu(A) > 0$ and $\mu(B) > 0$ there exists $n \geq 1$ such that $\mu(T^{-1}(A) \cap B) > 0$.

Proof. Exercise. See [113, Theorem 1.3, p. 22]. □

Let $A \in \mathcal{B}$. For $x \in X$ we could ask with what frequency do the elements of the orbit $\{x, Tx, T^2x, \ldots\}$ lie in the set A (equivalently, how often the orbit $\{x, Tx, T^2x, \ldots\}$ of x is in A). Since clearly, $T^n x \in A$ iff $\chi_A(T^n x) = 1$, it follows that the number of elements $\{x, Tx, T^2x, \ldots, T^{n-1}x\}$ in A is

$$\big|[0, n-1] \cap \{k \geq 0 \mid T^k x \in A\}\big| = \sum_{k=0}^{n-1} \chi_A(T^k x). \tag{5.44}$$

The relative number of elements of $\{x, Tx, T^2x, \ldots, T^{n-1}x\}$ in A (equivalently the average number of times that the first n points of the orbit of x are in A) is given by

$$\frac{\big|[0, n-1] \cap \{k \geq 0 \mid T^k x \in A\}\big|}{n} = \frac{1}{n}\sum_{k=0}^{n-1} \chi_A(T^k x) = S_n\chi_A(x). \tag{5.45}$$

Theorem 5.5.2. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS. The following are equivalent*

(i) T is ergodic.

(ii) For each $f \in L^1(X, \mathcal{B}, \mu)$, the time mean of f equals the space mean of f, i.e.:

$$\lim_{n\to\infty} S_n f = \int_X f \, d\mu \ \text{a.e.}.$$

(iii) Whenever $f \in L^p(X, \mathcal{B}, \mu)$ for $1 \leq p \leq \infty$,

$$\lim_{n\to\infty} S_n f = \int_X f \, d\mu \ \text{a.e.}.$$

(iv) For all $A \in \mathcal{B}$,

$$\lim_{n\to\infty} \frac{\big|[0, n-1] \cap \{k \geq 0 \mid T^k x \in A\}\big|}{n} = \mu(A) \ \text{almost for all } x \in X.$$

(or, equivalently $\lim_{n\to\infty} S_n\chi_A = \mu(A)$ a.e.)

Proof. (i)⇒(ii) By the Birkhoff Ergodic Theorem.

(ii)⇒(iii) Apply the fact that for $p \geq 1$, $L^p(X, \mathcal{B}, \mu) \subseteq L^1(X, \mathcal{B}, \mu)$.

(iii)⇒(iv) Apply (iii) with $f := \chi_A \in L^p(X, \mathcal{B}, \mu)$.

(iv)⇒(i) Let $A \in \mathcal{B}$ be such that $T^{-1}(A) = A$, hence $T^{-k}(A) = A$ for all $k \geq 1$. Then

$$S_n \chi_A = \frac{1}{n} \sum_{k=0}^{n-1} \chi_{T^{-k}(A)} = \chi_A.$$

By (iv), it follows that

$$\chi_A = \lim_{n \to \infty} S_n \chi_A = \mu(A) \text{ a.e..}$$

Hence, $\mu(A) \in \{0, 1\}$.

□

Theorem 5.5.3. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS and let $\mathcal{S}$ be a semialgebra that generates $\mathcal{B}$. The following are equivalent*

(i) T is ergodic.

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

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) = \mu(A)\mu(B). \tag{5.46}$$

(iii) For all $A, B \in \mathcal{S}$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) = \mu(A)\mu(B). \tag{5.47}$$

Proof. (i)⇒(ii) Assume that T is ergodic and let $A, B \in \mathcal{B}$. By Theorem 5.5.2.(iv), we have that $\lim_{n\to\infty} S_n \chi_A = \mu(A)$ a.e. Multiplying by χ_B gives $\lim_{n\to\infty} \chi_B S_n \chi_A = \mu(A)\chi_B$ a.e.. Since $\mu(A)\chi_B \in L^1(X, \mathcal{B}, \mu)$ and $\chi_B S_n \chi_A \in L^1(X, \mathcal{B}, \mu)$ for all $n \geq 1$, we can apply Lebesgue Dominated Convergence Theorem to conclude that

$$\lim_{n \to \infty} \int_X \chi_B S_n \chi_A \, d\mu = \int_X \mu(A)\chi_B \, d\mu = \mu(A)\mu(B).$$

By Proposition 5.0.4, we have that

$$\int_X \chi_B \cdot S_n \chi_A \, d\mu = \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B).$$

(ii)⇒(i) Let $A \in \mathcal{B}$ be such that $T^{-1}(A) = A$, hence $T^{-i}(A) = A$ for all $i \geq 0$. Applying (ii) with $B := A$ we get that

$$\mu(A)^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(A) = \mu(A).$$

It follows that $\mu(A) \in \{0, 1\}$. Thus, T is ergodic.

(ii) ⇔ *(iii)* Exercise. See [114, Theorem 1.17, p.41]. □

Proposition 5.5.4. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS. The following are equivalent*

(i) T is ergodic.

(ii) For each $f, g \in L^2(X, \mathcal{B}, \mu)$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \left\langle U_T^k f, g \right\rangle = \langle f, \mathbf{1}\rangle\langle \mathbf{1}, g\rangle. \tag{5.48}$$

where $\mathbf{1}$ is the constant function $X \to \mathbb{C}$, $x \mapsto 1$.

Proof. TO WRITE. □

Chapter 6

Density Ramsey Theory

For all Ramsey theorems, one can express (but not always prove) the corresponding density statements. While the main theme of **partition** Ramsey theory is to look for nontrivial patterns in one cell of an arbitrary finite partition, the typical **density** Ramsey theory statement concerns an appropriately defined notion of largeness: any large subset of a highly organized structure contains large, highly organized substructures. One basic property required from the notion of largeness is the following:

If A is large and $A = \bigcup_{i=1}^{r} C_i$, then at least one of the C_i is large.

6.1 Different notions of density

Definition 6.1.1. *Let $A \subseteq \mathbb{Z}$. For any real number x, let $A(x)$ denote the number of positive elements of A not exceeding x, that is*

$$A(x) = \sum_{a \in A,\, 1 \leq a \leq x} 1 = |A \cap [1, x]|. \tag{6.1}$$

The function $A : \mathbb{R} \to \mathbb{N}$ is called the ***counting function*** *on the set A.*

Let us recall that the **prime counting function**, denoted by $\pi(x)$, gives the number of primes less than or equal to a given real number x. Thus, $A(x) = \pi(x)$ if $A \subseteq \mathbb{Z}_+$ is the set of prime numbers.

Remark 6.1.2. *(i) For $x < 1$, $A(x) = 0$, as $[1, x] = \emptyset$.*

(ii) For $x \geq 1$, $A(x) = A([x]) \leq [x]$, hence $0 \leq \frac{A(x)}{x} \leq 1$.

Lemma 6.1.3. *Let $A, B \subseteq \mathbb{Z}$.*

(i) $(A \cup B)(n) \leq A(n) + B(n)$, with equality if $A \cap B = \emptyset$.

(ii) If A is finite, then A is constant eventually, i.e.: $A(n) = A(N)$ for all $n \geq N$, where $N = \max A$.

(iii) *If* $A = \{a+id \mid i \geq 0\}, a, d \in \mathbb{Z}_+$ *is an arithmetic progression, then* $A(n) = \left[\frac{n-a}{d}\right]+1$.

(iv) *If* $A \subseteq \mathbb{Z}_+$ *is the set of even integers, then* $A(n) = \left[\frac{n}{2}\right]$.

(v) *If* $A \subseteq \mathbb{Z}_+$ *is the set of odd integers, then* $A(n) = \left[\frac{n+1}{2}\right]$.

Proof. Exercise. □

6.1.1 Shnirel'man density

Shnirel'man density is an important additive measure of the size of a set of integers. We refer to [76, Ch.7] for details.

Definition 6.1.4. *The* ***Shnirel'man density*** *of the set* $A \subseteq \mathbb{Z}$*, denoted by* $\sigma(A)$*, is defined by*

$$\sigma(A) := \inf_{n\in\mathbb{Z}_+} \frac{A(n)}{n} = \inf_{n\in\mathbb{Z}_+} \frac{|A \cap [1,n]|}{n}. \tag{6.2}$$

Lemma 6.1.5. *Let* $A \subseteq \mathbb{Z}$.

(i) $0 \leq \sigma(A) \leq 1$ *and* $\sigma(A) = \sigma(A \cap \mathbb{Z}_+)$.

(ii) *If* $m \notin A$ *for some* $m \in \mathbb{Z}_+$*, then* $\sigma(A) \leq 1 - \frac{1}{m} < 1$.

(iii) $\sigma(A) = 1$ *if and only if* A *contains* $\mathbb{Z}_+$.

(iv) *If* A *is finite, then* $\sigma(A) = 0$.

(v) *If* $A \subseteq \mathbb{Z}_+$ *is the set of even integers, then* $\sigma(A) = 0$.

(vi) *If* $A \subseteq \mathbb{Z}_+$ *is the set of odd integers, then* $\sigma(A) = \dfrac{1}{2}$.

Proof. Exercise. □

From (ii) of the above lemma, we conclude that the Shnirel'man density is sensitive to the first values of a set. In particular, if $1 \notin A$, then $\sigma(A) = 0$, while if $2 \notin A$, then $\sigma(A) \leq 1/2$.

6.1.2 Asymptotic density

Definition 6.1.6. *Let* $A \subseteq \mathbb{Z}$.

(i) *The* ***upper (asymptotic) density*** *of* A*, denoted by* $\overline{d}(A)$*, is defined by*

$$\overline{d}(A) = \limsup_{n\to\infty} \frac{A(n)}{n} = \limsup_{n\to\infty} \frac{|A \cap [1,n]|}{n}. \tag{6.3}$$

(ii) The ***lower (asymptotic) density*** *of* A*, denoted by* $\underline{d}(A)$*, is defined by*

$$\underline{d}(A) = \liminf_{n\to\infty} \frac{A(n)}{n} = \liminf_{n\to\infty} \frac{|A \cap [1,n]|}{n}. \tag{6.4}$$

(iii) If $\overline{d}(A) = \underline{d}(A)$*, then the common value is denoted by* $d(A)$ *and it is called the* ***(asymptotic) density*** *of* A*.*

The asymptotic density is also called **natural density**. For convenience, we have defined density in terms of the interval $[1, n]$ rather than $[-n, n]$ in order to deal simultaneously with subsets of $\mathbb{Z}_+$ and $\mathbb{Z}$.

It is obvious that for all $A \subseteq \mathbb{Z}$,

$$0 \le \sigma(A) \le \underline{d}(A) \le \overline{d}(A) \le 1. \tag{6.5}$$

Lemma 6.1.7. *Let* $A \subseteq \mathbb{Z}$*.*

(i) If A *contains* $\mathbb{Z}_+$*, then* $d(A) = 1$*.*

(ii) Finite sets have density 0*.*

(iii) If $A = \{a + id \mid i \ge 0\}, a, d \in \mathbb{Z}_+$ *is an arithmetic progression, then* $d(A) = \frac{1}{d}$*. In particular, the set of positive odd (resp. odd) integers has density* $\frac{1}{2}$*.*

(iv) The set $A \subseteq \mathbb{Z}_+$ *of prime numbers has density* 0*.*

Proof. Exercise. □

Proposition 6.1.8.

(i) If $A, B \subseteq \mathbb{Z}$*, then* $\overline{d}(A \cup B) \le \overline{d}(A) + \overline{d}(B)$*.*

(ii) If $A, B \subseteq \mathbb{Z}$ *are such that* $\overline{d}(A) > 0$ *and* $\overline{d}(B) = 0$*, then* $\overline{d}(A \setminus B) > 0$*.*

(iii) $\overline{d}$ *is not countably subadditive:* $\overline{d}(\{n\}) = 0$ *for all* $n \in \mathbb{Z}_+$*, while* $\overline{d}(\mathbb{Z}_+) = 1$*.*

(iv) If $\overline{d}(A) > 0$ *and* $A = \bigcup_{i=1}^r C_i$*, then there exists* $i \in [1, r]$ *such that* $\overline{d}(C_i) \ge \dfrac{\overline{d}(A)}{r} > 0$*.*

Proof. (i) Exercise.
(ii) We have that $0 < \overline{d}(A) \subseteq \overline{d}(B) + \overline{d}(A \setminus B)$.
(iii) Exercise. □

Proposition 6.1.9. *Any syndetic set* $A \subseteq \mathbb{Z}_+$ *has positive lower density.*

Proof. Exercise. □

6.1.3 Banach density

It is the notion of positive upper Banach density which naturally appears in many questions and results of density Ramsey theory.

Definition 6.1.10. *Let $A \subseteq \mathbb{Z}_+$ (resp. $A \subseteq \mathbb{Z}$). The* ***upper Banach density*** *of A is defined by*

$$Bd^\star(A) := \limsup_{|I|\to\infty} \frac{|A\cap I|}{|I|}, \quad \text{where } I \text{ ranges over intervals of } \mathbb{Z}_+ \text{ (resp. of } \mathbb{Z}). \tag{6.6}$$

That is, there exists a sequence of intervals $(I_k)_{k\geq 1}$ in $\mathbb{Z}_+$ (resp. in $\mathbb{Z}$) such that $\lim_{k\to\infty} |I_k| = \infty$ and

$$\lim_{k\to\infty} \frac{|A\cap I_k|}{|I_k|} = Bd^\star(A), \tag{6.7}$$

and for any other sequence $(J_k)_{k\geq 1}$ in $\mathbb{Z}_+$ (resp. in $\mathbb{Z}$) with $\lim_{k\to\infty} |J_k| = \infty$,

$$\limsup_{k\to\infty} \frac{|A\cap J_k|}{|J_k|} \leq Bd^\star(A). \tag{6.8}$$

The **lower Banach density** of A is defined similarly and is denoted by $Bd_\star(A)$. If $Bd^\star(A) = Bd_\star(A)$, then the common value is denoted by $Bd(A)$ and it is called the **Banach density** of A. This, however, is a much rarer phenomenon than the equality of upper and lower density.

We have obviously that

$$0 \leq \overline{d}(A) \leq Bd^\star(A) \leq 1, \quad \text{and } 0 \leq Bd_\star(A) \leq \underline{d}(A) \leq 1. \tag{6.9}$$

Lemma 6.1.11. *(i) If A contains $\mathbb{Z}_+$, then $Bd^\star(A) = 1$.*

(ii) Finite sets have Banach density 0.

(iii) If $A, B \subseteq \mathbb{Z}$, then $Bd^\star(A\cup B) \leq Bd^\star(A) + Bd^\star(B)$.

(iv) $Bd^\star$ is not countably sub-additive: $Bd^\star(\{n\}) = 0$ for all $n \in \mathbb{Z}_+$, while $Bd^\star(\mathbb{Z}_+) = 1$.

(v) If $A = \{a + id \mid i \geq 0\}, a, d \in \mathbb{Z}_+$ is an arithmetic progession, then $Bd(A) = \frac{1}{d}$. In particular, the set of positive even (resp. odd) integers has Banach density $\frac{1}{2}$.

(vi) If $Bd^\star d(A) > 0$ and $A = \bigcup_{i=1}^r C_i$, then there exists $i \in [1, r]$ such that $Bd^\star(C_i) \geq \dfrac{Bd^\star(A)}{r} > 0$.

Proof. Exercise. □

6.2 A notion of density convergence

Let (x_n) be a sequence in $\mathbb{C}$ and $x \in \mathbb{C}$.

Definition 6.2.1. *If there exists a set $E \subseteq \mathbb{Z}^+$ of zero density such that $\lim_{\substack{n\to\infty \\ n\notin E}} a_n = 0$, then we say that (x_n) D-**converges** to x and we write $D-\lim_{n\to\infty} x_n = x$.*

Lemma 6.2.2. *Consider a bounded sequence $(a_n)_{n\geq 1}$ of real numbers. The following are equivalent*

(i) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0.$

(ii) $D-\lim_{n\to\infty} x_n = x.$

(iii) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} a_i^2 = 0.$

Proof. See [113, Theorem 1.8, p.40]. □

Proposition 6.2.3. *(i) Let $(a_n), (b_n)$ be sequences in $\mathbb{C}$. If $D-\lim_{n\to\infty} a_n = 0$ and (b_n) is bounded, then $D-\lim_{n\to\infty}(a_n b_n) = 0$.*

Proof. Easy exercise. □

6.3 Furstenberg Correspondence Principle

Theorem 6.3.1. *Let $A \subseteq \mathbb{Z}$ with $Bd^\star(A) > 0$. Then there exist an invertible MDS $(X, \mathcal{B}, \mu, T)$ (in which X is a compact metric space and T is a homeomorphism from X onto X) and a set $A' \in \mathcal{B}$ such that*

(1) $\mu(A') = Bd^\star(A)$ *and*

(2) For every finite subset $F \subseteq \mathbb{Z}$,

$$Bd^\star\left(\bigcap_{n\in F}(-n+A)\right) \geq \mu\left(\bigcap_{n\in F} T^{-n}(A')\right). \tag{6.10}$$

Proof. **Step 1** Let $W = \{0,1\}$ and consider the Bernoulli shift $(W^{\mathbb{Z}}, \mathcal{B}(W^{\mathbb{Z}}), \mu, T)$. The full 2-shift $(W^{\mathbb{Z}}, T)$ is an invertible TDS with $W^{\mathbb{Z}}$ being a compact metric space.

Let $\chi_A : \mathbb{Z} \to \{0,1\}$ be the characteristic function of A,

$$X := \overline{\{T^n \chi_A \mid n \in \mathbb{Z}\}} \tag{6.11}$$

be the orbit closure of χ_A, and consider the invertible subsystem (X, T_X) of the full shift $(W^{\mathbb{Z}}, T)$. We have that X is a compact metric space and $T_X : X \to X$, denoted also by T in the sequel, is a homeomorphism. If $\mathcal{B} := X \cap \mathcal{B}(W^{\mathbb{Z}})$, then $\mathcal{B}$ is the Borel σ-algebra on X, by C.5.6.

Let $\mathcal{A} = \mathcal{A}(\mathcal{C}_e)$ be the algebra on $W^{\mathbb{Z}}$ generated by the elementary cylinders, and $\mathcal{A}_X := X \cap \mathcal{A}$. Since elementary cylinders are clopen subsets of $W^{\mathbb{Z}}$ by Proposition 1.2.8.(iii), we can apply C.1.6 to conclude that the elements of $\mathcal{A}$ are clopen subsets of X. As a consequence, the elements of $\mathcal{A}_X$ are clopen subsets of X.

Since any elementary cylinder in $W^{\mathbb{Z}}$ is of the form C_n^0 or $C_n^1 = W^{\mathbb{Z}} \setminus C_n^0$, we get that $\mathcal{A} = \mathcal{A}(\mathcal{C}_e) = \mathcal{A}(\mathcal{C}_e^1)$ and $\mathcal{B}(W^{\mathbb{Z}}) = \sigma(\mathcal{C}_e) = \sigma(\mathcal{C}_e^1)$, where $\mathcal{C}_e^1 = \{C_n^1 \mid n \in \mathbb{Z}\}$.

Let

$$D_n \quad = \quad X \cap C_n^1 \text{ for } n \in \mathbb{Z} \qquad \text{and } \mathcal{D} := \{D_n \mid n \in \mathbb{Z}\} = X \cap \mathcal{C}_e^1$$

Applying C.1.8 and C.2.5, we get the following relations:

$$\mathcal{B} \quad = \quad X \cap \sigma(\mathcal{C}_e^1) = \sigma_X(X \cap \mathcal{C}_e^1) = \sigma_X(\mathcal{D}), \tag{6.12}$$

$$\mathcal{A}_X \quad = \quad X \cap \mathcal{A}(\mathcal{C}_e^1) = \mathcal{A}_X(X \cap \mathcal{C}_e^1) = \mathcal{A}_X(\mathcal{D}), \tag{6.13}$$

$$\sigma_X(\mathcal{A}_X) \quad = \quad \sigma_X(\mathcal{D}) = \mathcal{B}. \tag{6.14}$$

Furthermore, $T(D_n) = D_{n-1}$ and $T^{-1}(D_n) = D_{n+1}$ for all $n \in \mathbb{Z}$, so $T(\mathcal{D}) \subseteq \mathcal{D}$ and $T^{-1}(\mathcal{D}) \subseteq \mathcal{D}$. Since $\mathcal{B} = \sigma_X(\mathcal{D})$, we get by C.7.2 that both T and T^{-1} are measurable, by C.7.2.

Thus, we have already defined a measurable space $(X, \mathcal{B})$ with X compact metric space and a homeomorphism $T : X \to X$ such that T and T^{-1} are measurable. It remains to define a measure on $\mathcal{B}$.

Step 2 Let us define

$$\varphi : \mathcal{P}(X) \to \mathcal{P}(\mathbb{Z}), \quad \varphi(B) = \{n \in \mathbb{Z} \mid T^n \chi_A \in B\} = \{n \in \mathbb{Z} \mid \chi_A \in T^{-n}(B)\}.$$

Then φ has the following easy to verify properties:

(i) $\varphi(\emptyset) = \emptyset$ and $\varphi(X) = \mathbb{Z}$.

(ii) $B \subseteq C$ implies $\varphi(B) \subseteq \varphi(C)$.

(iii) $\varphi(\cup_{i \in I} B_i) = \bigcup_{i \in I} \varphi(B_i)$

(iv) $\varphi(\cap_{i \in I} B_i) = \bigcap_{i \in I} \varphi(B_i)$.

(v) If B and C are disjoint, then $\varphi(B) \cap \varphi(C) = \emptyset$.

Furthermore, for all $n \in \mathbb{Z}$

$$\varphi(D_n) = -n + A, \tag{6.15}$$

since

$$\begin{aligned}\varphi(D_n) &= \{k \in \mathbb{Z} \mid \chi_A \in T^{-k}(D_n)\} = \{k \in \mathbb{Z} \mid \chi_A \in D_{n+k}\} \\ &= \{k \in \mathbb{Z} \mid \chi_A(n+k) = 1\} = \{k \in \mathbb{Z} \mid n+k \in A\} \\ &= -n + A,\end{aligned}$$

and for all $B \subseteq X$,

$$\varphi(T^{-1}(B)) = -1 + \varphi(B), \tag{6.16}$$

since

$$\begin{aligned}\varphi(T^{-1}(B)) &= \{n \in \mathbb{Z} \mid T^n\chi_A \in T^{-1}(B)\} = \{n \in \mathbb{Z} \mid T^{n+1}\chi_A \in B\} \\ &= -1 + \{m \in \mathbb{Z} \mid T^m\chi_A \in B\} = -1 + \varphi(B).\end{aligned}$$

Since $Bd^\star(A) > 0$, we get a sequence $(I_k)_{k\in\mathbb{Z}_+}$ of intervals in $\mathbb{Z}$ such that

$$\lim_{k\to\infty} |I_k| = \infty \text{ and } \lim_{k\to\infty} \frac{|A \cap I_k|}{|I_k|} = Bd^\star(A). \tag{6.17}$$

Pick any non-principal ultrafilter $p \in \beta\mathbb{Z}_+$ and define

$$\nu : \mathcal{P}(X) \to [0,1], \quad \nu(B) = p\!-\!\lim \frac{|\varphi(B) \cap I_k|}{|I_k|}, \tag{6.18}$$

Since $0 \le x_k := \frac{|\varphi(B)\cap I_k|}{|I_k|} \le 1$ for all $k \in \mathbb{Z}_+$, we can apply Proposition 2.2.27.(i) to conclude that (x_k) has a unique p-limit satisfying $0 \le p\!-\!\lim x_k \le 1$. Thus, ν is well-defined. Furthermore,

$$\begin{aligned}\nu(X) &= p\!-\!\lim \frac{|\varphi(X) \cap I_k|}{|I_k|} = p\!-\!\lim \frac{|\mathbb{Z} \cap I_k|}{|I_k|} = p\!-\!\lim 1 &(6.19)\\ &= 1, \quad \text{by Proposition 2.2.26.(ii).} &(6.20)\end{aligned}$$

Claim 1 ν is T-invariant, i.e. $\nu\left(T^{-1}(B)\right) = \nu(B)$ for all $B \subseteq X$.

Proof: Let $B \subseteq X$. We have that

$$\nu\left(T^{-1}(B)\right) = p\!-\!\lim \frac{|\varphi\left(T^{-1}(B)\right) \cap I_k|}{|I_k|} = p\!-\!\lim \frac{|(-1+\varphi(B)) \cap I_k|}{|I_k|}$$

Let $y_k := \dfrac{|\varphi(B) \cap I_k|}{|I_k|}$ and $z_k := \dfrac{|(-1+\varphi(B)) \cap I_k|}{|I_k|}$. Since $|\varphi(B)\cap I_k|$ and $|(-1+\varphi(B))\cap I_k|$ differ at most by 1, it follows that $|x_k - y_k| \le \frac{1}{|I_k|}$ for all $k \in \mathbb{Z}_+$. Hence, by Proposition 2.2.27.(iv), we have that $p\!-\!\lim y_k = p\!-\!\lim z_k$. It follows that $\nu\left(T^{-1}(B)\right) = \nu(B)$. □

Claim 2 ν is finitely additive.

Proof: Let $B, C \subseteq X$ be such that $B \cap C = \emptyset$.

$$\begin{aligned}
\nu(B \cup C) &= p\text{-}\lim \frac{|\varphi(B \cup C) \cap I_k|}{|I_k|} = p\text{-}\lim \frac{|(\varphi(B) \cup \varphi(C)) \cap I_k|}{|I_k|} \\
&= p\text{-}\lim \frac{|(\varphi(B) \cap I_k) \cup (\varphi(C) \cap I_k)|}{|I_k|} = p\text{-}\lim \frac{|\varphi(B) \cap I_k| + |\varphi(C) \cap I_k|}{|I_k|} \\
&\quad \text{since } B \cap C = \emptyset \text{ implies } \varphi(B) \cap \varphi(C) = \emptyset \\
&= p\text{-}\lim \frac{|\varphi(B) \cap I_k|}{|I_k|} + p\text{-}\lim \frac{|\varphi(C) \cap I_k|}{|I_k|}, \quad \text{by Proposition 2.2.27.(iii)} \\
&= \nu(B) + \nu(C). \quad \square
\end{aligned}$$

Step 3 Let $\nu : \mathcal{A}_X \to [0, 1]$ be the restriction of ν to the algebra $\mathcal{A}_X$ on X. Then $\nu(X) = 1$, ν is finitely additive and $\mathcal{A}_X$ is an algebra of clopen sets of the compact metric space X. Hence, we can apply C.5.7 to conclude that ν is countably additive.

Apply now C.6.2 to conclude that there is a unique extension of ν to a probability measure μ on the σ-algebra $\sigma_X(\mathcal{A}_X) = \mathcal{B}$ To conclude that $(X, \mathcal{B}, \mu, T)$ is an invertible MDS, it is enough, by Lemma 3.0.11, to show that T is measure-preserving. Since ν is T-invariant and $\nu|_{\mathcal{A}_X} = \mu|_{\mathcal{A}_X}$, it follows that $\mu(T^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{A}_X$. Apply now Proposition 3.0.12 and the fact that $\mathcal{B} = \sigma_X(\mathcal{A}_X)$ to get that T is measure-preserving.

Step 4 It remains to find $A' \in \mathcal{B}$ satisfying the conclusion of the theorem. Let

$$A' := D_0 = \{\mathbf{x} \in X \mid x_0 = 1\}. \tag{6.21}$$

Then $A' \in \mathcal{A}_X$,

$$T^{-n}(A') = D_n, \quad \varphi(A') = A, \text{ by (6.15)}.$$

It follows that

$$\begin{aligned}
\mu(A') &= \nu(A') = p\text{-}\lim \frac{|\varphi(A') \cap I_k|}{|I_k|} = p\text{-}\lim \frac{|A \cap I_k|}{|I_k|} \\
&= Bd^\star(A), \quad \text{since } \lim_{k\to\infty} \frac{|A \cap I_k|}{|I_k|} = Bd^\star(A).
\end{aligned}$$

We prove finally (6.10). Let $F \subseteq \mathbb{Z}$ be finite and let $C := \bigcap_{n\in F}(-n + A)$. It follows that

$$\begin{aligned}
\mu\left(\bigcap_{n\in F} T^{-n}(A')\right) &= \mu\left(\bigcap_{n\in F} D_n\right) = \nu\left(\bigcap_{n\in F} D_n\right) \quad \text{since } \bigcap_{n\in F} D_n \in \mathcal{A}_X \\
&= p\text{-}\lim \frac{|\varphi(\bigcap_{n\in F} D_n) \cap I_k|}{|I_k|} = p\text{-}\lim \frac{|\bigcap_{n\in F} \varphi(D_n) \cap I_k|}{|I_k|} \\
&= p\text{-}\lim \frac{|\bigcap_{n\in F}(-n + A) \cap I_k|}{|I_k|} = p\text{-}\lim \frac{|\bigcap_{n\in F}(-n + A) \cap I_k|}{|I_k|} \\
&= p\text{-}\lim \frac{|C \cap I_k|}{|I_k|} \\
&\le Bd^\star(C), \quad \text{by the definition of the upper Banach density.}
\end{aligned}$$

□

Chapter 7

Mixing

In the sequel, $(X, \mathcal{B}, \mu, T)$ is a MDS.

Definition 7.0.2. *(i) T is **weak-mixing** if for all $A, B \in \mathcal{B}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \big|\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)\big| = 0. \tag{7.1}$$

*(ii) T is **(strong) mixing** if for all $A, B \in \mathcal{B}$,*

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B). \tag{7.2}$$

If T is weak (strong) mixing, we also say that the MDS $(X, \mathcal{B}, \mu, T)$ is **weak (strong) mixing**.

By Theorem 5.5.3, T is ergodic if and only if for all $A, B \in \mathcal{B}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) = \mu(A)\mu(B).$$

It is well known that for any sequence (a_n) of real numbers, we have that

$$\lim_{n\to\infty} a_n = 0 \Rightarrow \lim \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0 \Rightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} a_i = 0. \tag{7.3}$$

It follows that

$$T \text{ strong mixing} \Rightarrow T \text{ weak-mixing} \Rightarrow T \text{ ergodic}.$$

The following result gives a way of checking the mixing properties for examples by reducing the computations to a class of sets we can manipulate with. For example, it implies we need only consider cylinder sets when dealing with the mixing properties of shifts.

Proposition 7.0.3.
Let $(X, \mathcal{B}, \mu, T)$ be a MDS and let $\mathcal{S}$ be a semialgebra that generates $\mathcal{B}$. Then

(i) T is weak-mixing if and only if for all $A, B \in \mathcal{S}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left|\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)\right| = 0. \tag{7.4}$$

(ii) T is strong mixing if and only if for all $A, B \in \mathcal{S}$,

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B). \tag{7.5}$$

Proof. See [114, Theorem 1.17, p.41] or [83, Proposition 5.3, p. 58]. □

Example 7.0.4. (i) The Bernoulli shift is strong mixing.

(ii) No rotation on a compact group is weak mixing.

Proof. (i) Exercise.

(ii) See [113, Examples (4), p.48].

□

We refer to [83, Section 4.5] for examples of MDSs which are weakly mixing but not strong mixing.

Proposition 7.0.5. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS. If T is weak mixing, then T^n is weak mixing for all $n \geq 1$.*

Proof. Exercise. □

7.1 Density characterizations of weak mixing

Proposition 7.1.1. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS. The following are equivalent:*

(i) T is weak mixing

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

$$D\!-\!\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

(iii) For all $A, B \in \mathcal{B}$,

$$\frac{1}{n} \sum_{i=0}^{n-1} \left|\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)\right|^2 = 0.$$

Proof. Apply Proposition 6.2.2 for $a_n := \mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)$. □

Proposition 7.1.2. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS. The following are equivalent:*

(i) T is weak mixing.

(ii) For all $f, g \in L^2(X, \mathcal{B}, \mu)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \langle U_T^i f, g\rangle - \langle f, \mathbf{1}\rangle\langle \mathbf{1}, g\rangle \right| = 0.$$

(iii) For all $f, g \in L^2(X, \mathcal{B}, \mu)$,

$$D\!-\lim_{n\to\infty} \langle U_T^n f, g\rangle = \langle f, \mathbf{1}\rangle\langle \mathbf{1}, g\rangle.$$

Proof. (i)⇔(ii) See [113, Theorem 1.9, p. 42].
(ii)⇔(iii) Apply Proposition 6.2.2 for $a_n := \langle U_T^n f, g\rangle - \langle f, \mathbf{1}\rangle\langle \mathbf{1}, g\rangle$. □

7.2 A generalization of the von Neumann mean ergodic theorem

The main tool is an abstract version of the classical van der Corput's difference theorem (see, e.g., [65, p.25-26].

Lemma 7.2.1 (van der Corput Lemma)**.** *Let (x_n) be a bounded sequence in a Hilbert space H. If*

$$D\!-\lim_{h\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle x_n, x_{n+h}\rangle = 0,$$

then $\displaystyle\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} x_n \right\| = 0.$

Proof. TO WRITE. □

Theorem 7.2.2 (Weak mixing of order k)**.**
Let $(X, \mathcal{B}, \mu, T)$ be a weak mixing MDS. For all $f_1 \ldots, f_k \in L^\infty(X)$,

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{k} U_T^{in} f_i - \prod_{i=1}^{k} \int_X f_i \, d\mu \right\|_2 = 0.$$

Proof. The proof is by induction on k. When $k=1$ the result follows from von Neumann ergodic Theorem 5.3.7, since any weak mixing transformation is ergodic.

Suppose that the result has been established for k functions.

Step 1: Assume that $\int_X f_i\,d\mu = 0$ for some $i=1,\ldots,k+1$. Let $x_n := \prod_{i=1}^{k+1} U_T^{in} f_i$. We have to show that $\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N} x_n\right\|_2 = 0$. The idea is to apply van der Corput Lemma 7.2.1. Applying repeatedly the fact that $U_T(g_1\cdot g_2) = U_T(g_1)\cdot U_T(g_2)$ for all $g_1, g_2 \in \mathcal{M}_{\mathbb{C}}(X,\mathcal{B})$ (by Lemma 3.1.3.(ii)), we get for all $h, n \geq 1$

$$\begin{aligned}
\langle x_n, x_{n+h}\rangle &= \int_X \prod_{i=1}^{k+1}\left((U_T^{in} f_i)(U_T^{i(n+h)} f_i)\right) d\mu = \int_X \prod_{i=1}^{k+1} U_T^{in}\left(f_i U_T^{ih} f_i\right) d\mu \\
&= \int_X U_T^{n}\left(\prod_{i=1}^{k+1} U_T^{(i-1)n}\left(f_i U_T^{ih} f_i\right)\right) d\mu = \int_X \prod_{i=1}^{k+1} U_T^{(i-1)n}\left(f_i U_T^{ih} f_i\right) d\mu \\
&\quad \text{by Proposition 3.1.5} \\
&= \int_X \left(f_1 U_T^h f_1\right)\left(\prod_{i=1}^{k} U_T^{in}\left(f_{i+1} U_T^{(i+1)h} f_{i+1}\right)\right) d\mu \\
&= \left\langle f_1 U_T^h f_1, \prod_{i=1}^{k} U_T^{in}\left(f_{i+1} U_T^{(i+1)h} f_{i+1}\right)\right\rangle,
\end{aligned}$$

so that

$$\frac{1}{N}\sum_{n=1}^{N}\langle x_n, x_{n+h}\rangle = \left\langle f_1 U_T^h f_1, \frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{k-1} U_T^{in}\left(f_{i+1} U_T^{(i+1)h} f_{i+1}\right)\right\rangle$$

Let $g_i := f_{i+1} U_T^{(i+1)h} f_{i+1}$. Then $g_i \in L^\infty(X,\mathcal{B},\mu)$ for all $i=1,\ldots,k$, hence we can apply the induction hypothesis to obtain

$$\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{k} U_T^{in}\left(f_{i+1} U_T^{(i+1)h} f_{i+1}\right) - \prod_{i=1}^{k}\int_X f_{i+1} U_T^{(i+1)h} f_{i+1}\,d\mu\right\|_2 = 0.$$

It follows that

$$\begin{aligned}
\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle x_n, x_{n+h}\rangle &= \left\langle f_1 U_T^h f_1, \prod_{i=1}^{k}\int_X f_{i+1} U_T^{(i+1)h} f_{i+1}\,d\mu\right\rangle \\
&= \prod_{i=1}^{k+1}\int_X f_i U_T^{ih} f_i\,d\mu.
\end{aligned}$$

Let $j = 1, \dots, k+1$ be such that $\int_X f_j \, d\mu = 0$. By Proposition **??** T^j is weak mixing, so we can apply Proposition 7.1.2 to conclude that

$$D-\lim_{h\to\infty} \int_X f_j U_T^{jh} f_j \, d\mu = D-\lim_{h\to\infty} \left\langle U_T^{jh} f_j, f_j \right\rangle = \langle f_j, \mathbf{1}\rangle \langle \mathbf{1}, f_j\rangle = \left(\int_X f_j \, d\mu\right)^2 = 0.$$

Since $\displaystyle\int_X f_i U_T^{ih} f_i \, d\mu$ is bounded for all $i \neq j$, we can apply Proposition 6.2.3.(i) to conclude that

$$D-\lim_{h\to\infty} \frac{1}{N}\sum_{n=1}^{N} \langle x_n, x_{n+h}\rangle \;=\; D-\lim_{h\to\infty} \prod_{i=1}^{k} \int_X f_i U_T^{ih} f_i \, d\mu = 0.$$

Step 2: In general, let $c_i := \displaystyle\int_X f_i \, d\mu$ and take $g_i := f_i - c_i$, so that $\displaystyle\int_X g_i \, d\mu = 0$ for all $i = 1, \dots, k+1$. Using the identity

$$\prod_{i=1}^{P} a_i - \prod_{i=1}^{P} b_i = (a_1 - b_1)b_2 \dots b_P + a_1(a_2 - b_2)b_3 \dots b_P + \dots + \tag{7.6}$$

$$+a_1 \dots a_{P-1}(a_P - b_P) \tag{7.7}$$

$$= \sum_{i=1}^{P} \left(\prod_{j=1}^{i-1} a_j\right)(a_i - b_i)\left(\prod_{j=i+1}^{P} b_j\right), \tag{7.8}$$

We get that

$$\prod_{i=1}^{k+1} U_T^{in} f_i - \prod_{i=1}^{k+1} \int_X f_i \, d\mu = \sum_{i=1}^{k+1} \left(\prod_{j=1}^{i-1} U_T^{jn} f_j\right) U_T^{in} g_i \prod_{j=i+1}^{k+1} c_j = \sum_{i=1}^{k+1} \prod_{j=1}^{k+1} U_T^{jn} h_j^i,$$

where $h_j^i = f_j$ for $j = 1, \dots i-1$, $h_i^i = g_i$, $h_j^i = c_j$ for $j = i+1, \dots k+1$. Furthermore, $\displaystyle\int_X h_i^i = 0 \, d\mu$. For all $i = 1, \dots, k+1$, we can apply Step 1 for the functions h_j^i, $j = 1, \dots, k+1$ to get that

$$\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=1}^{N}\prod_{j=1}^{k+1} U_T^{jn} h_j^i \right\|_2 = 0.$$

Since

$$\left\| \frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{k+1} U_T^{in} f_i - \prod_{i=1}^{k+1}\int_X f_i \, d\mu \right\|_2 = \left\| \frac{1}{N}\sum_{n=1}^{N} \left(\prod_{i=1}^{k+1} U_T^{in} f_i - \prod_{i=1}^{k+1}\int_X f_i \, d\mu\right) \right\|_2$$

$$= \left\| \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{k+1}\prod_{j=1}^{k+1} U_T^{jn} h_j^i \right\|_2 \leq \sum_{i=1}^{k+1} \left\| \frac{1}{N}\sum_{n=1}^{N}\prod_{j=1}^{k+1} U_T^{jn} h_j^i \right\|_2,$$

it follows that

$$\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{k+1}U_T^{in}f_i-\prod_{i=1}^{k+1}\int_X f_i\,d\mu\right\|_2 \quad = \quad 0.$$

□

Chapter 8

Szemeredi Theorem

In 1975, E. Szemerédi proved the following theorem conjectured some forty years earlier by Erdös and Turan [30].

Theorem 8.0.3. *[102] Let $A \subseteq \mathbb{Z}$ with $\overline{d}(A) > 0$. Then A contains arbitrarily long arithmetic progressions.*

Let us denote with **(Sz)** the above formulation of Szemeredi theorem. It is obvious that **(Sz)** is equivalent with the following statement: For any $A \subseteq \mathbb{Z}$ with $\overline{d}(A) > 0$ and any $k \in \mathbb{Z}_+$, there exists $n \in \mathbb{Z}_+$ such that

$$A \cap (-n + A) \cap (-2n + A) \cap \ldots \cap (-(k-1)n + A) \neq \emptyset.$$

Definition 8.0.4. *We say that a MDS $(X, \mathcal{B}, \mu, T)$ has the **Szemeredi property** at level $k \geq 1$ if, for any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $n \geq 1$ such that*

$$\mu\left(A \cap T^{-n}(A) \cap T^{-2n}(A) \cap \ldots \cap T^{-(k-1)n}(A)\right) > 0. \tag{8.1}$$

Theorem 8.0.5. *Every MDS has the Szemeredi property at any level $k \geq 1$.*

This turns out to be equivalent with Szemeredi Theorem. That is why we label it **(Sz-ergodic)**.

In this chapter we shall prove the following result.

Proposition 8.0.6. (Sz-ergodic) *implies* **(Sz)**.

Proof. Let $A \subseteq \mathbb{Z}$ with $\overline{d}(A) > 0$ and $k \geq 1$. Since $0 < \overline{d}(A) \leq Bd^{\star}(A)$, we can apply Furstenberg correspondence principle (Theorem 6.3.1) to obtain a MDS $(X, \mathcal{B}, \mu, T)$ and a set $A' \in \mathcal{B}$ such that $\mu(A') = Bd^{\star}(A)$ and for every finite subset $F \subseteq \mathbb{Z}$,

$$Bd^{\star}\left(\bigcap_{n\in F}(-n + A)\right) \geq \mu\left(\bigcap_{n\in F} T^{-n}(A')\right). \tag{8.2}$$

Since $\mu(A') = Bd^\star(A) > 0$, we can apply **(Sz-ergodic)** to obtain the existence of $n \geq 1$ such that

$$\mu\left(A' \cap T^{-n}(A') \cap T^{-2n}(A') \cap \ldots \cap T^{-(k-1)n}(A')\right) > 0. \tag{8.3}$$

Apply now (8.2) with $F = \{0, n, 2n \ldots, (k-1)n\}$ to obtain that

$$Bd^\star\left(A \cap (-n+A) \cap (-2n+A) \cap \ldots \cap (-(k-1)n+A)\right) \;\geq\; \mu\left(\bigcap_{d=0}^{k-1} T^{-nd}(A')\right) > 0.$$

In particular, we get that $A \cap (-n+A) \cap (-2n+A) \cap \ldots \cap (-(k-1)n+A) \neq \emptyset$. □

8.1 Finitary versions

Let us consider the following finitary version of Szemeredi Theorem:

(Sz-finitary) For any $k \geq 1$ and $\varepsilon \in (0,1]$ there exists $N = N(k, \varepsilon) \geq 1$ such that if $m \geq N$ and $S \subseteq \{1, \ldots, m\}$ satisfies $|S| \geq \varepsilon m$ then S contains an arithmetic progression of length k.

Furthermore, we consider also the finitary version of **(Sz-ergodic)**:

(Sz-ergodic-finitary) For any $k \geq 1$ and $\varepsilon \in (0,1]$ there exist $M = M(k, \varepsilon) \geq 1$ and $\delta = \delta(k, \varepsilon) > 0$ such that for any MDS $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$ with $\mu(A) \geq \varepsilon$ there exists n with $1 \leq n \leq M$ satisfying $\mu\left(A \cap T^{-n}(A) \cap T^{-2n}(A) \cap \ldots \cap T^{-(k-1)n}(A)\right) \geq \delta$.

Proposition 8.1.1. (Sz)⇔(Sz-finitary)⇔(Sz-ergodic)⇔(Sz-ergodic-finitary)

Proof. **(Sz-ergodic)** implies **(Sz)** By Proposition 8.0.6.

(Sz-finitary)⇒ (Sz) Assume $\overline{d}(A) > 0$ and let $k \geq 1$. Since $\overline{d}(A) = \limsup_{n\to\infty} \frac{|A \cap [1,n]|}{n} > 0$, we get the existence of $\varepsilon > 0$ such that for every $n \geq 1$ there exists $m \geq n$ with $|A \cap [1,m]| > \varepsilon m$. Let $N = N(k, \varepsilon)$ and take $m \geq N$ such that $|A \cap [1,m]| > \varepsilon m$. We can apply now **(Sz-finitary)** to conclude that $S := A \cap [1,m]$ contains an arithmetic progression of length k. Hence, A contains an arithmetic progression of length k.

(Sz-ergodic-finitary)⇒(Sz-ergodic) follows immediately: Apply **(Sz-ergodic-finitary)** with $\varepsilon := \mu(A) > 0$.

(Sz-finitary)⇒(Sz-ergodic-finitary) Let $k \geq 1$, $\varepsilon \in (0,1]$, and $N = N(k, \varepsilon) \geq 1$ satisfying **(Sz-finitary)**. Define

$$M = N\left(k, \frac{\varepsilon}{2}\right). \tag{8.4}$$

Denote with $\mathcal{P}$ the set of all distinct arithmetic progressions of length k in $\{1, \ldots, M\}$, and let $J = |\mathcal{P}|$. Let us remark that $J \neq 0$, since the set $S := \{1, \ldots, M\}$ satisfies

$|S| = M \geq \frac{\varepsilon}{2}M$, so it contains an arithmetic progression of length k by **(Sz-finitary)**. Put

$$\delta = \frac{\varepsilon}{2J}. \tag{8.5}$$

Let now $(X, \mathcal{B}, \mu, T)$ be a MDS and $A \in \mathcal{B}$ be with $\mu(A) \geq \varepsilon$. Define

$$f : X \to \mathbb{R}, \quad f(x) = (S_M \chi_{T^{-1}(A)})(x) = \frac{1}{M}\sum_{i=0}^{M-1} \chi_{T^{-n-1}(A)}(x) = \frac{1}{M}\sum_{i=1}^{M} \chi_{T^{-n}(A)}(x), \tag{8.6}$$

and let

$$B \;=\; \{x \in X \mid f(x) \geq \varepsilon/2\}.$$

By Proposition 5.0.4 , we have that $\int_X f\, d\mu = \mu(A) \geq \varepsilon$. We get that

$$\begin{aligned} \varepsilon &\leq \int_X f\, d\mu = \int_B f\, d\mu + \int_{X\setminus B} f\, d\mu \leq \int_B 1\, d\mu + \int_{X\setminus B} \frac{\varepsilon}{2}\, d\mu, \text{ since } f \leq 1 \\ &= \mu(B) + \frac{\varepsilon}{2}(1 - \mu(B)) = \frac{\varepsilon}{2} + \left(1 - \frac{\varepsilon}{2}\right)\mu(B). \end{aligned}$$

It follows that $\left(1 - \frac{\varepsilon}{2}\right)\mu(B) \geq \frac{\varepsilon}{2}$, hence

$$\mu(B) > \varepsilon/2. \tag{8.7}$$

For all $x \in X$, let us denote

$$E_x = \{n \in \{1, \dots, M\} \mid x \in T^{-n}(A)\}.$$

We have that $|E_x| = Mf(x)$. For $x \in B$, it follows that $|E_x| \geq \frac{\varepsilon}{2}M$, hence we can apply **(Sz-finitary)** to conclude that E_x contains an arithmetic progression P_x of length k, which implies that $x \in \bigcap_{t \in P_x} T^{-t}(A)$. Thus, we have proved that

$$B \subseteq \bigcup_{x \in B} \bigcap_{t \in P_x} T^{-t}(A) \subseteq \bigcup_{P \in \mathcal{P}} \bigcap_{t \in P} T^{-t}(A).$$

Since $|\mathcal{P}| = J$, it follows that for some $P = \{a, a+n, \dots, a+(k-1)n\} \subseteq \{1, \dots, M\}$ we have that

$$\mu\left(\bigcap_{t \in P} T^{-t}(A)\right) \geq \frac{1}{J}\mu(B) \geq \frac{\varepsilon}{2J} = \delta.$$

Since T is measure-preserving,

$$\begin{aligned} \mu\left(\bigcap_{t \in P} T^{-t}(A)\right) &= \mu\left(T^{-a}(A) \cap T^{-a-n}(A) \cap \ldots \cap T^{-a-(k-1)n}(A)\right) \\ &= \mu\left(T^{-a}\left(A \cap T^{-n}(A) \cap \ldots \cap T^{-(k-1)n}(A)\right)\right) \\ &= \mu\left(A \cap T^{-n}(A) \cap \ldots \cap T^{-(k-1)}(A)\right). \end{aligned}$$

Thus, there exists $n \in \{1, \ldots, M\}$ such that $\mu\left(A \cap T^{-n}(A) \cap \ldots \cap T^{-(k-1)}(A)\right) \geq \delta$.

(Sz)⇒(Sz-finitary) Exercise. □

8.2 Uniform Szemeredi property

Definition 8.2.1. *We say that a MDS $(X, \mathcal{B}, \mu, T)$ has the **uniform Szemeredi property** at level $k \geq 1$ if, for any $A \in \mathcal{B}$ with $\mu(A) > 0$,*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(A \cap T^{-n}(A) \cap T^{-2n}(A) \cap \ldots \cap T^{-(k-1)n}(A)\right) > 0. \tag{8.8}$$

Remark 8.2.2. *If a MDS $(X, \mathcal{B}, \mu, T)$ has the uniform Szemeredi property at level $k \geq 1$, then it has the Szemeredi property at level k.*

Proof. Use the fact that for any sequence (a_n), $\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n > 0$ implies the existence of $n \geq 1$ such that $a_n > 0$. □

Proposition 8.2.3. *Let $k \geq 1$. The following are equivalent:*

(i) Every Borel (invertible) MDS has the uniform Szemeredi property at level k.

(ii) Every ergodic Borel (invertible) MDS has the uniform Szemeredi property at level k.

Proof. (ii)⇒(i). Let $(X, \mathcal{B}(X), \mu, T)$ be a Borel (invertible) MDS and $A \in \mathcal{B}$ with $\mu(A) > 0$. Apply the Ergodic Decomposition Theorem 4.1.2 to get a decomposition of μ in ergodic components μ_x, $x \in X$. Let

$$B := \{x \in X \mid \mu_x(A) > 0\}.$$

We have that $\mu(B) > 0$, since

$$0 < \mu(A) = \int_X \mu_x(A) d\mu(x) = \int_B \mu_x(A) d\mu(x) \leq \mu(B).$$

For almost all $x \in B$, we have that the Borel (invertible) MDS $(X, \mathcal{B}(X), \mu_x, T)$ is ergodic and $\mu_x(A) > 0$, so we can apply (ii) to get that

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu_x\left(\bigcap_{i=0}^{k-1} T^{-in}(A)\right) > 0.$$

It follows that

$$\begin{aligned}
\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu\left(\bigcap_{i=0}^{k-1} T^{-in}(A)\right) &= \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int_X \mu_x\left(\bigcap_{i=0}^{k-1} T^{-in}(A)\right) d\mu(x) \\
&= \liminf_{N\to\infty} \int_X \frac{1}{N}\sum_{n=1}^{N} \mu_x\left(\bigcap_{i=0}^{k-1} T^{-in}(A)\right) d\mu(x) \\
&\geq \int_X \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu_x\left(\bigcap_{i=0}^{k-1} T^{-in}(A)\right) d\mu(x) \\
&\quad \text{by Fatou's Lemma} \\
&= \int_B \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu_x\left(\bigcap_{i=0}^{k-1} T^{-in}(A)\right) d\mu(x) \\
&> 0,
\end{aligned}$$

since $\mu(B) > 0$ and $\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu_x\left(\bigcap_{i=0}^{k-1} T^{-in}(A)\right) > 0$ for almost all $x \in B$. □

8.3 Uniform Szemeredy property for compact systems

Definition 8.3.1. *Let $(X, \mathcal{B}, \mu, T)$ be a MDS. A function $f \in L^2(X, \mathcal{B}, \mu)$ is called* ***almost periodic*** *if $\overline{\{U_T^n f \mid n \geq 0\}}$ is compact in $L^2(X, \mathcal{B}, \mu)$. The MDS $(X, \mathcal{B}, \mu, T)$ is said to be* ***compact*** *if every $f \in L^2(X, \mathcal{B}, \mu)$ is almost periodic.*

Thus, $f \in L^2(X, \mathcal{B}, \mu)$ is almost periodic if and only if the orbit $\{U_T^n f \mid n \geq 0\}$ is totally bounded in the norm topology of $L^2(X, \mathcal{B}, \mu)$.

The archetypal example of a compact MDS is the rotation on the circle group.(see [47, The compact case, p. 2]

We shall use the notation $AP(X)$ for the set of almost periodic functions in $L^2(X, \mathcal{B}, \mu)$.

Proposition 8.3.2. *(i) If $f \in AP(X)$, then for any $\varepsilon > 0$ the set*

$$A_\varepsilon := \{n \geq 1 \mid \|U_T^n f - f\|_2 < \varepsilon\} \tag{8.9}$$

is syndetic.

(ii) $AP(X)$ is a closed linear subspace of $L^2(X, \mathcal{B}, \mu)$.

Proof. (i) See Seminar 9.4.

(ii) TO WRITE.

□

Theorem 8.3.3. *Let $(X, \mathcal{B}, \mu, T)$ be a compact MDS and $k \geq 1$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X \prod_{i=0}^{k} U_T^{in} f \, d\mu > 0.$$

for all $f \in L^\infty(X), f \geq 0$, but f not a.e. 0.

Proof. Let $a := \displaystyle\int_X f^{k+1} \, d\mu > 0$. We can assume w.l.o.g that $0 \leq f \leq 1$ a.e. Since U_T is positive, we get that $0 \leq U_T^{jn} f \leq 1$ a.e. for all $n \geq 1$, $j = 0, \ldots, k$. Using the identity (7.2):

$$\begin{aligned}
\prod_{i=0}^{P} a_i - \prod_{i=0}^{P} b_i &= (a_0 - b_0)b_1 \ldots b_P + a_0(a_1 - b_1)b_2 \ldots b_P + \ldots + \\
&\quad + a_0 \ldots a_{P-1}(a_P - b_P) \\
&= \sum_{i=0}^{P} \left(\prod_{j=0}^{i-1} a_j \right) (a_i - b_i) \left(\prod_{j=i+1}^{P} b_j \right),
\end{aligned}$$

we get that

$$\begin{aligned}
\left| \int_X \prod_{i=0}^{k} U_T^{in} f \, d\mu - a \right| &= \left| \int_X \left(\prod_{i=0}^{k} U_T^{in} f \, d\mu - \prod_{i=0}^{k} f \right) \, d\mu \right| \\
&= \left| \int_X \sum_{i=0}^{k} \left(\prod_{j=0}^{i-1} U_T^{jn} f \right) (U_T^{in} f - f) f^{k-i} \, d\mu \right| \\
&\leq \sum_{i=0}^{k} \int_X \left| \left(\prod_{j=0}^{i-1} U_T^{jn} f \right) (U_T^{in} f - f) f^{k-i} \right| \, d\mu \\
&\leq \sum_{i=0}^{k} \int_X |U_T^{in} f - f| \, d\mu = \sum_{j=0}^{k} \|U_T^{in} f - f\|_1 \\
&\leq \sum_{i=0}^{k} \|U_T^{in} f - f\|_2.
\end{aligned}$$

Choose $\varepsilon < \dfrac{a}{k(k+1)}$. By Proposition 8.3.2.(i), we have that the set

$$A_\varepsilon = \{n \geq 1 \mid \|U_T^n f - f\|_2 < \varepsilon\}$$

is syndetic. For all $n \in A_\varepsilon$, we have that

$$\|U_T^{(i+1)n} f - U_T^{in} f\|_2 = \|U_T^{in}(U_T^n f - f)\|_2 = \|U_T^n f - f\|_2 < \varepsilon \quad \text{for all } i = 0, \ldots, k,$$

so that, by the triangle inequality,

$$\|U_T^{in} f - f\|_2 \leq \sum_{l=0}^{i-1} \|U_T^{(l+1)n} f - U_T^{ln} f\|_2 \leq i\varepsilon \leq k\varepsilon$$

for all $i = 0, \ldots, k$. It follows that for all $n \in A_\varepsilon$,

$$\left| \int_X \prod_{i=0}^{k} U_T^{in} f \, d\mu - a \right| \leq \sum_{i=0}^{k} \|U_T^{in} f - f\|_2 < k(k+1)\varepsilon,$$

hence

$$\int_X \prod_{i=0}^{k} U_T^{in} f \, d\mu \geq a - k(k+1)\varepsilon > 0.$$

Let $b := a - k(k+1)\varepsilon > 0$. It follows that

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X \prod_{i=0}^{k} U_T^{in} f \, d\mu \geq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1, n\in A_\varepsilon}^{N} b = b \liminf_{N\to\infty} \frac{|A_\varepsilon \cap [1,N]|}{N} = b\underline{d}(A_\varepsilon) > 0,$$

since A_ε is syndetic, hence it has positive lower density, by Proposition 6.1.9. □

As immediate consequences, we get that

Corollary 8.3.4. *Any compact MDS has the uniform Szemeredi property at level k for all $k \in \mathbb{Z}_+$.*

Proof. Let $k \in \mathbb{Z}_+$ and $A \in \mathcal{B}$ be such that $\mu(A) > 0$. We can take $f = \chi_A$ in the above theorem, since $0 \leq \chi_A \leq 1$, and $\chi_A(x) = 1$ for all $x \in A$, so χ_A is not a.e. 0. We get that

$$\begin{aligned} 0 &< \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X \prod_{i=0}^{k} U_T^{in} \chi_A \, d\mu = \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X \prod_{i=0}^{k} \chi_{T^{-in}A} \, d\mu \\ &= \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X \chi_{A\cap T^{-n}(A)\cap\ldots\cap T^{-kn}(A)} \, d\mu = \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap \ldots \cap T^{-kn}(A)) \\ &\leq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}(A) \cap \ldots \cap T^{-(k-1)n}(A)). \end{aligned}$$

□

8.4 Uniform Szemeredy property for weak-mixing systems

Proposition 8.4.1. *Let $(X, \mathcal{B}, \mu, T)$ be a weak mixing MDS. For all $k \geq 1$ and all $A \in \mathcal{B}$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(A \cap T^{-n}(A) \cap T^{-2n}(A) \cap \ldots \cap T^{-(k-1)n}(A)\right) = (\mu(A))^{k+1}.$$

Proof. Apply Theorem 7.2.2 with $f_i := \chi_A$, $i = 1, \ldots, k$. Since

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{k} U_T^{in} \chi_A = \frac{1}{N} \sum_{n=1}^{N} \chi_{\bigcap_{i=1}^{k} T^{-in}(A)}, \text{ and } \prod_{i=1}^{k} \int_X \chi_A \, d\mu = (\mu(A))^k,$$

we get that

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} \chi_{\bigcap_{i=1}^{k} T^{-in}(A)} - (\mu(A))^k \right\|_2 = 0.$$

It follows that

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigcap_{i=1}^{k} T^{-in}(A)\right) &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X \chi_{\bigcap_{i=1}^{k} T^{-in}(A)} \, d\mu \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left\langle \chi_{\bigcap_{i=1}^{k} T^{-in}(A)}, \mathbf{1} \right\rangle \\
&= \left\langle \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \chi_{\bigcap_{i=1}^{k} T^{-in}(A)}, \mathbf{1} \right\rangle \\
&= \left\langle (\mu(A))^k, \mathbf{1} \right\rangle = (\mu(A))^{k+1}.
\end{aligned}$$

Since

$$\begin{aligned}
\mu\left(\bigcap_{i=1}^{k} T^{-in}(A)\right) &= \mu\left(T^{-n}\left(\bigcap_{i=0}^{k-1} T^{-in}(A)\right)\right) \\
&= \mu\left(\bigcap_{i=0}^{k-1} T^{-in}(A)\right),
\end{aligned}$$

the conclusion follows. ☐

Corollary 8.4.2. *Any weak mixing MDS $(X, \mathcal{B}, \mu, T)$ has the uniform Szemeredi property at level k for all $k \in \mathbb{Z}_+$.*

Proof. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. By Proposition 8.4.1, we have that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(A \cap T^{-n}(A) \cap T^{-2n}(A) \cap \ldots \cap T^{-(k-1)n}(A)\right) > 0.$$

☐

8.5 Proof of Roth Theorem

In the sequel, $(X, \mathcal{B}, \mu, T)$ is an invertible MDS. Let $AP_{\mathbb{R}}(X)$ be the set of almost periodic functions in $L^2_{\mathbb{R}}(X, \mathcal{B}, \mu)$, and

$$\mathcal{H}_{wm} := \{f \in L^2_{\mathbb{R}}(X, \mathcal{B}, \mu) \mid D-\lim_{n\to\infty} \int_X g U_T^n f \, d\mu = 0 \text{ for all } g \in L^2_{\mathbb{R}}(X, \mathcal{B}, \mu)\}. \tag{8.10}$$

Theorem 8.5.1 (Koopman-von Neumann)**.** $L^2_{\mathbb{R}}(X, \mathcal{B}, \mu) = AP_{\mathbb{R}}(X) \oplus \mathcal{H}_{wm}$.

Proof. TO WRITE. □

Theorem 8.5.2. *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic invertible MDS and $A \in \mathcal{B}$ with $\mu(A) > 0$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}(A) \cap T^{-2n}(A)) > 0.$$

Proof. We have that $\chi_A \in L^2(X)$, hence we can apply the Koopman-von Neumann Theorem 8.5.1 to get $f \in AP_{\mathbb{R}}(X)$ and $g \in \mathcal{H}_{wm}$ such that $\chi_A = f + g$. It is clear that $\chi_A \not\in \mathcal{H}_{wm}$, as the condition (8.10) does not hold for $f := \chi_A$ and $g := 1_X$, hence $f \neq 0$. Since χ_A is the closest function to χ_A in L^2-norm, we get that $0 \leq f \leq 1$. Furthermore, $|g| = |\chi_A - f| \leq 2$.

We have that

$$\begin{aligned}
\mu(A \cap T^{-n}(A) \cap T^{-2n}(A)) &= \int \chi_{A\cap T^{-n}(A)\cap T^{-2n}(A)} = \int_X \chi_A U_T^n(\chi_A) U_T^{2n}(\chi_A) \, d\mu \\
&= \int_X (f+g) U_T^n(f+g) U_T^{2n}(f+g) \, d\mu = \int_X f U_T^n f U_T^{2n} f \, d\mu + \\
&\quad + \int g U_T^n f U_T^{2n} f \, d\mu + \int (f+g) U_T^n f U_T^{2n} g \, d\mu + \\
&\quad + \int (f+g) U_T^n g U_T^{2n} g \, d\mu + \int (f+g) U_T^n g U_T^{2n} f \, d\mu.
\end{aligned}$$

Since f is almost periodic, $f \geq 0$, but f not a.e. 0, we can apply Theorem 8.3.3 for $k = 2$ to conclude that

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f U_T^n f U_T^{2n} f \, d\mu + > 0. \tag{8.11}$$

Claim 1: $\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} U_T^n f U_T^{2n} g \right\|_2 = \lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} U_T^n g U_T^{2n} g \right\|_2 = \lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} U_T^n g U_T^{2n} f \right\|_2 =$ 0.

Proof: We use van der Corput Lemma 7.2.1. We give the proof for $x_n := U_T^n f U_T^{2n} g$, the other two cases being similar. We have that (x_n) is bounded, since f, g are bounded. Furthermore, for all $n, h \geq 1$,

$$\begin{aligned}
\langle x_n, x_{n+h}\rangle &= \int_X U_T^n f U_T^{2n} g U_T^{n+h} f U_T^{2n+2h} g \, d\mu = \int_X U_T^n \left(f U_T^n g U_T^h f U_T^{n+2h} g \right) d\mu \\
&= \int_X f U_T^n g U_T^h f U_T^{n+2h} g \, d\mu \quad \text{by Proposition 3.1.5} \\
&= \int_X \left(f U_T^h f \right) U_T^n (g U_T^{2h} g) \, d\mu,
\end{aligned}$$

hence for all $h \geq 1$,

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \langle x_n, x_{n+h}\rangle &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \int_X \left(f U_T^h f \right) U_T^n (g U_T^{2h} g) \, d\mu \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \langle U_T^n (g U_T^{2h} g), f U_T^h f \rangle \\
&= \left(\int_X f U_T^h f \, d\mu \right) \left(\int_X g U_T^{2h} g \, d\mu \right),
\end{aligned}$$

since T is ergodic, so we can apply Proposition 5.5.4. As $g \in \mathcal{H}_{wm}$, we have in particular that $D- \lim_{h\to\infty} \int_X g U_T^{2h} g \, d\mu = 0$. Since $f \leq 1$, we have that $\int_X f U_T^h f \, d\mu \leq 1$ for all $h \geq 1$. Thus, we can apply Proposition 6.2.3.(i) to conclude that

$$D- \lim_{h\to\infty} \left(\int_X f U_T^h f \, d\mu \right) \left(\int_X g U_T^{2h} g \, d\mu \right) = 0.$$

Finally, van der Corput Lemma gives us $\left\| \frac{1}{N} \sum_{n=1}^N x_n \right\|_2 = 0$, that is

$$\left\| \frac{1}{N} \sum_{n=1}^N U_T^n f U_T^{2n} g \right\|_2 = 0. \quad \square$$

It follows then that

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \int_X (f+g) U_T^n f U_T^{2n} g \, d\mu &= \lim_{N\to\infty} \int_X (f+g) \frac{1}{N} \sum_{n=1}^N U_T^n f U_T^{2n} g \, d\mu \\
&= \lim_{N\to\infty} \left\langle f+g, \frac{1}{N} \sum_{n=1}^N U_T^n f U_T^{2n} g \right\rangle = 0.
\end{aligned}$$

We get similarly that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \int_X (f+g) U_T^n g U_T^{2n} g \, d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \int_X (f+g) U_T^n g U_T^{2n} f \, d\mu = 0.$$

It remains $\int_X gU_T^n fU_T^{2n} f\, d\mu$. Since T is invertible, we get that

$$\int_X gU_T^n fU_T^{2n} f\, d\mu = \int_X U_T^{-2n}(gU_T^n fU_T^{2n} f)\, d\mu = \int_X fU_T^{-n} fU_T^{-2n} g\, d\mu.$$

Claim 2: $\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=1}^N U_T^{-n} fU_T^{-2n} g \right\|_2 = 0.$

Proof: It is similar with the proof of the first claim. Let $x_n := U_T^{-n} fU_T^{-2n} g$. Then for all $n, h \geq 1$,

$$\begin{aligned}\langle x_n, x_{n+h}\rangle &= \int_X U_T^{-n} fU_T^{-2n} gU_T^{-n-h} fU_T^{-2n-2h} g\, d\mu = \int_X U_T^{-2n}\left(U_T^n fgU_T^{n-h} fU_T^{-2h} g\right)\, d\mu \\ &= \int_X \left(gU_T^{-2h} g\right) U_T^n (fU_T^{-h} f)\, d\mu.\end{aligned}$$

hence for all $h \geq 1$,

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N \langle x_n, x_{n+h}\rangle &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N \int_X \left(gU_T^{-2h} g\right) U_T^n (fU_T^{-h} f)\, d\mu \\ &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N \left\langle U_T^n (fU_T^{-h} f), gU_T^{-2h} g\right\rangle \\ &= \left(\int_X fU_T^{-h} f\, d\mu\right)\left(\int_X gU_T^{-2h} g\, d\mu\right) \\ &= \left(\int_X fU_T^{-h} f\, d\mu\right)\left(\int_X U_T^{2h}(gU_T^{-2h} g)\, d\mu\right) \\ &= \left(\int_X fU_T^{-h} f\, d\mu\right)\left(\int_X gU_T^{2h} g\, d\mu\right).\end{aligned}$$

Since $D-\lim_{h\to\infty} \int_X gU_T^{2h} g\, d\mu = 0$ and $\int_X fU_T^{-h} f\, d\mu \leq 1$ for all $h \geq 1$, we get that

$$D-\lim_{h\to\infty} \left(\int_X fU_T^{-h} f\, d\mu\right)\left(\int_X gU_T^{2h} g\, d\mu\right) = 0,$$

and apply van der Corput Lemma to obtain that $\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=1}^N x_n \right\|_2 = 0.$ □

As a consequence of the above claim,

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N \int_X gU_T^n fU_T^{2n} f\, d\mu &= \lim_{N\to\infty} \int_X g\frac{1}{N}\sum_{n=1}^N U_T^n fU_T^{2n} f\, d\mu \\ &= \lim_{N\to\infty} \left\langle g, \frac{1}{N}\sum_{n=1}^N U_T^n fU_T^{2n} f\right\rangle = 0.\end{aligned}$$

Thus, $\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T^{-n}(A) \cap T^{-2n}(A))$. □

As an immediate consequence, it follows that

Corollary 8.5.3. *Any ergodic invertible MDS $(X, \mathcal{B}, \mu)$ has the uniform Szemerédi property at level $k = 3$.*

Combining the above Corollary and Proposition 8.2.3, we get that

Corollary 8.5.4. *Any Borel invertible MDS $(X, \mathcal{B}, \mu)$ has the uniform Szemerédi property at level $k = 3$.*

Finally, we obtain Roth theorem.

Theorem 8.5.5 (Roth theorem)**.** *[92]*
Let $A \subseteq \mathbb{Z}$ with $\overline{d}(A) > 0$. Then A contains infinitely many long arithmetic arithmetic progressions $a, a+n, a+2n$ of length three, with $a \in \mathbb{Z}$ and $n \geq 1$.

Proof. Let $(X, \mathcal{B}, \mu, T)$ be the Borel invertible MDS and $A' \in \mathcal{B}$ associated to A by the Furstenberg correspondence principle. Applying Corollary 8.5.4, we get that $(X, \mathcal{B}, \mu, T)$ has the Szemerédi property at level $k = 3$. Thus, there exists $n \geq 1$ such that

$$\mu\left(A' \cap T^{-n}(A') \cap T^{-2n}(A')\right) > 0.$$

It follows that

$$Bd^{\star}\left(A \cap (-n + A) \cap (-2n + A)\right) \;\;\geq\;\; \mu\left(A' \cap T^{-n}(A') \cap T^{-2n}(A')\right) > 0.$$

In particular, we get that $A \cap (-n + A) \cap (-2n + A) \neq \emptyset$. □

Part III

Appendices

Appendix A

Set theory

Proposition A.0.6 (Zorn's Lemma).
Let $(X, \leq)$ be a nonempty partially ordered set. Assume every chain (i.e. totally ordered subset) has an upper bound (resp. a lower bound). Then X has a maximal element (resp., minimal element).

Let $T : X \to X$. For any $n \geq 1$, $T^n : X \to X$ is the composition of T n-times. For $n \geq 1$ and $A \subseteq X$, we shall use the notation

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

If T is bijective with inverse T^{-1}, then the inverse of T^n is $(T^{-1})^n$, the composition of T^{-1} n-times. We shall denote it with T^{-n}. Thus,

$$T^{-n} = \left(T^{-1}\right)^n = (T^n)^{-1}. \tag{A.2}$$

Lemma A.0.7. *Let $T : X \to X$ and $A \subseteq X$.*

(i) If $T(A) \subseteq A$, then $T^{n+1}(A) \subseteq T^n(A) \subseteq A$ for all $n \geq 0$.

(ii) If $T(A) = A$, then $T^n(A) = A$ for all $n \geq 0$.

(iii) $T^{-n-1}(A) = T^{-1}(T^{-n}(A)) = T^{-n}(T^{-1}(A))$.

(iv) If $T^{-1}(A) \subseteq A$, then $T^{-n-1}(A) \subseteq T^{-n}(A) \subseteq A$ for all $n \geq 0$.

(v) If $T^{-1}(A) = A$, then $T(A) \subseteq A$.

(vi) If $T^{-1}(A) = A$, then $T^{-n}(A) = A$ for all $n \geq 0$.

Lemma A.0.8. *Let $T : X \to X$ be bijective and $A \subseteq X$.*

(i) $T(A) = A$ if and only if $T^{-1}(A) = A$.

(ii) If $T(A) = A$, then $T^n(A) = A$ for all $n \in \mathbb{Z}$.

A.1 Collections of sets

In the sequel, X is a nonempty set and $\mathcal{C}$ is a collection of subsets of X.

Definition A.1.1. *$\mathcal{C}$ is said to* **cover** *X, or to be a* **cover** *or a* **covering** *of X, if every point in X is in one of the sets of $\mathcal{C}$, i.e. $X = \bigcup \mathcal{C}$.*

Given any cover $\mathcal{C}$ of X, a **subcover** *of $\mathcal{C}$ is a subset of $\mathcal{C}$ that is still a cover of X.*

Definition A.1.2. *$\mathcal{C}$ is said to have the* **finite intersection property** *if for every finite subcollection $\{C_1, \ldots, C_n\}$ of $\mathcal{C}$, the intersection $C_1 \cap \ldots \cap C_n$ is nonempty.*

Remark A.1.3. *If X has a finite cover $X = \bigcup_{i=1}^{n} A_i$, then we can always construct a cover $X = \bigcup_{i=1}^{n} B_i$ of X such that $m \leq n$, $B_i \subseteq A_i$, and $B_i \cap B_j = \emptyset$ for all $i \neq j$. Just take $B_i := A_i \setminus \bigcup_{j \neq i} A_j$.*

For any nonempty subset A of X, we denote

$$\mathcal{C} \cap A = \{C \cap A \mid C \in \mathcal{C}\}. \tag{A.3}$$

A.1.1 Sequences of sets

Let X be a nonempty set and $(E_n)_{n\geq 1}$ be a sequence of subsets of X.

Definition A.1.4. *(i) The* ***limit superior*** *of (E_n) is defined by*

$$\limsup_{n\to\infty} E_n := \bigcap_{n\geq 1} \bigcup_{i\geq n} E_i. \tag{A.4}$$

(ii) The ***limit inferior*** *of (E_n) is defined by*

$$\liminf_{n\to\infty} E_n := \bigcup_{n\geq 1} \bigcap_{i\geq n} E_i. \tag{A.5}$$

Alternative names are **superior (inferior) limit** or **upper (lower) limit**.

Definition A.1.5. *If $\limsup_{n\to\infty} E_n = \liminf_{n\to\infty} E_n$, we say that the sequence $(E_n)_{\geq 1}$* ***converges*** *to the set $\lim_{n\to\infty} E_n := \limsup_{n\to\infty} E_n = \liminf_{n\to\infty} E_n$ and call $\lim_{n\to\infty} E_n$ its* ***limit***.

Definition A.1.6. *The sequence $(E_n)_{n\geq 1}$ is said to be*

(i) ***increasing*** *if $E_n \subseteq E_{n+1}$ for each n;*

(ii) ***decreasing*** *if* $E_n \supseteq E_{n+1}$ *for each* n*;*

(iii) ***monotone*** *if it is either decreasing or increasing.*

Proposition A.1.7. *(i)* $\limsup_{n\to\infty} E_n$ *is the set of those elements which are in* E_n *for infinitely many* n*.*

(ii) $\liminf_{n\to\infty} E_n$ *is the set of those elements which are in all but a finite number of the sets* E_n*.*

(iii) $\liminf_{n\to\infty} E_n \subseteq \limsup_{n\to\infty} E_n$*.*

(iv) If (E_n) *is increasing, then* $\lim_{n\to\infty} E_n = \bigcup_{n\geq 1} E_n$*.*

(v) If (E_n) *is decreasing, then* $\lim_{n\to\infty} E_n = \bigcap_{n\geq 1} E_n$*.*

(vi) If $E_1, E_2, \ldots$ *are pairwise disjoint, then* $\lim_{n\to\infty} E_n = \emptyset$*.*

Proof. See [110, Claim 1, p.43]. □

Proposition A.1.8. *Let* $(E_n)_{n\geq 1}$ *be a sequence of subsets of* X *and* $f : X \to \mathbb{R}$*. Then*

$$\lim_{n\to\infty} \chi_{\cup_{i=1}^n E_i} f = \chi_{\cup_{i\geq 1} E_i} f \tag{A.6}$$

Proof. Let

$$B_n := \bigcup_{i=1}^n E_i, \quad B := \bigcup_{i=1}^\infty E_i, \quad g_n := \chi_{B_n} f, \quad g := \chi_B f.$$

Let $x \in X$. We have two cases:

(i) $x \in B$. Then $g(x) = f(x)$ and there exists $N \geq 1$ such that $x \in E_N$. It follows that $x \in B_n$ for all $n \geq N$, hence $g_n(x) = f(x)$ for all $n \geq N$. In particular, $\lim_{n\to\infty} g_n(x) = f(x) = g(x)$.

(ii) $x \notin B$. Then $g(x) = 0$ and $x \notin E_n$ for any $n \geq 1$. It follows that $x \notin B_n$ for any $n \geq 1$, hence $g_n(x) = 0$ for all $n \geq 1$. In particular, $\lim_{n\to\infty} g_n(x) = 0 = g(x)$.

□

A.1.2 Monotone classes

Definition A.1.9. *A nonempty collection* $\mathcal{M}$ *of subsets of a set* X *is called a* ***monotone class*** *if for every monotone sequence* $(E_n)_{n\geq 1}$*,*

$$E_n \in \mathcal{M} \text{ for all } n \text{ implies } \lim_{n\to\infty} E_n \in \mathcal{M}.$$

Since the intersection of any family of monotone classes is a monotone class, we can speak of the monotone clas generated by any given collection of subsets of X.

Appendix B

Topology

In the sequel, spaces X, Y, Z are nonempty topological spaces.

Definition B.0.10. *A point x in X is said to be an* ***isolated point*** *of X if the one-point set $\{x\}$ is open in X.*

Definition B.0.11. *Let X, Y be topological spaces and $f : X \to Y$.*

(i) f is said to be an **open** *map if for each open set U of X, the set $f(U)$ is open in Y.*

(ii) f is said to be a **closed** *map if for each closed set F of X, the set $f(F)$ is closed in Y.*

B.1 Closure, interior and related

Let A be a subset of X.

Definition B.1.1. *The* **closure** *of A, denoted by $\overline{A}$, is defined as the intersection of all closed subsets of X that contain A.*

Definition B.1.2. *The* **interior** *of A, denoted by A°, is the union of all open subsets of X that are contained in A.*

Proposition B.1.3. *(i) If U is an open set that intersects $\overline{A}$, then U must intersect A.*

(ii) If X is a Hausdorff space without isolated points, then given any nonempty open set U of X and any finite subset S of X, there exists a nonempty open set V contained in U such that $S \cap \overline{V} = \emptyset$.

Proof. See [74, proof of Theorem 27.7, p.176]. □

Definition B.1.4. *A subset A of X is* **dense** *in X if $\overline{A} = X$.*

Proposition B.1.5. *Let $A \subseteq X$. The following are equivalent:*

(i) A is dense in X.

(ii) A meets every nonempty open subset of X.

(iii) A meets every nonempty basis open subset of X.

(iv) the complement of A has empty interior.

Definition B.1.6. *A subset A of a topological space X is called* **nowhere dense** *if its closure $\overline{A}$ has empty interior.*

Hence, a closed subset is nowhere dense if and only if it has nonempty interior.

B.2 Hausdorff spaces

Definition B.2.1. *X is said to be* **Hausdorff** *if for each pair x, y of distinct points of X, there exist disjoint open sets containing x and y, respectively.*

Proposition B.2.2. *(i) Every finite subset of a Hausdorff topological space is closed.*

(ii) Any subspace of a Hausdorff space is Hausdorff.

Proof. (i) See [74, Theorem 17.8, p.99].

(ii) See [67, Proposition 3.4, p.41-42].

(iii) See [74, Ex. 13, p.101].

□

B.3 Bases and subbases

Definition B.3.1. *Let X be a set. A* **basis** *(for a topology) on X is a collection $\mathcal{B}$ of subsets of X (called* **basis elements***) satisfying the following conditions:*

(i) Every element is in some basis element; in other words, $X = \bigcup_{B\in\mathcal{B}} B$.

(ii) If $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 \cap B_2$, there exists a basis element $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

Let $\mathcal{B}$ be basis on a set X, and define

$$\mathcal{T} := \text{the collection of all unions of elements of } \mathcal{B}.$$

Then $\mathcal{T}$ is a topology on X, called the **topology generated by** $\mathcal{B}$. We also say that $\mathcal{B}$ **is a basis for** $\mathcal{T}$.

Another way of describing the topology generated by a basis is given in the following. Given a set X and a collection $\mathcal{B}$ of subsets of X, we say that a subset $U \subseteq X$ satisfies the **basis criterion** with respect to $\mathcal{B}$ if for every $x \in U$, there exists $B \in \mathcal{B}$ such that $x \in B \subseteq U$.

Proposition B.3.2. *Let $\mathcal{B}$ be a basis on a set X and $\mathcal{T}$ be the topology generated by $\mathcal{B}$. Then $\mathcal{T}$ is precisely the collection of all subsets of X that satisfy the basis criterion with respect to $\mathcal{B}$.*

Proof. See [67, Lemma 2.10, p.27-28]. □

Proposition B.3.3. *Suppose X is a topological space, and $\mathcal{B}$ is a collection of open subsets of X. If every open subset of X satisfies the basis criterion with respect to $\mathcal{B}$, then $\mathcal{B}$ is a basis for the topology of X.*

Proof. See [67, Lemma 2.11, p.29]. □

Definition B.3.4. *A* **subbasis** *(for a topology) on X is a collection of subsets of X whose union equals X. The* **topology generated by the subbasis** *$\mathcal{S}$ is defined to be the collection $\mathcal{T}$ of all unions of finite intersections of elements of $\mathcal{S}$.*

If $\mathcal{S}$ is a subbasis on X and $\mathcal{B}$ is the collection of all finite intersections of elements of $\mathcal{S}$, then $\mathcal{B}$ is a basis on X and $\mathcal{T}$ is the topology generated by $\mathcal{B}$.

B.4 Continuous functions

A function $f : X \to Y$ is said to be **continuous** if for each open subset V of Y, the set $f^{-1}(V)$ is open in X.

Remark B.4.1. *If the topology of Y is given by a basis (resp. a subbasis), then to prove continuity of f it suffices to show that the inverse image of every* **basis element** *(resp.* **subbasis element***) is open.*

Proof. See [74, p.103]. □

Proposition B.4.2. *Let $f : X \to Y$. The following are equivalent*

(i) f is continuous.

(ii) For every closed subset B of Y, the set $f^{-1}(B)$ is closed in X.

(iii) For every subset A of X, $f(\overline{A}) \subseteq \overline{f(A)}$.

(iv) For each $x \in X$ and each open neighborhood V of $f(x)$, there is an open neighborhood U of x such that $f(U) \subseteq V$.

Proof. See [74, Theorem 18.1, p.104]. □

Proposition B.4.3. *Let X, Y, Z be topological spaces.*

(i) (Inclusion) If A is a subspace of X, then the inclusion function $j : A \to X$ is continuous.

(ii) *(Composition) If* $f : X \to Y$ *and* $g : Y \to Z$ *are continuous, then the map* $g \circ f$ *is continuous.*

(iii) *(Restricting the domain) If* $f : X \to Y$ *is continuous and* A *is a subspace of* X*, then the restricted function* $f\mid_A : A \to Y$ *is continuous.*

(iv) *(Restricting or expanding the range) Let* $f : X \to Y$ *be continuous. If* Z *is a subspace of* Y*, containing the image set* $f(X)$ *of* f*, then the function* $g : X \to Z$ *obtained by restricting the range of* f *is continuous. If* Z *is a space having* Y *as a subspace, then the function* $h : X \to Z$*, obtained by expanding the range of* f *is continuous.*

(v) *(Local formulation of continuity) The map* $f : X \to Y$ *is continuous if* X *can be written as the union of open sets* $U_i (i \in I)$ *such that* $f\mid_{U_i}$ *is continuous for each* $i \in I$*.*

Proof. See [74, Theorem 18.2, p.108]. □

B.4.1 Homeomorphisms

Definition B.4.4. *A mapping* $f : X \to Y$ *is called a* **homeomorphism** *if* f *is bijective and both* f *and its inverse* f^{-1} *are continuous.*

If $f : X \to X$ is a homeomorphism, then $f^n : X \to X$ is also a homeomorphism for all $n \in \mathbb{Z}$.

Definition B.4.5. *A continuous map* $f : X \to Y$ *is a* **local homeomorphism** *if every point* $x \in X$ *has a neighborhood* $U \subseteq X$ *such that* $f(U)$ *is an open subset of* Y *and* $f|_U : U \to f(U)$ *is a homeomorphism.*

Proposition B.4.6. *Let* $f : X \to Y$ *be bijective. The following properties of* f *are equivalent*

(i) f *is a homeomorphism.*

(ii) f *is continuous and open.*

(iii) f *is continuous and closed.*

(iv) $f(\overline{A}) = \overline{f(A)}$ *for each* $A \subseteq X$*.*

(v) f *is a local homeomorphism.*

Proof. See [26, Theorem 12.2, p.89] and [67, Ex. 2.8.(d), p.24]. □

Proposition B.4.7. *Every local homeomorphism is an open map.*

Proof. See [67, Ex. 2.8.(a), p.24]. □

B.5 Metric topology and metrizable spaces

Let (X, d) be a metric space. Given $x \in X$ and $r > 0$,

$$\begin{aligned} B_r(x) &= \{y \in X \mid d(x,y) < r\} \text{ is the } \textbf{open ball} \text{ with center } x \text{ and radius } r, \text{ while} \\ \overline{B}_r(x) &= \{y \in X \mid d(x,y) \le r\} \text{ is the } \textbf{open ball} \text{ with center } x \text{ and radius } r. \end{aligned}$$

Proposition B.5.1. *The collection*

$$\mathcal{B} := \{B_r(x) \mid x \in X, r > 0\}$$

is a basis for a topology on X.

Proof. See [74, p.119]. □

The topology generated by $\mathcal{B}$ is called the **metric topology (induced by d)**.

Remark B.5.2. *It is easy to see that the set $\{B_{2^{-k}}(x) \mid x \in X, k \in \mathbb{N}\}$ is also a basis for the metric topolgy.*

Example B.5.3. (i) Let X be a discrete metric space. Then the induced metric topology is the discrete topology.

(ii) Let $(\mathbb{R}, d)$ be the set of real numbers with the natural metric $d(x, y) = |x - y|$. Then the induced metric topology is the standard topology on $\mathbb{R}$.

(iii) Let $(\mathbb{C}, d)$ be the set of complex numbers with the natural metric $d(z_1, z_2) = |z_1 - z_2|$.

(iv) Let $\mathbb{R}^n (n \ge 1)$ and define the **euclidean metric** on $\mathbb{R}^n$ by

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}$$
$$\text{for all } \mathbf{x} = (x_1, x_2 \ldots, x_n),\ \mathbf{y} = (y_1, y_2 \ldots, y_n).$$

The metric space $(\mathbb{R}^n, d)$ is called the **euclidean n-space**.

Definition B.5.4. *Let (X, d) be a metric space and $\emptyset \ne A \subseteq X$.*

(i) A is said to be **bounded** *if there exists $M \ge 0$ such that $d(x, y) \le M$ for all $x, y \in A$.*

(ii) If A is bounded, the **diameter** *of A is defined by*

$$diam(A) = \sup\{d(x, y) \mid x, y \in A\}. \tag{B.1}$$

Let (X, d) be a metric space. Define

$$\overline{d} : X \times X \to [0, \infty), \quad \overline{d}(x, y) = \min\{d(x, y), 1\} \tag{B.2}$$

Proposition B.5.5. *$\overline{d}$ is a metric on X that induces the same topology as d.*

Proof. See [74, Theorem 20.1, p.121]. □

The metric $\overline{d}$ is called the **standard bounded metric** corresponding to d. Thus, $(X, \overline{d})$ is bounded.

Definition B.5.6. *If X is a topological space, X is said to be* **metrizable** *if there exists a metric d on X that induces the topology of X.*

Thus, a metric space is a metrizable topological space together with a specific metric d that gives the topology of X.

Proposition B.5.7. *Let X be a metrizable space.*

(i) X is Hausdorff.

(ii) If $A \subseteq X$ and $x \in X$, then $x \in \overline{A}$ if and only if there is a sequence of points of A converging to x.

Proof. (i) is easy to see.

(ii) See [74, Lemma 21.2, p.129-130].

□

Proposition B.5.8 (Continuity)**.** *Let $f : X \to Y$; let X and Y be metrizable with metrics d_X and d_Y. The following are equivalent*

(i) f is continuous.

(ii) Given $x \in X$ and given $\varepsilon > 0$ there exists $\delta > 0$ such that for all $y \in X$,

$$d_X(x, y) < \delta \quad \Rightarrow \quad d_Y(f(x), f(y)) < \varepsilon.$$

(iii) Given $x \in X$, for every sequence (x_n) in X,

$$\lim_{n\to\infty} x_n = x \quad \Rightarrow \quad \lim_{n\to\infty} f(x_n) = f(x).$$

B.6 Disjoint unions

Let X, Y be topological spaces. Consider the **disjoint union** $X \sqcup Y$ of the sets X, Y. Thus, the points in $X \sqcup Y$ are given by taking all the points of X together with all the points of Y, and thinking of all these points as being distinct. So if the sets X and Y overlap, then each point in the intersection occurs twice in the disjoint union $X \sqcup Y$. We can therefore think of X as a subset of $X \sqcup Y$ and we can think of Y as a subset of $X \sqcup Y$, and these two subsets do not intersect.

Define a topology on $X \sqcup Y$ by

$$\mathcal{T} = \{A \cup B \mid A \text{ open in } X,\ B \text{ open in } Y\}.$$

It is easy to see that both X and Y are clopen subsets of $X \sqcup Y$.

Remark B.6.1. *Formally,* $X \sqcup Y = \{(x,1) \mid x \in X\} \cup \{(y,2) \mid y \in Y\}$, $j_1 : X \to X \sqcup Y$, $j_1(x) = (x,1)$ *and* $j_2 : Y \to X \sqcup Y$, $j_2(y) = (y,2)$ *are the canonical embeddings, and*

$$\mathcal{T} = \{j_1(A) \cup j_2(B) \mid A \text{ open in } X,\ B \text{ open in } Y\}.$$

Proposition B.6.2. *(i)* $X \sqcup Y$ *is Hausdorff if and only if both* X *and* Y *are Hausdorff.*

(ii) For any topological space Z*, a map* $f : X \sqcup Y \to Z$ *is continuous if and only if its components* $f_1 : X \to Z$, $f_2 : Y \to Z$ *are continuous.*

Proof. See [23, Theorems 5.31, 5.35, 5.36, p.68-70]. □

B.7 Product topology

Let $(X_i)_{i\in I}$ be an indexed family of nonempty topological spaces and $\pi_i : \prod_{i\in I} X_i \to X_i$ be the projections.

Definition B.7.1. *The* **product topology** *is the smallest topology on* $\prod_{i\in I} X_i$ *for which all the projections* $\pi_i\,(i \in I)$ *are continuous. In this topology,* $\prod_{i\in I} X_i$ *is called a* **product space**.

Let us define, for $i \in I$

$$\begin{aligned} \mathcal{S}_i &:= \{\pi_i^{-1}(U) \mid U \text{ is open in } X_i\} \\ &= \left\{\prod_{j\in I} U_j \mid U_i \text{ is open in } X_i \text{ and } U_j = X_j \text{ for } j \neq i\right\} \end{aligned}$$

and let $\mathcal{S}$ denote the union of these collections,

$$\mathcal{S} := \bigcup_{i\in I} \mathcal{S}_i. \tag{B.3}$$

Then S is a subbasis for the product topology on $\prod_{i\in I} X_i$.
Furthermore, if we define

$$\begin{aligned} \mathcal{B} := \Bigg\{ &\prod_{i\in I} U_i \mid U_i \text{ is open in } X_i \text{ for each } i \in I \text{ and } U_i = X_i \\ &\text{for all but finitely many values of } i \in I \Bigg\}, \end{aligned}$$

then $\mathcal{B}$ is the basis generated by $\mathcal{S}$ for the product topology .

Proposition B.7.2. *(i) Suppose that the topology on each space* X_i *is given by a basis* $\mathcal{B}_i$*. Then the collection* $\mathcal{B} := \Big\{ \prod_{i\in I} B_i \mid B_i \in \mathcal{B}_i$ *for finitely many indices* $i \in I$ *and* $B_i = X_i$ *for the remaining indices* $\Big\}$ *is a basis for the product topology.*

(ii) *Suppose that the topology on each space* X_i *is given by a subbasis* $\mathcal{C}_i$*. Then the collection* $\mathcal{C} := \bigcup_{i\in I}\{\pi_i^{-1}(U) \mid U \in \mathcal{C}_i\}$ *is a subbasis for the product topology.*

Proof. (i) See [74, Theorem 19.2, p.116].

(ii) See [26, 1.2, p.99].

□

Proposition B.7.3.

(i) *For any topological space* Y*, a map* $f : Y \to \prod_{i\in I} X_i$ *is continuous if and only if each of its components* $f_i : Y \to X_i$, $f_i = \pi_i \circ f$ *is continuous.*

(ii) *If each* X_i *is Hausdorff, then* $\prod_{i\in I} X_i$ *is Hausdorff.*

(iii) *Let* (x^n) *be a sequence in* $\prod_{i\in I} X_i$ *and* $x \in \prod_{i\in I} X_i$*. Then* $\lim_{n\to\infty} x^n = x$ *if and only if* $\lim_{n\to\infty} x_i^n = x_i$ *for all* $i \in I$*, where* $x_i^n := \pi_i(x^n), x_i := \pi_i(x)$*.*

Proof. (i) See [74, Theorem 19.6, p.117].

(ii) See [74, Theorem 19.4, p.116].

(iii) See [74, Exercise 6, p.118].

□

Proposition B.7.4. *Let* $(f_i : X_i \to Y_i)_{i\in I}$ *be a family of functions and*

$$\prod_{i\in I} f_i : \prod_{i\in I} X_i \to \prod_{i\in I} Y_i, \quad \prod_{i\in I} f_i((x_i)_{i\in I}) = (f_i(x_i))_{i\in I}$$

be the product function. If each f_i *is continuous (resp. a homeomorphism), then* $\prod_{i\in I} f_i$ *is continuous (resp. a homeomorphism).*

Proof. See [26, Theorem 2.5, p.102]. □

B.7.1 Metric spaces

Proposition B.7.5. *Let* $(X_1, d_1), \ldots, (X_n, d_n)$ *be metric spaces. Then*

$$d : \prod_{i=1}^{n} X_i \times \prod_{i=1}^{n} X_i \to [0,\infty), \quad d(x,y) := \max_{i=1,\ldots,n} d_i(x_i, y_i) \tag{B.4}$$

is a metric that induces the product topology on $\prod_{i=1}^{n} X_i$*.*

Proof. See [74, Ex 3, p. 133]. □

Proposition B.7.6. *Any countable product of metric spaces is metrizable.*

Proof. See, for example, [58, Theorem 14, p. 122]. □

B.8 Quotient topology

Definition B.8.1. *Let X and Y be topological spaces and $p : X \to Y$ be a surjective map. The map p is said to be a* **quotient map** *provided a subset U of Y is open if and only if $p^{-1}(U)$ is open.*

The condition is stronger than continuity; some mathematicians call it "*strong continuity*". An equivalent condition is to require that a subset F of Y is closed if and only if $p^{-1}(F)$ is closed.

Now we show that the notion of quotient map can be used to construct a topology on a set.

Definition B.8.2. *Let X be a topological space, Y be any set and $p : X \to Y$ be a surjective map. There is exactly one topology $\mathcal{Q}$ on X relative to which p is a quotient map; it is called the* **quotient topology** *induced by p.*

The topology $\mathcal{Q}$ is of course defined by

$$\mathcal{Q} := \{U \subseteq Y \mid p^{-1}(U) \text{ is open in } X\}. \tag{B.5}$$

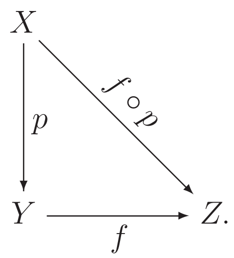
It is easy to check that $\mathcal{Q}$ is a topology. Furthermore, the quotient topology is the largest topology on Y for which p is continuous

Proposition B.8.3. *If $p : X \to Y$ is a surjective continuous map that is either open or closed, then p is a quotient map.*

Proposition B.8.4 (Characteristic property of quotient maps)**.**
Let X and Y be topological spaces and $p : X \to Y$ be a surjective map. The following are equivalent:

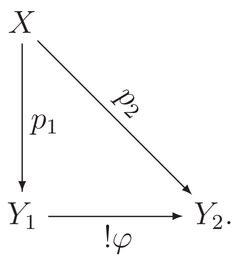
(i) p is a quotient map;

(ii) for any topological space Z and any map $f : Y \to Z$, f is continuous if and only if the composite map $f \circ p$ is continuous:



Proof. See [67, Theorem 3.29, p.56] and [67, Theorem 3.31, p.57]. □

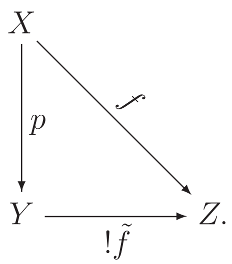
Proposition B.8.5 (Uniqueness of quotient spaces)**.**
Suppose $p_1 : X \to Y_1$ *and* $p_2 : X \to Y_2$ *are quotient maps that make the same identifications (i.e.,* $p_1(x) = p_1(z)$ *if and only if* $p_2(x) = p_2(z)$*). Then there is a unique homeomorphism* $\varphi : Y_1 \to Y_2$ *such that* $\varphi \circ p_1 = p_2$*.*



Proof. See [67, Corollary 3.32, p.57-58]. □

Proposition B.8.6. (Passing to the quotient) *Suppose* $p : X \to Y$ *is a quotient map,* Z *is a topological space and* $f : X \to Z$ *is a map that is constant on the fibers of* p *(i.e.* $p(x) = p(z)$ *implies* $f(x) = f(z)$*). Then there exists a unique map* $\tilde{f} : Y \to Z$ *such that* $f = \tilde{f} \circ p$*.*

The induced map $\tilde{f}$ *is continuous if and only if* f *is continuous;* $\tilde{f}$ *is a quotient map if and only if* f *is a quotient map.*



Proof. [67, Corollary 3.30, p.56], [74, Theorem 22.2, p.142]. □

The most common source of quotient maps is the following construction. Let $\equiv$ be an equivalence relation on a topological space X. For each $x \in X$ let $[x]$ denote the equivalence class of x, and let $X/\equiv$ denote the set of equivalence classes. Let $\pi : X \to X/\equiv$ be the natural projection sending each element of X to its equivalence class. Then $X/\equiv$ together with the quotient topology induced by π is called **the quotient space of** X **modulo** $\equiv$.

One can think of $X/\equiv$ as having been obtained by "identifying" each pair of equivalent points. For this reason, the quotient space $X/\equiv$ is often called an **identification space**, or a **decomposition space** of X.

We can describe the topology of $X/\equiv$ in another way. A subset U of $X/\equiv$ is a collection of equivalence classes, and the set $p^{-1}(U)$ is just the union of the equivalence classes belonging to U. Thus, the typical open set of $X/\equiv$ is a collection of equivalence classes whose **union** is an open set of X.

Any equivalence relation on X determines a partition of X, that is a decomposition of X into a collection of disjoint subsets whose union is X. Hence, alternatively, a quotient space can be defined by explicitly giving a partition of X. Thus, let $X^\star$ be a partition of X into and $\pi : X \to X^\star$ be the surjective map that carries each point of X to the unique element of $X^\star$ containing it. Then $X^\star$ together with the quotient topology induced by π is called also a **quotient space of** X.

Whether a given quotient space is defined in terms of an equivalence relation or a partition is a matter of convenience.

B.9 Complete regularity

Definition B.9.1. *[74, p. 211]*
A topological space X is **completely regular** *if it satisfies the following:*

(i) One-point sets are closed in X.

(ii) For each point $x_0 \in X$ and each closed set A not containing x_0, there is a continuous function $f : X \to [0,1]$ such that $f(x_0) = 1$ and $f(A) = \{0\}$.

B.10 Compactness

Definition B.10.1. *An* **open cover** *of X is a collection of open sets that cover X.*

Definition B.10.2. *A topological space X is said to be* **compact** *if every open cover $\mathcal{A}$ of X contains a finite subcover of X.*

Proposition B.10.3 (Equivalent characterizations)**.**
Let X be a topological space. The following are equivalent:

(i) X is compact.

(ii) For every collection $\mathcal{C}$ of nonempty closed sets in X having the finite intersection property, the intersection $\bigcap \mathcal{C}$ of all the elements of $\mathcal{C}$ is nonempty.

Proof. See [74, Theorem 26.9, p.169]. □

Corollary B.10.4. *If $\mathcal{C}$ is a chain (i.e. totally ordered by inclusion) of nonempty closed subsets of a compact space X, then the intersection $\bigcap \mathcal{C}$ is nonempty.*

Proof. It is easy to see that $\mathcal{C}$ has the finite intersection property. □

As an immediate consequence, we get

Corollary B.10.5. *If $(C_n)_{n\geq 0}$ is a decreasing sequence of nonempty closed subsets of a compact space X, then the intersection $\bigcap\limits_{n\geq 0} C_n$ is nonempty.*

Proposition B.10.6.

(i) Any finite topological space is compact.

(ii) Every closed subspace of a compact space is compact.

(iii) Every compact subspace of a Hausdorff space is closed.

(iv) The product of finitely many compact spaces is compact.

(v) $X \sqcup Y$ is a compact space if and only if both X and Y are compact spaces.

(vi) The image of a compact space under a continuous map is compact.

Proof. (i) Obviously.

(ii) See [74, Theorem 26.2, p.165].

(iii) See [74, Theorem 26.3, p.165].

(iv) See [74, Theorem 26.7, p.167].

(v) See [74, Exercise 3, p.171].

(vi) See [74, Theorem 26.5, p.166].

□

Proposition B.10.7. *Let X be a compact space.*

(i) If $x \in X$ and U is an open neighborhood of x, then there exists an open neighborhood V of x such that $\overline{V} \subseteq U$.

Proposition B.10.8. *Let X be a compact space. Then for any disjoint open cover $(U_i)_{i\in I}$ of X we have that $U_i \neq \emptyset$ for a finite number of i. In particular, if $(U_n)_{n\geq 1}$ is a countable disjoint open cover of X, then there exists $N \geq 1$ such that $U_n = \emptyset$ for all $n \geq N$.*

Proof. Let $(U_i)_{i\in I}$ be a disjoint cover of X. Since X is compact, we have that $X = U_{i_1} \cup \ldots U_{i_n}$ for some $i_1, \ldots, i_n \in I$. Let $i \not\in \{i_1, \ldots, i_n\}$. Since $U_i \cap U_{i_k} = \emptyset$ for all $k = 1, \ldots, n$, it follows that $U_i \cap X = \emptyset$, hence we must have $U_i = \emptyset$. □

Theorem B.10.9 (Tychonoff Theorem).
An arbitrary product of compact spaces is compact in the product topology.

Proof. See [74, Theorem 37.3, p.234]. □

Theorem B.10.10 (Heine-Borel Theorem).
A subspace A of the euclidean space $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. See [74, Theorem 27.3, p.173]. □

Theorem B.10.11. *Let X be a compact Hausdorff space. The following are equivalent:*

(i) X is metrizable.

(ii) X is **second-countable**, *that is X has a countable basis for its topology.*

Proof. See [74, Ex. 3, p.218]. □

B.10.1 Sequential compactness

Definition B.10.12. *A topological space X is* **sequentially compact** *if every sequence of points of X has a convergent subsequence.*

Proposition B.10.13. *If X is metrizable, then X is compact if and only if it is sequentially compact.*

Proof. See [74, Theorem 28.2, p.179]. □

B.10.2 Total boundedness

Definition B.10.14. *A metric space (X, d) is said to be* **totally bounded** *if for every $\varepsilon > 0$ there is a finite cover of X by ε-balls.*

Proposition B.10.15. *A metric space (X, d) is compact if and only if it is complete and totally bounded.*

Proof. See [74, Theorem 45.1, p.276]. □

B.10.3 Stone-Čech compactification

Definition B.10.16. *A* **compactification** *of a topological space X is a compact Hausdorff space Y containing X as a subspace such that $\overline{X} = Y$. Two compactifications Y_1 and Y_2 of X are said to be* **equivalent** *if there is a homeomorphism $h : Y_1 \to Y_2$ such that $h(x) = x$ for every $x \in X$.*

Proposition B.10.17. *Let X be a completely regular space. There exists a compactification βX of X having the following properties:*

(i) βX satisfies the following **extension property***: Given any continuous map $f : X \to C$ of X into a compact Hausdorff space C, the map f extends uniquely to a continuous map $\tilde{f} : \beta X \to C$.*

(ii) Any other compactification Y of X satisfying the extension property is equivalent with βX.

Proof. See [74, Theorem 38.4, p.240] and [74, Theorem 38.5, p.240]. □

βX is called the **Stone-Čech compactification** of X.

Proposition B.10.18. *Let X and Y be completely regular spaces. Then any continuous mapping $f : X \to Y$ extends uniquely to a continuous function $\beta f : \beta X \to \beta Y$.*

B.11 Baire category

Definition B.11.1. *[97, 20.6, p. 532] Let X be a topological space. A set $A \subseteq X$ is* **meager**, *or of the* **first category of Baire**, *if it is the union of countably many nowhere dense sets.*

A set that is not meager is called **nonmeager**, *or of the* **second category of Baire**.

Thus, every set is either of first or second category.

Definition B.11.2. *[97, 20.6, p. 532] A set A is* **residual** *(or* **comeager** *or* **generic***) if $X \setminus A$ is meager.*

Lemma B.11.3. *Let X be a topological space.*

(i) A is meager iff A is contained in the union of countably many closed sets having empty interiors.

(ii) A is comeager iff A contains the intersection of countably many open dense sets.

Definition B.11.4. *A topological space X is said to be a* **Baire space** *if the following condition holds:*

Given any countable collection $(F_n)_{n\geq 1}$ of closed sets each of which has empty interior, their union $\bigcup_{n\geq 1} F_n$ has empty interior.

Proposition B.11.5 (Equivalent characterizations)**.** *Let X be a topological space. The following are equivalent:*

(i) X is a Baire space.

(ii) Given any countable collection $(G_n)_{n\geq 1}$ of open dense subsets of X, their intersection $\bigcap_{n\geq 1} G_n$ is also dense in X.

(iii) Any residual subset of X is dense in X.

(iv) Any meager subset of X has empty interior.

(v) Any nonempty open subset of X is nonmeager.

Proof. See [97, 20.15, p. 537]. □

An immediate consequence of Proposition B.11.5.(iii) is the following

Corollary B.11.6. *Any residual subset of a Baire space is nonempty.*

We may think of the meager sets as "small" and the residual sets as "large". Although "large" is a stronger property than "nonempty", in some situations the most conveninet way to prove that some set A is nonempty is by showing the set is "large". That is one way in which the above corollary is used.

The most important result about Baire spaces is

Theorem B.11.7 (Baire Category Theorem)**.** *If X is a compact Hausdorff space or a complete metric space, then X is a Baire space.*

Proof. See [74, Theorem 48.2, p. 296]. □

B.12 Covering maps

Definition B.12.1. *Let $p : Y \to Y$ be a continuous surjective map. The open set U of Y is said to be* **evenly covered** *by p if the inverse image $p^{-1}(U)$ can be written as the union of disjoint open sets V_α in X such that for each α, the restriction of p to V_α is a homeomorphism of V_α onto U. The collection (V_α) will be called a partition of $p^{-1}(U)$ into* **slices***.*

Definition B.12.2. *Let $p : Y \to Y$ be a continuous surjective map. If every point of Y has an open neighborhood U that is evenly covered by p. then p is called a* **covering map***, and Y is said to be a* **covering space** *of X.*

Lemma B.12.3. *Any covering map is a local homeomorphism, but the converse does not hold.*

Proof. See [74, Example 2, p.338]. □

Proposition B.12.4. *The map*

$$\varepsilon : \mathbb{R} \to \mathbb{S}^1, \quad \varepsilon(t) = e^{2\pi i t}. \tag{B.6}$$

is a covering map.

Proof. See [74, Theorem 53.3, p.339] or [67, Lemma 8.5, p.183]. □

Appendix C

Measure Theory

C.1 Set systems

C.1.1 Semirings

Definition C.1.1. *A collection $\mathcal{S}$ of subsets of X is called a **semiring on/over** X if*

(i) $\emptyset \in \mathcal{S}$.

(ii) If $A, B \in \mathcal{S}$, then $A \cap B \in \mathcal{S}$.

(iii) If $A, B \in \mathcal{S}$, $A \subseteq B$, then there exist disjoint $C_1, \ldots, C_n \in \mathcal{S}$ such that $B \setminus A = C_1 \cup \ldots \cup C_n$.

C.1.2 Algebras and semialgebras

Definition C.1.2. *A collection $\mathcal{S}$ of subsets of X is called a **semialgebra on** X if*

(i) $\emptyset \in \mathcal{S}$.

(ii) If $A, B \in \mathcal{S}$, then $A \cap B \in \mathcal{S}$.

(iii) If $A \in \mathcal{S}$, then there exist pairwise disjoint subsets $C_1, \ldots, C_n \in \mathcal{S}$ such that $X \setminus A = C_1 \cup \ldots \cup C_n$.

Lemma C.1.3. *Any semialgebra is a semiring.*

Proof. Let $\mathcal{S}$ be a semialgebra and $A, B \in \mathcal{S}, A \subseteq B$. There are then $C_1, \ldots, C_n$ pairwise disjoint such that $X \setminus A = C_1 \cup \ldots \cup C_n$. It follows that

$$B \setminus A = B \cap (X \setminus A) = B \cap (C_1 \cup \ldots \cup C_n) = (B \cap C_1) \cup \ldots (B \cap C_n).$$

□

Definition C.1.4. *A collection $\mathcal{A}$ of subsets of X is called an* ***algebra*** *or a* ***field*** *on X if*

(i) $X \in \mathcal{A}$.

(ii) If $A, B \in \mathcal{A}$, *then* $A \cup B \in \mathcal{A}$.

(iii) If $A \in \mathcal{A}$, *then* $X \setminus A \in \mathcal{A}$.

The intersection of any family of algebras on a set X is again an algebra on X.

Definition C.1.5. *Let $\mathcal{C}$ be a collection of subsets of X. The* ***algebra generated by $\mathcal{C}$ on*** *X, denoted by $\mathcal{A}(\mathcal{C})$, is the intersection of all algebras in X containing $\mathcal{C}$.*

Proposition C.1.6. *Let $\mathcal{C}$ be a collection of subsets of X. Then*

$$\mathcal{A}(\mathcal{C}) = \text{ the class of sets of the form } \bigcup_{i=1}^{m} \bigcap_{j=1}^{n_i} A_{ij}, \tag{C.1}$$

where for each (i, j) pair either A_{ij} or $X \setminus A_{ij}$ is in $\mathcal{C}$, and where $\bigcap_{j=1}^{n_1} A_{1j}, \ldots, \bigcap_{j=1}^{n_m} A_{mj}$ are pairwise disjoint.

Proof. See [110, Ex. 10, p.13]. □

Proposition C.1.7. *Let $\mathcal{S}$ be a semialgebra on X. Then*

$$\mathcal{A}(\mathcal{S}) = \text{ the class of sets of the form } \bigcup_{i=1}^{m} A_i, \tag{C.2}$$

where each $A_i \in \mathcal{S}$ and $A_1, \ldots, A_m$ are pairwise disjoint.

Proof. See [114, Theorem 0.1, p.4] □

Proposition C.1.8. *Let $\mathcal{C}$ be a collection of subsets of X. For any nonempty subset B of X,*

$$\mathcal{A}(\mathcal{C}) \cap B = \mathcal{A}_B(\mathcal{C} \cap B),$$

where $\mathcal{A}_B(\mathcal{C} \cap B)$ denotes the algebra generated by $\mathcal{C} \cap B$ in B.

Proof. □

C.2 σ-algebras

Definition C.2.1. *A collection $\mathcal{B}$ of subsets of X is said to be a* ***σ-algebra on*** *X if*

(i) $X \in \mathcal{B}$.

(ii) If $A \in \mathcal{B}$, *then* $X \setminus A \in \mathcal{B}$.

(iii) If $(\{A_n\})_{n\geq 1}$ *is a sequence in* $\mathcal{B}$, *then* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$.

The pair $(X, \mathcal{B})$ is called a ***measurable space****, and the sets in $\mathcal{B}$ are called the* ***measurable sets****.*

Proposition C.2.2. *Let $\mathcal{B}$ be a σ-algebra on X.*

(i) If $A_1, \ldots, A_n \in \mathcal{B}$, *then* $\bigcup_{k=1}^{n} A_k, \bigcap_{k=1}^{n} A_k \in \mathcal{B}$.

(ii) If $A, B \in \mathcal{B}$, *then* $A \setminus B \in \mathcal{B}$.

(iii) If $(A_n)_{n\geq 1}$ *is a sequence of sets in* $\mathcal{B}$, *then*

(a) $\bigcap_{n\geq 1} A_n \in \mathcal{B}$.

(b) $\limsup_{n\to\infty} A_n, \liminf_{n\to\infty} A_n \in \mathcal{B}$. *In particular, if* $\lim_{n\to\infty} A_n$ *exists, then* $\lim_{n\to\infty} A_n \in \mathcal{B}$.

Proof. See [110, Section 1.3, p.9]. □

Thus σ-algebras are closed under the application of countably many of the standard set manipulations. The standard set operations are union, intersection, complementation, difference, and symmetric difference, and all of these can be expressed in terms of unions and complements. Thus, when one works with a collection of sets in a σ-algebra , one will never by using at most countably many set operations on these sets produce a set outside the σ-algebra.

C.2.1 Generated σ-algebras

Proposition C.2.3. *If* $(\mathcal{B}_i)_{i\in I}$ *is a family of σ-algebras on X, then* $\bigcap_{i\in I} \mathcal{B}_i$ *is a σ-algebra on X.*

Definition C.2.4. *Let $\mathcal{C}$ be a collection of subsets of a set X. The* ***σ-algebra generated by $\mathcal{C}$ on*** *X, denoted by $\sigma(\mathcal{C})$, is the intersection of all algebras in X containing $\mathcal{C}$.*

Proposition C.2.5. *Let $\mathcal{C}$ be a collection of subsets of X. Then*

(i) If $\mathcal{C} \subseteq \mathcal{D} \subseteq \sigma(\mathcal{C})$, *then* $\sigma(\mathcal{D}) = \sigma(\mathcal{C})$.

(ii) If $\mathcal{C}$ is finite, then $\sigma(\mathcal{C}) = \mathcal{A}(\mathcal{C})$.

(iii) $\sigma(\mathcal{C}) = \sigma(\mathcal{A}(\mathcal{C}))$.

(iv) For any nonempty subset B of X,

$$\sigma(\mathcal{C}) \cap B = \sigma_B(\mathcal{C} \cap B),$$

where $\sigma_B(\mathcal{C} \cap B)$ denotes the σ-algebra generated by $\mathcal{C} \cap B$ in B.

Proof. See [110, Ex. 9, p.13] and [110, Ex. 17, p.14]. □

Given $\mathcal{C}$ a

The following result is called **Halmos Monotone Class theorem** and is very useful.

Proposition C.2.6. *Let $\mathcal{A}$ be an algebra on X. Then $\sigma(\mathcal{A})$ coincides with the monotone class generated by $\mathcal{A}$. Hence, if a monotone class contains $\mathcal{A}$, then it contains $\sigma(\mathcal{A})$.*

Proof. See [110, Ex. 21, p.14-15] or [50, Theorem B. p.27]. □

C.3 Countable products of probability spaces

Let $(X_n, \mathcal{B}_n, \mu_n)$, $n \in \mathbb{Z}$ be probability spaces. Their **direct product** is defined as follows.

Let $X = \prod_{n\in\mathbb{Z}} X_n$. We shall denote with boldface letters $\mathbf{x} = (x_n)_{n\in\mathbb{Z}}, \mathbf{y}, \mathbf{z}, \ldots$ the elements of X. For every $n \in \mathbb{Z}$, let

$$\pi_n : X \to X_n, \quad \pi_n(\mathbf{x}) = x_n \tag{C.3}$$

be the nth-projection.

An **elementary measurable rectangle** is a set of the form

$$R_n^A = \pi_n^{-1}(A) = \{\mathbf{x} \in X \mid x_n \in A\}, \quad \text{where } n \in \mathbb{Z},\ A \in \mathcal{B}_n.$$

A **measurable rectangle** is a set of the form

$$R_{n_1,\ldots,n_t}^{A_1,\ldots,A_t} = \{\mathbf{x} \in X \mid x_{n_i} \in A_i \text{ for all } i = 1, \ldots, t\} = \bigcap_{i=1}^{t} R_{n_i}^{A_i},$$

where $t \geq 1$, $n_1 < n_2 < \ldots < n_t \in \mathbb{Z}$, and $A_i \in \mathcal{B}_{n_i}$ for all $i = 1, \ldots, t$.

The **product σ-algebra**, denoted by $\bigotimes_{n\in\mathbb{Z}} \mathcal{B}_n$, is the σ-algebra generated by the set of all measurable rectangles. We write

$$\left(X, \mathcal{B} = \bigotimes_{n\in\mathbb{Z}} \mathcal{B}_n\right) = \prod_{n\in\mathbb{Z}} (X_n, \mathcal{B}_n). \tag{C.4}$$

There is a unique probability measure μ on $(X, \mathcal{B})$ such that

$$\mu(R_{n_1,\ldots,n_t}^{A_1,\ldots,A_t}) = \prod_{i=1}^{t} \mu_{n_i}(A_i). \tag{C.5}$$

We write $\mu = \bigotimes_{n\in\mathbb{Z}} \mu_n$ and call it the **product** of $\mu_n, n \in \mathbb{Z}$.

Then $(X, \mathcal{B}, \mu)$ is a probability space, called the **direct product** of probability spaces $(X_n, \mathcal{B}_n, \mu_n),\ n \in \mathbb{Z}$. We write

$$\left(X, \mathcal{B} = \bigotimes_{n\in\mathbb{Z}} \mathcal{B}_n, \mu = \bigotimes_{n\in\mathbb{Z}} \mu_n\right) = \prod_{n\in\mathbb{Z}} (X_n, \mathcal{B}_n, \mu_n). \tag{C.6}$$

C.4 Set functions

A **set function** is a function defined on a nonempty collection of sets. In the sequel, $\mathcal{C}$ is a collection of sets containing the empty set $\emptyset$ and $\mu : \mathcal{C} \to [0, \infty]$.

Definition C.4.1. *(i) μ is called **finitely additive** if $\mu(\emptyset) = 0$ and*

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i) \tag{C.7}$$

for all $n \geq 1$ and all pairwise disjoint sets $A_1, \ldots, A_n \in \mathcal{C}$ such that $\bigcup_{i=1}^{n} A_i \in \mathcal{C}$.

*(ii) μ is called **finitely subadditive** if*

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mu(A_i) \tag{C.8}$$

for all $n \geq 1$ and all sets $A_1, \ldots, A_n \in \mathcal{C}$ such that $\bigcup_{i=1}^{n} A_i \in \mathcal{C}$.

*(iii) μ is called **countably additive** if $\mu(\emptyset) = 0$ and*

$$\mu\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mu(A_n) \tag{C.9}$$

for all sequences $(A_n)_{n\geq 1}$ of pairwise disjoint sets in $\mathcal{C}$ such that $\bigcup_{n\geq 1} A_n \in \mathcal{C}$.

(iv) μ *is called* ***countably subadditive*** *if*

$$\mu\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} \mu(A_n) \tag{C.10}$$

for all sequences $(A_n)_{n\geq 1}$ *in* $\mathcal{C}$ *such that* $\bigcup_{n\geq 1} A_n \in \mathcal{C}$.

Definition C.4.2. μ *is* σ***-finite*** *if there exists a sequence* $(A_n)_{n\geq 1}$ *of members of* $\mathcal{C}$ *such that* $X = \bigcup_{n\geq 1} A_n$ *and* $\mu(A_n) < \infty$ *for all* $n \geq 1$.

C.5 Measure spaces

Definition C.5.1. *Let* $(X, \mathcal{B})$ *be a measurable space. A* ***measure*** *on* $\mathcal{B}$ *is a countably additive set function* $\mu : \mathcal{B} \to [0, \infty]$.

Definition C.5.2. *A* ***measure space*** *is a triple* $(X, \mathcal{B}, \mu)$, *where* $(X, \mathcal{B})$ *is a measurable space and* μ *is a measure on* $\mathcal{B}$.

Definition C.5.3. *Let* $(X, \mathcal{B}, \mu)$ *be a measure space.*

(i) If μ *is* σ*-finite, then* $(X, \mathcal{B}, \mu)$ *is called a* σ***-finite measure space***.

(ii) μ *is* ***finite*** *if* $\mu(X) < \infty$. *In this case,* $(X, \mathcal{B}, \mu)$ *is called a* ***finite measure space***.

(iii) μ *is a* ***probability measure*** *if* $\mu(X) = 1$. *In this case,* $(X, \mathcal{B}, \mu)$ *is called a* ***probability space***.

Proposition C.5.4. *Let* $(X, \mathcal{B}, \mu)$ *be a measure space, and* $A, B \in \mathcal{B}$.

(i) μ *is finitely additive.*

(ii) $A \subseteq B$ *implies* $\mu(A) \leq \mu(B)$. *Furthermore, if* $\mu(A) < \infty$ *or* $\mu(B) < \infty$, *then* $\mu(B \setminus A) = \mu(B) - \mu(A)$.

(iii) μ *is countably subadditive and finitely subadditive.*

(iv) $\mu(A \Delta B) = 0$ *if and only if* $\mu(A) = \mu(B) = \mu(A \cap B)$.

(v) If $\mu(A) = 0$, *then* $\mu(A \cup B) = \mu(A)$, $\mu(A \Delta B) = 0$ *and* $\mu(B \setminus A) = \mu(B)$.

Proof. (i) See [110, (M4), p.37].

(ii) See [110, (M5), p.38].

(iii) See [110, (M7), p.40].

(iv) See [110, Ex. 10(b), p.41].

(v) See [110, Ex. 10(c),(d), p.41].

□

Proposition C.5.5. *Let $(X, \mathcal{B}, \mu)$ be a finite measure space.*

(i) For all $A, B \in \mathcal{B}$, $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B$.

(ii) For every sequence $(A_n)_{n\geq 1}$ in $\mathcal{B}$ such that $\lim\limits_{n\to\infty} A_n$ exists, we have that $\mu\left(\lim\limits_{n\to\infty} A_n\right) = \lim\limits_{n\to\infty} \mu(A_n)$.

Proof. (i) See [110, (M6), p.39].

(ii) See [110, (M12), p.48].

□

C.5.1 Dirac probability measure

Let $(X, \mathcal{B})$ be a measurable space. Each $x \in X$ defines a measure $\delta_x : \mathcal{B} \to [0, 1]$ by

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

The measure δ_x is called the **Dirac probability measure** on X defined by $x \in X$.

C.5.2 Measures in topological spaces

Let X be a topological space.

The **Borel σ-algebra** on X, denoted by $\mathcal{B}(X)$, is the σ-algebra generated by the open sets of X. By a **Borel** (probability) measure on X we shall understand a probability (measure) $\mu : \mathcal{B}_X \to [0, 1]$. By a **Borel (probability) space** we mean a probability space $(X, \mathcal{B})X), \mu)$, where μ is a Borel (probability) measure on X.

If X and Y are Borel spaces, a measurable mapping $T : X \to Y$ is called **Borel measurable**.

Proposition C.5.6. *Let $Y \subseteq X$. Then $\mathcal{B}(Y) = Y \cap \mathcal{B}(X)$.*

Proof. See [80, Theorem 1.9, p.5]. □

Proposition C.5.7. *[81, Proposition 2.3.4] Let X be a compact space and let $\mathcal{A}$ be an algebra of clopen subsets of X. Then any finitely additive set function $\mu : \mathcal{A} \to [0, \infty]$ is countably additive.*

Proof. Let $(A_n)_{n\geq 1}$ be a sequence of disjoint sets in $\mathcal{A}$ such that $A = \bigcup_{n\geq 1} A_n$ is in $\mathcal{A}$. We have to show that $\mu\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

Since A is compact as a closed subset of the compact space X, by B.10.8, we get $N \geq 1$ such that $A_n = \emptyset$ (hence $\mu(A_n) = 0$) for all $n \geq N$.

Using the fact that μ is finitely additive, it follows that

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mu(A_n) = \sum_{n\geq 1} \mu(A_n). \tag{C.11}$$

□

C.6 Extensions of measures

Let $\mathcal{C}$ be a collection of subsets of X containing $\emptyset$ and $\mu : \mathcal{C} \to [0, \infty]$ be a set function such that $\mu(\emptyset) = 0$. Define $\mu^\star : \mathcal{P}(X) \to [0, \infty]$ by

$$\mu^\star(A) = \inf\left\{\sum_{n=1}^{\infty} \mu(A_n) \mid A_1, A_2 \ldots \in \mathcal{C}, A \subseteq \bigcup_{n\geq 1} A_n\right\}. \tag{C.12}$$

If it happens to be the case that there is no sequence of sets in $\mathcal{C}$ whose union contains A, we define $\mu^\star(A) = \infty$.

Let $\tilde{\mu} : \sigma(\mathcal{C}) \to [0, \infty]$ be the restriction of $\mu^\star$ to $\sigma(\mathcal{C})$.

Theorem C.6.1 (Carathéodory Extension Theorem)**.**
Let $\mathcal{S}$ be a semiring on X and $\mu : \mathcal{S} \to [0, \infty]$ be finitely additive and countably subadditive.

(i) $\tilde{\mu}$ is a measure on $\sigma(\mathcal{C})$ that extends μ, i.e. $\tilde{\mu}(A) = \mu(A)$ for all $A \in \mathcal{S}$.

(ii) If μ is σ-finite on $\mathcal{S}$, then $\tilde{\mu}$ is the unique measure on $\sigma(\mathcal{S})$ extending μ. Furthermore, $\tilde{\mu}$ is also σ-finite.

Proof. See [110, p. 75] and [110, Claim 3, p.85]. □

Theorem C.6.2. *Let $\mathcal{A}$ be an algebra on X and $\mu : \mathcal{A} \to [0, \infty]$ be countably additive. Then*

(i) $\tilde{\mu}$ is a measure on $\sigma(\mathcal{A})$ that extends μ.

(ii) If μ is σ-finite on $\mathcal{A}$, then $\tilde{\mu}$ is the unique measure on $\sigma(\mathcal{A})$ extending μ. Furthermore, $\tilde{\mu}$ is also σ-finite.

(iii) If $\mu(X) = 1$, then $\tilde{\mu}$ is a probability measure.

Proof. See [110, Exercise 6, p. 81] or [114, Theorem 0.3, p.4]. □

C.7 Measurable mappings

Definition C.7.1. *Let $(X, \mathcal{B})$, $(Y, \mathcal{C})$ be measurable spaces. A mapping $T : X \to Y$ is said to be* ***measurable*** *if $T^{-1}(\mathcal{C}) \subseteq \mathcal{B}$.*

We should write $T : (X, \mathcal{B}) \to (Y, \mathcal{C})$ since the measurability property depends on $\mathcal{B}, \mathcal{C}$.

Proposition C.7.2. *Let $(X, \mathcal{B})$ and $(Y, \mathcal{C})$ be measurable spaces.*

(i) Let $T : X \to Y$. The following are equivalent

  (a) T is measurable.

  (b) $T^{-1}(A) \in \mathcal{B}$ for every each $A \in \mathcal{A}$, where $\mathcal{A}$ is a collection of subsets of Y that generates $\mathcal{C}$.

Proof. See [110, (MF1'), p.206]. □

Proposition C.7.3. *Let $(X, \mathcal{B})$ and $(Y, \mathcal{C})$ be measurable spaces.*

(i) $1_X : X \to X$ is measurable.

(ii) The composition of measurable functions is measurable.

Notation C.7.4. *Let $(X, \mathcal{B})$ be a measurable space.*

(i) $\mathcal{M}_{\mathbb{C}}(X, \mathcal{B})$ is the set of all complex-valued measurable functions $f : (X, \mathcal{B}) \to (\mathbb{C}, \mathcal{B}(\mathbb{C}))$.

(ii) $\mathcal{M}_{\mathbb{R}}(X, \mathcal{B})$ is the set of all real-valued measurable functions $f : (X, \mathcal{B}) \to (\mathbb{C}, \mathcal{B}(\mathbb{R}))$.

Proposition C.7.5. *Let $(X, \mathcal{B})$ be a measurable space and $f : X \to \mathbb{C}$. The following are equivalent*

(i) f is measurable.

(ii) both its real and imaginary parts are measurable.

Proof.
See [93, 1.9 (a),(b), p.11]. □

Proposition C.7.6. *Let $(X, \mathcal{B})$ be a measurable space.*

(i) $A \subseteq X$ is measurable if and only if its characteristic function $\chi_A : X \to \mathbb{R}$ is measurable.

(ii) If $f : X \to \mathbb{C}$ is measurable, then so is $|f|$.

(iii) If $f, g : X \to \mathbb{C}$ are measurable, then so are $f + g$ and fg.

(iv) If $f : X \to \mathbb{C}$ is measurable and $c > 0$, and $g : X \to \mathbb{R}$ is defined by $g(x) = |f(x)|^c$, then g is measurable.

(v) *If* $f, g : X \to \mathbb{C}$ *are measurable, then* $\{x \in X \mid f(x) > g(x)\}$, $\{x \in X \mid f(x) \geq g(x)\}$, $\{x \in X \mid f(x) = g(x)\}$, $\{x \in X \mid f(x) \neq g(x)\}$ *are measurable.*

(vi) *If* $f_n : X \to \mathbb{R}$ *is measurable for* $n \geq 1$*, then* $\sup_{n\geq 1} f_n, \inf_{n\geq 1} f_n, \limsup_{n\to\infty} f_n, \liminf_{n\to\infty} f_n$ *are measurable. If* $\lim_{n\to\infty} f_n$ *exists, then* $\lim_{n\to\infty} f_n$ *is measurable.*

(vii) *If* $f, g : X \to \mathbb{R}$ *are measurable, then* $\max\{f, g\}$, $\min\{f, g\}$ *are also measurable.*

(viii) *If* $f : X \to \mathbb{R}$ *is measurable, then*

$$f^+, f^- : X \to \mathbb{R}, \quad f^+(x) = \max\{f(x), 0\}, \quad f^-(x) = -\min\{f(x), 0\} \tag{C.13}$$

are measurable.

Proof. (i) See [93, 1.9 (d), p.11] or [110, (MF3), p.167].

(ii) See [93, 1.9 (b), p.11].

(iii) See [93, 1.9 (c), p.11].

(iv) See [110, (MF7).(c), p.172].

(v) See [110, (MF8), p.173].

(vi) See [110, (MF11), p.180].

(vii) See [93, Corollaries, p.15].

(viii) See [93, Corollaries, p.15].

□

The nonnegative functions f^+, f^- are called the **positive and negative parts** of f. We have that

$$f = f^+ - f^-, \quad |f| = f^+ + f^-. \tag{C.14}$$

C.7.1 Almost everywhere

Definition C.7.7. *Let* $f, g : X \to \mathbb{C}$ *be measurable. We say that* f *and* g *are* ***equal almost everywhere****, and write* $f = g$ *a.e., if* $\mu\{x \in X \mid f(x) \neq g(x)\} = 0$.

C.7.2 Simple functions

Let $(X, \mathcal{B})$ be a measurable space.

Definition C.7.8. *A function $s : X \to \mathbb{C}$ is said to be **simple** if it has finitely many different values.*

Proposition C.7.9. *Let $s : X \to \mathbb{C}$ be a simple function, $s(X) = \{c_1, \dots, c_n\}$, and denote $A_i = \{x \in X \mid s(x) = c_i\}$ for all $i = 1, \dots, n$. Then*

(i) $s = \sum_{i=1}^{n} c_i \chi_{A_i}$.

(ii) s is measurable if and only if $A_1, \dots, A_n$ are measurable.

Proof. See [93, p.15] or [110, (MF16), p.185]. □

Theorem C.7.10. *Let $f : X \to \mathbb{R}$ be measurable with $f \geq 0$. There exists a sequence $(s_n)_{n\geq 1}$ of measurable simple functions $s_n : X \to \mathbb{R}$ such that*

(i) $0 \leq s_1 \leq s_2 \leq \dots \leq f$.

(ii) $\lim_{n\to\infty} s_n(x) = f(x)$ for all $x \in X$.

Proof. See [93, Theorem 1.17, p.15]. □

C.8 Integration

Let $(X, \mathcal{B}, \mu)$ be a measure space.

C.8.1 Simple functions

If $s : X \to C$ is a measurable simple function of the form

$$s = \sum_{i=1}^{n} c_i \chi_{A_i},$$

where $c_1, \dots, c_n$ are the distinct values of s, and if $E \in \mathcal{B}$, we define

$$\int_E s \, d\mu = \sum_{i=1}^{n} c_i \mu(A_i \cap E). \tag{C.15}$$

The convention $0 \cdot \infty = 0$ is used here; it may happen that $c_i = 0$ for some i and that $\mu(A_i \cap E) = \infty$.

C.8.2 Nonnegative functions

Suppose that $f : X \to \mathbb{R}$ is measurable and $f \geq 0$. For any $E \in \mathcal{B}$ we define

$$\int_E f\, d\mu = \sup_{s \in S_f} \int_E s\, d\mu, \tag{C.16}$$

where S_f is the set of all simple measurable functions s such that $0 \leq s \leq f$.

The left member of (C.16) is called the **Lebesgue integral** of f over E, with respect to the measure μ. It is a number in $[0, \infty]$.

Proposition C.8.1 (Equivalent definition)**.**
For every $E \in \mathcal{B}$,

$$\int_E f\, d\mu \quad = \quad \lim_{n\to\infty} \int_E s_n\, d\mu,$$

where (s_n) is any increasing sequence of measurable simple functions in $\mathcal{S}_f$ such that $\lim_{n\to\infty} s_n(x) = f(x)$ for all $x \in X$.

Proof. [110, Ex.10, p.230-231]. □

Proposition C.8.2. *Let $f, g : X \to \mathbb{R}$ be measurable and nonnegative, $A, B \subseteq X$ be measurable.*

(i) If $f \leq g$, then $\int_A f\, d\mu \leq \int_A g\, d\mu$.

(ii) If $A \subseteq B$, then $\int_A f\, d\mu \leq \int_B f\, d\mu$.

Proof. See [93, 1.24, p.20]. □

Proposition C.8.3 (Fatou's Lemma)**.**
If $f_n : X \to [0, \infty)$ is measurable for $n \geq 1$, then

$$\int_X \liminf_n f_n\, d\mu \leq \liminf_{n\to\infty} \int_X f_n\, d\mu \tag{C.17}$$

Proof. See [93, Theorem 1.28, p.21]. □

C.8.3 Complex-valued functions

Definition C.8.4. *We define $L^1(X, \mathcal{B}, \mu)$ (or $L^1(\mu)$) to be the collection of all measurable functions $f : X \to \mathbb{C}$ for which*

$$\int_X |f|\, d\mu < \infty. \tag{C.18}$$

Remark that $|f| : X \to \mathbb{R}$ is a nonnegative measurable function, hence the above integral is defined.

The members of $L^1(X, \mathcal{B}, \mu)$ are called the **Lebesgue integrable** functions (with respect to μ).

Definition C.8.5. *If $f = u + iv$, where u and v are real measurable functions on X, and if $f \in L^1(X, \mathcal{B}, \mu)$, we define for every measurable subset E of X,*

$$\int_E f\, d\mu = \int_E u^+\, d\mu - \int_E u^-\, d\mu + i\left(\int_E v^+\, d\mu - \int_E v^-\, d\mu\right) \tag{C.19}$$

Here u^+, u^- (resp. v^+, v^-) are the positive and negative parts of u (resp. v). These four functions are measurable, real, and nonnegative; hence the four integrals on the right of (C.19) exist. Furthermore, $u^+ \leq |u| \leq |f|$, etc., so that each of these four integrals is finite. Thus $\int_E f\, d\mu$ is a complex number.

Proposition C.8.6. *Let $f, g : X \to \mathbb{C}$ and $\alpha, \beta \in \mathbb{C}$.*

(i) f is integrable if and only if $|f|$ is integrable.

(ii) If $f, g \in L^1(X, \mathcal{B}, \mu)$, then $(\alpha f + \beta g) \in L^1(X, \mathcal{B}, \mu)$, and

$$\int_X (\alpha f + \beta g)\, d\mu = \alpha \int_X f\, d\mu + \beta \int_X g\, d\mu. \tag{C.20}$$

(iii) If $f \in L^1(X, \mathcal{B}, \mu)$, then

$$\left|\int_X f\, d\mu\right| \leq \int_X |f|\, d\mu. \tag{C.21}$$

Proof. (i) Obviously, by definition.

(ii) See [93, Theorem 1.32, p.25].

(iii) See [93, Theorem 1.33, p.26].

□

Theorem C.8.7 (Lebesgue's Dominated Convergence Theorem)**.**
Let $(f_n)_{n\geq 1}$ denote a sequence of complex measurable functions on X such that

(i) $f(x) = \lim_{n\to\infty} f_n(x)$ exists for every $x \in X$.

(ii) there is $g \in L^1(X, \mathcal{B}, \mu)$ with $|f_n(x)| \leq g(x)$ for every $n \geq 1$, and every $x \in X$.

Then

(i) $f \in L^1(X, \mathcal{B}, \mu)$,

(ii) $\lim_{n\to\infty} \int_X |f_n - f|\, d\mu = 0$, and

(iii) $\lim_{n\to\infty} \int_X f_n\, d\mu = \int_X f\, d\mu$.

Proof. See [93, Theorem 1.34, p.26]. □

C.8.4 Real-valued functions

Definition C.8.8. *We define $L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$ to be the collection of all measurable functions $f : X \to \mathbb{R}$ for which*

$$\int_X |f|\, d\mu < \infty. \tag{C.22}$$

Proposition C.8.9. *Let $f, g \in L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)$ and $A \subseteq X$ be measurable.*

(i) If $f = g$ a.e. on A, then $\int_A f\, d\mu = \int_A g\, d\mu$.

(ii) If $f \leq g$ a.e. on A, then $\int_A f\, d\mu \leq \int_A g\, d\mu$.

(iii) If $\mu(A) = 0$ or $f = 0$ a.e. on A, then $\int_A f\, d\mu = 0$.

(iv) If $E_1, E_2, \ldots, E_n$ are pairwise disjoint measurable sets, then

$$\int_{\cup_{i=1}^n E_i} f\, d\mu = \sum_{i=1}^n \int_{E_i} f\, d\mu.$$

In particular, $\int_X f\, d\mu = \int_E f\, d\mu + \int_{X \setminus E} f\, d\mu$.

(v) If $(E_n)_{n\geq 1}$ is an increasing sequence of measurable sets and $E = \bigcup_{n\geq 1} E_n$, then

$$\int_E f\, d\mu = \lim_{n\to\infty} \int_{E_n} f\, d\mu. \tag{C.23}$$

Proof. (i) See [110, (G5), p.236].

(ii) See [110, (G6), p.237].

(iii) See [110, (G1), p.237].

(iv) See [110, (G2), p.237].

(v) See Seminar 6.

□

C.9 L^p-spaces

In the sequel, $(X, \mathcal{B}, \mu)$ is a measure space. If $0 < p < \infty$ and if $f : X \to \mathbb{C}$ is a measurable function, define

$$\|f\|_p = \left(\int_X |f|^p\, d\mu\right)^{1/p}. \tag{C.24}$$

Let

$$L^p(X, \mathcal{B}, \mu) = \{f : X \to C \mid f \text{ is measurable and } \int_X |f|^p\, d\mu < \infty\}. \tag{C.25}$$

We call $\|f\|_p$ the L^p-norm of f. We shall denote with $L^p_{\mathbb{R}}(X,\mathcal{B},\mu)$ the real-valued members of $L^p(X,\mathcal{B},\mu)$.

We shall identify two functions $f,g \in L^p(X,\mathcal{B},\mu)$ if they are equal almost everywhere and use the same notation $L^p(X,\mathcal{B},\mu)$ for the quotient set. Thus, $L^p(X,\mathcal{B},\mu)$ is a space whose elements are equivalence classes of functions.

Theorem C.9.1 (Riesz-Fischer Theorem)**.**
For every $1 \le p < \infty$, $(L^p(X,\mathcal{B},\mu), \|\cdot\|_p)$ is a complex Banach space, and $(L^p_{\mathbb{R}}(X,\mathcal{B},\mu), \|\cdot\|_p)$ is a real Banach space.

Proof. See [93, Thm. 3.11, p.70] or, for the real case, [110, p.303]. □

Proposition C.9.2. *$L^2(X,\mathcal{B},\mu)$ is a complex Hilbert space, with the scalar product*

$$\langle f,g\rangle = \int_X f\overline{g}\,d\mu. \tag{C.26}$$

$L^2_{\mathbb{R}}(X,\mathcal{B},\mu)$ is a real Hilbert space.

Proof. See [93, p.78]. □

C.9.1 L^∞

Let $(X,\mathcal{B},\mu)$ be a measure space.

Definition C.9.3. *Let $f: X \to \mathbb{R}$ be measurable. The **essential supremum** of f on X is defined by*

$$\begin{aligned} \textit{ess sup} f &:= \inf\{M \ge 0 \mid |f| \le M \ \textit{a.e.}\} && \text{(C.27)}\\ &= \inf\{\alpha \ge 0 \mid \mu(\{x \in X \mid |f(x)| > \alpha\}) = 0\}. && \text{(C.28)}\end{aligned}$$

It can be seen that both sets in the definition of the ess sup f coincide and hence have the same infimum. Note that $+\infty$ is in both sets, hence the infima above are never taken over the empty set.

Definition C.9.4. *Let $L^\infty(X,\mathcal{B},\mu)$ denote the collection of all $f \in \mathcal{M}_{\mathbb{R}}(X,\mathcal{B})$ with ess sup$f < \infty$. A function $f \in L^\infty(X,\mathcal{B},\mu)$ is called **essentially bounded**.*

Theorem C.9.5. *Define $\|f\|_\infty :=$ ess sup$f$ for all $f \in L^\infty(X,\mathcal{B},\mu)$. Then $(L^\infty(X,\mathcal{B},\mu), \|\cdot\|_\infty)$ is a Banach space.*

Proof. See [110, p.314]. □

C.9.2 Containment relations

Let $(X, \mathcal{B}, \mu)$ be a measure space. It is natural to ask whether there are any containment relations between L^p and L^q, where p and q are distinct positive numbers. It is easy to construct situations where $0 < p < q < \infty$, but $L^p \nsubseteq L^q$ and $L^q \nsubseteq L^p$. See [110, Exercise 1, p.319].

Proposition C.9.6. *Assume that $\mu(X) < \infty$ and let $0 < p < q \leq \infty$. Then*

(i) $L^q \subseteq L^p$.

(ii) If $f \in L^p$ (and hence $f \in L^q$), then

$$\|f\|_p \leq \|f\|_q \cdot \mu(X)^{\frac{1}{p}-\frac{1}{q}}. \tag{C.29}$$

In particular, if $\mu(X) = 1$, then $\|f\|_p \leq \|f\|_q$.

Proof. See [110, Claim 1, p.316]. □

C.10 Modes of convergence

Let $(X, \mathcal{B}, \mu)$ be a measure space and let (f_n) be a sequence in $\mathcal{M}_{\mathbb{C}}(X, \mathcal{B})$. Also let $f \in \mathcal{M}_{\mathbb{C}}(X, \mathcal{B})$.

Definition C.10.1. *We consider the following notions of convergence:*

(i) (f_n) ***converges to*** f ***a.e.*** *if* $\lim_{n\to\infty} f_n(x) = f(x)$ *a.e..*

(ii) (f_n) ***converges to*** f ***in measure*** *if* $\lim_{n\to\infty} \mu(\{x \in X \mid |f_n(x) - f(x)| > \varepsilon) = 0$ *for all* $\varepsilon > 0$. *We will also say that* $f_n \to f$ *in* L^0 *or* $f_n \xrightarrow{\mu} f$.

(iii) (f_n) ***converges to*** f ***in*** L^p *if* $f \in L^p, f_n \in L^p$ *for all* n *and* $\lim_{n\to\infty} \|f_n - f\|_p = 0$.

We define similarly the corresponding notions of Cauchy sequences.

Proposition C.10.2 (Relations between modes of convergence)**.** *[25, Section 10.2]*

(i) Convergence in measure implies almost everywhere convergence for some subsequence.

(ii) For all $0 < p < \infty$, L^p-convergence implies convergence in measure. The converse is not true.

(iii) Neither L^p-convergence nor convergence μ-a.e. implies the other.

Proof. (i) See [110, Claim 1, p.189].

(ii) See [110, Claim 2, p.331] and [110, Exercise 2, p.340].

(iii) See [110, Exercises 3,4, p.340].

□

Proposition C.10.3 (Relations between modes of convergence-finite measure)**.** *Assume that* $\mu(X) < \infty$*. Then*

(i) Almost everywhere convergence implies convergence in measure.

(ii) For all $0 < p < q \leq \infty$*,* L^q*-convergence implies* L^p*-convergence.*

Proof. (i) See [110, Claim 3, p.191].

(ii) By C.9.6.(ii).

□

C.11 Probability measures on compact metric spaces

Appendix D

Topological groups

References for topological groups are, for example, [73] or [52].

Definition D.0.1. *Let G be a set that is a group and also a topological space. Suppose that*

(i) the mapping $(x, y) \mapsto xy$ of $G \times G$ onto G is continuous.

(ii) the mapping $x \mapsto x^{-1}$ of G onto G is continuous.

Then G is called a **topological group**.

Definition D.0.2. *A* **compact group** *is a topological group whose topology is compact Hausdorff.*

Example D.0.3. (i) Every group is a topological group when equipped with the discrete topology.

(ii) All finite groups are compact groups with their discrete topology.

(iii) The additive group $\mathbb{R}$ of real numbers is a Hausdorff topological group which is not compact.

(iv) More generally, the additive group of the euclidean space R^n is a Hausdorff topological group.

(v) The multiplicative group $\mathbb{R}^\star = \mathbb{R} \setminus \{0\}$ with the induced topology is a topological group.

(vi) The multiplicative group $\mathbb{C}^\star = \mathbb{C} \setminus \{0\}$ of nonzero complex numbers with the induced topology is a topological group.

(vii) The unit circle $\mathbb{S}^1 = \{z \in \mathbb{C} \mid |z| = 1\}$ with the group operation being multiplication is a compact group, called the **circle group**.

In the sequel, G is a topological group. For every $a \in G$, let us define the maps

$$L_a : G \to G,\ L_a(x) = ax, \qquad R_a : G \to G,\ R_a(x) = xa.$$

L_a is called the **left translation** by a, while R_a is the **right translation** by a.

Proposition D.0.4. *Left and right translations are homeomorphisms of G. Thus, for all $a \in G$, $(L_a)^{-1} = L_{a^{-1}}$ and $(R_a)^{-1} = R_{a^{-1}}$.*

Appendix E

Miscellaneous

Theorem E.0.5 (Prime Number Theorem)**.**

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\ln(x)} = 1.$$

Let $C(X)$ be the set of all continuous functions $f : X \to \mathbb{R}$

E.0.1 The limit superior and inferior

Proposition E.0.6. *Let $(x_n), (y_n)$ be bounded sequences in $\mathbb{R}$. Then the following hold:*

(i) $\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n$

(ii) (x_n) *is convergent if and only if* $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$*. In this case,* $\lim_{n\to\infty} x_n$ *is the common value.*

(iii) If $c \geq 0$*, then*

$$\liminf_{n\to\infty}(cx_n) = c \liminf_{n\to\infty} x_n \quad \textit{and} \quad \limsup_{n\to\infty}(cx_n) = c \limsup_{n\to\infty} x_n$$

(iv) If $c \leq 0$*, then*

$$\liminf_{n\to\infty}(cx_n) = c \limsup_{n\to\infty} x_n \quad \textit{and} \quad \limsup_{n\to\infty}(cx_n) = c \liminf_{n\to\infty} x_n$$

(v)

$$\begin{aligned} \liminf_{n\to\infty} x_n + \liminf_{n\to\infty} y_n &\leq \liminf_{n\to\infty}(x_n + y_n) \\ &\leq \liminf_{n\to\infty} x_n + \limsup_{n\to\infty} y_n \\ &\leq \limsup_{n\to\infty}(x_n + y_n) \\ &\leq \limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n. \end{aligned}$$

(vi) If $\lim_{n\to\infty} y_n = y$*, then*

$$\liminf_{n\to\infty}(x_n + y_n) = \liminf_{n\to\infty} x_n + y, \quad \limsup_{n\to\infty}(x_n + y_n) = \limsup_{n\to\infty} x_n + y.$$

(vii) $\limsup_{n\to\infty}(x_n y_n) \leq \limsup_{n\to\infty} x_n \cdot \limsup_{n\to\infty} y_n.$

(viii) If (x_n) *and* (y_n) *are nonnegative sequences and* $\lim_{n\to\infty} x_n = x$*, then*

$$\liminf_{n\to\infty}(x_n y_n) = x \cdot \liminf_{n\to\infty} y_n, \quad \limsup(x_n y_n) = x \cdot \limsup_{n\to\infty} y_n. \tag{E.1}$$

(ix) If $x_n \leq y_n$ *eventually (i.e. for all* n *from some* N *on), then*

$$\liminf_{n\to\infty} x_n \leq \liminf_{n\to\infty} y_n, \quad \limsup_{n\to\infty} x_n \leq \limsup_{n\to\infty} y_n.$$

Proof. See, for example, [2, Section 14]. □

Proposition E.0.7. *Let* (a_n) *be a sequence of real numbers. Then*

$$\lim_{n\to\infty} a_n = 0 \Rightarrow \lim \frac{1}{n}\sum_{i=0}^{n-1} |a_i| = 0 \Rightarrow \frac{1}{n}\sum_{i=0}^{n-1} a_i = 0. \tag{E.2}$$

Bibliography

[1] J. Avigad, P. Gerhardy, H. Towsner, Local stability of ergodic averages, Trans. Amer. Math. Soc. 362 (2010), 261-288.

[2] R. G. Bartle, The elements of Real Analysis, second edition, John Wiley & Sons, 1967.

[3] J. E. Baumgartner, A short proof of Hindman's theorem. J. Combinatorial Theory (Series A) 17 (1974), 384-386.

[4] V. Bergelson, Sets of recurrence of $\mathbb{Z}^m$-actions and properties of sets of differences in $\mathbb{Z}^m$, J. London Math. Soc. 31 (1985), 295-304.

[5] V. Bergelson, A density statement generalizing Schur's theorem, Journal of Combinatorial Theory, Series A 43 (1986), 338-343.

[6] V. Bergelson, Ergodic Ramsey Theory, Contemporary Mathematics 65, Amer. Math. Soc., 1987, 63-87.

[7] V. Bergelson, Ergodic Ramsey theory—an update. Ergodic theory of Z^d actions (Warwick, 1993–1994), 1–61, London Math. Soc. Lecture Note Ser., 228, Cambridge Univ. Press, Cambridge, 1996.

[8] V. Bergelson, Minimal idempotents and ergodic Ramsey theory, etc???.

[9] V. Bergelson, The multifarious Poincare recurrence theorem, in: M. Foreman, A. Kechris, A. Louveau, B. Weiss, Descriptive set theory and dynamical systems, London Math. Soc. Lecture Series 277, 2000, 31-57.

[10] V. Bergelson, Ergodic theory and Diophantine problems, in: Topics in symbolic dynamics and applications (Temuco, 1997), 167–205, London Math. Soc. Lecture Note Ser., 279, Cambridge Univ. Press, Cambridge, 2000.

[11] V. Bergelson, Combinatorial and Diophantine applications of ergodic theory. Appendix A by A. Leibman and Appendix B by Anthony Quas and Máté Wierdl. Handbook of dynamical systems. Vol. 1B, 745–869, Elsevier B. V., Amsterdam, 2006.

[12] V. Bergelson, Bernard Host, Randall McCutcheon, Francois Parrea, Aspects of uniformity in recurrence, Colloq. Math. 84/85 (2000), 549-576.

[13] V. Bergelson, A. Leibman, Polynomial extensions of van der Waerden's and Szemerdi's theorems, J. Amer. Math. Soc. 9 (1996), no. 3, 725–753.

[14] V. Bergelson, R. McCutcheon, Recurrence for semigroup actions and a non-commutative Schur theorem, Topological dynamics and applications (Minneapolis, MN, 1995), 205–222, Contemp. Math., 215, Amer. Math. Soc., Providence, RI, 1998.

[15] V. Bergelson, Benjamin Weiss, Translation properties of sets of positive upper density, Proc. Amer. Math. Soc. 94 (1985), 371-376.

[16] E. R, Berlekamp, A construction for partitions which avoid long arithmetic progressions, Canad. Math. Bull. 11 (1968), 409-414.

[17] G. Birkhoff, The mean ergodic theorem. Duke Math. J. 5 (1939), no. 1, 19-20.

[18] A. Błaszczyk, S. Plewik, S. Turek, Topological multidimensional van der Waerden theorem, Comment. Math. Univ. Carolin. 30 (1989), no. 4, 783–787.

[19] M. Brin, G. Stuck, Introduction to dynamical systems, Cambridge University Press, 2002.

[20] J.A. Clarkson, Uniformly convex spaces. Trans. Amer. Math. Soc. 40 (1936), 396-414.

[21] L.W. Cohen, On the mean ergodic theorem, Annals of Math. 41 (1940), 505-509.

[22] Fred H. Croom, Principles of Topology, Saunders College Publishing, 1989.

[23] M. D. Crossley, Essential Topology, Springer, 2005.

[24] J. Diestel, Geometry of Banach spaces. Selected topics, Lecture Notes in Mathematics 485, Springer, 1975.

[25] B. Driver, Analysis tools with applications, draft, 2003.

[26] J. Dugundji, Topology, Allyn and Bacon, Inc., Boston, 1968.

[27] P. Ehrenfest, T. Ehrenfest, The conceptual foundations of the statistical approach in mechanics, Cornell University Press, Ithaca, New York, 1957.

[28] M. Einsiedler, T. Ward, Ergodic Theory with a view towards Number Theory, Graduate Texts in Mathematics 259, 2011.

[29] P. Erdös, R. Rado, Combinatorial theorems on classification of subsets of a given set, Proc. London Math. Soc. 2 (1952), 417-439.

[30] P. Erdös, P. Turán, On Some Sequences of Integers, J. London Math. Soc. 11 (1936), 261-264.

[31] D. Evans, D. Searles, The fluctuation theorem, Advances in Physics 51 (2002), 1529-1585.

[32] B. Farkas, T. Eisner, M. Haase, R. Nagel, Ergodic Theory - An Operator-theoretic approach, 12th International Internet Seminar, February 16, 2009.

[33] N. Frantzikinakis, R. McCutcheon, Ergodic theory: recurrence, arXiv:0705.0033v2 [math.DS]; to appear in the Encyclopedia of Complexity and System Science.

[34] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerdi on arithmetic progressions. J. Analyse Math. 31 (1977), 204–256.

[35] A. Friedman, Foundations of Modern Analysis, Dover, 1982.

[36] H. Furstenberg, Recurrence in ergodic theory and combinatorial number theory. M. B. Porter Lectures. Princeton University Press, Princeton, N.J., 1981.

[37] H. Furstenberg, Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, J. Analyse Math. 34 (1978), 275–291.

[38] H. Furstenberg, Y. Katznelson, D. Ornstein, The ergodic theoretical proof of Szemer edi's theorem, Bull. Amer. Math. Soc. (N.S.) 7 (1982), no. 3, 527–552.

[39] H. Furstenberg, B. Weiss, Topological dynamics and combinatorial number theory, J. Analyse Math 34 (1978), 61–85.

[40] A. M. Garcia, A simple proof of E. Hopf's maximal ergodic theoem, J. Math. Mech. 14 (1965), 381-382.

[41] W. T. Gowers, A new proof of Szemerédi theorem, GAFA 11 (2001), 465-588.

[42] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, Ann. of Math. (2) 166 (2007), no. 3, 897–946.

[43] R. L. Graham, Rudiments of Ramsey Theory, Amer. Math. Soc., 1981.

[44] R. L. Graham, B. Rotschild, Numbers in Ramsey Theory, Surveys in Combinatorics 1987, London Math. Soc. Lecture Notes 123 (1987), 111-153.

[45] R. L. Graham, B. Rotschild, J. H. Spencer, Ramsey Theory, John-Wiley & Sons, 1980.

[46] R. L. Graham, B. Rotschild, J. H. Spencer, Ramsey Theory. Second Edition, John-Wiley & Sons, 1990.

[47] B. Green, Ergodic Theory. Part III, Lecture Notes, 2008.

[48] B. Green, T. Tao, The primes contain arbitrarily long arithmetic progressions. Ann. of Math. (2) 167 (2008), no. 2, 481–547.

[49] P. Halmos, Invariant measures, Ann. of Math. 48 (1947), 735-754.

[50] P. Halmos, Measure Theory, Springer, 1950.

[51] A. W. Hales, R. I. Jewett, Regularity and positional games, Trans. Amer. Math. Soc. 106 (1963), 222–229.

[52] E. Hewitt, K. A. Ross, Abstract harmonic analysis, Volume I, 2nd edition, Springer, 1979.

[53] D. Hilbert, Über die Irreduzibilität ganzer rationaler Funktionen mit ganzzahlingen Koeffizienten, J. Reine Angew. Math. 110 (1892), 104-129.

[54] N. Hindman, Finite sums from sequences within cells of a partition of $\mathbb{N}$, J. Combinatorial Theory (Series A) 17 (1974), 1-11

[55] N. Hindman, D. Strauss, Algebra in the Stone-Cech compactification, De Gruyter Expositions in Mathematics 27, Walter De Gruyter, 1998.

[56] E. Hopf, The general temporally discrete Markoff process, J. Rat. Mech. Anal. 3 (1954), 13-45.

[57] K. Jacobs, Lecture notes on ergodic theory, Aarhus Universitet, 1962/63.

[58] J. L. Kelley, General Topology, Springer, 1955.

[59] J. Ketonen, R. Solovay, Rapidly growing Ramsey functions, Ann. of Math. 113 (1981), 267-314.

[60] A. Khintchin, Three pearls of number theory, Fraylock Press, New York, 1948.

[61] W.A. Kirk, B. Sims (eds.), Handbook of metric fixed point theory, Kluwer, Dordrecht-Boston-London, 2001.

[62] U. Kohlenbach, L. Leuştean, A quantitative Mean Ergodic Theorem for uniformly convex Banach spaces, Ergodic Theory Dynam. Systems 29 (2009), 1907-1915.

[63] B. O. Koopman, J. von Neumann, Dynamical systems of continuous spectra, Proc. Nat. Acad. Sci. U.S.A. 18 (1932), 255-263.

[64] U. Krengel, Ergodic Theorems, Walter de Gruyter, 1985.

[65] L. Kuipers, H. Niederreiter, Uniform distribution of sequences, Wiley, New York, 1974.

[66] B. M. Landman, A. Robertson, Ramsey Theory on the Integers, Student Mathematical Library 24, AMS, 2004.

[67] J. M. Lee, Introduction to topological manifolds, Graduate Texts in Mathematics 202, Springer, 2000.

[68] D. Lind, B. Marcus, An inntroduction to symbolic dynamics and coding, Cambridge University Press, 1999.

[69] A. Magyar, Roth Theorem - the ergodic approach, notes.

[70] P. Matet, Shelah's proof of the Hales-Jewett theorem revisited. European J. Combin. 28 (2007), no. 6, 1742–1745.

[71] I. J. Maddox, Elements of functional analysis, Cambridge University Press, 1970.

[72] R. McCutcheon, Elemental Methods in Ergodic Ramsey Theory, Lecture Notes in Mathematics 1722, Springer, Berlin, 1999.

[73] R. Miculescu, A. Mihail. O introducere in teoria grupurilor topologice, Ed. Universităţii din Bucureşti, 2009.

[74] J. R. Munkres, Topology, second edition, Prentice Hall, Inc., 2000.

[75] B. Nagle, V. Rödl, M. Schacht, The counting lemma for regular k-uniform hypergraphs, Random Structures Algorithms 28 (2006), no. 2, 113–179.

[76] M. B. Nathanson, Additive number theory. The classical bases, GTM, Springer, 1996.

[77] J. von Neumann, Proof of the quasi-ergodic hypothesis, Proceedings of the National Academy of Sciences USA 18 (1932), 263-266.

[78] J. von Neumann, Zur Operatorenmethode in der klassischen Mechanik, Annals of Mathematics 4 (1970), 119-128.

[79] W. Parry, Topics in ergodic theory, Cambridge Tracts in Mathematics 75, Cambridge University Press, 1981.

[80] K. R. Parthasarathy, Probability measures on metric spaces, Academic Press, 1967.

[81] K. R. Parthasarathy, Introduction to Probability and Measure, Texts and Readings in Mathematics 33, Hindustan Book Agency, New Delhi, 2005.

[82] R.E. Powell, S.M. Shah, Summability theory and applications, van Nostrand Reingold, 1975.

[83] K. Petersen, Ergodic Theory, Ca,bridge studies in advanced mathematics 2, Cambridge University Press, 1983.

[84] D. H. J. Polymath, A new proof of the density Hales-Jewett theorem, draft in progress, June 25, 2009.

[85] M. Policott, Michiko Yuri, Dynamical systems and ergodic theory, Londom Mathematical Society Student Texts 40, Cambridge University Press, 1998.

[86] R. Rado, Studien zur Kombinatorik, Math. Z. 36 (1933), 242-280.

[87] R. Rado, Notes on combinatorial analysis, Proc. London Math. Soc. 48 (1943), 122-160.

[88] R. Raimi, Translation properties of finite partitions of the positive integers, Fund. Math. 61 (1968), 253-256.

[89] F. Ramsey, On a problem of formal logic, Proc. London Math. Soc. 30 (1930), 264-286.

[90] V. Rödl, J. Skokan, Regularity lemma for k-uniform hypergraphs, Random Structures Algorithms 25 (2004), no. 1, 1–42.

[91] V. Rödl, J. Skokan, Applications of the regularity lemma for uniform hypergraphs, Random Structures Algorithms 28 (2006), no. 2, 180–194.

[92] K. Roth, On certain sets of integers, J. London Math Soc. 28 (1953), 104-109.

[93] W. Rudin, Real and complex analysis. Third edition, Tata McGraw-Hill, 1986.

[94] W. Rudin, Functional Analysis. Second Edition, McGraw-Hill, 1991.

[95] D. J. Rudolph, Fundamentals of measurable dynamic. Ergodic theory on Lebesgue spaces, Clarendon Press, Oxford, 1990.

[96] A. Sárközy, A. On difference sets of sequences of integers. III. Acta Math. Acad. Sci. Hungar. 31 (1978), no. 3-4, 355–386.

[97] E. Schechter, Handbook of Analysis and Its Foundations, Academic Press, 1977.

[98] I. Schur, Über die Kongruenz $x^m + y^m = z^m (mod\, p)$, Jahresber. Deutsch. Math.-Verein 25 (1916), 114-117.

[99] S. Shelah, Primitive recursive bounds for van der Waerden numbers. J. Amer. Math. Soc. 1 (1988), no. 3, 683–697.

[100] A. J. Sieradski, An introduction to topology and homotopy, PWS-KENT Publishing Company, 1992.

[101] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, Acta Math. Acad. Sci. Hungar. 20 (1969), 89-104.

[102] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, Acta Arith. 27 (1975), 199-245.

[103] L. Sklar, Philosophy of Statistical Mechanics, The Stanford Encyclopedia of Philosophy (Summer 2004 edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/sum2004/entries/statphys-statmech/.

[104] C. T. Taam, Notices Amer. Math. Soc. 6 (1959), 813.

[105] T. Tao, Arithmetic Ramsey Theory, notes.

[106] T. Tao, A variant of the hypergraph removal lemma, J. Combin. Theory Ser. A 113 (2006), no. 7, 1257–1280.

[107] T. Tao, Soft analysis, hard analysis, and the finite convergence principle, essay posted May 23, 2007, available at http://terrytao.wordpress.com/.

[108] T. Tao, Norm convergence of multiple ergodic averages for commuting transformations, Ergodic Theory and Dynamical Systems 28 (2008), 657-688.

[109] T. Tao, Ergodic Theory, in: Poincare's Legacies, Part I: pages from year two of a mathematical blog, AMS, 2009.

[110] E. M. Vestrup, The theory of measure and integration, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, 2003.

[111] B. van der Waerden, Beweis einer Baudetschen Vermutung, Nieuw. Arch. Wiskunde 19 (1927), 212-216.

[112] M. Walters, Combinatorial proof of the polynomial van der Waerden theorem and the polynomial Hales-Jewett theorem, J. London Math. Soc. 61 (2000), 1-12.

[113] P. Walters, Ergodic theory - introductory lectures, Lecture Notes in Mathematics 458, Springer, 1975.

[114] P. Walters, An introduction to ergodic theory, Graduate Texts in Mathematics 79, Springer, 1982.

[115] N. Wiener, The ergodic theorem, Duke Math. J. 5 (1939), 1-18.

[116] F. B. Wright, The recurrence theorem, Amer. Math. Monthly 68 (1961), 247-248.

[117] K. Yosida, Functional Analysis, Springer, 1965.

[118] K. Yosida, S. Kakutani, Birkhoff's ergodic theorem and the maximal ergodic theorem, Proc. Imp. Acad. Tokyo 16 (1940), 280-284.

[119] E. Zermelo, Über einen Satz der Dynamik und die mechanische Wärmetheorie, Annalen der Physik 57 (1896), 485-494; english translation On a Theorem of Dynamics and the Mechanical Theory of Heat, in: S. G. Brush(ed.), Kinetik Theory, Oxford, 1966, II, 208-217.