B. Farkas, T. Eisner, M. Haase, R. Nagel

Ergodic Theory — An Operator-theoretic Approach

– 12th International Internet Seminar –

February 26, 2009

Contents

1 What is ergodic theory? 1

2 Topological Dynamical Systems 7
2.1 Examples 8
2.2 Basic constructions 10
2.3 Topological Transitivity 12
Exercises 16

3 Minimality and Recurrence 17
3.1 Minimality 18
3.2 Topological Recurrence 20
3.3 Recurrence in Extensions 23
Exercises 25

4 The Space $\mathrm{C}(K)$ and the Induced Operator 27
4.1 The space $\mathrm{C}(K)$ as a commutative C^*-algebra 28
4.2 The Theorem of Gelfand–Naimark 31
4.3 The Induced Operator 33
Exercises 35

5 Measure-Preserving Systems 37
5.1 Examples 39
5.2 Measures on Compact Spaces 42
5.3 Haar Measure and Rotations 44
Exercises 45

6 Recurrence and Ergodicity 47
6.1 The Measure Algebra and Invertible MDS 47
6.2 Recurrence 49
6.3 Ergodicity 53
Exercises 57

7 The Banach Lattice L^p and the Induced Operator . . . 59
7.1 The Space $\mathrm{L}^p(\Omega, \Sigma, \mu)$ as a Banach Lattice . . . 60
7.2 The Induced Operator and Ergodicity . . . 65
7.3 Isomorphism of Measure Preserving Systems . . . 67
Exercises . . . 69

8 The Mean Ergodic Theorem I . . . 71
8.1 Von Neumann's Mean Ergodic Theorem . . . 72
8.2 Mean Ergodic Operators on Banach Spaces . . . 75
8.3 Examples . . . 78
Supplement: Powers and Convex Combinations of Mean Ergodic Operators . . . 79
Exercises . . . 81

9 The Mean Ergodic Theorem II . . . 83
9.1 More on Ergodicity . . . 83
9.2 From Topological to Measure-Preserving Systems . . . 87
9.3 Application: Equidistribution . . . 91
Exercises . . . 94

10 The Pointwise Ergodic Theorem . . . 97
10.1 Pointwise Ergodic Operators . . . 98
10.2 Banach's Principle and Maximal Inequalities . . . 100
10.3 Applications . . . 103
Exercises . . . 108

11 Compact Semigroups and Groups . . . 111
11.1 Compact Semigroups . . . 112
11.2 Compact Groups . . . 115
11.3 The Character Group . . . 116
11.4 The Pontryagin Duality Theorem . . . 119
11.5 Application: Ergodic Rotations and Kronecker's Theorem . . . 121
Exercises . . . 127

12 The Jacobs–de Leeuw–Glicksberg Decomposition . . . 129
12.1 Compact Group Actions on Banach Spaces . . . 133
12.2 Almost Weakly Stable Vectors . . . 135

13 Dynamical Systems with Discrete Spectrum . . . 143
13.1 Monothetic Groups . . . 144
13.2 Minimal TDSs with Discrete Spectrum . . . 146
13.3 Ergodic MDSs with Discrete Spectrum . . . 148
13.4 Examples . . . 150
Exercises . . . 152

14 Mixing Dynamical Systems 155
14.1 Strong mixing 156
14.2 Weak mixing 159
14.3 Weak mixing of all orders 163
Exercises 167

15 A Glimpse on Arithmetic Progressions 169
15.1 The Furstenberg Correspondence Principle 171
15.2 A Multiple Ergodic Theorem 173
15.3 The Furstenberg–Sárközy Theorem 178

A Topology 181

B Measure and Integration Theory 191

C Functional Analysis 209

References 223

Lecture 1
What is ergodic theory?

Ergodic Theory is a recent mathematical discipline and its name, in contrast to, e.g., number theory, does not explain its subject. However, its origin can be described quite precisely.

It was around 1880 when L. Boltzmann, J. C. Maxwell and others tried to explain thermodynamical phenomena by mechanical models and their underlying mathematical principles. In this context, L. Boltzmann [Boltzmann (1885)] coined the word *Ergode* (as a special kind of *Monode*)[1]

> *"Monoden, welche nur durch die Gleichung der lebendigen Kraft beschränkt sind, will ich als* **Ergoden** *bezeichnen."*

A few years later (in 1911) P. and T. Ehrenfest [Ehrenfest (1912)] wrote

> *"... haben Boltzmann und Maxwell eine Klasse von mechanischen Systemen durch die folgende Forderung definiert:*
> *Die einzelne ungestörte Bewegung des Systems führt bei unbegrenzter Fortsetzung schließlich durch jeden Phasenpunkt hindurch, der mit der mitgegebenen Totalenergie verträglich ist. – Ein mechanisches System, das diese Forderung erfüllt, nennt Boltzmann ein* **ergodisches System***."*

The assumption that certain systems are "ergodic" is then called "Ergodic Hypothesis". Leaving the original problem behind, "Ergodic Theory" set out on its fascinating journey into mathematics and arrived at quite unexpected destinations.

Before we, too, undertake this journey, let us explain the original problem without going too deep into the underlying physics. We start with an (ideal) gas contained in a box and represented by d (frictionless) moving particles. Each particle is described by six coordinates (three for position, three for velocity), so the situation of the gas (better: the *state* of the system) is given by a point $\omega \in \mathbb{R}^{6d}$. Clearly, not all points in $\mathbb{R}^{6d}$ can be attained by our gas in the box, so we restrict our considerations to the set Ω of all *possible* states and call it the *state space* of our system. We now observe

[1] For the still controversial discussion "On the origin of the notion 'Ergodic Theory'" see the excellent article by M. Mathieu [Mathieu (1988)] but also the remarks by G. Gallavotti in [Gallavotti (1975)].

that our system changes while time is running, i.e., the particles are moving (in the box) and therefore a given state (= point in Ω) also "moves" (in Ω). This motion (in the box, therefore in Ω) is governed by Newton's laws of mechanics and then by Hamilton's differential equations. The solutions to these equations determine a map

$$\varphi : \Omega \longrightarrow \Omega$$

in the following way: If our system, at time $t = 0$, is in the state $\omega_0 \in \Omega$, then at time $t = 1$ it will be in a new state ω_1, and we define φ by $\varphi(\omega_0) := \omega_1$. As a consequence, at time $t = 2$ the state ω_0 becomes

$$\omega_2 := \varphi(\omega_1) = \varphi^2(\omega_0)$$

and

$$\omega_n := \varphi^n(\omega_0)$$

at time $t = n \in \mathbb{N}$. We call $\{\varphi^n(\omega_0) \mid n \in \mathbb{N}_0\}$ the **orbit** of ω_0. In this way, the physical motion of the system of particles becomes a "motion" of the points in the state space. The motion of all states within one time unit is given by the map φ. For these objects we introduce the following terminology.

Definition 1.1. A pair (Ω, φ) consisting of a state space Ω and a map $\varphi : \Omega \longrightarrow \Omega$ is called a **dynamical system**. [2]

The mathematical goal now is not so much to determine φ but rather to find interesting properties of φ. Motivated by the underlying physical situation, the emphasis is on "long term" properties of φ, i.e., properties of φ^n as n gets large.

First objection. In the physical situation it is not possible to determine exactly the given initial state $\omega_0 \in \Omega$ of the system or any of its later states $\varphi^n(\omega_0)$.

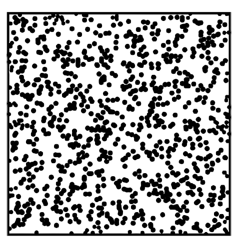

Fig. 1.1 Try to determine the exact state of the system for only $n = 1000$ gas molecules!

To overcome this objection we introduce "observables", i.e., functions $f : \Omega \longrightarrow \mathbb{R}$ assigning to each state $\omega \in \Omega$ the value $f(\omega)$ of a measurement, for instance of

[2] This is a mathematical model for the philosophical principle of determinism and we refer to the article by G. Nickel in [Engel and Nagel (2000)] for more on this interplay between mathematics and philosophy.

the temperature. The motion in time ("evolution") of the states described by the map $\varphi : \Omega \longrightarrow \Omega$ is then reflected by a map T_φ of the observables defined as

$$f \longmapsto T_\varphi f := f \circ \varphi.$$

This change of perspective is not only physically justified, it also has an enormous mathematical advantage:

> The set of all observables $\{f : \Omega \longrightarrow \mathbb{R}\}$ has a *vector space* structure and the map $T_\varphi = (f \longmapsto f \circ \varphi)$ becomes a *linear operator* on this vector space.

So instead of looking at the orbits $\varphi^n(\omega_0)$ of the state map φ we study the orbit $\{(T_\varphi)^n f \mid n \in \mathbb{N}_0\}$ of an observable f under the linear operator T_φ. This change of perspective allows the use of operator theoretic tools and is the basis of our approach.

Returning to our description of the motion of the particles in a box and keeping in mind realistic experiments, we should make another objection.

Second objection. The motion of our system happens so quickly that we will not be able to determine the states

$$\omega_0, \varphi(\omega_0), \varphi^2(\omega_0), \ldots$$

nor their measurements

$$f(\omega_0), T_\varphi f(\omega_0), T_\varphi^2 f(\omega_0), \ldots$$

at time $t = 0, 1, 2, \ldots$.

Reacting on this objection we slightly change our perspective and instead of the above measurements we look at the averages over time

$$\frac{1}{N}\sum_{n=0}^{N-1} T_\varphi^n f(\omega_0) = \frac{1}{N}\sum_{n=0}^{N-1} f(\varphi^n(\omega_0))$$

and their limit

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T_\varphi^n f(\omega_0),$$

called the *time mean* (of the observable f in the state ω_0). This seems to be a good idea, but it is still based on the knowledge of the states $\varphi^n(\omega_0)$ and their measurements $f(\varphi^n(\omega_0))$, so the first objection remains valid. At this point, Boltzmann asked for more.

Third objection. The time mean $\lim_N \frac{1}{N}\sum_{n=0}^{N-1} f(\varphi^n(\omega_0))$ should not depend on the initial state $\omega_0 \in \Omega$.

Boltzmann even suggested what the time mean should be. Indeed, for his system there exists a canonical probability measure μ on the state space Ω for which he claimed the validity of the so-called

Ergodic Hypothesis. *For each initial state $\omega_0 \in \Omega$ and each (reasonable) observable $f : \Omega \longrightarrow \mathbb{R}$, it is true that "time mean equals space mean", i.e.,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\varphi^n(\omega_0)) = \int_\Omega f\, d\mu.$$

Up to now our arguments are quite vague and based on some physical intuition (that one may or may not have). However, we arrived at the point where mathematicians can take over and build a mathematical theory. To do so we propose the following steps.

1) Make reasonable assumptions on the basic objects such as the *state space* Ω, the dynamics given by the map $\varphi : \Omega \longrightarrow \Omega$ and the *observables* $f : \Omega \longrightarrow \mathbb{R}$.
2) Prove theorems that answer the following questions.
   (i) Under which assumptions and in which sense does the limit appearing in the time mean exist?
   (ii) What are the best assumptions to guarantee that the ergodic hypothesis holds.

Historically, these goals could be achieved only after the necessary tools had been created. Fortunately, between 1900 and 1930 new mathematical theories emerged such as topology, measure theory and functional analysis. Using the tools from these new theories, G. D. Birkhoff and J. von Neumann proved independently in 1931 the so called *individual ergodic theorem* [Birkhoff (1931)] and the *mean ergodic theorem* [von Neumann (1932b)] and thereby established ergodic theory as a new and independent mathematical discipline.

Now, a valuable mathematical theory should offer more than precise definitions (see Step 1) and beautiful theorems (see Step 2). Clearly you also expect that applications can be made to the original problem from physics. But something very fascinating happened, something that E. Wigner [Wigner (1960)] called "the unreasonable effectiveness of mathematics". While Wigner noted this "effectiveness" with regard to the natural sciences, we shall see that ergodic theory is effective in completely different fields of mathematics such as, e.g., number theory. This phenomenon was unconsciously anticipated by A. Borel's theorem on normal numbers from 1909 or by the equidistribution theorem of H. Weyl [Weyl (1916)]. Both results are now considered to be a part of ergodic theory.

But not only classical results can now be proved using ergodic theory. Confirming a conjecture of 1936 by P. Erdős and P. Turán, B. Green and T. Tao created a mathematical sensation by proving the following theorem using, among others, ergodic-theoretic techniques and results.

Theorem 1.2 (Green–Tao, 2004). *The primes $\mathbb{P}$ contain arbitrarily long arithmetic progressions, i.e., for every $k \in \mathbb{N}$ there exist $a \in \mathbb{P}$ and $n \in \mathbb{N}$ such that*

$$a, a+n, a+2n, \ldots, a+(k-1)n \in \mathbb{P}.$$

The Green-Tao theorem had a precursor first proved by Szemerédi and then reproved by Fürstenberg using purely ergodic-theoretic tools.

Theorem 1.3 (Szemerédi, 1975). *If a set $A \subset \mathbb{N}$ has upper density*

$$\overline{d}(A) := \overline{\lim}_{n\to\infty} \frac{\#(A \cap \{1,\ldots,n\})}{n} > 0,$$

then it contains arbitrarily long arithmetic progressions.

While a complete proof of this theorem will remain beyond the reach of this course, we shall at least indicate some of its tools.

As a warm-up before the real work, we give you a simple exercise.

Problem. Consider the unit circle $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$ and take a point $1 \neq a \in \mathbb{T}$. What can we say about the behaviour of the barycenters b_n of the polygons formed by the points $1, a, a^2, \ldots, a^{n-1}$ as n tends to infinity?

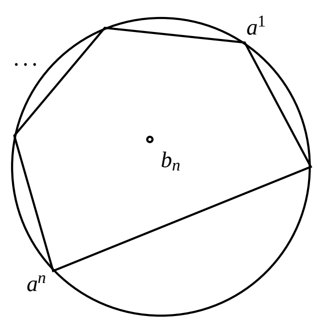

Fig. 1.2 Polygon

While this problem can be solved using only very elementary mathematics, we will show later that its solution is just a (trivial) special case of von Neumann's Ergodic Theorem. We hope you are looking forward to seeing this theorem and more.

Exercises

1. Translate the above quotes from German into English.

2. Read the various ethymological explanations on the name "ergodic". Which explanation convinces you most?

3. Imagine all particles being in the left half of the box at time $t = 0$. What happens after a while? Explain why the ergodic hypothesis is reasonable in this situation.

4. Interpret the "observables" $f : \Omega \longrightarrow \mathbb{R}_+$ as probability densities of the states with respect to a probability measure μ. Why should φ preserve this measure?

5. Why can the "motion" φ in Ω be called deterministic? Explain why (for our gas in the box) it is essential to work in an $6d$-dimensional space instead of the physical 3-dimensional space.

6. Find more (algebraic) properties of the linear operator T_φ.

Lecture 2
Topological Dynamical Systems

In Lecture 1 we introduced a dynamical system as consisting of a set Ω and a self-map $\varphi : \Omega \longrightarrow \Omega$. However, in concrete situations one usually has some additional structure on the set Ω, e.g., a *topology* and/or a *measure*, and then the map φ is continuous and/or measurable. We shall study *measure theoretical dynamical systems* in later lectures and start with an introduction into *topological dynamics*.

As a matter of fact, this requires some familiarity with elementary (point-set) topology as discussed for instance in [Kuratowski (1966)], [Kelley (1975)] or [Willard (2004)]. For the convenience of the reader we have collected some basic definitions and results in Appendix A. The most important notion is that of a compact space, but the reader unfamiliar with general topology may think of "compact metric spaces" whenever "compact topological spaces" appear.

Definition 2.1. The pair $(K;\varphi)$ is a **topological dynamical system** (TDS for short) if K is a non-empty compact space and $\varphi : K \longrightarrow K$ is continuous; it is called an **invertible TDS** if φ is invertible, i.e., a homeomorphism.

An invertible TDS $(K;\varphi)$ defines two "one-sided" TDSs, namely the **forward system** $(K;\varphi)$ and the **backward system** $(K;\varphi^{-1})$. Let $(K;\varphi)$ be a TDS and $x \in K$. The **forward orbit** of x is

$$\mathrm{orb}_+(x) := \{\varphi^n(x) \,:\, n \in \mathbb{N}_0\};$$

and for invertible TDS, the **(total) orbit** of x is $\mathrm{orb}(x) := \{\varphi^n(x) \,:\, n \in \mathbb{Z}\}$. We shall write

$$\overline{\mathrm{orb}_+}(x) := \overline{\{\varphi^n(x) \,:\, n \in \mathbb{N}_0\}} \quad \text{and} \quad \overline{\mathrm{orb}}(x) := \overline{\{\varphi^n(x) \,:\, n \in \mathbb{Z}\}}$$

for the closure of the forward and the total orbit, respectively. (We agree that $\mathbb{N}_0 = \{0,1,2,\dots\}$, the set of non-negative integers, and $\mathbb{N} = \mathbb{N}_0 \setminus \{0\}$.)

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 spaces and reserve the term TDS for this special situation.

Before proving any results, we present examples of dynamical systems. Some of them are "fundamental" in a sense yet to make precise and will frequently reappear in the rest of the course.

2.1 Examples

Example 2.2 (Finite state space). Take $d \in \mathbb{N}$ and consider the finite set $K := \{1,\dots,d\}$, with the discrete topology. Then K is obviously compact and every map $\varphi : K \longrightarrow K$ is continuous. The TDS $(K;\varphi)$ is invertible if and only if φ is injective (or surjective).

Example 2.3 (Finite-dimensional contractions). Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$ and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ be linear and contractive with respect to the chosen norm, i.e., $\|Tx\| \le \|x\|$, $x \in \mathbb{R}^d$. Then the unit ball $K := \{x \in \mathbb{R}^d : \|x\| \le 1\}$ is compact, and $\varphi := T|_K$ is a continuous self-map of K.

As a concrete situation we choose the maximum-norm $\|x\|_\infty = \max\{|x_1|,\dots|x_d|\}$, $x \in \mathbb{R}^d$, and the linear operator T given by a **row-substochastic** matrix $(t_{ij})_{i,j}$, i.e.,

$$t_{ij} \ge 0 \quad \text{and} \quad \sum\nolimits_{k=1}^{d} t_{ik} \le 1 \qquad (1 \le i,j \le d).$$

If we let $\mathbf{1} := (1,\dots,1)$, then $\|x\|_\infty \le 1$ is equivalent to $|x| := (|x_1|,\dots,|x_d|) \le \mathbf{1}$. Hence

$$|Tx| \le T\,|x| \le \|x\|_\infty\, T\mathbf{1} \le \|x\|_\infty\, \mathbf{1}$$

for every $x \in \mathbb{R}^d$, since T is row-substochastic. So T is contractive.

Example 2.4 (Infinite-dimensional contractions). Let X be a Banach space and let $T \in \mathscr{L}(X)$ be a bounded linear operator on X. If T is a contraction, it leaves the closed unit-ball $B_X := \{x \in X : \|x\| \le 1\}$ of X invariant, but B_X will not be compact in the norm-topology unless X is finite-dimensional.

However, the *adjoint* operator T' on the dual space X' is also a contraction, and hence leaves the dual unit ball $K := \{x' \in X' : \|x'\| \le 1\}$ invariant. This is a compact set with respect to the weak*-topology (by the Banach-Alaoglu theorem) and since T' is also continuous with respect to this topology, $\varphi := T'|_K$ defines a TDS $(K;\varphi)$. Note that in the case that X is separable, K is actually a compact metrisable space.

Example 2.5 (Shift). Take $k \in \mathbb{N}$ and consider the set

$$K := \mathscr{W}_k^+ := \{0,1,\dots,k-1\}^{\mathbb{N}_0}$$

of infinite sequences within the set $\{0,\dots,k-1\}$. In this context, the set $\{0,\dots,k-1\}$ is often called an **alphabet**, its elements being the **letters**. So the elements of K

are the infinite **words** composed of these letters. If we endow the alphabet with the discrete metric, it becomes a compact metric space, and so does K, when we take the product topology on K (Tychonoff's theorem, see Section A.5 and also Exercise 2.1). On K we consider the (left) **shift** τ defined by

$$\tau : K \longrightarrow K, \quad (x_n)_{n\in\mathbb{N}_0} \longmapsto (x_{n+1})_{n\in\mathbb{N}_0}.$$

Then $(K;\tau)$ is a TDS, called the **one-sided shift**. If we consider two-sided sequences instead, that is, $\mathscr{W}_k := \{0,1,\dots,k-1\}^{\mathbb{Z}}$ with τ defined analogously, we obtain an invertible TDS $(K;\tau)$, called the **two-sided shift**.

Example 2.6 (Addition mod 1). Consider the interval $K := [0,1)$, and define a new metric d on it by

$$d(x,y) := \min\{|x-y|, 1-|x-y|\} \qquad (x,y\in[0,1)).$$

In Exercise 2.2 you are asked to show that this is indeed a metric, continuous with respect to the standard one, and making K a compact metric space. For a real number $x\in\mathbb{R}$ we write

$$x \pmod 1 := x - \lfloor x \rfloor,$$

where $\lfloor x \rfloor := \max\{n\in\mathbb{Z} : n \le x\}$ is the greatest integer than less or equal to x. Now, given $\alpha\in[0,1)$ we define **addition mod 1** by α by

$$\varphi_\alpha(x) := x+\alpha \pmod 1 \;=\; x+\alpha-\lfloor x+\alpha\rfloor \qquad (x\in[0,1)).$$

In Exercise 2.2 it will be shown that φ_α is continuous with respect to the metric d, hence constituting a TDS $([0,1);\varphi_\alpha)$.

Example 2.7 (Rotations on the torus). Let $\mathbb{T} := \{z\in\mathbb{C} : |z| = 1\}$ be the unit circle, also called the (one-dimensional) **torus**. Take $a\in\mathbb{T}$ and define $\varphi_a : \mathbb{T}\longrightarrow\mathbb{T}$ by

$$\varphi_a(z) := a\cdot z \quad \text{for all } z\in\mathbb{T}.$$

Since $\mathbb{T}$ is a closed and bounded subset of $\mathbb{C}$, it is compact, hence $(\mathbb{T};\varphi_a)$ is a TDS, called the **rotation** by a.

Example 2.8 (Rotation on compact groups). A more abstract version of the previous example is the following: A group $(G,\cdot)$ is called a **topological group** if it is a topological space and the mappings

$$(g,h) \longmapsto g\cdot h \qquad G\times G\longrightarrow G$$

and

$$g \longmapsto g^{-1} \qquad G\longrightarrow G$$

are continuous. A topological group is a **compact group** if G is compact. The torus $\mathbb{T}$ is the most important example of a compact group.

Let now G be any compact group, fix $h\in G$, and define the (left) **rotation**

$$\varphi_h(g) := h\cdot g \quad \text{for all } g\in G.$$

Again, $(G;\varphi_h)$ is a TDS. Note that we do not require G to be commutative, hence we have to distinguish between left and right rotation. If the group is commutative, one also speaks of *translation* instead of rotation.

2.2 Basic constructions

Of course one can study dynamical systems individually, each at a time. However, it is also reasonable to look for a good “structure theory” answering questions like:

1) When are two TDSs “essentially equal”?
2) What are the “basic” building blocks of TDSs?
3) How can we build up a complex system from those basic ones?

It will eventually turn out that Examples 2.5 and 2.8 provide some answers to question 2). In this section we consider some of the basic notions related to 1) and 3).

A **homomorphism** between two TDSs $(K_1;\varphi_1)$, $(K_2;\varphi_2)$ is a continuous map $\Psi : K_1 \longrightarrow K_2$ such that the following diagram commutes:

$$\begin{tikzcd} K_1 \arrow[r, "\varphi_1"] \arrow[d, "\Psi"'] & K_1 \arrow[d, "\Psi"] \\ K_2 \arrow[r, "\varphi_2"] & K_2 \end{tikzcd}$$

which means that $\Psi \circ \varphi_1 = \varphi_2 \circ \Psi$. Two TDSs $(K_1;\varphi_1)$, $(K_2;\varphi_2)$ are called **isomorphic** if there is a homeomorphism $\Psi : K_1 \longrightarrow K_2$ which makes the previous diagram commutative. In this case Ψ is called a TDS **isomorphism**. A self-isomorphism $\Psi : K \longrightarrow K$ is called an **automorphism** of the TDS.

Example 2.9. Consider the rotation $(\mathbb{T};\varphi_a)$ by $a \in \mathbb{T}$ (Example 2.7) and the addition mod 1 $([0,1),d;\varphi_\alpha)$ by $\alpha \in [0,1)$ (Example 2.6). If $a = e^{2\pi i\alpha}$ then these two systems are isomorphic, see Exercise 2.2.

Let $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ be two TDSs. The **product** TDS $(K;\varphi)$ is defined by

$$\begin{aligned} &K := K_1 \times K_2 \\ &\varphi := \varphi_1 \times \varphi_2 : K \longrightarrow K,\ \ \varphi_1 \times \varphi_2(x,y) := (\varphi_1(x), \varphi_2(x)). \end{aligned}$$

It is invertible whenever both $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ are.

A more sophisticated way of constructing TDSs is that of **group extensions**. Let G be a compact group, $(K;\varphi)$ a TDS and $\Phi : K \longrightarrow G$ a continuous mapping. We take the product $H := K \times G$ as a new state space and define

$$\psi(x,g) := (\varphi(x), \Phi(x)g) \quad \text{for} \quad (x,g) \in K \times G.$$

Clearly, $(H;\psi)$ is a TDS and if $(K;\varphi)$ is invertible so is $(H;\psi)$. The TDS $(H;\psi)$ is called the **group extension** of $(K;\varphi)$ **along Φ**. In this case, the right-multiplication ρ_h in the second coordinate by an element $h \in G$, defined by $\rho_h : H \longrightarrow H$, $\rho_h(x,g) = (x,gh)$ is an automorphism of the group extension $(H;\psi)$. Indeed, it is a homeomorphism of H onto itself, and we have

$$\rho_h(\psi(x,g)) = \rho_h(\varphi(x),\Psi(x)g) = (\varphi(x),\Psi(x)gh) = \psi(x,gh) = \psi(\rho_h(x,g)),$$

for all $(x,g) \in K \times G$, so ρ_h and ψ commute.

Examples 2.10. 1) The n**-torus**, which is the n-fold direct product $G = \mathbb{T} \times \cdots \times \mathbb{T} = \mathbb{T}^n$ is a compact topological group. The rotation by $a = (a_1, a_2, \ldots, a_n) \in G$ is the n-fold product of the TDSs $(\mathbb{T};\varphi_{a_i})$, $i = 1,\ldots,n$.

2) The product $(\mathbb{T}\times\mathbb{T};\varphi_{a_1}\times\varphi_{a_2})$ may be seen as a group extension of $(\mathbb{T};\varphi_{a_1})$ by taking $\Phi(x) = a_2$ for all $x \in \mathbb{T}$.

3) Consider the TDS $(\mathbb{T};\varphi_a)$ with $a \in \mathbb{T}$, $G = \mathbb{T}$ and Φ the identity map on $\mathbb{T}$. Then for the corresponding group extension $(H;\psi)$ we have

$$\psi(x,y) = (ax,xy) \qquad (x,y \in \mathbb{T}).$$

Let $(K_1;\varphi_1)$, $(K_2;\varphi_2)$ be two TDSs such that $K_1 \cap K_2 = \emptyset$, and set $K := K_1 \cup K_2$. To define a topology (see A.2) in K, we set

$$\mathscr{O} := \{A \cup B : A \in \mathscr{O}_1, B \in \mathscr{O}_2\},$$

where $\mathscr{O}_1$ and $\mathscr{O}_2$ are the families open sets in K_1 respectively K_2. Then $\mathscr{O}$ as the family of opens sets defines a topology on K, which is the inductive topology (see A.4) with respect to the canonical embeddings $J_1 : K_1 \longrightarrow K$, $J_2 : K_2 \longrightarrow L$. Further we define the map

$$\varphi(x) = \begin{cases} \varphi_1(x) & \text{if } x \in K_1, \\ \varphi_2(x) & \text{if } x \in K_2. \end{cases}$$

The pair $(K;\varphi)$ is a TDS, called the **disjoint union** of the TDSs $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$, which is invertible if and only if $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ are both invertible.

We finally turn to the notion of a *subsystem* of a TDS $(K;\varphi)$.

Definition 2.11. A set $A \subset K$ is called **invariant** (under φ, φ-invariant) if $\varphi(A) \subset A$, and it is called **strongly invariant** or **bi-invariant** if $A = \varphi^{-1}(A)$.

We shall return to these important notions in Lecture 3. Here we just show how invariant sets give rise to further TDSs. Given a non-empty closed invariant set $A \subset K$ one can restrict φ to A and obtains a new TDS $(A;\varphi|_A)$, called a **subsystem** of the original system. If $(K;\varphi)$ is an invertible TDS and A is bi-invariant, then $(A;\varphi|_A)$ is an *invertible subsystem*.

Example 2.12. If $(K;\varphi)$ is a disjoint union of $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$, then both are subsystems of $(K;\varphi)$.

Example 2.13 (Subshifts). Consider the shift $(\mathscr{W}_k^+;\tau)$ on the alphabet $\{0,1,\dots,k-1\}$, (Example 2.5). We try to determine, in a systematic way, all of its subsystems called **subshifts**. For this purpose we need some further notions. An n-**block** of an infinite word $x\in\mathscr{W}_k^+$ is a finite sequence $y\in\{0,1,\dots,k-1\}^n$ which occurs in x at some position. For this relation we write $y\in x$. Now take an arbitrary set

$$B\subset\bigcup_{n\in\mathbb{N}_0}\{0,1,\dots,k-1\}^n,$$

and consider it as the family of **excluded blocks**. From this we define

$$\mathscr{W}_k^B:=\left\{x\in\mathscr{W}_k^+ : \text{there is no block } y\in B \text{ with } y\in x\right\}.$$

It is easy to see that $\mathscr{W}_k^B$ is a closed τ-invariant subset, hence, if non-empty, it gives rise to a subsystem $(\mathscr{W}_k^B;\tau|_{\mathscr{W}_k^B})$. We claim that *each subsystem of* $(\mathscr{W}_k^+;\tau)$ *arises in this way.* Let $(F;\tau|_F)$ be a subsystem and consider the set of finite sequences that are not present in any of the words in F:

$$B:=\left\{y : y \text{ is a finite sequence and } y\notin x \text{ for all } x\in F\right\}.$$

We have $F\subset\mathscr{W}_k^B$ by definition. If $x\notin F$, then there is an open rectangle $U\subset\mathscr{W}_k^+$ with $x\in U$ and $U\cap F=\emptyset$. By taking a possibly smaller rectangle we can assume that U has the form

$$U=\left\{z\in\mathscr{W}_k^+ : z_i=j_0,\, z_{i+1}=j_1,\dots,z_{i+n}=j_n\right\}$$

for some $n\in\mathbb{N}_0$ and $j_0,j_1,\dots,j_n\in\{0,1,\dots,k-1\}$. This means that $y=(j_0,j_1,\dots,j_n)$ is an $(n+1)$-block of x and it does not occur in any of the words in F (since $F\cap U=\emptyset$), hence $y\in B$ by definition. But then x cannot belong to $\mathscr{W}_k^B$, so the equality $F=\mathscr{W}_k^B$ follows.

Particularly important are those subshifts $(F;\tau|_F)$ that arise from a finite set B as $F=\mathscr{W}_k^B$. These are called **subshifts of finite type**. The subshift is called of **type** n if there is an excluded block-system B containing only sequences not longer than n, i.e., $B\subset\bigcup_{i\le n}\{0,1,\dots,k-1\}^i$.

2.3 Topological Transitivity

We are primarily interested how a particular state of the system evolves in time, i.e., in the behaviour of the orbit $\varphi^n(x)$ as n tends to infinity. A first question to ask is how φ mixes the points of K as it is applied over and over again. Will two points that are close to each other initially, stay close even after a long time? Will a point return to its original position (at least very near to it)? Will a certain point x never leave a certain region or will it come arbitrarily close to any other given point of K?

We first investigate the phenomenon described in the last statement, which is certainly equivalent to saying that the (forward) orbit of x is dense in K. Let us give a name to that.

Definition 2.14. Let $(K;\varphi)$ be a TDS. A point $x \in K$ is called **(forward) transitive** if its forward orbit $\mathrm{orb}_+(x)$ is dense in K. If there is at least one forward transitive point, then the TDS $(K;\varphi)$ is called **topologically (forward) transitive**.

Analogously, a point $x \in K$ in an invertible TDS $(K;\varphi)$ is called **transitive** if its total orbit $\mathrm{orb}(x)$ is dense, and the invertible TDS $(K;\varphi)$ is called **topologically transitive** if there exists at least one transitive point.

We shall often use the term "topologically transitive" for non-invertible systems by which we actually mean "topologically forward transitive". This shall not cause any ambiguity.
As indicated above, this property expresses that if we start at the point x we can reach, at least approximately, any other point in K after some time. The following result tells us that this is indeed a mixing-property: Arbitrary two open regions are mixed with each other under the action of φ in finitely many steps.

Proposition 2.15. *Let $(K;\varphi)$ be a TDS and consider the following assertions.*

a) *$(K;\varphi)$ is topologically forward transitive, i.e., there is a point $x \in K$ with forward orbit-closure $\overline{\mathrm{orb}_+}(x) = K$.*
b) *For all $\emptyset \neq U, V$ open sets in K there is $n \in \mathbb{N}$ with $\varphi^n(U) \cap V \neq \emptyset$.*
c) *For all $\emptyset \neq U, V$ open sets in K there is $n \in \mathbb{N}$ with $\varphi^{-n}(U) \cap V \neq \emptyset$.*

Then b) and c) are equivalent; b) implies a) if K is metrisable; and a) implies b) if K has no isolated points.

Proof. The equivalence of b) and c) is easy.

a) $\Rightarrow$ b): Suppose that $x \in K$ has dense forward orbit and that U, V are non-empty open subsets of K. Then certainly $\varphi^k(x) \in U$ for some $k \in \mathbb{N}_0$. Consider the open set $W := V \setminus \{x, \varphi(x), \ldots, \varphi^k(x)\}$. If K has no isolated points, W cannot be empty, and hence $\varphi^m(x) \in W$ for some $m > k$. Now, $\varphi^m(x) = \varphi^{m-k}(\varphi^k(x)) \in \varphi^{m-k}(U) \cap V$.

c) $\Rightarrow$ a): Suppose that K is metrisable. Then there is a countable base $\{U_n : n \in \mathbb{N}\}$ for the topology on K (see Section A.8). For each $n \in \mathbb{N}$ consider the open set

$$G_n := \bigcup_{k \in \mathbb{N}_0} \varphi^{-k}(U_n).$$

By assumption b), G_n intersects non-trivially every non-empty open set, and hence is dense in K. By Baire's Category Theorem A.4 the set $\bigcap_{n \in \mathbb{N}} G_n$ is non-empty (it is even dense). Any x in this intersection has dense forward orbit, the claim is thus proved.

Remarks 2.16. 1) In general a) does not imply b) even if K is metrisable: Take $K := \mathbb{N} \cup \{\infty\}$, and $\varphi : K \longrightarrow K$, $\varphi(n) = n+1$, $\varphi(\infty) = \varphi(\infty)$. The point $1 \in K$ has dense forward orbit, but for $U = \{2\}$, $V = \{1\}$ condition b) fails to hold.

2) The proof shows that if K is metrisable without isolated points the set of points with dense forward orbit is either empty or a dense G_δ, see Appendix A.10.

For invertible TDSs we obtain the following characterisation of topological transitivity.

Proposition 2.17. *Let $(K;\varphi)$ be an invertible TDS, with K compact metrisable. Then the following properties are equivalent.*

a) *$(K;\varphi)$ is topologically transitive, i.e., there is a point $x \in K$ with dense orbit.*
b) *For all $\emptyset \neq U, V$ open sets in K there is $n \in \mathbb{Z}$ with $\varphi^n(U) \cap V \neq \emptyset$.*

The proof is almost verbatim the same as for Proposition 2.15. Now we turn to examples:

Example 2.18 (Rotations on compact groups). Let us look on a left rotation $(G;\varphi_h)$, with a compact group G and $h \in G$. We claim that *if $(G;\varphi_g)$ is topologically transitive, then actually all points have dense (forward) orbit.*

Proof. Since every right rotation $\psi_g : x \longmapsto xg$ is a homeomorphism of G commuting with φ_h, we have

$$\varphi_h^k(g) = h^k g = (h^k 1)g = \psi_g(\varphi_h^k(1)) \qquad (k \geq 0).$$

Hence $\overline{\mathrm{orb}}_+(g) = \psi_g(\overline{\mathrm{orb}}_+(1))$, for any $g \in G$. So there is a point with dense forward orbit in G if and only if $\mathrm{orb}_+(1)$ is dense, and if this is so, then every point has dense orbit.

A corollary is that whenever the group rotation $(G;\varphi)$ is topologically transitive, there cannot be a non-trivial closed invariant set A in G, for otherwise any point $x \in A$ would not have a dense orbit. We can exploit this fact in the following examples.

Example 2.19 (Kronecker's theorem, easy case). Let us consider the special case of the torus $G = \mathbb{T}$. We claim that *the rotation $(\mathbb{T};\varphi_a)$ is topologically transitive if and only if $a \in \mathbb{T}$ is not a root of unity*.

Proof. If $a^{n_0} = 1$ for some $n_0 \in \mathbb{N}$, then $\{z \in \mathbb{T} : z^{n_0} = 1\}$ is closed and φ-invariant, so by the remark after Example 2.18, $(G;\varphi_h)$ cannot be transitive. For the other implication, we show that $\mathrm{orb}_+(1) = \{1, a, a^2, \ldots\}$ is dense in $\mathbb{T}$. Take $\varepsilon > 0$. Since by assumption $a^{n_1} \neq a^{n_2}$ for $n_1 \neq n_2$, there exist $l < k \in \mathbb{N}$ such that $0 < |a^l - a^k| < \varepsilon$ (use the pigeonhole principle), and hence

$$|a^{(k-l)n} - a^{(k-l)(n+1)}| = |1 - a^{k-l}| = |a^l - a^k| < \varepsilon \quad \text{ for all } n \in \mathbb{N}.$$

Since the set of "segments"

$$\{(a^{(k-l)n}, a^{(k-l)(n+1)}) : n \in \mathbb{N}\}$$

covers $\mathbb{T}$, we proved that there is at least one power of a in every ε-segment of $\mathbb{T}$.

Example 2.20. The product of two topologically transitive systems need not to be topologically transitive: Consider $a_1 = e^{i}$, $a_2 = e^{2i}$, and the product system $(\mathbb{T}^2; \varphi_{a_1} \times \varphi_{a_2})$. Then $M = \{(x,y) \in \mathbb{T}^2 : x^2 = y\}$ is a non-trivial, closed $\varphi_{a_1} \times \varphi_{a_2}$-invariant set.

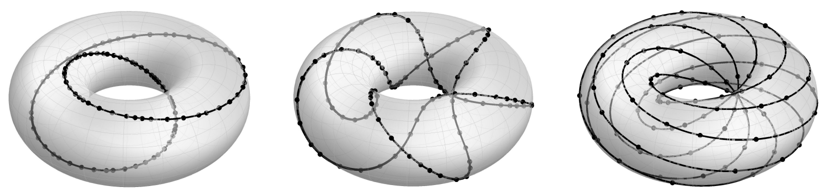

Fig. 2.1 The first 100 iterates of a point under the rotation in Example 2.20 and its orbit closure. The same for the rotation ratios 2 : 5 and 8 : 5.

Because of these two examples, it is interesting to characterise the transitivity of the rotations on the n-torus for $n > 1$.

Theorem 2.21 (Kronecker). *Let $a = (a_1, \ldots, a_n) \in \mathbb{T}^n$ and consider the rotation $\varphi_a : \mathbb{T}^n \longrightarrow \mathbb{T}^n$. Then the TDS $(\mathbb{T}^n; \varphi_a)$ is topologically transitive if and only if $a_1, a_2, \ldots, a_n$ are linearly independent in the $\mathbb{Z}$-module $\mathbb{T}$, i.e., whenever $a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n} = 1$ for $k_1, k_2, \ldots, k_n \in \mathbb{Z}$, then it follows that $k_1 = k_2 = \cdots = k_n = 0$.*

We do not give the proof here, but shall return to it later. We hope to have stimulated your curiosity! We close this lecture by the following characterisation of topological transitivity of subshifts.

Example 2.22 (Subshifts of type 2**).** We can associate to any subshift of type 2 a $k \times k$-matrix $A = (a_{ij})$, called the **transition matrix** of the subshift:

$$a_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ occurs in a word in } F, \\ 0 & \text{otherwise.} \end{cases}$$

Proposition. *Let $(F; \tau|_F)$ be a subshift of type 2. The following assertions are equivalent.*

a) *$(F, \tau|_F)$ is topologically transitive.*
b) *Its transition matrix is **irreducible**, i.e., for all $1 \leq i, j \leq k$ some power A^n of A has strictly positive entry in the i^{th} row and j^{th} column.*

For a proof we refer to 17.9 in [Denker et al. (1976)].

Exercises

1. Find a metric inducing the product topology on $\mathscr{W}_k^+ = \{0,1,\ldots,k-1\}^{\mathbb{N}_0}$ (see Example 2.5). Show that such a metric cannot be τ-invariant, or equivalently, that there is no metric d on K making τ an isometry.

2. (See Examples 2.6 and 2.7) Show that the function $d(x,y) := \min(|x-y|, 1-|x-y|)$ is a metric on $[0,1)$ which turns it into a compact space. Prove that

$$\Psi : [0,1) \longrightarrow \mathbb{T}, \quad x \longmapsto \mathrm{e}^{2\pi \mathrm{i} x}$$

is a homeomorphism. Fix $\alpha \in [0,1)$ and set $a := \mathrm{e}^{2\pi \mathrm{i}\alpha}$; show that ψ_α, the addition mod 1 by α, is continuous with respect to the metric d, and show that Ψ induces an isomorphism of the dynamical systems $([0,1);\psi_\alpha)$ and $(\mathbb{T};\varphi_a)$.

3. Let $(K;\varphi)$ be a TDS, and let $A \subset K$. Show that if A is invariant then $\overline{A}$ is invariant, too. Show that the intersection of arbitrary (bi-)invariant sets is (bi-)invariant.

4. Let the **doubling map** be defined as $\varphi(x) = 2x \pmod 1$ for $x \in [0,1]$. Show that φ is continuous with respect to the metric introduced in Example 2.6. Show that $([0,1);\varphi)$ is transitive.

5. Consider the **tent map** $\varphi(x) := 1 - |2x-1|$, $x \in [0,1]$. Find a point with dense forward orbit. Prove that for every non-empty closed sub-interval I of $[0,1]$, there is $n \in \mathbb{N}$ with $\varphi^n(I) = [0,1]$.

6. Consider the function $\psi : [0,1] \longrightarrow [0,1]$, $\psi(x) := 4x(1-x)$. Show that $([0,1];\psi)$ and $([0,1];\varphi)$ are isomorphic, where φ is the tent map from Exercise 2.5.

7. Let $k \in \mathbb{N}$ and consider the full shift $(\mathscr{W}_k^+;\tau)$ over the alphabet $\{0,\ldots,k-1\}$. Let $x,y \in \mathscr{W}_k^+$. Show that $y \in \overline{\mathrm{orb}}_+(x)$ if and only if for every $n \geq 0$ there is $k \in \mathbb{N}$ such that

$$y_0 = x_k,\, y_1 = x_{k+1},\, \ldots,\, y_n = x_{k+n}.$$

Show that there is $x \in \mathscr{W}_k^+$ with dense forward-orbit.

8. Provide an example of a topologically transitive invertible TDS such that neither the forward nor the backward system is topologically transitive.

Lecture 3
Minimality and Recurrence

In this lecture, we study the existence of non-trivial subsystems in TDSs and the intrinsically connected phenomenon of *regularly recurrent* points. It was G.D. Birkhoff who discovered this and wrote in [Birkhoff (1912)]:

> THÉORÈME III. — *La condition nécessaire et suffisante pour qu'un mouvement positivement et négativement stable M soit un mouvement récurrent est que pour tout nombre positif ε, si petit qu'il soit, il existe un intervalle de temps T, assez grand pour que l'arc de la courbe représentative correspondant à tout intervalle égal à T ait des points distants de moins de ε de n'importe quel point de la courbe tout entière.*

and then

> THÉORÈME IV. — *L'ensemble des mouvements limites oméga M', de tout mouvement positivement stable M, contient au moins un mouvement récurrent.*

Our aim is to prove these results, today known as Birkhoff's Recurrence Theorem, and to study the connected notions. To start with, let us recall from Lecture 2 the notion of an *invariant set*. Given a TDS $(K;\varphi)$, a set $A \subset K$ is called φ*-invariant* if $\varphi(A) \subset A$; if even $A = \varphi^{-1}(A)$ holds we call it *bi-invariant*. If we have a non-trivial closed invariant set A we can consider the corresponding *subsystem* $(A;\varphi|_A)$.

It was Exercise 2.3 in Lecture 2 to show that the closure of an invariant set is invariant and the intersection of arbitrarily many (bi)-invariant sets is again (bi)-invariant. It is also easy to see that for every point $x \in K$ the forward orbit $\mathrm{orb}_+(x)$ and hence its closure are both invariant. So we have many possibilities to construct a subsystem: Simply pick a point $x \in K$ and consider $F := \overline{\mathrm{orb}}_+(x)$. Then $(F;\varphi|_F)$ is a subsystem, which coincides with $(K;\varphi)$ whenever $x \in K$ is a *transitive point*. Can we always find non-transitive points, hence non-trivial subsystems?

Before answering this question let us spend a few words on the invertibility of our TDSs and the difference between invariant and bi-invariant sets. If the TDS is invertible, then A is bi-invariant if and only if $\varphi(A) = A$, and in an invertible TDS the closure of a bi-invariant set is bi-invariant and every total orbit $\mathrm{orb}(x)$, is bi-invariant. The next lemma shows that every TDS contains a surjective subsystem.

Lemma 3.1. *Let $(K;\varphi)$ be a TDS and let $\emptyset \neq A \subset K$ be closed and invariant. There is a closed set $\emptyset \neq B \subset A$ such that $\varphi(B) = B$.*

Proof. Since $A \subset K$ is invariant,

$$A \supset \varphi(A) \supset \varphi^2(A) \supset \cdots \supset \varphi^n(A) \quad \text{holds for all } n \in \mathbb{N},$$

and all these sets are compact and non-empty, since A is closed and φ is continuous. Thus the set $B := \bigcap_{n\in\mathbb{N}} \varphi^n(A)$ is again non-empty and compact, and satisfies $\varphi(B) \subset B$. We now prove that actually $\varphi(B) = B$. Let $x \in B$ be arbitrary. Then for each $n \in \mathbb{N}$ we have $x \in \varphi^n(A)$, i.e., $\varphi^{-1}\{x\} \cap \varphi^{n-1}(A) \neq \emptyset$. Hence these sets form a decreasing sequence of compact non-empty sets, and therefore their intersection $\varphi^{-1}\{x\} \cap B$ is not empty either.

We now return to the above question.

3.1 Minimality

TDSs without non-trivial subsystems deserve a special name.

Definition 3.2. A TDS $(K;\varphi)$ is called **minimal** if there are no non-trivial closed φ-invariant sets in K. This means that if $A \subset K$ is closed and $\varphi(A) \subset A$, then $A = \emptyset$ or $A = K$.

By Lemma 3.1, in a minimal system $(K;\varphi)$ one must have $\varphi(K) = K$, i.e. φ is surjective. So an invertible TDS is minimal if and only if it does not contain non-trivial closed *bi-invariant* subsets.

Example 3.3. a) If $(K;\varphi)$ is a TDS and $(K_1;\varphi_1)$, $(K_2;\varphi_2)$ are minimal subsystems, then either $K_1 \cap K_2 = \emptyset$ or $K_1 = K_2$.

b) Minimality is an isomorphism invariant, i.e. if two TDSs are isomorphic and one of them is minimal, so is the other.

c) More generally, if $(K;\varphi)$ and $(H;\psi)$ are TDSs such that there is a surjective homomorphism $\Psi : K \longrightarrow H$ (in this case $(H;\psi)$ is called a **factor** of the TDS $(K;\varphi)$), then $(H;\psi)$ is minimal whenever $(K;\varphi)$ is minimal. In particular, if a product TDS is minimal so are each of its components (the converse is not true, see for example Theorem 2.21 in Lecture 2).

Since minimality means that there are no non-trivial subsystems, *"irreducible"* appears to be the adequate term. However, the term *"minimal"* is generally used in topological dynamics.

The following is an easy characterisation of minimality by the forward orbits of the points in K.

Proposition 3.4. *Let $(K;\varphi)$ be a TDS. The following assertions are equivalent.*

(i) $(K;\varphi)$ *is minimal;*

(ii) $\mathrm{orb}_+(x)$ *is dense in K for all* $x \in K$*;*

(iii) $K = \bigcup_{n\in\mathbb{N}_0} \varphi^{-n}(U)$ *for every open set* $\emptyset \neq U \subset K$.

Proof. This is Exercise 3.3.

By this proposition, every minimal system is forward transitive. Furthermore, we observed in Lecture 2 that for a group rotation $(G;\varphi_h)$, G a compact group and $h \in G$, either all points have dense orbit or there is no point with this property. In combination with Proposition 3.4 this yields the following.

Proposition 3.5. *Let G be a compact group and $h \in G$. Then the TDS $(G;\varphi_h)$ is minimal if and only if it is topologically transitive. Moreover, this happens exactly when* $\mathrm{orb}_+(1)$ *is dense (where* 1 *denotes the unit element in G).*

There are systems which are far from being minimal. As an example, consider the full shift $(\mathscr{W}_k^+;\tau)$ on words from a k-letter alphabet. Then every constant sequence is shift-invariant and constitutes a one-point minimal subsystem. It is actually a general fact that one always finds a subsystem that is minimal.

Theorem 3.6. *In every $(K;\varphi)$ there exists a non-empty φ-invariant, closed subset $F \subset K$ such that the subsystem $(F;\varphi|_F)$ is minimal.*

Proof. Let $\mathscr{M}$ be the family of all non-empty closed φ-invariant subsets of K. Then, of course, $K \in \mathscr{M}$, so $\mathscr{M}$ is non-empty. Further, it is ordered by the inclusion relation $\subset$. The partially ordered set $(\mathscr{M},\subset)$ satisfies the chain-condition of Zorn's Lemma, so there is a minimal element $F \in \mathscr{M}$ with respect to this partial order. It is obvious that $(F;\varphi|_F)$ is minimal (in the sense of Definition 3.2).

Beside compact groups there is another important class of TDSs in which minimality and transitivity coincide.

Proposition 3.7. *Let $(K;\varphi)$ be an isometric TDS, i.e. assume that there is a metric d inducing the topology of K such that φ is an isometry with respect to d. Then $(K;\varphi)$ is minimal if and only if it is topologically transitive.*

Proof. Suppose that $x_0 \in K$ has dense forward orbit and pick $y \in K$. It suffices to prove that $x_0 \in \overline{\mathrm{orb}_+(y)}$. Let $\varepsilon > 0$ be arbitrary. Using denseness of $\mathrm{orb}_+(x_0)$ we find $m \in \mathbb{N}_0$ such that $d(\varphi^m(x_0),y) \leq \varepsilon$. The sequence $(\varphi^{mn}(x_0))_n$ has a convergent subsequence, and hence by using that φ is an isometry we obtain

$$d(x_0,\varphi^{m(n_{k+1}-n_k)}(x_0)) = d(\varphi^{mn_k}(x_0),\varphi^{mn_{k+1}}(x_0)) \to 0 \quad \text{for } k \to \infty.$$

This yields that there is $n \geq m$ with $d(x_0,\varphi^n(x_0)) < \varepsilon$. We can write

$$\begin{aligned} d(\varphi^{(n-m)}(y),x_0) &\leq d(\varphi^{n-m}(y),\varphi^n(x_0)) + d(\varphi^n(x_0),x_0) \\ &= d(y,\varphi^m(x_0)) + d(\varphi^n(x_0),x_0) < 2\varepsilon. \end{aligned}$$

The assertion follows.

For isometric systems we have the following decomposition into minimal subsystems.

Corollary 3.8 ("Structure theorem" for isometric systems). *An isometric system is a (possibly) infinite disjoint union of minimal subsystems.*

Proof. The statement is equivalent to the fact that every point in K is contained in a unique minimal system (by Example 3.3 different minimal subsystems must be disjoint). But for any $x \in K$ we have that $(\overline{\mathrm{orb}}_+(x); \varphi|_{\overline{\mathrm{orb}}_+(x)})$ is topologically transitive, hence minimal by Proposition 3.7.

Example 3.9. Consider the rotation φ_a, $a \in \mathbb{T}$, on the closed unit disc

$$\mathbb{D} = \{z \in \mathbb{C} : |z| \leq 1\}, \quad \varphi_a : z \longmapsto az.$$

The system $(\mathbb{D}; \varphi_a)$ is isometric, but is not minimal. If a is not a root of unity, then any of the subsystems $(F_r; \varphi_a|_{F_r})$ with $F_r := \{z \in \mathbb{C} : |z| = r\}$, $r \in [0,1]$ is minimal, and we have $\mathbb{D} = \bigcup_{r \in [0,1]} F_r$ (disjoint union).

3.2 Topological Recurrence

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

For the rotation on the circle $\mathbb{T}$ we have observed in Example 2.19 two mutually exclusive phenomena:

1) For rational rotation every orbit is *periodic*.
2) For irrational rotation every orbit is *dense*.

In neither case, however, it matters at which point $x \in \mathbb{T}$ we start, the iterates return to a neighbourhood of x from time to time (in case of a rational rotation even exactly to the same point). Given a TDS $(K; \varphi)$ we can classify the points according to this behaviour of their orbits.

Definition 3.10. A point $x \in K$ is called

a) **recurrent** if for every open neighbourhood U of x, there is $m \in \mathbb{N}$ such that $\varphi^m(x) \in U$.
b) **almost periodic** if for every open neighbourhood U of x, the set of **return times**
$$\{m \in \mathbb{N} : \varphi^m(x) \in U\}$$
has bounded gaps. (A subset $M \subset \mathbb{N}$ has bounded gaps if there is $N \in \mathbb{N}$ such that $M \cap [n, n+N] \neq \emptyset$ for every $n \in \mathbb{N}$.)
c) **periodic** if there is $m \in \mathbb{N}$ such that $\varphi^m(x) = x$.

All these properties are invariant under automorphisms (this we shall use frequently). Clearly, a periodic point is almost periodic, and the latter is recurrent. None of the converse implications is true in general (see Example 3.11). Further, a recurrent point is even **infinitely recurrent**, i.e. it returns infinitely often to all of its neighbourhoods (see Exercise 3.5). It is also obvious that all these properties of a point $x \in K$ are decided in $\overline{\mathrm{orb}}_+(x)$ only. So we can restrict the system to the forward orbit-closure and assume topological transitivity. In this context it is very helpful to use the notation

$$\mathrm{orb}_{>0}(x) := \left\{ \varphi^n(x) : n \in \mathbb{N} \right\} = \mathrm{orb}_+(\varphi(x)).$$

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

For an irrational rotation on the torus all points are almost periodic but not periodic. Here are some more examples.

Example 3.11. Consider the full shift $(\mathscr{W}_k^+; \tau)$ in a k-alphabet.

a) An $x \in \mathscr{W}_k^+$ is recurrent if every finite block of x occurs infinitely often in x.
b) The point x is almost periodic if it is recurrent and the gaps between two occurrences of a given block y of x is bounded.
c) A concrete example for a non-periodic but almost periodic point is given in Exercise 3.9.
d) By the above, a non-almost periodic but recurrent point is easy to find: For the sake of simplicity suppose $k = 2$ and consider all the finite words formed from the alphabet $\{0, 1\}$, and enumerate them according to a lexicographical ordering. Now write these words into one infinite word $x \in \mathscr{W}_2^+$:

$$0\,|\,1\,|\,00\,|\,01\,|\,10\,|\,11\,|\,000\,|\,001\,|\,010\,|\,011\,|\,100\,|\,101\,|\,110\,|\,111\,|\,0000|\cdots.$$

All blocks of x occur as sub-block of later finite blocks, hence they are repeated infinitely often, and hence x is recurrent. On the other hand, since there are arbitrary long sub-words consisting of the letter 0 only, we see that the block 1 does not appear with bounded gaps.

In the TDS $(\mathbb{T}; \varphi_a)$ all points are recurrent, see Example 2.19. Not only in groups, but also in minimal systems this is so. Indeed, if K is minimal and $x \in K$, then $\overline{\mathrm{orb}}_{>0}(x)$ is a non-trivial, closed invariant set, hence it coincides with K and thus contains x. One can push this result a little further.

Proposition 3.12 (G.D. Birkhoff). *Every point in a minimal TDS $(K; \varphi)$ is almost periodic.*

Proof. Assume $(K; \varphi)$ to be minimal and take $x \in K$ and $U \subset K$ a non-empty open neighbourhood of x. Since every point in K is transitive, we have $\bigcup_{n \in \mathbb{N}_0} \varphi^{-n}(U) = K$. By compactness, we find $n_1 < n_2 < \cdots < n_k \in \mathbb{N}_0$ with

$$K = \varphi^{-n_1}(U) \cup \varphi^{-n_2}(U) \cup \cdots \cup \varphi^{-n_k}(U).$$

For each $n \in \mathbb{N}$ we have $\varphi^n(x) \in \varphi^{-n_i}(U)$, i.e. $\varphi^{n+n_i}(x) \in U$, for some $1 \le i \le k$. So the set of return times $\{m \in \mathbb{N}_0 : \varphi^m(x) \in U\}$ has gaps of length at most n_k.

Combining Theorem 3.6 and the previous result we obtain the following.

Theorem 3.13 (G.D. Birkhoff). *Let $(K;\varphi)$ be a TDS. Then there is an $x_0 \in K$ which is almost periodic (hence recurrent).*

As a corollary we obtain the following property of the translation on $[0,1)$.

Proposition 3.14. *Let $\alpha \in [0,1)$ and $I \subset [0,1)$ an interval containing α in its interior. Then the set*

$$\{n \in \mathbb{N} : n\alpha - \lfloor n\alpha \rfloor \in I\}$$

has bounded gaps.

Proof. Consider $([0,1);\varphi_\alpha)$, the translation mod 1 (see Example 2.6). If α is rational, the orbit is even periodic, while for irrational α the system is minimal hence α is an almost periodic point.

Example 2.18 suggests that it is actually the group structure, more precisely the "uniform-looking" of the neighbourhoods in a group, that accounts for the recurrence phenomenon in the above TDS.

Proposition 3.15. *Let G be a compact group, $h \in G$ and φ_h the corresponding left-rotation. Then each $g \in G$ is recurrent and even almost periodic.*

Proof. By Proposition 3.13, there is at least one $g_0 \in G$ which is almost periodic. Now, for every $g \in G$ the right multiplication $\rho_g : G \longrightarrow G$, $\rho_g(x) = xg$, is an automorphism of $(G;\varphi_h)$. Since automorphisms map almost periodic points to almost periodic points, $g_0 g$ is almost periodic, for every $g \in G$. Replacing g by $g_0^{-1}g$ in this argument proves the claim.

It is an interesting question for which x the forward orbit-closure $\overline{\mathrm{orb}}_+(x)$ defines a minimal subsystem. If this is so, x has to be almost periodic by Proposition 3.12. The converse holds, too.

Proposition 3.16. *Let $x \in K$ be an almost periodic point in a TDS $(K;\varphi)$, and set $F := \overline{\mathrm{orb}}_+(x)$. Then $(F;\varphi|_F)$ is a minimal subsystem.*

Proof. The set $F = \overline{\mathrm{orb}}_+(x)$ is closed and invariant. We have to prove that for all $y \in F$ the forward orbit $\mathrm{orb}_+(y)$ is dense in F. It suffices to show that $x \in \overline{\mathrm{orb}}_+(y)$. Let U be a closed neighbourhood of x (see Exercise 3.10.a). Since x is almost periodic, the set $\{n \in \mathbb{N} : \varphi^n(x) \in U\}$ has gaps with maximal length $m \in \mathbb{N}$. This shows that

$$\mathrm{orb}_+(x) \subset \bigcup_{i=1}^{m+1} \varphi^{-i}(U)$$

and hence $y \in \overline{\mathrm{orb}}_+(x) \subset \bigcup_{i=1}^{m+1} \varphi^{-i}(U)$. This implies $\varphi^i(y) \in U$ for some $1 \le i \le m+1$, and therefore $x \in \overline{\mathrm{orb}}_+(y)$.

This immediately yields a sufficient condition for a topologically transitive system to be minimal.

Corollary 3.17. a) *Let $(K;\varphi)$ be a forward topologically transitive TDS with $x_0 \in K$ a forward transitive point. If x_0 is almost periodic, then the system is minimal.*

b) *In an isometric TDS every point is almost periodic.*

Proof. a) follows from Proposition 3.16. For b) we have to take into account that for $F := \overline{\mathrm{orb}}_+(x)$, the subsystem $(F;\varphi|_F)$ is forward topologically transitive. If the system is isometric, then it is even minimal by Proposition 3.7. In a minimal system, however, every point, hence x, is almost periodic.

3.3 Recurrence in Extensions

Also in group extensions (see Lecture 2, page 10) there are many recurrent points.

Proposition 3.18. *Let $(K;\varphi)$ be a TDS, G a compact group and $(H;\psi)$ the group extension along some $\Phi : K \longrightarrow G$. If $x_0 \in K$ is a recurrent point (such points exists by Theorem 3.13), then $(x_0,g) \in H$ is recurrent in H for all $g \in G$.*

Proof. It suffices to prove the assertions for $g = 1 \in G$. Indeed, for every $g \in G$ the map $\rho_g : H \longrightarrow H$, $\rho_g(x,h) = (x,hg)$, is an automorphism of $(H;\psi)$, and hence maps recurrent points to recurrent points. For every $n \in \mathbb{N}$ we have

$$\psi^n(x_0,1) = \big(\varphi^n(x_0), \Phi(\varphi^{n-1}(x_0)) \cdots \Phi(x_0)\big).$$

Since the projection $\pi : H \longrightarrow K$ is continuous, we have $\pi(\overline{\mathrm{orb}}_{>0}(x_0,1)) = \overline{\mathrm{orb}}_{>0}(x_0)$ (see Exercise 3.10.b). The point x_0 is recurrent by assumption, so $x_0 \in \overline{\mathrm{orb}}_{>0}(x_0)$, and therefore for some $h \in G$ we have $(x_0,h) \in \overline{\mathrm{orb}}_{>0}(x_0,1)$. Multiply again from the right in the second coordinate by h and obtain by continuity that $(x_0,h^2) \in \overline{\mathrm{orb}}_{>0}(x_0,h)$. Clearly we have $\overline{\mathrm{orb}}_{>0}(x_0,h) \subseteq \overline{\mathrm{orb}}_{>0}(x_0,1)$ and therefore $(x_0,h^2) \in \overline{\mathrm{orb}}_{>0}(x_0,1)$. Inductively we obtain that $(x_0,h^n) \in \overline{\mathrm{orb}}_{>0}(x_0,1)$ for all $n \in \mathbb{N}$. By Proposition 3.15 we know that 1 is recurrent in $(G;\varphi_h)$, so if V is a neighbourhood of 1, then $h^n \in V$ for some $n \in \mathbb{N}$. This means that $(x_0,h^n) \in U \times V$ for U any neighbourhood of x_0. Thus

$$(x_0,1) \in \overline{\big\{(x_0,h),(x_0,h^2),\dots\big\}},$$

and therefore $(x_0,1) \in \overline{\mathrm{orb}}_{>0}(x_0,1)$ by the above. This means that $(x_0,1)$ is a recurrent point in $(H;\psi)$.

An analogous result is true for almost periodic points.

Proposition 3.19. *Let $(K;\varphi)$ be a TDS, G a compact group and $(H;\psi)$ the group extension along a given $\Phi : K \longrightarrow G$. If $x_0 \in K$ is an almost periodic point, then $(x_0,g) \in H$ is almost periodic in H for all $g \in G$.*

Proof. The set $\overline{\mathrm{orb}}_+(x_0)$ is minimal by Proposition 3.16, so by passing to a sub-system we can assume that $(K;\varphi)$ is minimal. Now let $(H';\psi|_{H'})$ be a minimal sub-system in $(H;\psi)$ (see Proposition 3.6). The projection $\pi : H \longrightarrow K$ is a homomorphism from $(H;\psi)$ to $(K;\varphi)$, so the image $\pi(H')$ is a φ-invariant subset in K, and therefore must be equal to K. Let x_0 be an almost periodic point in $(K;\varphi)$, and $h \in G$ such that $(x_0,h) \in H'$. By the Proposition 3.12 the point (x_0,h) is almost periodic. For arbitrary $g \in G$ the right multiplication with $h^{-1}g$ in the second coordinate maps (x_0,h) to (x_0,g) and is an automorphism of $(H;\psi)$ (see Lecture 2, page 10). Hence (x_0,g) is almost periodic for every $g \in G$.

A first application of the above is to Diophantine approximations. It is elementary that any real number can be approximated by rational numbers, but how well and, at the same time, how simple (keeping the denominator small) this approximation can be, is a hard question. The most basic answer, Dirichlet's Theorem, is given in Exercise 3.8. Another, not so elementary result, which we prove by using the above recurrence result, is the following.

Corollary 3.20. *Let $\alpha \in \mathbb{R}$ and $\varepsilon > 0$ be given. Then there exists $n \in \mathbb{N}$, $k \in \mathbb{Z}$ such that*

$$\left|n^2\alpha - k\right| \le \varepsilon.$$

Proof. Consider the TDS $([0,1);\varphi_\alpha)$ from Example 2.6, and recall that with the metric $d(x,y) := \min\{|x-y|, 1-|x-y|\}$ the space $[0,1)$ endowed with modulo 1 addition becomes a compact group being homeomorphic to $\mathbb{T}$. We consider a group extension similar to Example 2.10.3 but now in the mod 1 setting. Define the function $\Phi(x) = 2x+\alpha$ and the group extension by $[0,1)$ (modulo 1). Then $\psi(x,y) = (x+\alpha, 2x+\alpha+y)$, and by Proposition 3.18 $(0,0)$ is a recurrent point in $([0,1)\times[0,1);\psi)$. We have $\psi(0,0) = (\alpha,\alpha)$, $\psi^2(0,0) = \psi(\alpha,\alpha) = (2\alpha,4\alpha),\ldots,\psi^n(0,0) = (n\alpha,n^2\alpha)$ (use induction!). The recurrence of $(0,0)$ implies that for any $0 < \varepsilon$ we have $d(0,n^2\alpha - \lfloor n^2\alpha\rfloor) < \varepsilon$ for some $n \in \mathbb{N}$, hence the assertion follows.

Later we will see more applications of ergodic theory to number theory, and particulary to Diophantine approximations. Just with the same technique, i.e. by using the recurrence of points in appropriate group extensions, we can obtain a result on Diophantine inequalities.

Proposition 3.21. *Let $p \in \mathbb{R}[x]$ be a polynomial of degree $k \in \mathbb{N}$ with $p(0) = 0$. For every $\varepsilon > 0$ there is $n \in \mathbb{N}$ and $m \in \mathbb{N}_0$ with*

$$\left|p(n) - m\right| < \varepsilon.$$

Proof. Start from a polynomial $p_k(x) := p(x)$ and define $p_{k-i}(x) := p_{k-i+1}(x+1) - p_{k-i+1}(x)$ for $i = 1,\ldots,k$. Each polynomial p_i has degree i. So p_0 is a constant α. Consider the TDS $([0,1);\varphi_\alpha)$ (addition mod 1, see Example 2.6) and the following tower of group extensions. Set $H_1 := [0,1)$, $\psi_1 := \varphi_\alpha$, and for $2 \le i \le k$ define

$$H_i = H_{i-1} \times [0,1), \quad \Phi_i : H_{i-1} \longrightarrow [0,1),\ \Phi_i((x_1,x_2,\ldots,x_{i-1})) = x_{i-1}$$

hence for the group extension $(H_i;\psi_i)$ of $(H_{i-1};\psi_{i-1})$ along Φ_i we have

$$\psi_i(x_1,x_2,\dots,x_i) = (x_1+\alpha, x_1+x_2, x_2+x_3,\dots,x_i+x_{i-1}).$$

We can apply Proposition 3.18 and obtain that by starting from a recurrent point x_1 in $(H_1;\psi_1)$ every point in $(x_1,x_2,\dots,x_i)$ in $(H_i;\psi_i)$, $1\le i\le k$ is recurrent. In particular, the point $p_1(0)$ just as any other point in $[0,1)$ is recurrent in the TDS $([0,1);\varphi_\alpha)$, hence so is the point $(p_1(0),p_2(0),\dots,p_k(0))\in H_k$. By the definition of ψ_k we have

$$\begin{aligned}\psi_k\big((p_1(0),p_2(0),\dots,p_k(0))\big) &= \big(p_1(0)+\alpha, p_2(0)+p_1(0),\dots,p_k(0)+p_{k-1}(0)\big)\\ &= \big(p_1(0)+p_0(0), p_2(0)+p_1(0),\dots,p_k(0)+p_{k-1}(0)\big)\\ &= \big(p_1(1),p_2(1),\dots,p_k(1)\big),\end{aligned}$$

and therefore

$$\psi_k^n\big((p_1(0),p_2(0),\dots,p_k(0))\big) = \big(p_1(n),p_2(n),\dots,p_k(n)\big).$$

By looking at the last component we see that for some $n\in\mathbb{N}$ the point $p_k(n)=p(n)$ comes arbitrarily close to $p_k(0)=p(0)=0$. The assertion is proved.

We refer to [Furstenberg (1981)] for more results on Diophantine approximations that can be obtained by means of simple arguments using topological recurrence.

Exercises

$*$ **1.** Translate the quotation from [Birkhoff (1912)] and explain its meaning.

2. Consider the **tent map** $\varphi(x) := 1-|2x-1|$, $x\in[0,1]$. In Exercise 2.3 we showed the topological transitivity of $([0,1];\varphi)$. Is this TDS also minimal?

3. Prove Proposition 3.4.

4. Take $\mathbb{A} := \{0,1\}^{\mathbb{N}}$ with the product topology. We define an addition on $\mathbb{A}$ by adding pointwise modulo 2, with carry to the right. Show that with respect to this addition, $\mathbb{A}$ becomes a compact group (the **dyadic integers**). Let $\mathbf{1} := (1,0,0\dots)$ and $\varphi(x) := x+\mathbf{1}$. Show that the system $(\mathbb{A};\varphi)$ is minimal. This system is called the **dyadic adding machine**.

5. Let $(K;\varphi)$ be a TDS and let $x_0\in K$ be a recurrent point. Show that x_0 is *infinitely recurrent*, i.e. for every neighbourhood U of x_0 one has

$$x_0\in\bigcap_{n\in\mathbb{N}}\bigcup_{m\ge n}\varphi^{-n}(U)$$

6. Let $(K;\varphi)$ be a TDS and let $x_0 \in K$ be a recurrent point. Show that x_0 is also recurrent in the TDS $(K;\varphi^m)$ for all $m \in \mathbb{N}$. (Hint: Use a group extension by the cyclic group $\mathbb{Z}_m$.)

7. Consider the sequence $2, 4, 8, 16, 32, 64, 128, \ldots, 2^n, \ldots$. It seems that 7 does not appear as leading digit in this sequence. Show however that actually 7, just as any other digit, occurs infinitely often. As a test try out 2^{46}.

8.[Dirichlet's Theorem] Let $\alpha \in \mathbb{R}$ be irrational and $n \in \mathbb{N}$ be fixed. Show that there are $p_n \in \mathbb{Z}$, $q_n \in \mathbb{N}$ such that $1 \le q_n \le n$ and

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{nq_n}.$$

(Hint: Divide the interval $[0,1)$ into n subintervals of length $1/n$, and use the pigeonhole principle.) Show that $q_n \to +\infty$ for $n \to \infty$.

$*$ **9.**[Morse sequence] Consider $\mathscr{W}_2^+ = \{0,1\}^{\mathbb{N}}$ and define the following recursion on finite words over the alphabet $\{0,1\}$. We start with $f_1 = 01$. Then we replace every occurrence of letter 0 by the block 0110 and every occurrence of the letter 1 by 1001. We repeat this procedure at each step. For instance:

$$f_1 = 0|1 \quad f_2 = 0110|1001 \quad f_3 = 0110|1001|1001|0110|1001|0110|0110|1001.$$

We can now consider the finite words f_i, $i \in \mathbb{N}$ as elements of $\mathscr{W}_2^+$ by setting the "missing coordinates" as 0. Show that the sequence (f_n) converges to some $f \in \mathscr{W}_2^+$. Show that this f is almost periodic but not periodic in the shift TDS $(\mathscr{W}_2^+;\tau)$.

$*$ **10.** Let K be a compact space.

a) Let $x \in K$ and U an open neighbourhood of x. Show that there is an open neighbourhood V of x with $\overline{V} \subset U$.

b) Let $f : K \longrightarrow L$ be continuous, where L is Hausdorff and $A \subset K$. Show that $f(\overline{A}) = \overline{f(A)}$.

Lecture 4
The Space $\mathrm{C}(K)$ and the Induced Operator

To be linear or not to be linear: that is the question!

Rainer Nagel[1]

In the previous two lectures we introduced the concept of a topological dynamical system and certain central notions such as minimality, recurrence and transitivity. However, a deeper study of dynamical systems requires a change of perspective: instead of the state space transformation $\varphi : K \longrightarrow K$ we now consider its **induced operator** $T = T_\varphi$ defined by

$$T_\varphi f := f \circ \varphi$$

for scalar-valued functions f on K. On the space of all functions from K to $\mathbb{C}$ there is a rich structure: one can add, multiply, take absolute values or the complex conjugate. All these operations are defined pointwise, that is, we have

$$\begin{aligned} &(f+g)(x) := f(x)+g(x), \quad (\lambda f)(x) := \lambda f(x), \\ &(fg)(x) := f(x)g(x), \qquad \overline{f}(x) := \overline{f(x)}, \qquad |f|(x) := |f(x)| \end{aligned}$$

for $f,g : K \longrightarrow \mathbb{C}$, $\lambda \in \mathbb{C}$, $x \in K$. Therefore, the operator $T = T_\varphi$ commutes with all these operations, meaning that

$$\begin{aligned} &T(f+g) = Tf + Tg, \quad T(\lambda f) = \lambda (Tf), \\ &T(fg) = (Tf)(Tg), \qquad \overline{Tf} = T\overline{f}, \qquad |Tf| = T\,|f| \end{aligned}$$

for all $f,g : K \longrightarrow \mathbb{C}$, $\lambda \in \mathbb{C}$ (Exercise 1.6). In particular, the operator T is linear. Denoting by $\mathbf{1}$ the function taking everywhere the value 1, we also have $T\mathbf{1} = \mathbf{1}$.

Now, since φ is continuous, the operator T_φ leaves invariant the space $\mathrm{C}(K)$ of all $\mathbb{C}$-valued continuous functions on K; i.e., $f \circ \varphi$ is continuous whenever f is. The major part of this chapter is devoted to showing that actually the induced operator T_φ on $\mathrm{C}(K)$ *contains all information* about φ, that is, φ can be recovered from T_φ. In order to achieve this, we shall have a more detailed look on the space $\mathrm{C}(K)$.

[1] G.B. Shaw allegedly has said: “I often quote myself. It adds spice to my conversation.”

4.1 The space $\mathrm{C}(K)$ as a commutative C^*-algebra

Since a continuous image of a compact space is compact, each $f \in \mathrm{C}(K)$ is bounded, and so

$$\|f\|_\infty := \sup\{|f(x)| : x \in K\}$$

is finite. The map $\|\cdot\|_\infty$ is a norm on $\mathrm{C}(K)$ turning it into a complex *Banach space* (see also Appendix A.7). Beside the linear structure, $\mathrm{C}(K)$ is closed under the multiplication of functions, and we have

$$\|fg\|_\infty \le \|f\|_\infty \|g\|_\infty \quad \text{and} \quad \|\mathbf{1}\|_\infty = 1$$

for $f, g \in \mathrm{C}(K)$. This means that $\mathrm{C}(K)$ is a *commutative unital Banach algebra*. Since $\mathrm{C}(K)$ is also invariant under complex conjugation and

$$\|f\overline{f}\|_\infty = \|f\|_\infty^2 \qquad (f \in \mathrm{C}(K))$$

holds, we conclude that $\mathrm{C}(K)$ is a commutative unital *C^*-algebra*. (See Appendix C.1 for these definitions.)

In our study of $\mathrm{C}(K)$ we shall rely on three classical theorems which we state here without proofs. The first shows that the space $\mathrm{C}(K)$ is not poor.

Lemma 4.1 (Urysohn). *Let A, B be disjoint closed subsets of a compact space K. Then there exists a continuous function $f : K \longrightarrow [0,1]$ such that $f(A) \subset \{0\}$ and $f(B) \subset \{1\}$.*

If K is metrisable with metric d, then a function f as in Urysohn's Lemma is given by

$$f(x) = \frac{d(x,A)}{d(x,A) + d(x,B)} \qquad (x \in K),$$

where for a subset C one writes $d(x,C) := \inf\{d(x,y) : y \in C\}$. For the proof in the general case we refer to [Rudin (1987), 2.12]; cf. also Lemma A.3 in Appendix A. Urysohn's lemma says in particular that $\mathrm{C}(K)$ "separates the points" of K, that is, for every two distinct points of K there is a continuous function on K that takes different values on the given points.

Our second classical result is essentially an application of Urysohn's Lemma, see [Rudin (1987), Thm. 20.4] for a proof.

Theorem 4.2 (Tietze). *Let K be a compact space, let $A \subset K$ be closed and let $f \in \mathrm{C}(A)$. Then there is $h \in \mathrm{C}(K)$ such that $h|_A = f$.*

Our last standard result that we take from the literature is the following:

Theorem 4.3 (Stone–Weierstrass).

(Real version) *Let $\mathscr{A}$ be a real subalgebra of $\mathrm{C}(K;\mathbb{R})$ containing the constant functions and separating the points of K. Then $\mathscr{A}$ is dense in $\mathrm{C}(K;\mathbb{R})$.*

(Complex version) *Let $\mathscr{A}$ be a complex conjugation-invariant subalgebra of $\mathrm{C}(K)$ containing the constant functions and separating the points of K. Then $\mathscr{A}$ is dense in* $\mathrm{C}(K)$.

See [Rudin (1976), p.159] and [Yosida (1978), p. 9] for proofs.

Our aim in the rest of this section is to show that the compact space K can be reconstructed entirely from the Banach algebraic structure of $\mathrm{C}(K)$ (see Theorem 4.5 below).

Let $\mathscr{A}$ be a commutative Banach algebra. A linear subspace $I \subset \mathscr{A}$ is called an **algebra ideal** if $f \in I, g \in \mathscr{A}$ implies $fg \in I$. The norm closure of an ideal is again an ideal, since the multiplication is continuous. Clearly, if an ideal I contains the unit element e of $\mathscr{A}$ then $I = \mathscr{A}$. If $\mathrm{e} \notin I$, the ideal is called **proper**.

There is a simple way to construct closed ideals of $\mathrm{C}(K)$. Take a closed subset $A \subset K$; the set

$$I_A := \left\{ f \in \mathrm{C}(K) \,:\, f \equiv 0 \text{ on } A \right\}$$

clearly is a closed ideal. The next theorem tells that *each* closed ideal is of this type.

Theorem 4.4. *Let $I \subset \mathrm{C}(K)$ be a closed algebra ideal. Then there is a closed subset $A \subset K$ such that $I = I_A$.*

Proof. We use the notation $[\, f > 0\,] := \{x \in K \,:\, f(x) > 0\}$ and, analogously, $[\, f = 0\,]$ for the set of zeroes of a function f. Define

$$A := \left\{ x \in K \,:\, f(x) = 0 \;\forall f \in I \right\} = \bigcap\nolimits_{f \in I} [\, f = 0\,].$$

Evidently, A is closed and $I \subset I_A$. Fix $f \in I_A$, let $\varepsilon > 0$ and define $A_\varepsilon := [\,|f| \geq \varepsilon\,]$. Since f is continuous and vanishes on A, A_ε is a closed subset of $K \setminus A$. Hence for each point $x \in A_\varepsilon$ one can find a function $f_x \in I$ such that $f_x(x) \neq 0$. By multiplying with $\overline{f_x}$ and a positive constant we may assume that $f_x \geq 0$ and $f_x(x) > 1$. The collection of open sets $([\, f_x > 1\,])_{x \in A_\varepsilon}$ covers A_ε. Since A_ε is compact, this cover has a finite subcover. So there are $0 \leq f_1, \ldots, f_k \in I$ such that

$$A_\varepsilon \subset [\, f_1 > 1\,] \cup \cdots \cup [\, f_k > 1\,].$$

Let $g := f_1 + \cdots + f_k \in I$. Then $0 \leq g$ and $[\,|f| \geq \varepsilon\,] \subset [\, g \geq 1\,]$. Define

$$g_n := \frac{nf}{1+ng} g \in I.$$

Then
$$|g_n - f| = \frac{|f|}{1+ng} \leq \max\left\{ \varepsilon, \frac{\|f\|_\infty}{1+n} \right\}$$

on K. (One has $g \geq 1$ on A_ε and $|f| < \varepsilon$ on $K \setminus A_\varepsilon$.) Hence for n large enough we have $\|g_n - f\|_\infty \leq \varepsilon$. Since ε was arbitrary and I is closed, we obtain $f \in I$ as desired.

An ideal I in a unital Banach algebra $\mathscr{A}$ is called a **maximal ideal** if $I \neq \mathscr{A}$, and for every ideal J satisfying $I \subset J \subset \mathscr{A}$ either $J = I$ or $J = \mathscr{A}$. A maximal ideal is always closed (Exercise 4.1). If $\mathscr{A} = \mathrm{C}(K)$ then the maximal ideals are easy to spot: since each (closed) ideal is of the form I_A for some "vanishing-set" $A \subset K$, the ideal becomes larger as A shrinks. Hence $I = I_A$ is maximal if and only if $A = \{x\}$ for some $x \in K$, i.e., the vanishing-set is a point. So in $\mathrm{C}(K)$ the points of K are in one-to-one correspondence with the maximal ideals of $\mathrm{C}(K)$. In this way we recover the *set* K from the algebraic structure of $\mathrm{C}(K)$.

To recover also the topology, we have to work a little more. Let $\mathscr{A}, \mathscr{B}$ be unital Banach algebras with units e, e$'$, respectively. A linear mapping $T : \mathscr{A} \longrightarrow \mathscr{B}$ is called a **unital algebra homomorphism** if

$$T\mathrm{e} = \mathrm{e}' \quad \text{and} \quad T(fg) = (Tf)(Tg)$$

for all $f, g \in \mathscr{A}$. An example for this is the restriction mapping

$$R_A : \mathrm{C}(K) \longrightarrow \mathrm{C}(A), \qquad f \longmapsto f|_A,$$

where $A \subset K$ is closed. Note that the ideal I_A is exactly the kernel of R_A and R_A is surjective, by Tietze's theorem.

An important special case of this is when A consists of one single point $x \in K$ only. Identifying functions on $\{x\}$ with $\mathbb{C}$ we see that restriction amounts to point evaluation. Hence we define the **Dirac functional** δ_x by

$$\langle f, \delta_x \rangle := f(x) \qquad (f \in \mathrm{C}(K)).$$

Then $\delta_x \in \mathrm{C}(K)'$ is a unital algebra homomorphism from $\mathrm{C}(K)$ into $\mathbb{C}$, satisfying $\|\delta_x\| = 1$. Its kernel is the maximal ideal $\ker \delta_x = I_{\{x\}}$.

The Banach–Alaoglu theorem (Theorem C.4) asserts that the dual unit ball

$$M := \left\{ \nu \in \mathrm{C}(K)' : \|\nu\| \leq 1 \right\}$$

is a compact set with respect to the weak*-topology. (The weak*-topology is the coarsest topology on the dual making all functionals $\mu \longmapsto \langle f, \mu \rangle$, $f \in \mathrm{C}(K)$, continuous. See Appendix C.4.) Now, consider the mapping

$$\Phi : K \longrightarrow M, \qquad \Phi(x) := \delta_x \qquad (x \in K).$$

By Urysohn's theorem Φ is injective, and by the definition of the topology on M, Φ is continuous. Since the topology on M is Hausdorff and K is compact, Φ is a homeomorphism onto its image, the set of all Dirac functionals (cf. Appendix A.8). The next result tells us in particular that the Dirac functionals are exactly the unital complex algebra homomorphisms. Hence we have successfully recovered the compact space K from the algebraic structure of $\mathrm{C}(K)$.

Theorem 4.5. *Let K be a compact space and let $\psi : \mathrm{C}(K) \longrightarrow \mathbb{C}$ be a unital algebra homomorphism. Then there is a unique $x \in K$ such that $\psi = \delta_x$. Moreover, the map*

$$\Phi : K \longrightarrow \{\delta_x : x \in K\}, \quad x \longmapsto \delta_x$$

is a homeomorphism, when the image is endowed with the weak**-topology as a subset of* $\mathrm{C}(K)'$.

Proof. Only the first assertion remains to be proved. The uniqueness follows directly from Urysohn's lemma. For the existence, let us first prove that ψ is continuous. Fix $f \in \mathrm{C}(K)$ and take $\lambda \in \mathbb{C}$ such that $|\lambda| > \|f\|_\infty$. Then $\lambda - f$ has no zeroes, hence $(\lambda - f)^{-1} \in \mathrm{C}(K)$. Consequently

$$1 = \psi\mathbf{1} = \psi\big((\lambda - f)^{-1}(\lambda - f)\big) = \psi\big((\lambda - f)^{-1}\big)\,\psi(\lambda - f)$$

and thus $\lambda - \psi(f) = \psi(\lambda - f) \neq 0$. This shows that $|\psi(f)| \leq \|f\|_\infty$, and as f was arbitrary, ψ is bounded with norm equal to one. In particular, its kernel $I := \ker\psi$ is closed, and it is an algebra ideal since ψ is a homomorphism. By Theorem 4.4, we have $I = I_A$ for some closed subset $A \subset K$. However, since the range $\operatorname{ran}\psi$ of ψ is one-dimensional, by Urysohn's lemma A cannot contain more than one point. This means that there is $x \in K$ such that $\ker\psi = \ker\delta_x$, and this space together with $\mathbf{1}$ generates $\mathrm{C}(K)$. Since ψ and δ_x coincide on $\mathbf{1}$, too, they must coincide everywhere, i.e., $\psi = \delta_x$.

4.2 The Theorem of Gelfand–Naimark

Before we turn to induced operators we shall briefly continue with some abstract ideas. However, if you find this material too difficult, just skip it at the moment and return to it when we use it later.

Our investigations will lead us into the so-called *Gelfand theory* and — unfortunately — we shall not be able to provide full proofs of all the statements. However, we hope to make the results at least plausible. Extensive treatment of Gelfand theory can be found in [Rudin (1987), Chapter 18], [Rudin (1991), Part III] and most other books on functional analysis.

We have seen above that $\mathrm{C}(K)$ is a commutative unital C^*-algebra. Our aim is to show the converse:

Theorem 4.6 (Gelfand–Naimark).
Let $\mathscr{A}$ be an arbitrary commutative unital C^*-algebra. Then there is a compact space K and an isometric $*$-isomorphism $\Phi : \mathscr{A} \longrightarrow \mathrm{C}(K)$. The space K is unique up to homeomorphism.

By an isometric $*$-isomorphism of C^*-algebras we mean a Banach algebra homomorphism that is bijective and isometric and satisfies $\Phi(x^*) = \overline{\Phi(x)}$ for each $x \in \mathscr{A}$.

The idea behind the proof is to mimic what we did in the case $\mathscr{A} = \mathrm{C}(K)$. Recall that we found K back from $\mathrm{C}(K)$ by looking at $\mathbb{C}$-valued unital algebra homomorphisms.

Definition 4.7. Let $\mathscr{A}$ be a commutative unital complex Banach algebra. Then the set

$$\Gamma(\mathscr{A}) := \{\psi : \mathscr{A} \longrightarrow \mathbb{C} : \psi \text{ is a unital algebra homomorphism}\}$$

is called the **Gelfand (representation) space** of $\mathscr{A}$.

The crucial result about this notion is the following.

Lemma 4.8. *Let $\mathscr{A}$ be a unital commutative complex Banach algebra. If $\psi \in \Gamma(\mathscr{A})$ then ψ is continuous with $\|\psi\| \le 1$, and $\ker\psi$ is a maximal ideal. Conversely, if $I \subset \mathscr{A}$ is a maximal ideal, then it is closed and there is a unique $\psi \in \Gamma(\mathscr{A})$ such that $I = \ker\psi$.*

Proof. The first statement is proved in Exercises 4.1 and 4.2. For the second statement note that I must be closed by Exercise 4.1. The uniqueness is easy but the existence rests on the Gelfand–Mazur theorem, which would lead us too far away. See for instance [Rudin (1987), Thm. 18.17] for a concise exposition.

One of the important consequences of Lemma 4.8 is that $\Gamma(\mathscr{A})$ cannot be empty, since by an application of Zorn's Lemma, every proper ideal I of $\mathscr{A}$ is contained in a maximal one. Another consequence is that $\Gamma(\mathscr{A})$ is a subset of the dual unit ball of $\mathscr{A}$, i.e., the unit ball of the Banach space dual $\mathscr{A}'$.

As in the case of $\mathscr{A} = \mathrm{C}(K)$ above, let us consider $\mathscr{A}'$ endowed with its weak*-topology, i.e, the coarsest topology that makes all evaluation mappings

$$\mathscr{A}' \longrightarrow \mathbb{C}, \quad \psi \longmapsto \psi(x), \qquad x \in \mathscr{A},$$

continuous (see Appendix C.4). Then $\Gamma(\mathscr{A})$ is a weakly* closed subset of $\mathscr{A}'$ (see Exercise 4.3). By the Banach-Alaoglu theorem (Theorem C.4) the set $\Gamma(\mathscr{A})$ is compact. For $x \in \mathscr{A}$ we consider the function

$$\hat{x} : \Gamma(\mathscr{A}) \longrightarrow \mathbb{C}, \qquad \hat{x}(\psi) := \psi(x) \qquad (\psi \in \Gamma(\mathscr{A})).$$

By definition of the topology of $\Gamma(\mathscr{A})$, $\hat{x}$ is continuous. Hence the map

$$\Phi : \mathscr{A} \longrightarrow \mathrm{C}(\Gamma(\mathscr{A})), \qquad x \longmapsto \hat{x}$$

is well-defined and obviously a unital algebra homomorphism.

To conclude the proof of the Gelfand–Naimark theorem it remains to show that Φ commutes with the involutions (1), has dense range (2), and is isometric (3). The proof of (1) is elementary. Given (1), then (2) follows from the Stone–Weierstrass theorem, since $\operatorname{ran}\Phi$ is a conjugation-invariant subalgebra of $\mathrm{C}(\Gamma(\mathscr{A}))$, containing the constants and (trivially) separating the points. The proof of (3) rests on the so called spectral radius formula in Banach algebras. Details can be found in [Rudin (1991), Thm. 11.18].

4.3 The Induced Operator

We now return to our original setting of a TDS $(K;\varphi)$ with its induced operator $T = T_\varphi$. Actually, we shall be a little more general at first, i.e., we shall consider possibly different compact spaces K, L and a mapping $\varphi : L \longrightarrow K$. Obviously we can form the induced operator T_φ on the associated function spaces. The following lemma says that φ is continuous if and only if $T_\varphi(\mathrm{C}(K)) \subset \mathrm{C}(L)$.

Lemma 4.9. *Let K, L be compact spaces, and let $\varphi : L \longrightarrow K$ be a mapping. Then φ is continuous if and only if $f \circ \varphi$ is continuous for all $f \in \mathrm{C}(K)$.*

Proof. Clearly, if φ is continuous then also $f \circ \varphi$ is continuous for every $f \in \mathrm{C}(K)$. Conversely, if this condition holds then $\varphi^{-1}[\,|f| > 0\,] = [\,|f \circ \varphi| > 0\,]$ is open in L for every $f \in \mathrm{C}(K)$. To conclude that φ is continuous, it suffices to show that these sets *form a base of the topology of K* (see Appendix A.2, page 182). Now, these sets are clearly open. On the other hand, if $U \subset K$ is open and $x \in U$ is a point then by Urysohn's lemma one can find a function $f \in \mathrm{C}(K)$, such that $0 \le f \le 1$, $f(x) = 1$ and $f \equiv 0$ on $K \setminus U$. But then $x \in [\,|f| > 0\,] \subset U$.

If $\varphi : L \longrightarrow K$ is continuous, the operator $T = T_\varphi$ is a unital algebra homomorphism from $\mathrm{C}(K)$ to $\mathrm{C}(L)$. The next result shows that actually *every* unital algebra homomorphism is such an induced operator. For the case of a TDS, i.e., for $K = L$, the result shows in particular that the state space mapping φ is uniquely determined by its induced operator T_φ.

Theorem 4.10. *Let K, L be (non-empty) compact spaces and let $T : \mathrm{C}(K) \longrightarrow \mathrm{C}(L)$ be linear. Then the following assertions are equivalent:*

(i) *T is a unital algebra homomorphism.*

(ii) *There is a continuous mapping $\varphi : L \longrightarrow K$ such that $T = T_\varphi$.*

In this case, φ in (ii) is uniquely determined and T is bounded, with norm $\|T\| = 1$.

Proof. Urysohn's lemma yields that φ as in (ii) is uniquely determined, and it is clear from (ii) that $\|T\| = 1$. For the proof of the implication (i) $\Rightarrow$ (ii) take $f \in \mathrm{C}(K)$ and $y \in L$. Then

$$T'\delta_y := \delta_y \circ T : \mathrm{C}(K) \longrightarrow \mathbb{C}, \quad f \longmapsto (Tf)(y)$$

is a unital algebra homomorphism. By Theorem 4.5 there is a unique $x =: \varphi(y)$ such that $T'\delta_y = \delta_{\varphi(y)}$. This means that $(Tf)(y) = f(\varphi(y))$ for all $y \in L$, i.e., $Tf = f \circ \varphi$ for all $f \in \mathrm{C}(K)$. By Lemma 4.9, φ is continuous, whence (ii) is established.

Theorem 4.10 with $K = L$ shows that *no information is lost* when looking at T_φ in place of φ itself. On the other hand, one has all the tools from linear analysis — in particular spectral theory — to study the linear operator T_φ. Our aim is to see how properties of the TDS $(K;\varphi)$ reflect in properties of the operator T_φ. Here is a first result in this direction.

Lemma 4.11. *Let K, L be compact spaces, and let $\varphi : L \longrightarrow K$ be continuous, with induced operator $T = T_\varphi : \mathrm{C}(K) \longrightarrow \mathrm{C}(L)$. Then the following assertion holds:*

a) *φ is surjective if and only if T is injective; and in this case, T is isometric.*

b) *φ is injective if and only if T is surjective.*

Proof. This is Exercise 4.4.

We now return to TDSs and their invariant sets.

Lemma 4.12. *Let $(K;\varphi)$ be a TDS with induced operator $T = T_\varphi$, and let $A \subset K$ be a closed subset. Then A is φ-invariant if and only if the ideal I_A is T-invariant.*

Proof. Suppose that A is φ-invariant and $f \in I_A$. If $x \in A$ then $\varphi(x) \in A$ and hence $(Tf)(x) = f(\varphi(x)) = 0$ since f vanishes on A. Thus Tf vanishes on A, too, i.e., $Tf \in I_A$. Conversely, suppose that I_A is T-invariant. If $x \notin A$ then by Urysohn's lemma there is $f \in I_A$ such that $f(x) \neq 0$. By hypothesis, $Tf = f \circ \varphi$ vanishes on A, hence $x \notin \varphi(A)$. This implies that $\varphi(A) \subset A$.

Using Theorem 4.5, we obtain the following characterisation of minimality.

Corollary 4.13. *Let $(K;\varphi)$ be a TDS and $T = T_\varphi$ its induced operator on $\mathrm{C}(K)$. Then the TDS $(K;\varphi)$ is minimal if and only if no non-trivial closed algebra ideal of $\mathrm{C}(K)$ is invariant under T.*

One of the most important objects in the study of T_φ is its **fixed space**

$$\mathrm{Fix}(T_\varphi) := \big\{ f \in \mathrm{C}(K) \,:\, T_\varphi f = f \big\}.$$

The fixed space is the eigenspace corresponding to the eigenvalue 1, hence a spectral notion. It is never trivial since $T_\varphi \mathbf{1} = \mathbf{1}$, so it contains the constant functions. Clearly, if $f \in \mathrm{Fix}(T_\varphi)$ then f is constant on each forward orbit $\mathrm{orb}_+(x)$, and, since f is continuous, even on the closure of each forward orbit. This observation leads to a spectral implication of topological transitivity.

Lemma 4.14. *Let $(K;\varphi)$ be a TDS with induced operator $T = T_\varphi$ on $\mathrm{C}(K)$. If $(K;\varphi)$ is topologically transitive, then $\mathrm{Fix}(T)$ is one-dimensional.*

Proof. As already remarked, if $x \in K$ and $f \in \mathrm{Fix}(T)$, then $f(\varphi^n(x)) = (T^n f)(x) = f(x)$ is independent of $n \geq 0$, and hence f is constant on $\overline{\mathrm{orb}}_+(x)$. Consequently, if there is a point with dense forward orbit, then every $f \in \mathrm{Fix}(T)$ must be a constant.

Exercise 4.5 shows that the converse statement fails.

An eigenvalue λ of T of modulus 1 is called a **peripheral eigenvalue**. The set of peripheral eigenvalues

$$\sigma_p(T) \cap \mathbb{T} := \left\{\lambda \in \mathbb{T} : \exists 0 \neq f \in \mathrm{C}(K) \text{ with } Tf = \lambda f\right\}$$

is called the **peripheral point spectrum** of T, cf. Appendix C.9. We shall see later that this set is of importance for the asymptotic behaviour of the iterates of T. In the case of a topologically transitive system, the peripheral point spectrum of the induced operator has a particularly nice property.

Theorem 4.15. *Let $(K;\varphi)$ be a TDS with induced operator $T = T_\varphi$. If $\mathrm{Fix}(T)$ is one-dimensional, then the peripheral point spectrum of T is a subgroup of $\mathbb{T}$, and each eigenvalue is simple.*

Proof. Let $\lambda \in \mathbb{T}$ be an eigenvalue of T, and let $f \in \mathrm{C}(K)$ be a corresponding eigenvector with $\|f\|_\infty = 1$. Then

$$|f| = 1\,|f| = |\lambda f| = |Tf| = T\,|f|\,.$$

Since $\mathrm{Fix}(T)$ is one-dimensional it follows that $|f| = \mathbf{1}$, i.e., f is *unimodular*. Now let λ, μ be peripheral eigenvalues of T, with normalised eigenvectors f, g. Then f, g are both unimodular and with $h := f\overline{g}$ we have

$$Th = T(f\overline{g}) = Tf\;\overline{Tg} = \lambda f\;\overline{\mu g} = \lambda\overline{\mu}h.$$

Hence h is an unimodular eigenvector for $\lambda\overline{\mu}$. This shows that the peripheral point spectrum of T is a subgroup of $\mathbb{T}$. If we let $\lambda = \mu$ in the above argument, we see that h is a unimodular fixed vector of T, that is, $h = \mathbf{1}$. This shows that every eigenspace corresponding to a peripheral eigenvalue is one-dimensional.

Exercises

1. Let $\mathscr{A}$ be a unital Banach algebra with unit element e. Show that if $q \in \mathscr{A}$ such that $\|q\| < 1$, then $\mathrm{e} - q$ is invertible and its inverse is given by the Neumann series

$$(\mathrm{e} - q)^{-1} = \sum_{n=0}^{\infty} q^n.$$

Conclude that if $I \subset \mathscr{A}$ is an ideal with $\mathrm{e} \notin I$, then $\mathrm{e} \notin \overline{I}$. Conclude that each maximal ideal is closed. Conclude also that if $|\lambda| > \|x\|$ then $\lambda\mathrm{e} - x$ is invertible.

2. Let $\mathscr{A}$ be a commutative unital complex Banach algebra and let $\psi : \mathscr{A} \longrightarrow \mathbb{C}$ be a unital algebra homomorphism. Show that $\ker\psi$ is a maximal ideal and ψ is continuous satisfying $|\psi(x)| \leq \|x\|$ for all $x \in \mathscr{A}$. (Hint: Use Exercise 4.1.)

3. Let $\mathscr{A}$ be a commutative unital complex Banach algebra. Show that the Gelfand space $\Gamma(\mathscr{A})$ is closed in $\mathscr{A}'$ with respect to the weak* topology.

4. Prove Lemma 4.11.

5. Let $K := [0,1]$ and $\varphi(x) := x^2$, $x \in K$. Show that the system $(K;\varphi)$ is not topologically transitive, but the fixed space of the induced operator is one-dimensional. Determine the peripheral point spectrum of T_φ.

6. Let $(K;\varphi)$ be a TDS with induced operator $T = T_\varphi$. For $m \in \mathbb{N}$ let $P_m := \{x \in K : \varphi^m(x) = x\}$. Let $\lambda \in \mathbb{T}$ be a peripheral eigenvalue of T, with eigenfunction f. Show that either $\lambda^m = 1$ or f vanishes on P_m. Determine the peripheral point spectrum of the induced operator for the one-sided shift $(\mathscr{W}_k^+;\tau)$, $k \geq 1$.

$*$ **7.** (This is an exercise for the functional analysis specialists.) Show that C(K) is separable if and only if K is metrisable.

Lecture 5
Measure-Preserving Systems

Haste still pays haste, and leisure answers leisure;
Like doth quit like, and measure still for measure.

William Shakespeare[1]
Measure for Measure, Act V, Scene I

In the previous lectures we looked at TDSs $(K;\varphi)$ and their properties, but now turn to dynamical systems that preserve some probability measure on the state space. We shall first motivate this slight change of perspective.

As explained in Lecture 1, "ergodic theory" as a mathematical discipline has its roots in the development of statistical mechanics, in particular in the attempts of Boltzmann, Maxwell and others to derive the second law of thermodynamics from mechanical principles. Central to this theory is the concept of (thermodynamical) *equilibrium*. In topological dynamics, an equilibrium state is a state of rest of the dynamical system itself. However, this is different in the case of a thermodynamical equilibrium, which is an emergent (=macro) phenomenon, while on the microscopic level the molecules show plenty of activity. What is at rest here is rather of a statistical nature.

To clarify this, consider the example from Lecture 1, an ideal gas in a box. What could it mean that it is in equilibrium? Now, since the internal (micro-)states of this system are so manifold and the time scale of their changes is so much smaller than the time scale of our observations, a measurement on the gas has the character of a *random experiment*: the outcome appears to be random, although the underlying dynamics is deterministic. To wait a moment with the next experiment is like shuffling a deck again before drawing the next card; and the "equilibrium" hypothesis — still intuitive — just means that the experiment can be repeated at any time under the same "statistical conditions", i.e., the distribution of an observable $f : \Omega \longrightarrow \mathbb{R}$ (modelling our experiment and now viewed as a random variable) does not change with time.

[1] The Works of William Shakespeare, Volume I., Jacob Tonson, M DCC XXV.
http://www.ub.uni-bielefeld.de/diglib/shakespeare/works/six/

If $\varphi : \Omega \longrightarrow \Omega$ describes the change of the system in one unit of time and if μ denotes the (assumed) probability measure on Ω, then time-invariance of the distribution of f simply means

$$\mu[f > \alpha] = \mu[f \circ \varphi > \alpha] \qquad (\alpha \in \mathbb{R}).$$

Having this for a sufficiently rich class of observables f is equivalent to

$$\mu(\varphi^{-1}A) = \mu(A)$$

for all $A \subset \Omega$ in the underlying σ-algebra. Thus we see how the "equilibrium hypothesis" translates into the existence of a probability measure μ which is *invariant* under the dynamics φ.

We now leave physics and intuitive reasoning and return to "solid" mathematics. The reader is assumed to have some background in abstract measure and integration theory, but some definitions and results are collected in Appendix B.

A **measurable space** is a pair (Ω, Σ), where Ω is a set and Σ is a σ-algebra of subsets of Ω. Given a measurable space (Ω, Σ), a mapping $\varphi : \Omega \longrightarrow \Omega$ is called measurable if

$$[\,\varphi \in A\,] := \varphi^{-1}(A) = \big\{x \in \Omega \,:\, \varphi(x) \in A\big\} \in \Sigma \qquad (A \in \Sigma).$$

Given a measure μ on Σ, its **push-forward measure** (or: image measure) $\varphi_*\mu$ is defined by

$$(\varphi_*\mu)(A) := \mu[\,\varphi \in A\,] \qquad (A \in \Sigma).$$

Definition 5.1. Let (Ω, Σ, μ) be a measure space. A measurable mapping $\varphi : \Omega \longrightarrow \Omega$ is called **measure-preserving** if $\varphi_*\mu = \mu$, i.e.,

$$\mu[\,\varphi \in A\,] = \mu(A) \qquad (A \in \Sigma).$$

In this case μ is called an **invariant measure** for φ (φ-invariant, invariant under φ).

Standard arguments from measure theory show that for the φ-invariance of a measure μ it is sufficient to have $\mu[\,\varphi \in A\,] = \mu(A)$ for all A belonging to a *generator* $\mathcal{E}$ of the σ-algebra Σ. Furthermore, it is also equivalent to

$$\int_\Omega (f \circ \varphi)\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu$$

for all $f \in \mathfrak{M}_+(\Omega)$ (the positive measurable functions on Ω).

We can now introduce our main objects of interest.

Definition 5.2. A quadruple $(\Omega, \Sigma, \mu; \varphi)$ is called a **measure-preserving dynamical system** (MDS) if (Ω, Σ, μ) is a probability space, $\varphi : \Omega \longrightarrow \Omega$ is measurable and μ is φ-invariant.

We reserve the notion of MDS for probability spaces. However, some results remain true for general or at least for σ-finite measure spaces.

5.1 Examples

1. The Baker's Transformation

On $\Omega = [0,1] \times [0,1]$ we define the map

$$\varphi : \Omega \longrightarrow \Omega, \quad \varphi(x,y) := \begin{cases} (2x, y/2), & 0 \le x < 1/2; \\ (2x-1, (y+1)/2), & 1/2 \le x \le 1. \end{cases}$$

Lebesgue measure is invariant under this transformation since

$$\begin{aligned} &\int_{[0,1]^2} f(\varphi(x,y))\,\mathrm{d}(x,y) \\ &\quad = \int_0^1 \int_0^{1/2} f(2x, y/2)\,\mathrm{d}x\,\mathrm{d}y + \int_0^1 \int_{1/2}^1 f(2x-1, (y+1)/2)\,\mathrm{d}x\,\mathrm{d}y \\ &\quad = \int_0^{1/2} \int_0^1 f(x,y)\,\mathrm{d}x\,\mathrm{d}y + \int_{1/2}^1 \int_0^1 f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &\quad = \int_0^1 \int_0^1 f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{[0,1]^2} f(x,y)\,\mathrm{d}(x,y) \end{aligned}$$

for every positive measurable function on $[0,1]^2$. The mapping φ is called the **baker's transformation**, a name explained by Figure 5.1.

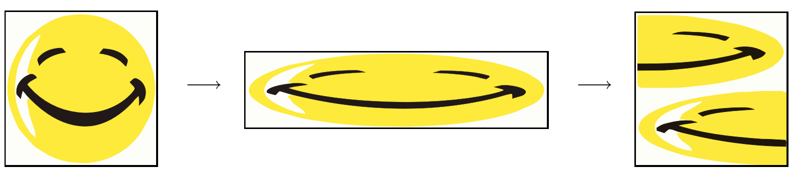

Fig. 5.1 The baker's transformation deforms a square like a baker does with puff pastry when kneading.

2. The Doubling Map

Let $\Omega = [0,1]$ and consider the doubling map $\varphi : \Omega \longrightarrow \Omega$ defined as

$$\varphi(x) := 2x \pmod 1 = \begin{cases} 2x, & 0 \le x < 1/2; \\ 2x-1, & 1/2 \le x \le 1 \end{cases}$$

(cf. Exercise 2.4). Lebesgue measure is invariant under φ since

$$\int_0^1 f(\varphi(x))\,\mathrm{d}x = \int_0^{1/2} f(2x)\,\mathrm{d}x + \int_{1/2}^1 f(2x-1)\,\mathrm{d}x$$
$$= \frac{1}{2}\int_0^1 f(x)\,\mathrm{d}x + \frac{1}{2}\int_0^1 f(x)\,\mathrm{d}x = \int_0^1 f(x)\,\mathrm{d}x$$

for every positive measurable function f on $[0,1]$.

3. The Tent Map

Let $\Omega = [0,1]$ and consider the tent map $\varphi : \Omega \longrightarrow \Omega$ given by

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

(cf. Exercise 2.5). Lebesgue measure is invariant under φ, see Exercise 5.2.

4. The Gauss Map

Consider $\Omega = [0,1)$ and define the **Gauss map** $\varphi : \Omega \longrightarrow \Omega$ by

$$\varphi(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor \qquad (0 < x < 1), \quad \varphi(0) := 0.$$

It is easy to see that

$$\varphi(x) = \frac{1}{x} - n \quad \text{if} \quad x \in \left(\frac{1}{n+1}, \frac{1}{n}\right],\ n \in \mathbb{N},$$

and

$$\varphi^{-1}\{y\} = \left\{\frac{1}{y+n} : n \in \mathbb{N}\right\}$$

for every $y \in [0,1)$. Exercise 5.3 shows that the measure

$$\mu := \frac{\mathrm{d}x}{x+1}$$

on $[0,1)$ is invariant for φ. The Gauss map links ergodic theory with number theory via *continued fractions*, see [Silva (2008), p. 154] and [Baker (1984), pp. 44–46].

5. Bernoulli Shifts

Fix $k \in \mathbb{N}$, consider the finite space $L := \{0, \ldots, k-1\}$ and form the product space

$$\Omega := \prod_{n \in \mathbb{N}_0} L = \{0, \ldots, k-1\}^{\mathbb{N}_0} = \mathscr{W}_k^+$$

(cf. Example 2.5) with the product σ-algebra $\Sigma := \bigotimes_{n \geq 0} \mathscr{P}(L)$. On (Ω, Σ) we consider the left shift τ. To see that τ is a measurable mapping, note that the *cylinder sets*, i.e., sets of the form

$$A := A_0 \times A_1 \times \cdots \times A_{n-1} \times L \times \ldots \tag{5.1}$$

with $n \in \mathbb{N}$, $A_0, \ldots, A_{n-1} \subset L$, generate Σ. Then

$$[\tau \in A] = L \times A_0 \times A_1 \times \cdots \times A_{n-1} \times L \times \ldots$$

is again contained in Σ.

There are many shift invariant probability measures on $\Omega = \mathscr{W}_k^+$, and we just construct one. (For another one see the next paragraph.) Fix a probability vector $p = (p_0, \ldots, p_{k-1})$ and consider the associated measure $\nu = \sum_{j=0}^{k-1} p_j \delta_{\{j\}}$ on L. Then let $\mu := \bigotimes_{n \geq 0} \nu$ be the infinite product measure on Σ, defined on cylinder sets A as in (5.1) by

$$\mu(A) = \nu(A_0)\nu(A_1) \cdots \nu(A_{n-1}).$$

(It is a standard result in measure theory that there indeed exists a unique measure μ on Σ satisfying this requirement. Its construction rests on Lemma B.6 and Theorem B.5.) The product measure μ is shift-invariant, because for cylinder sets A as in (5.1) we have

$$\mu[\tau \in A] = \nu(L)\nu(A_0) \ldots \nu(A_{n-1}) = \nu(A_0) \ldots \nu(A_{n-1}) = \mu(A),$$

since $\nu(L) = 1$. This MDS $(\Omega, \Sigma, \mu; \tau)$ is called the **Bernoulli shift** $B(p_0, \ldots, p_{k-1})$.

The Bernoulli shift is a special case of a more general construction. Namely, fix a probability space (Ω', Σ', ν) and consider the infinite product $\Omega := \prod_{n \geq 0} \Omega'$ with the product measure $\mu := \bigotimes_{n \geq 0} \nu$ on the product σ-algebra $\Sigma := \bigotimes_{n \geq 0} \Sigma'$. Then the shift τ is measurable and μ is τ-invariant.

Obviously, all this can also be done with two-sided shifts.

6. Markov Shifts

We consider a slight generalisation of Example 5. Let $L = \{0, \ldots, k-1\}$, $\Omega = L^{\mathbb{N}_0}$ and Σ as in Example 5. Let $P := (p_{ij})$ be a row stochastic $k \times k$-matrix, i.e.,

$$p_{ij} \geq 0 \quad (0 \leq i, j \leq k-1), \qquad \sum_{j=0}^{k-1} p_{ij} = 1 \quad (0 \leq i \leq k-1).$$

For every probability vector $p := (p_0, \dots, p_{k-1})$ we construct the **Markov measure** μ on $\mathscr{W}_k^+$ requiring that on the special cylinder sets

$$A := \{j_0\} \times \{j_1\} \cdots \times \{j_n\} \times \prod_{m>n} L \qquad (n \geq 1, j_0, \dots, j_n \in L)$$

one has

$$\mu(A) := p_{j_0} p_{j_0 j_1} \dots p_{j_{n-1} j_n}. \tag{5.2}$$

It is again standard measure theory based on Theorem B.5 to show that there exists a unique measure μ on (Ω, Σ) satisfying this requirement. Now with A as above one has

$$\mu[\tau \in A] = \mu\left(L \times \{j_0\} \times \{j_1\} \cdots \times \{j_n\} \times \prod_{m>n} L\right) = \sum_{j=0}^{k-1} p_j p_{j j_0} \dots p_{j_{n-1} j_n}.$$

Hence the measure μ is invariant under the left shift if and only if

$$p_{j_0} p_{j_0 j_1} \dots p_{j_{n-1} j_n} = \sum_{j=0}^{k-1} p_j p_{j j_0} \dots p_{j_{n-1} j_n}$$

for all choices of parameters $n \geq 0, 0 \leq j_1, \dots j_n < k$. This is true if and only if it holds for $n = 0$ (sum over the other indices!), i.e.,

$$p_l = \textstyle\sum_{j=0}^{k-1} p_j p_{jl};$$

and this means that $pP = p$, i.e., p is a left fixed vector of P. Such a left fixed probability vector indeed exists, by Perron's theorem. We shall prove this in a later lecture when we treat the so-called *mean ergodic theorem*.

If P is a row-stochastic $k \times k$-matrix, p is a left fixed probability vector for P and $\mu = \mu(P, p)$ is the unique probability measure on $\mathscr{W}_k^+$ with (5.2), then the MDS $(\mathscr{W}_k^+, \Sigma, \mu(P, p); \tau)$ is called the **Markov shift** associated with (P, p). If P is the matrix whose rows are all equal to p, then μ is just the product measure and the Markov system is the same as the Bernoulli system $B(p_0, \dots, p_{k-1})$.

As in the case of Bernoulli shifts, Markov shifts can be generalised to arbitrary probability spaces. One needs the notion of a probability kernel and the Ionescu–Tulcea theorem (Theorem B.15).

5.2 Measures on Compact Spaces

This section connects our discussion about topological dynamics with invariant measures. Unfortunately — since this is not a course on measure theory — we cannot provide all details here. Interested readers may consult [Bogachev (2007)] or [Bauer (1981)].

Let K be a compact space. There are two natural σ-algebras on K: (1) the **Borel algebra** $\mathfrak{B}(K)$, generated by the family of open set, and (2) the **Baire algebra** $\mathfrak{B}_0(K)$, which is the smallest σ-algebra that makes all continuous functions measurable. Equivalently,

$$\mathfrak{B}_0(K) = \sigma\big\{[f > 0] \,:\, 0 \le f \in \mathrm{C}(K)\big\}.$$

Clearly $\mathfrak{B}_0(K) \subset \mathfrak{B}(K)$, and by the proof of Lemma 4.9 the open Baire sets form a base for the topology of K. Exercise 5.6 shows that $\mathfrak{B}_0(K)$ is generated by the compact G_δ-subsets of K. If K is metrisable, then $\mathfrak{B}_0(K)$ and $\mathfrak{B}(K)$ coincide, but in general this is false.[An example is in [Bogachev (2007), II, 7.1.3].] (complex) measure μ on $\mathfrak{B}_0(K)$ (resp. $\mathfrak{B}(K)$) is called a **Baire measure** (resp. **Borel measure**). Every finite positive Baire measure is **regular** in the sense that for every $B \in \mathfrak{B}_0(K)$ one has

$$\begin{aligned}\mu(B) &= \sup\big\{\mu(A) \,:\, A \in \mathfrak{B}_0(K), A \text{ compact}, A \subset B\big\}\\ &= \inf\big\{\mu(O) \,:\, O \in \mathfrak{B}_0(K), O \text{ open}, B \subset O\big\}.\end{aligned}$$

(This is a standard argument involving Dynkin systems (Theorem B.2); for a different proof see [Bogachev (2007), II, 7.1.8].) The regularity of μ combined with Urysohn's lemma shows that $\mathrm{C}(K)$ is dense in $\mathrm{L}^p(K, \mathfrak{B}_0(K), \mu)$, $1 \le p < \infty$.

Let us denote the complex Baire measures on K by $\mathrm{M}(K)$. It is a Banach space with respect to the total variation norm $\|\mu\|_{\mathrm{M}} := |\mu|(K)$ (see Appendix B.12). Every $\mu \in \mathrm{M}(K)$ defines a linear functional on $\mathrm{C}(K)$ via

$$\langle f, \mu \rangle := \int_K f \, \mathrm{d}\mu \qquad (f \in \mathrm{C}(K)).$$

Since $|\langle f, \mu\rangle| \le \int_K |f| \, \mathrm{d}|\mu| \le \|f\|_\infty \|\mu\|_{\mathrm{M}}$, we have $\langle \cdot, \mu\rangle \in \mathrm{C}(K)'$, the Banach space dual of $\mathrm{C}(K)$. The following important theorem states the converse.

Theorem 5.3 (Riesz' Representation Theorem). *Let K be a compact space. Then the mapping*

$$\mathrm{M}(K) \longrightarrow \mathrm{C}(K)', \qquad \mu \longmapsto \langle \cdot, \mu \rangle$$

is an isometric isomorphism. The finite Baire measure μ is positive if and only if $\langle f, \mu \rangle \ge 0$ for all $f \in \mathrm{C}(K)$ with $f \ge 0$.

We do not give the proof here, see for example [Rudin (1987), 2.14] or [Lang (1993), IX.2]. Justified by the Riesz theorem, we shall identify $\mathrm{M}(K) = \mathrm{C}(K)'$ and often write μ in place of $\langle \cdot, \mu \rangle$. The Riesz representation theorem directly implies the following very useful statement.

Lemma 5.4. *Let K be a compact space and let $\varphi : K \longrightarrow K$ be Baire measurable. Then a finite positive Baire measure μ is φ-invariant if and only if*

$$\int_K (f \circ \varphi) \, \mathrm{d}\mu = \int_K f \, \mathrm{d}\mu$$

An example is in [Bogachev (2007), II, 7.1.3]. A *of all* $f \in \mathrm{C}(K)$.

Remark 5.5. Let us remark that the usual form of Riesz' theorem — for instance in [Rudin (1987)] — involves *regular Borel measures* instead of Baire measures. This stronger form implies in particular that each finite positive Baire measure has a unique extension to a regular Borel measure. It is sometimes convenient to use this extension, and we shall do it without further reference. However, it is advantageous to work with Baire measures in general since regularity is automatic and the Baire algebra goes nicely with infinite products (see Exercise 5.6). From a functional analytic point of view there is anyway no difference between a Baire measure and its regular Borel extension, since the associated L^p-spaces coincide.

Given a TDS $(K;\varphi)$, the question is now natural to ask whether there exist Baire probability measures on K which are invariant under φ. The following classical theorem says that the answer is indeed yes.

Theorem 5.6 (Krylov–Bogoljubov). *Let* $(K;\varphi)$ *be a TDS. Then there is at least one* φ*-invariant Baire probability measure on* K.

Note that under the identification $\mathrm{M}(K) = \mathrm{C}(K)'$ from above, the φ-invariance of μ just means $T'_\varphi(\mu) = \mu$, i.e., μ is a fixed point of the adjoint of the induced operator. One possible proof of Theorem 5.6 builts on the Markov–Kakutani fixed point theorem [Rudin (1991), 5.23], and we shall probably see this in detail in connection with the *mean ergodic theorem*. As a consequence of Theorem 5.6 we state that — by choosing some invariant measure μ — every TDS $(K;\varphi)$ gives rise to an MDS $(K, \mathfrak{B}_0(K), \mu; \varphi)$.

5.3 Haar Measure and Rotations

Let G be a compact topological group. For $a \in G$ consider the left rotation by a, defined as $\varphi_a(g) = ag$. A nontrivial finite Baire measure on G that is invariant under all rotations $\{\varphi_a : a \in G\}$ is called a **Haar measure**, and it can be shown that such a Haar measure is essentially unique, i.e., for any two Haar measures μ, ν on G there is a positive constant α such that $\mu = \alpha\nu$. Moreover, if μ is a Haar measure and $\emptyset \neq U \subset G$ is open, then $\mu(U) > 0$. Finally — even when the group is not commutative — a Haar measure is automatically *right-invariant*, i.e., invariant under right rotations, and *inversion-invariant*, i.e., invariant under the inversion mapping $g \longmapsto g^{-1}$ of the group. (See [Hewitt and Ross (1979), Thm. 15.13] for proofs of these statements.)

It is a fundamental result of topological group theory that there exists a Haar measure on each compact group. (For a functional-analytic proof see [Rudin (1991), Thm. 5.14], cf. also [Hewitt and Ross (1979), IV.15] or [Lang (1993), XII.3].) However, in many concrete cases one does not need to rely on this abstract result because the Haar measure can be described explicitly. For example, the Haar measure $\mathrm{d}z$ on $\mathbb{T}$ is given by

$$\int_{\mathbb{T}} f(z)\,\mathrm{d}z := \int_0^1 f(e^{2\pi i t})\,\mathrm{d}t$$

for measurable $f : \mathbb{T} \longrightarrow [0,\infty]$. The Haar measure on a finite discrete group is the counting measure. In particular, the Haar probability measure on the cyclic group $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z} = \{0,1,\ldots,n-1\}$ is given by

$$\int_{\mathbb{Z}_n} f(g)\,\mathrm{d}g = \frac{1}{n}\sum_{j=0}^{n-1} f(j)$$

for every $f : \mathbb{Z}_n \longrightarrow \mathbb{C}$. For the Haar measure of the dyadic integers see Exercise 5.4.

If G,H are compact groups with Haar measures μ,ν, respectively, then the product measure $\mu \otimes \nu$ is a Haar measure on the product group $G \times H$. The same holds for infinite products: Let $(G_n)_{n\in\mathbb{N}}$ be a family of compact groups, and let μ_n denote the unique Haar probability measure on G_n, $n \in \mathbb{N}$. Then the product measure $\bigotimes_n \mu_n$ is the unique Haar probability measure on the product group $\prod_n G_n$.

Exercises

1.[Boole transformation] Show that the Lebesgue measure on $\mathbb{R}$ is invariant under the map $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$, $\varphi(x) = x - 1/x$. This map is called the *Boole transformation*. Define the modified Boole transformation $\psi : \mathbb{R} \longrightarrow \mathbb{R}$

$$\psi(x) := \frac{1}{2}\varphi(x) = \frac{1}{2}\left(x - \frac{1}{x}\right) \qquad (x \neq 0).$$

Show that the (finite!) measure $\mu := \mathrm{d}x/(1+x^2)$ is invariant under ψ.

2. Show that the Lebesgue measure is invariant under the tent map (Example 3). Consider the continuous mapping $\varphi : [0,1] \longrightarrow [0,1]$, $\varphi(x) := 4x(1-x)$. Show that the (finite!) measure $\mu := \mathrm{d}x/2\sqrt{x(1-x)}$ on $[0,1]$ is invariant under φ.

3. Consider the Gauss transformation

$$\varphi(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor \qquad (0 < x < 1), \quad \varphi(0) := 0,$$

on $[0,1)$ and show that the measure $\mu := \mathrm{d}x/(x+1)$ is invariant for φ.

4. Consider the dyadic adding machine $(\mathbb{A};\varphi)$, where $\mathbb{A} := \{0,1\}^{\mathbb{N}}$ is the compact group of dyadic integers and $\varphi = (x \longmapsto x + \mathbf{1})$ is addition with $\mathbf{1} = (1,0,\ldots)$ as in Exercise 3.4. Show that the product measure

$$\mu = \bigotimes_{n\in\mathbb{N}} \frac{1}{2}(\delta_0 + \delta_1)$$

is invariant under φ and conclude that μ is the Haar measure on the compact group $\mathbb{A}$. (Hint: For $x \in \mathbb{A}$ denote by $[x]_n \in \{0,1\}^n$ the projection onto the first n coordinates; prove by induction on $n \in \mathbb{N}$ that $\mu\{x : [x+\mathbf{1}]_n = (j_1,\dots,j_n)\} = (1/2)^n$ for every $(j_1,\dots,j_n) \in \{0,1\}^n$.)

5. Let G be a compact *abelian* group, and fix $k \in \mathbb{N}$. Consider the map $\varphi_k : G \longrightarrow G$, $\varphi_k(g) = g^k$, $g \in G$. (This is a continuous group homomorphism, since G is abelian.) Show that if φ_k is surjective then the Haar measure is invariant under φ_k. (Hint: one can use the uniqueness of the Haar measure.)

$*$ **6.** Let K be a compact space with its Baire algebra $\mathfrak{B}_0(K)$.

a) Show that $\mathfrak{B}_0(K)$ is generated by the compact G_δ-sets.

b) Show that $\mathfrak{B}_0(K) = \mathfrak{B}(K)$ if K is metrisable.

c) Show that if $A \subset O$ with $A \subset K$ closed and $O \subset K$ open, there exists $A \subset A' \subset O' \subset O$ with $A', O' \in \mathfrak{B}_0(K)$, A' closed and O' open.

d) Let $(K_n)_{n\in\mathbb{N}}$ be a sequence of nonempty compact spaces and let $K := \prod_n K_n$ be their product space. Show that $\mathfrak{B}_0(K) = \bigotimes_n \mathfrak{B}_0(K_n)$.

Lecture 6
Recurrence and Ergodicity

> Siehe, wir wissen, was Du lehrst: dass alle Dinge ewig wiederkehren und wir selber mit, und dass wir schon ewige Male dagewesen sind, und alle Dinge mit uns.[1]
>
> Friedrich Nietzsche
> Also sprach Zarathustra[2]

Having introduced the notion of measure-preserving systems (MDS) and having seen many examples of such systems we now turn to their systematic study. In particular we shall define *invertibility* of an MDS, prove the classical *recurrence theorem* of Poincaré and introduce the central notion of an *ergodic system*.

6.1 The Measure Algebra and Invertible MDS

In a measure space (Ω, Σ, μ) we consider null-sets to be *negligible*. That means that we identify sets $A, B \in \Sigma$ that are (μ-)**essentially equal**, i.e., which satisfy $\mu(A \triangle B) = 0$ (here $A \triangle B = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference of A, B). To be more precise, we define the relation $\sim$ on Σ by

$$A \sim B \quad \stackrel{\text{Def.}}{\iff} \quad \mu(A \triangle B) = 0 \quad \iff \quad \mathbf{1}_A = \mathbf{1}_B \ \mu\text{-almost everywhere.}$$

Then $\sim$ is an equivalence relation on Σ and the corresponding set $\Sigma/{\sim}$ of equivalence classes is called the corresponding **measure algebra**. For a set $A \in \Sigma$ let us temporarily write $[A]$ for its equivalence class. It is an easy but tedious exercise to show that the set-theoretic relations and (countable) operations on Σ induce corresponding operations on the measure algebra via

[1] Behold, we know what thou teachest: that all things eternally return, and ourselves with them, and that we have already existed times without number, and all things with us.
[2] Teil III, Der Genesende. From: Werke II, hrsg. v. Karl Schlechta, Darmstadt 1997, S. 466. Translation from: http://www.geocities.com/thenietzschechannel/zarapt3.htm#con

$$[A] \subset [B] \overset{\text{Def.}}{\iff} A \subset B, \quad [A] \cap [B] := [A \cap B], \quad [A] \cup [B] := [A \cup B], \quad \ldots .$$

Moreover, the measure μ induces a (σ-additive) map

$$\Sigma/\!\sim \longrightarrow [0,\infty], \qquad [A] \longmapsto \mu(A).$$

As for equivalence classes of functions, we normally do not distinguish notationally between a set A and its equivalence class $[A]$ in $\Sigma/\!\sim$.

If the measure is finite, the measure algebra can be turned into a complete metric space with metric given by

$$d_\mu(A,B) := \mu(A \triangle B) = \mu((A \setminus B) \cup (B \setminus A)) = \|\mathbf{1}_A - \mathbf{1}_B\|_1 \qquad (A, B \in \Sigma/\!\sim).$$

The set-theoretic operations as well as the measure μ itself are continuous with respect to this metric. (See also Exercise 6.1 and Appendix B). Frequently, the σ-algebra Σ is *generated* by an algebra $\mathcal{E}$, i.e., $\Sigma = \sigma(\mathcal{E})$ (cf. Appendix B.1). This property has an important topological implication.

Lemma 6.1 (Approximation). *Let (Ω, Σ, μ) be a finite measure space and let $\mathcal{E} \subset \Sigma$ be an algebra of subsets such that $\sigma(\mathcal{E}) = \Sigma$. Then $\mathcal{E}$ is dense in the measure algebra, i.e., for every $A \in \Sigma$ and $\varepsilon > 0$ there is $E \in \mathcal{E}$ such that $d_\mu(A,E) < \varepsilon$.*

Proof. This is just Lemma B.22 from Appendix B. Its proof is standard measure theory using Dynkin systems. □

Suppose that $\varphi : \Omega \longrightarrow \Omega$ is a measurable mapping and consider the push-forward measure $\varphi_*\mu$. If, for every $A \in \Sigma$, $\mu(A) = 0$ implies $\mu(\varphi^{-1}A) = (\varphi_*\mu)(A) = 0$ (i.e., if $\varphi_*\mu$ is *absolutely continuous* with respect to μ), the mapping φ induces a mapping φ^* on the measure algebra defined by

$$\varphi^* : \Sigma/\!\sim \longrightarrow \Sigma/\!\sim, \quad [A] \longmapsto \left[\varphi^{-1}A\right].$$

Note that φ^* commutes with all the set-theoretic operations, i.e.,

$$\varphi^*(A \cap B) = \varphi^*A \cap \varphi^*B, \quad \varphi^* \bigcup_n A_n = \bigcup_n \varphi^* A_n \quad \ldots .$$

All this is applicable to the situation when μ is φ-invariant. In this case and when the measure is finite, the induced mapping φ^* on $\Sigma/\!\sim$ is an isometry, since

$$d_\mu(\varphi^*A, \varphi^*B) = \mu([\varphi \in A] \triangle [\varphi \in B]) = \mu[\varphi \in (A \triangle B)] = \mu(A \triangle B) = d_\mu(A,B)$$

for $A, B \in \Sigma$. We are now able to define invertibility of an MDS.

Definition 6.2. An MDS $(\Omega, \Sigma, \mu; \varphi)$ is called **invertible** if the induced map φ^* on $\Sigma/\!\sim$ is bijective.

Exercise 6.2 gives some sufficient criteria for invertibility.

Examples 6.3. The baker's transformation is invertible (use Exercise 6.2). Every group rotation is invertible. The tent map and the doubling map are not invertible (Exercise 6.3). A two-sided Bernoulli system is invertible but a one-sided is not (Exercise 6.4).

6.2 Recurrence

Recall that a point x in a TDS is recurrent if it returns eventually to each of its neighbourhoods. In the measure-theoretic setting pointwise notions are meaningless, due to the presence of null-sets. So, given an MDS $(\Omega, \Sigma, \mu; \varphi)$ in place of points of Ω we have to use sets of positive measure, i.e., "points" of the measure algebra $\Sigma/\!\sim$. We adopt that view with the following definition.

Definition 6.4. Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS, and let $A \in \Sigma$.

a) A set $A \in \Sigma$ is called **recurrent** if almost every point of A returns to A after some time. Equivalently,
$$A \subset \bigcup_{n\geq 1} \varphi^{*n} A \tag{6.1}$$
in the measure algebra.

b) A set $A \in \Sigma$ is **infinitely recurrent** if almost every point of A returns to A infinitely often. Equivalently,
$$A \subset \bigcap_{n\geq 1} \bigcup_{k\geq n} \varphi^{*k} A \tag{6.2}$$
in the measure algebra.

Here is an important characterisation.

Lemma 6.5. *Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS. Then the following statements are equivalent.*

(i) *Every $A \in \Sigma$ is recurrent.*

(ii) *Every $A \in \Sigma$ is infinitely recurrent.*

(iii) *For every $\emptyset \neq A \in \Sigma/\!\sim$ there exists $n \in \mathbb{N}$ such that $A \cap \varphi^{*n} A \neq \emptyset$.*

Proof. (i)$\Longleftrightarrow$(ii): Take $A \in \Sigma$ and apply φ^* to (6.1) to obtain $\varphi^* A \subset \bigcup_{n\geq 2} \varphi^{*n} A$. Putting this back into (6.1) we obtain
$$A \subset \bigcup\nolimits_{n\geq 1} \varphi^{*n} A = \varphi^* A \cup \bigcup\nolimits_{n\geq 2} \varphi^{*n} A \subset \bigcup\nolimits_{n\geq 2} \varphi^{*n} A.$$

Continuing in this manner, we see that $\bigcup_{n\geq k} \varphi^{*n} A$ is independent of $k \geq 1$ and this leads to (ii). The implication (ii)$\Longrightarrow$(i) is clear.
(i)$\Longrightarrow$(iii): To obtain a contradiction suppose that $A \cap \varphi^{*n} A = \emptyset$ for all $n \geq 1$. Then intersecting with A in (6.1) yields

$$A = A \cap \bigcup_{n\geq 1} \varphi^{*n} A = \bigcup_{n\geq 1} A \cap \varphi^{*n} A = \emptyset.$$

(iii)$\Longrightarrow$(i): Let $A \in \Sigma$ and consider the set $B := \left(\bigcap_{n\geq 1} \varphi^{*n} A^{\mathrm{c}}\right) \cap A$. Then, for every $n \in \mathbb{N}$

$$B \cap \varphi^{*n} B \subset \varphi^{*n} A^{\mathrm{c}} \cap \varphi^{*n} A = \emptyset$$

in the measure algebra. Now (iii) implies that $B = \emptyset$, i.e., (6.1). □

As a consequence of this lemma we obtain the famous recurrence theorem of H. Poincaré.

Theorem 6.6 (Poincaré). *Every MDS $(\Omega, \Sigma, \mu; \varphi)$ is (infinitely) recurrent, i.e., every set $A \in \Sigma$ is infinitely recurrrent.*

Proof. Let $A \in \Sigma/\sim$ be such that $A \cap \varphi^{*n} A = \emptyset$ for all $n \geq 1$. Thus for $n > m \geq 0$ we have

$$\varphi^{*m} A \cap \varphi^{*n} A = \varphi^{*m}(A \cap \varphi^{*(n-m)} A) = \emptyset.$$

This means that the sets $(\varphi^{*n} A)_{n \in \mathbb{N}_0}$ are (essentially) disjoint. (A set A with this property is called *wandering*.) On the other hand, all sets $\varphi^{*n} A$ have the same measure, and since μ is finite, this measure must be zero. Hence $A = \emptyset$, and the MDS $(\Omega, \Sigma, \mu; \varphi)$ is infinitely recurrent, by the implication "(iii)$\Longrightarrow$(i)" of Lemma 6.5.

□

Remark. Poincaré's theorem is false for measure-preserving transformations on infinite measure spaces. Just consider $\Omega = \mathbb{R}$, the shift $\varphi(x) = x + 1$ $(x \in \mathbb{R})$ and $A = [0, 1]$.

Poincaré's recurrence theorem has caused some irritation among scholars since its results may seem to be counterintuitive. To make this understandable, consider once again our example from Lecture 1, the ideal gas in a container. Suppose that we start observing the system after having collected all gas molecules in the left half of the box (e.g., by introducing a wall first, pumping all the gas to the left and then removing the wall), we expect that the gas diffuses in the whole box and eventually is distributed uniformly within it. It seems unplausible to expect that after some time the gas molecules again return by their own motion entirely to the left half of the box. However, since the initial state (all molecules in the left half) has (quite small but nevertheless) positive probability, the Poincaré theorem states that this will happen, see Figure 6.1.

Let us consider a more mathematical example which goes back to the Ehrenfests [Ehrenfest (1912)], cited after [Petersen (1989), p.35]. Suppose that we have n balls, numbered from 1 to n, distributed somehow over two urns. In each step, we determine randomly a number k between 1 and n, take the ball with number k out of the urn where it is at that moment, and put it into the other urn. Initially we start with all n balls contained in box 1.

The mathematical model of this experiment is that of a Markov shift over $L := \{0, \ldots, n\}$. The "state sequence" $(j_0, j_1, j_2 \ldots)$ records the number $j_m \in L$ of balls in

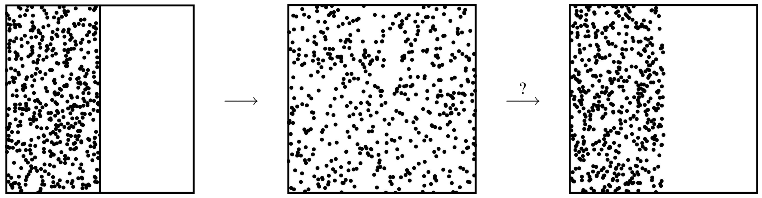

Fig. 6.1 What happens after removing the wall?

urn 1 after the m-th step. To determine the transition matrix P note that if there are $i \geq 1$ balls in urn 1, then the probability to decrease its number to $i-1$ is i/n; and if $i < n$ then the probability to increase the number from i to $i+1$ is $1-i/n$. Hence we have $P = (p_{ij})_{0 \leq i,j \leq n}$ with

$$p_{ij} = \begin{cases} 0 & \text{if} \quad |i-j| \neq 1 \\ \frac{i}{n} & \text{if} \quad j = i-1 \\ \frac{n-i}{n} & \text{if} \quad j = i+1 \end{cases} \qquad (0 \leq i, j \leq n).$$

The a priori probability to have j balls in urn 1 is certainly

$$p_j := 2^{-n} \binom{n}{j}$$

and it is easy to show that $p = (p_0, \ldots, p_n)$ is indeed a fixed probability row vector, i.e., that $pP = p$ (Exercise 6.5). (Since P is irreducible — cf. the proposition after Example 2.22 —, Perron's theorem implies that p is actually the *unique* fixed probability vector.)

Now, starting with all balls contained in urn 1 is exactly the event $A := \{(j_m)_{m \geq 0} \in L^{\mathbb{N}_0} : j_0 = n\}$. Clearly $\mu(A) = 1/2^n > 0$, whence Poincaré's theorem tells us that in almost all sequences $x \in A$ the number n occurs infinitely often.

If the number of balls n is large, this result may look again counterintuitive. However, we shall show that we will have to wait a *very* long time until the system comes back to A the first time. To make this precise, we return to the general setting.

Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS, fix $A \in \Sigma$ and define

$$B_0 := \bigcap\nolimits_{n \geq 1} \varphi^{*n} A^{\mathrm{c}}, \quad B_1 := \varphi^* A, \quad B_n := \varphi^{*n} A \cap \bigcap\nolimits_{k=1}^{n-1} \varphi^{*k} A^{\mathrm{c}} \qquad (n \geq 2).$$

The sequence $(B_n)_{n \in \mathbb{N}}$ is the "disjointification" of $(\varphi^{*n} A)_{n \in \mathbb{N}}$, and $(B_n)_{n \in \mathbb{N}_0}$ is a disjoint decomposition of Ω: the points from B_0 never reach A and the points from B_n reach A for the first time after n steps. Recurrence of A just means that $A \subset \bigcup_{n \geq 1} B_n$, i.e., $B_0 \subset A^{\mathrm{c}}$ (almost everywhere). If we let

$$A_n := B_n \cap A \qquad (n \in \mathbb{N})$$

then $(A_n)_{n\geq 1}$ is an essentially disjoint decomposition of A. We note the following technical lemma for later use.

Lemma 6.7. *Let $(\Omega,\Sigma,\mu;\varphi)$ be an MDS, and let $A,B\in\Sigma$. Then for $n\geq 1$*

$$\mu(B)=\sum_{k=1}^{n}\mu\left(A\cap\bigcap_{j=1}^{k-1}\varphi^{*j}A^c\cap\varphi^{*k}B\right) \quad + \quad \mu\left(\bigcap_{k=0}^{n-1}\varphi^{*k}A^c\cap\varphi^{*n}B\right). \tag{6.3}$$

Proof. Using the φ-invariance of μ we write

$$\mu(B)=\mu(\varphi^*B)=\mu(A\cap\varphi^*B)+\mu(A^{\mathrm{c}}\cap\varphi^*B)$$

and this is (6.3) with $n=1$. Doing the same again with the second summand yields

$$\begin{aligned}\mu(B)&=\mu(\varphi^*B)=\mu(A\cap\varphi^*B)+\mu(A^{\mathrm{c}}\cap\varphi^*B)\\&=\mu(A\cap\varphi^*B)+\mu(\varphi^*A^{\mathrm{c}}\cap\varphi^{*2}B)\\&=\mu(A\cap\varphi^*B)+\mu(A\cap\varphi^*A^{\mathrm{c}}\cap\varphi^{*2}B)+\mu(A^{\mathrm{c}}\cap\varphi^*A^{\mathrm{c}}\cap\varphi^{*2}B)\end{aligned}$$

and this is (6.3) for $n=2$. An induction argument concludes the proof. □

Suppose that $\mu(A)>0$. On the set A we consider the induced σ-algebra

$$\Sigma_A:=\big\{B\subset A : B\in\Sigma\big\}\subset\mathscr{P}(A)$$

and thereon the induced probability measure μ_A defined by

$$\mu_A:\Sigma_A\longrightarrow[0,1],\quad \mu_A(B)=\frac{\mu(B)}{\mu(A)}\quad (B\in\Sigma_A).$$

Then $\mu_A(B)$ is the **conditional probability** for B given A. Define the function

$$n_A:A\longrightarrow\mathbb{N},\quad n_A(x)=n\quad\text{if}\quad x\in A_n\qquad(n\geq 1).$$

The function n_A is called the **time of first return** to A. Clearly, n_A is Σ_A-measurable. The following theorem describes its expected value (with respect to μ_A).

Theorem 6.8. *Let $(\Omega,\Sigma,\mu;\varphi)$ be an MDS, and let $A\in\Sigma$ with $\mu(A)>0$. Then*

$$\int_A n_A\,\mathrm{d}\mu_A=\frac{\mu\left(\bigcup_{n\geq 0}\varphi^{*n}A\right)}{\mu(A)}. \tag{6.4}$$

Proof. We specialise $B=\Omega$ in Lemma 6.7. Note that

$$A\cap\bigcap_{j=1}^{k-1}\varphi^{*j}A^{\mathrm{c}}=\bigcup_{j=k}^{\infty}A_j\qquad(k\geq 1)$$

since A is recurrent, by Poincaré's theorem. Hence by (6.3) in Lemma 6.7 we obtain

$$\mu(\Omega) = \sum_{k=1}^{n} \sum_{j=k}^{\infty} \mu(A_j) + \mu\left(\bigcap_{k=0}^{n-1} \varphi^{*k} A^{\mathrm{c}}\right).$$

Letting $n \to \infty$ yields

$$1 = \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} \mu(A_j) + \mu\left(\bigcap_{n=0}^{\infty} \varphi^{*n} A^{\mathrm{c}}\right),$$

and interchanging the order of summation we arrive at

$$\mu\left(\bigcup_{n=0}^{\infty} \varphi^{*n} A\right) = \mu(\Omega) - \mu\left(\bigcap_{n=0}^{\infty} \varphi^{*n} A^{\mathrm{c}}\right) = \sum_{j=1}^{\infty} j\mu(A_j) = \mu(A) \int_A n_A \,\mathrm{d}\mu_A.$$

Dividing by $\mu(A)$ proves the claim. □

Let us return to our ball experiment. If we start with 100 balls, all contained in urn 1, the probability of A is $\mu(A) = 2^{-100}$. If we knew that

$$\bigcup_{n=0}^{\infty} \tau^{*n} A = \Omega \tag{6.5}$$

then we could conclude from (6.4) in Lemma 6.8 that the expected waiting time for returning to A is $1/\mu(A) = 2^{100}$ steps. Clearly, our lifetimes and that of our children would not suffice for the waiting, even if we do one step every millisecond.

However, this reasoning builds on (6.5), a property that is not yet established. This will be done in the next section.

6.3 Ergodicity

Ergodicity is the analogue of minimality in measurable dynamics. Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS and let $A \in \Sigma$. As in the topological case, we call A **invariant** if $A \subset [\varphi \in A]$. Since A and $[\varphi \in A]$ have the same measure, and μ is finite, $A \sim [\varphi \in A]$, i.e., $A = \varphi^* A$ in the measure algebra. The same reasoning applies if $[\varphi \in A] \subset A$, and hence we have established the following lemma.

Lemma 6.9. *Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS and $A \in \Sigma$. Then the following assertions are equivalent.*

(i) *A is invariant, i.e., $A \subset \varphi^* A$.*

(ii) *A is strictly invariant, i.e., $A = \varphi^* A$.*

(iii) *$\varphi^* A \subset A$.*

(iv) *A^{c} is (strictly) invariant.*

The following notion is analogous to minimality in the topological case.

Definition 6.10. An MDS $(\Omega, \Sigma, \mu; \varphi)$ is called **ergodic** if every invariant set is essentially equal either to $\emptyset$ or to Ω.

Since invariant sets are the fixed points of φ^* in the measure algebra, a system is ergodic if and only if φ^* on $\Sigma/\sim$ has only the trivial fixed points $\emptyset$ and Ω.

In contrast to minimality in the topological case (Theorem 3.6), an MDS need not have ergodic subsystems: just consider $\Omega = [0,1]$ with the Lebesgue measure and $\varphi = \mathrm{I}$, the identity. Another difference to the topological situation is that the presence of a nontrivial invariant set A does not only lead to a *restriction* of the system (to A), but to a decomposition $X = A \cup A^{\mathrm{c}}$. So ergodicity could also be termed *indecomposability*.

The following result characterises ergodicity.

Lemma 6.11. *Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS. Then the following statements are equivalent.*

(i) *The MDS $(\Omega, \Sigma, \mu; \varphi)$ is ergodic.*

(ii) *For every $\emptyset \neq A \in \Sigma/\sim$ one has*

$$\bigcap_{n \geq 0} \bigcup_{k \geq n} \varphi^{*k} A = \Omega.$$

(iii) *For every $\emptyset \neq A \in \Sigma/\sim$ one has*

$$\bigcup_{n \geq 0} \varphi^{*n} A = \Omega.$$

(iv) *For every $\emptyset \neq A, B \in \Sigma/\sim$ there is $n \geq 1$ such that*

$$\varphi^{*n} A \cap B \neq \emptyset.$$

Proof. (i)$\Longrightarrow$(ii): For $A \in \Sigma$ and $n \geq 0$ the set

$$A^{(n)} := \bigcup_{k \geq n} \varphi^{*k} A$$

satisfies $\varphi^* A^{(n)} \subset A^{(n)}$ and hence is an invariant set. If $A \neq \emptyset$ in the measure algebra then $A^{(n)} \neq \emptyset$ too, and by ergodicity, $A^{(n)} = \Omega$ for every $n \geq 0$. Taking the intersection yields (ii).
(ii)$\Longrightarrow$(iii): This follows from

$$\Omega = \bigcap_{n \geq 0} \bigcup_{k \geq n} \varphi^{*k} A \subset \bigcup_{k \geq 0} \varphi^{*k} A \subset \Omega.$$

(iii)$\Longrightarrow$(iv): By hypothesis, $A^{(0)} = \Omega$, and hence $A^{(1)} = \varphi^* A^{(0)} = \varphi^* \Omega = \Omega$, too. Now suppose that $B \cap \varphi^{*n} A = \emptyset$ for all $n \geq 1$. Then

$$\emptyset = \bigcup_{n \geq 1} B \cap \varphi^{*n} A = B \cap \bigcup_{n \geq 1} \varphi^{*n} A = B \cap A^{(1)} = B \cap \Omega = B.$$

(iv)$\Longrightarrow$(i): Let $A \in \Sigma$ be invariant. Then $\varphi^{*n} A = A$ for every $n \in \mathbb{N}$ and hence for $B := A^{\mathrm{c}}$ we have

$$B \cap \varphi^{*n} A = B \cap A = A^{\mathrm{c}} \cap A = \emptyset$$

for every $n \geq 1$. By hypothesis (iiv), this implies that $A = \emptyset$ or $A^{\mathrm{c}} = B = \emptyset$ in $\Sigma/\sim$. □

The implication (i)⟹(ii) of Lemma 6.11 says that in an ergodic system each set of positive probability is visited infinitely often by almost every point. Note also that (iv) expresses an analogue of condition b) in Proposition 2.15 for MDSs.

Let us now look at some examples. The implication (iv)⟹(i) of Lemma 6.11 combined with the following result shows that a Bernoulli shift is ergodic.

Proposition 6.12. *Let $(\mathscr{W}_k^+, \Sigma, \mu; \tau) = B(p_0, \ldots, p_{k-1})$ be a Bernoulli shift. Then*

$$\lim_{n\to\infty} \mu(\tau^{*n} A \cap B) = \mu(A)\mu(B)$$

for all $A, B \in \Sigma$.

Proof. We use the notation of Example 5 on page 41. Let $\mathcal{E}$ denote the algebra of cylinder sets on $\mathscr{W}_k^+ = L^{\mathbb{N}_0}$. If $B \in \mathcal{E}$ then $B = B_0 \times \prod_{k \geq n_0} L$ for some $n_0 \in \mathbb{N}$ and $B_0 \subset L^{n_0}$. Then $\tau^{*n} A = L^n \times A$ and

$$\mu(\tau^{*n} A \cap B) = \mu(B_0 \times A) = \mu(B)\mu(A)$$

if $n \geq n_0$, since $\mu = \bigotimes_{n \in \mathbb{N}_0} \nu$ is a product measure. If $A \in \Sigma$ is general, by the Approximation Lemma 6.1 one can find a sequence $(B_m)_m \subset \mathcal{E}$ of cylinder sets such that $d_\mu(B_m, B) \to 0$ as $m \to \infty$. Since

$$|\mu(\tau^{*n} A \cap B) - \mu(\tau^{*n} A \cap B_m)| \leq d_\mu(B, B_m)$$

(Exercise 6.1), one has $\mu(\tau^{*n} A \cap B) \to \mu(\tau^{*n} A \cap B_m)$ as $m \to \infty$, uniformly in $n \in \mathbb{N}$. Hence one can interchange limits to obtain

$$\lim_n \mu(\tau^{*n} A \cap B) = \lim_m \lim_n \mu(\tau^{*n} A \cap B_m) = \lim_m \mu(A)\mu(B_m) = \mu(A)\mu(B)$$

as claimed. □

Proposition 6.12 actually says that the Bernoulli shift is *strongly mixing*, a property that will occupy us in a later lecture.

Here are other examples of ergodic systems.

Proposition 6.13. *The following MDSs are ergodic:*

1) *A minimal rotation on a compact group.*
2) *A Markov shift with an irreducible transition matrix P.*

We postpone the proofs of these statements to later lectures.

The next result tells that in an ergodic system the average time of first return to a (nontrivial) set is inverse proportional to its measure.

Corollary 6.14 (Kac). *Let $(\Omega, \Sigma, \mu; \varphi)$ be an ergodic MDS and $A \in \Sigma$ with $\mu(A) > 0$. Then for the* **expected return time** *to A one has*

$$\int_A n_A \, \mathrm{d}\mu_A = \frac{1}{\mu(A)}.$$

Proof. Since the system is ergodic, the implication (i)$\Longrightarrow$(iii) of Lemma 6.11 shows that $\Omega = \bigcup_{n \geq 0} \varphi^{*n} A$. Hence the claim follows from Lemma 6.7. □

Let us return to our ball experiment. The transition matrix P described previously (on page 51) is indeed irreducible, whence by Proposition 6.13.b the corresponding Markov shift is ergodic. Therefore we can apply Kac's theorem concluding that indeed the expected return time to our initial state A is 2^n, so quite large — 2^{100} — for our choice $n = 100$.

Supplement: The Derived Transformation

The following construction is of some importance in ergodic theory. However, we shall not immediately use it and note it here for completeness. One may skip this section on a first reading.

Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS, and let $A \in \Sigma$ with $\mu(A) > 0$. Define the **derived transformation** φ_A by

$$\varphi_A : A \longrightarrow A, \quad \varphi_A(x) = \varphi^{n_A(x)}(x) \qquad (x \in A).$$

That means that $\varphi_A \equiv \varphi^n$ on A_n, $n \in \mathbb{N}$.

Theorem 6.15. *Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS and let $A \in \Sigma$ be a set of positive measure. Then the derived transformation φ_A is measurable with respect to Σ_A and preserves the induced measure μ_A.*

Proof. Take $B \in \Sigma, B \subset A$. Then

$$[\varphi_A \in B] = \bigcup\nolimits_{n \geq 1} A_n \cap [\varphi^n \in B],$$

which shows that φ_A is indeed Σ_A-measurable. To see that φ_A preserves μ, we use Lemma 6.7. Note that since $B \subset A$,

$$A \cap \bigcap_{j=1}^{k-1} \varphi^{*j} A^{\mathrm{c}} \cap \varphi^{*k} B = A_k \cap \varphi^{*k} B \qquad (k \geq 1)$$

and

$$\bigcap_{k=0}^{n} \varphi^{*k} A^{\mathrm{c}} \cap \varphi^{*n} B \subset A^{\mathrm{c}} \cap B_n.$$

Since μ is a probability measure, $\sum_n \mu(B_n) < \infty$ and hence $\mu(A^c \cap B_n) \to 0$ as $n \to \infty$. Letting $n \to \infty$ in (6.3) we obtain

$$\mu(B) = \sum_{k=1}^{\infty} \mu(A_k \cap \varphi^{*k} B) = \sum_{k=1}^{\infty} \mu(A_k \cap [\,\varphi_A \in B]) = \mu[\,\varphi_A \in B\,],$$

and that was to be proved. □

Exercises

1. Let (Ω, Σ, μ) be a finite measure space. Show that

$$d_\mu(A,B) := \mu(A \triangle B) = \|\mathbf{1}_A - \mathbf{1}_B\|_{\mathrm{L}^1} \qquad (A,B \in \Sigma/\sim).$$

defines complete metric on the measure algebra $\Sigma/\sim$. Show further that

a) $d_\mu(A \cap B, C \cap D) \le d_\mu(A,C) + d_\mu(B,D)$;

b) $d_\mu(A^c, B^c) = d_\mu(A,B)$;

c) $d_\mu(A \setminus B, C \setminus D) \le d_\mu(A,C) + d_\mu(B,D)$;

d) $d_\mu(A \cup B, C \cup D) \le d_\mu(A,C) + d_\mu(B,D)$;

e) $|\mu(A) - \mu(B)| \le d_\mu(A,B)$.

(In particular, the mappings

$$(A,B) \longmapsto A \cap B, A \setminus B, A \cup B \quad \text{and} \quad A \longmapsto A^c,\, \mu(A)$$

are continuous with respect to d_μ). Show also that

$$(A_n)_n \subset \Sigma/\sim, \quad A_n \nearrow A \quad \Longrightarrow \quad d_\mu(A_n, A) \searrow 0.$$

2. Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS. Show that the following assertions are equivalent:

(i) There is a measurable mapping $\psi : \Omega \longrightarrow \Omega$ such that $\psi \circ \varphi = \varphi \circ \psi = \mathrm{I}_\Omega$ μ-almost everywhere.

(ii) There is a μ-null set N such that φ maps $\Omega \setminus N$ bijectively onto itself, with measurable inverse.

Show moreover that (i) and (ii) imply that $(\Omega, \Sigma, \mu; \varphi)$ is an invertible MDS.

(Remark: The following example shows that invertibility of an MDS does not in general imply (i): $\Omega = \{0,1\}$, $\Sigma = \{\emptyset, \Omega\}$, $\varphi(x) = 1$ $(x = 0,1)$. However, invertiblity is eqquivalent to (i) if the measure space is a so-called standard Borel space. More about this in a forthcoming lecture.)

3. Show that the tent map and the doubling map are not invertible.

4. Show that if $k \geq 2$, the one-sided Bernoulli shift $B(p_0, \ldots, p_{k-1})$ on $\mathscr{W}_k^+$ is not invertible.

5. Show that $pP = p$, where P, p are the transition matrix and the probability vector defined above in connection with the ball experiment (see page 51).

6.[Recurrence in random literature] The book "Also sprach Zarathustra" by F. Nietzsche consists of roughly 680 000 characters, including blanks. Suppose that Snoopy — known as an immortal philosopher — is typing randomly on his typewriter, which has 90 symbols. Show that he will almost surely type Nietzsche's book infinitely often (thereby proving correct one of Nietzsche's most mysterious theories).

Fig. 6.2 Philosopher at work.

Show also that if Snoopy had been typing since eternity, he almost surely already would *have typed* the book infinitely often. (Nietzsche knew about this in principle, as the quote in the beginning of this lecture shows. However, he almost surely did not know Snoopy.)

Lecture 7
The Banach Lattice L^p and the Induced Operator

> ... Mi sono comportato da ostinato, inseguendo una parvenza di ordine, quando dovevo sapere bene che non vi è ordine, nell'universo. [...] L'ordine che la nostra mente immagine è come una rete, o una scala, che si costruisce per raggiungere qualcosa....[1,2]
>
> Umberto Eco
> Il Nome della Rosa[3]

Let (Ω, Σ, μ) be a measure space and suppose that we are given a measurable map $\varphi : \Omega \longrightarrow \Omega$. The **induced operator** $T = T_\varphi$ then maps measurable functions f to measurable functions $T_\varphi f = f \circ \varphi$. If, in addition, φ respects null sets, i.e., we have

$$\mu(A) = 0 \quad \Longrightarrow \quad \mu[\varphi \in A] = 0 \qquad (A \in \Sigma),$$

then

$$f = g \quad \mu\text{-almost everywhere} \quad \Longrightarrow \quad T_\varphi f = T_\varphi g \quad \mu\text{-almost everywhere}$$

for any pair f, g of scalar valued functions. Hence T_φ acts actually on *equivalence classes* of measurable functions (modulo equality μ-almost everywhere) via

$$T_\varphi [f] := \big[T_\varphi f \big] = [f \circ \varphi].$$

It is an easy but tedious exercise to show that all the common operations for functions induce corresponding operations on equivalence classes, so we usually do not distinguish between functions and their equivalence classes. Moreover, the induced operator T_φ commutes with all these operations:

---

[1] I behaved stubbornly, pursuing a semblance of order, when I should have known well that there is no order in the universe. [...] The order that our mind imagines is like a net, or like a ladder, built to attain something.

[2] Actually, the text continues: "... But afterward you must throw the ladder away, because you discover that, even if it was useful, it was meaningless [...] The only truths that are useful are instruments to be thrown away...."

[3] Il Nome Della Rosa, Bompiani, 2004, page 495.

$$T_\varphi(f+g) = T_\varphi f + T_\varphi g, \quad T_\varphi(\lambda f) = \lambda T_\varphi f, \quad \left|T_\varphi f\right| = T_\varphi\,|f| \quad \ldots.$$

Note that

$$T_\varphi \mathbf{1}_A = \mathbf{1}_A \circ \varphi = \mathbf{1}_{[\varphi \in A]} = \mathbf{1}_{\varphi^* A} \qquad (A \in \Sigma).$$

So T_φ acts on equivalence classes of characteristic functions as φ^* acts on the measure algebra $\Sigma/\!\sim$ (see Lecture 6.1).

Suppose now in addition that φ is measure-preserving, i.e., $\varphi_*\mu = \mu$. Then for any measurable function f and $1 \leq p < \infty$ we obtain (by Appendix B.6)

$$\left\|T_\varphi f\right\|_p^p = \int_\Omega |f \circ \varphi|^p \,\mathrm{d}\mu = \int_\Omega |f|^p \,\mathrm{d}(\varphi_*\mu) = \int_\Omega |f|^p \,\mathrm{d}\mu = \|f\|_p^p.$$

This shows that

$$T_\varphi : \mathrm{L}^p(\Omega, \Sigma, \mu) \longrightarrow \mathrm{L}^p(\Omega, \Sigma, \mu) \qquad (1 \leq p < \infty)$$

is a linear isometry. The same is true for $p = \infty$ (Exercise 7.1).

In Lecture 4 we have associated to a TDS $(K;\varphi)$ the commutative unital C^*-algebra $\mathrm{C}(K)$ and the induced operator $T = T_\varphi$ acting on it. We cannot take the same road with MDSs since, apart from the case $p = \infty$, the space $\mathrm{L}^p(\Omega, \Sigma, \mu)$ is in general not closed under multiplication and hence is not a Banach algebra. Since the measure μ is not determined by the L^∞-space, we have to use L^p-spaces for $1 \leq p < \infty$ and therefore need a new structural element preserved by the induced operator. It turns out that it is the *lattice structure*.

7.1 The Space $\mathrm{L}^p(\Omega, \Sigma, \mu)$ as a Banach Lattice

In this section we look at the space $\mathrm{L}^p(\Omega, \Sigma, \mu)$ with respect to its underlying order structure. Since we do not expect order-theoretic notions to be common knowledge, we shall introduce the main abstract concepts, mostly without formal definitions but exemplified by the L^p-spaces. If we ever talk about abstract Banach lattices, the reader may without harm think of the L^p-spaces for simplicity. For a very detailed account we refer to the monograph [Schaefer (1974)]. The reader familiar with the Banach lattice structure of L^p-spaces and positive operators on them, may skip this part and proceed directly to Section 7.2.

A **lattice** is a partially ordered set $(X, \leq)$ such that

$$x \vee y := \sup\{x, y\} \quad \text{and} \quad x \wedge y := \inf\{x, y\}$$

exist for all $x, y \in X$. (Here $\sup A$ denotes the least upper bound or **supremum** of the set $A \subset X$, if it exists. Likewise, $\inf A$ is the greatest lower bound, or **infimum** of A.) A lattice is called **complete** if every non-empty subset has a supremum and an infimum. A subset $D \subset X$ of a lattice $(X, \leq)$ is called $\vee$**-stable** ($\wedge$**-stable**) if $a, b \in D$

implies $a \vee b \in D$ ($a \wedge b \in D$). If X is a lattice and $A \subset X$ is a non-empty set, then A has a supremum if and only if the $\vee$-stable set

$$\big\{a_1 \vee a_2 \ldots \vee a_n : n \in \mathbb{N},\, a_1,\ldots,a_n \in A\big\}$$

has a supremum, and in this case these suprema coincide. If $(X,\leq)$ and $(Y,\leq)$ are lattices, then a map $\Theta : X \longrightarrow Y$ such that $\Theta(x \vee y) = (\Theta x) \vee (\Theta y)$ and $\Theta(x \wedge y) = (\Theta x) \wedge (\Theta y)$ is called a **homomorphism** of lattices. If Θ is bijective, then also Θ^{-1} is a homomorphism, and Θ is called a **lattice isomorphism**.

The prototype of a complete lattice is the extended real line $\overline{\mathbb{R}} := [-\infty,\infty]$ with the usual order. For a measure space (Ω,Σ,μ) we denote by

$$L^0 = L^0(\Omega;\overline{\mathbb{R}}) := L^0(\Omega,\Sigma,\mu;\overline{\mathbb{R}})$$

the set of all equivalence classes of measurable functions $f : \Omega \longrightarrow [-\infty,\infty]$, and define a partial order by

$$[f] \leq [g] \quad \overset{\text{Def.}}{\Longleftrightarrow} \quad f \leq g \quad \mu\text{-almost everywhere.}$$

The ordered set L^0 is a lattice, since for each two elements $f,g \in L^0$ the supremum and infimum is given by

$$(f \vee g)(x) = \max\big\{f(x),g(x)\big\}, \quad (f \wedge g)(x) = \min\big\{f(x),g(x)\big\} \qquad (x \in \Omega).$$

(To be very precise, this means that the supremum $f \vee g$ is represented by the pointwise supremum of representatives of the equivalence classes f and g.) Analogously we see that in L^0 the supremum (infimum) of every *sequence* exists and is represented by the pointwise supremum of representatives. (Exercise 7.10 shows that the lattice $L^0(\Omega;\overline{\mathbb{R}})$ is complete if the measure is finite.)

Now fix $1 \leq p < \infty$ and abbreviate $L^p(\Omega;\mathbb{R}) := L^p(\Omega,\Sigma,\mu;\mathbb{R})$. Then

$$f,g \in L^p(\Omega;\mathbb{R}) \quad \Longrightarrow \quad f \vee g, f \wedge g \in L^p(\Omega;\mathbb{R}).$$

This shows that the space $L^p(\Omega;\mathbb{R})$ is a **sublattice** of L^0, i.e., it is closed under the lattice operations $\vee$ and $\wedge$. Furthermore, $L^p(\Omega;\mathbb{R})$ is also a real vector space and we have

$$f \leq g \quad \Longrightarrow \quad f + h \leq g + h, \quad c \cdot f \leq c \cdot g \tag{7.1}$$

for all $f,g,h \in L^p(\Omega;\mathbb{R})$ and $c \geq 0$. This means that $L^p(\Omega;\mathbb{R})$ is a **vector lattice**. In any vector lattice one can define

$$|f| := f \vee (-f), \quad f^+ := f \vee 0, \quad f^- := (-f) \vee 0,$$

which — in our space $L^p(\Omega;\mathbb{R})$ — all coincide with the respective pointwise operations. Then we have

$$f = f^+ - f^-, \quad |f| = f^+ + f^-, \tag{7.2a}$$
$$|f+g| \le |f| + |g|, \quad |cf| = |c| \cdot |f| \quad (c \in \mathbb{R}), \tag{7.2b}$$
$$f \vee g = \tfrac{1}{2}(f+g+|f-g|), \quad f \wedge g = \tfrac{1}{2}(f+g-|f-g|), \tag{7.2c}$$
$$(f \vee g) + h = (f+h) \vee (g+h), \quad (f \wedge g) + h = (f+h) \wedge (g+h) \tag{7.2d}$$
$$|f \vee g - f_1 \vee g_1| \le |f - f_1| + |g - g_1|, \tag{7.2e}$$
$$|f \wedge g - f_1 \wedge g_1| \le |f - f_1| + |g - g_1|, \tag{7.2f}$$
$$\big||f| - |g|\big| \le |f - g| \tag{7.2g}$$

for all $f, f_1, g, g_1 \in \mathrm{L}^p(\Omega;\mathbb{R})$. These identities and inequalities are easy to establish in our special case, but they actually hold in any vector lattice. An element $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$ is called **positive** if $f \ge 0$. We denote by

$$\mathrm{L}^p_+ = \mathrm{L}^p_+(\Omega,\Sigma,\mu) := \big\{ f \in \mathrm{L}^p(\Omega,\Sigma,\mu) \,:\, f \ge 0 \big\}$$

the **positive cone**. Note that (7.2a) implies that $\mathrm{L}^p(\Omega;\mathbb{R}) = \mathrm{L}^p_+ - \mathrm{L}^p_+$.

By definition of the norm on $\mathrm{L}^p(\Omega,\Sigma,\mu)$ we have

$$|f| \le |g| \quad \Longrightarrow \quad \|f\|_p \le \|g\|_p$$

for all $f, g \in \mathrm{L}^p(\Omega;\mathbb{R})$. Since the space $\mathrm{L}^p(\Omega;\mathbb{R})$ is also a Banach space, this means that it is a **Banach lattice**. Here the lattice operations

$$f \longmapsto |f|\,, f^+, f^-, \quad (f,g) \longmapsto f \vee g, f \wedge g$$

are continuous, as follows from (7.2e)–(7.2g).

If $1 \le p < \infty$ the space $\mathrm{L}^p(\Omega,\Sigma,\mu)$ has **order continuous norm**. This means that for each decreasing sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{L}^p_+(\Omega,\Sigma,\mu)$, $f_n \ge f_{n+1}$, one has

$$\inf_n f_n = 0 \quad \Longrightarrow \quad \|f_n\|_p \to 0.$$

This is a direct consequence of the Monotone Convergence Theorem (see Appendix B.10) by considering the sequence $(f_1^p - f_n^p)_{n\in\mathbb{N}}$. Actually, the Monotone Convergence Theorem accounts also for the following statement.

Theorem 7.1. *Let (Ω,Σ,μ) be a measure space and let $1 \le p < \infty$. Let $\mathscr{F} \subset \mathrm{L}^p_+(\Omega,\Sigma,\mu)$ be a $\vee$-stable set such that*

$$s := \sup\big\{ \|f\|_p \,:\, f \in \mathscr{F} \big\} < \infty.$$

Then $f := \sup \mathscr{F}$ exists in the Banach lattice $\mathrm{L}^p(\Omega;\mathbb{R})$ and there exists an increasing sequence $(f_n)_n \subset \mathscr{F}$ such that $\sup_n f_n = f$ and $\|f_n - f\|_p \to 0$. In particular, $\sup \mathscr{F} \in \mathscr{F}$ if $\mathscr{F}$ is closed.

Proof. Take a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathscr{F}$ with $\|f_n\|_p \to s$. By passing to the sequence $(f_1 \vee f_2 \vee \ldots \vee f_n)_{n\in\mathbb{N}}$ we may suppose that $(f_n)_n$ is increasing. Define $f := \lim_n f_n$

pointwise almost everywhere. By the Monotone Convergence Theorem we obtain $\|f\|_p = s$ hence $f \in L^p$. Since $\inf_n(f - f_n) = 0$, by order continuity of the norm it follows that $\|f - f_n\|_p \to 0$. If h is any upper bound for $\mathscr{F}$ then $f_n \le h$ for each n, and then $f \le h$. Hence it remains to show that f is an upper bound of $\mathscr{F}$. Take an arbitrary $g \in \mathscr{F}$. Then $f_n \vee g \nearrow f \vee g$. Since $f_n \vee g \in \mathscr{F}$, $\|f_n \vee g\|_p \le s$ for each $n \in \mathbb{N}$. The Monotone Convergence Theorem implies that $\|f \vee g\|_p = \lim_n \|f_n \vee g\|_p \le s$. But $f \vee g \ge f$ and hence

$$\|(f \vee g)^p - f^p\|_1 = \|f \vee g\|_p^p - \|f\|_p^p \le s^p - s^p = 0.$$

This shows that $f = f \vee g \ge g$, as desired. □

Remark 7.2. In the previous proof we have used actually that $|f| \le |g|$, $|f| \ne |g|$ implies $\|f\|_1 < \|g\|_1$, a property telling that L^1 has **strictly monotone norm**. Each L^p-space $(1 \le p < \infty)$ has strictly monotone norm.

A Banach lattice is called **order complete** if any non-empty subset which has an upper bound has even a least upper bound. Notice that in an order complete Banach lattice the analogous statement is true for lower bounds, and, in particular, an order bounded sublattice in an order complete Banach lattice is a complete lattice. Now Theorem 7.1 has an important consequence.

Corollary 7.3. *Let (Ω,Σ,μ) be a measure space and let $1 \le p < \infty$. Then the Banach lattice $L^p(\Omega;\mathbb{R})$ is order complete.*

Proof. Let $\mathscr{F}' \subset L^p(\Omega;\mathbb{R})$ and suppose that there exists $F \in L^p(\Omega;\mathbb{R})$ with $f \le F$ for all $f \in \mathscr{F}'$. We may suppose without loss of generality that $\mathscr{F}'$ is $\vee$-stable. Pick $g \in \mathscr{F}'$ and consider the set $g \vee \mathscr{F}' := \{g \vee f : f \in \mathscr{F}'\}$, which is again $\vee$-stable and has the same set of upper bounds as $\mathscr{F}'$. Now $\mathscr{F} := (g \vee \mathscr{F}') - g$ is $\vee$-stable by (7.2d), consists of positive elements and is dominated by $F - g \ge 0$. In particular it satisfies the conditions of Theorem 7.1. Hence it has a supremum h. It is then obvious that $h + g$ is the supremum of $\mathscr{F}'$. By passing to $-\mathscr{F}'$ one can prove that a non-empty family in L^p that is *bounded from below* has an infimum in L^p. □

Remark 7.4 (Complex Banach lattices). For the purpose of spectral theory (which plays an important role in the study of induced operators) it is essential to work with the complex Banach spaces $L^p(\Omega,\Sigma,\mu) = L^p(\Omega,\Sigma,\mu;\mathbb{C})$. Any function $f \in L^p(\Omega,\Sigma,\mu)$ can be written uniquely as

$$f = \operatorname{Re} f + \mathrm{i} \operatorname{Im} f$$

with real-valued functions $\operatorname{Re} f$, $\operatorname{Im} f \in L^p(\Omega;\mathbb{R})$. Hence $L^p(\Omega,\Sigma,\mu)$ decomposes as

$$L^p(\Omega,\Sigma,\mu) = L^p(\Omega,\Sigma,\mu;\mathbb{R}) \oplus \mathrm{i}L^p(\Omega,\Sigma,\mu;\mathbb{R})$$

with $L^p(\Omega,\Sigma,\mu;\mathbb{R})$ a real Banach lattice, closed within $L^p(\Omega,\Sigma,\mu)$. Furthermore, the absolute value mapping $|\cdot|$ has an extension to $L^p(\Omega,\Sigma,\mu)$ that satisfies

$$|f| = \sup\{\operatorname{Re}(cf) : c \in \mathbb{C},\, |c| = 1\}, \tag{7.3}$$

cf. Exercise 7.9. This gives a key for the general definition of a complex Banach lattice (see [Schaefer (1974), page 133]).

Recall that algebra ideals play an important role in the study of $C(K)$. In the Banach lattice setting there is an analogous notion.

Definition 7.5. Let (Ω, Σ, μ) be a measure space and let $1 \le p \le \infty$. A linear subspace $I \subset L^p(\Omega, \Sigma, \mu)$ is called a (vector) **lattice ideal** if

$$f, g \in L^p(\Omega, \Sigma, \mu),\ |f| \le |g|, \quad g \in I \quad \Longrightarrow \quad f \in I.$$

If I is a lattice ideal then $f \in I$ if and only if $|f| \in I$ if and only if $\mathrm{Re}\, f, \mathrm{Im}\, f \in I$. By (7.2c) a real part of a lattice ideal is also a sublattice, i.e., if $f, g \in I$ then $\mathrm{Re}\, f \vee \mathrm{Re}\, g, \mathrm{Re}\, f \wedge \mathrm{Re}\, g \in I$. Immediate examples of closed lattice ideals in $L^p(\Omega, \Sigma, \mu)$ are obtained from measurable sets $A \in \Sigma$ by a construction similar to the TDS case (cf. page 29):

$$I_A := \{f : |f| \wedge \mathbf{1}_A = 0\} = \{f : |f| \wedge \mathbf{1} \le \mathbf{1}_{A^c}\} = \{f : A \subset [f = 0]\}.$$

Then I_A is indeed a closed lattice ideal, and for $A = \emptyset$ and $A = \Omega$ we recover $L^p(\Omega, \Sigma, \mu)$ and $\{0\}$, the two trivial lattice ideals. The following characterisation tells that actually all closed lattice ideals in $L^p(\Omega, \Sigma, \mu)$ arise by this construction.

Theorem 7.6. *Let (Ω, Σ, μ) be a finite measure space and $1 \le p < \infty$. Then each closed lattice ideal $I \subset L^p(\Omega, \Sigma, \mu)$ has the form I_A for some $A \in \Sigma$.*

Proof. Let $I \subset L^p(\Omega, \Sigma, \mu)$ be a closed lattice ideal. The set

$$J := \{f \in I : 0 \le f \le \mathbf{1}\}$$

is non-empty, closed, $\vee$-stable and has upper bound $\mathbf{1} \in L^p(\Omega, \Sigma, \mu)$ (since $\mu(\Omega)$ is finite). Therefore, by Corollary 7.3 it has even a least upper bound $g \in J$. It follows that $0 \le g \le \mathbf{1}$ and thus $h := g \wedge (\mathbf{1} - g) \ge 0$. Since $0 \le h \le g$ and $g \in I$, the ideal property yields $h \in I$. Since I is a subspace, $g + h \in I$. But $h \le \mathbf{1} - g$, so $g + h \le \mathbf{1}$, and this yields $h + g \in J$. Thus $h + g \le g$, i.e., $h \le 0$. All in all we obtain $g \wedge (\mathbf{1} - g) = h = 0$, hence g must be a characteristic function $\mathbf{1}_{A^c}$ for some $A \in \Sigma$.

We claim that $I = I_A$. To prove the inclusion "$\subset$" take $f \in I$. Then $|f| \wedge \mathbf{1} \in J$ and therefore $|f| \wedge \mathbf{1} \le g = \mathbf{1}_{A^c}$. This means that $f \in I_A$. To prove the converse inclusion take $f \in I_A$. It suffices to show that $|f| \in I$ hence we may suppose that $f \ge 0$. Then $f_n := f \wedge (n\mathbf{1}) = n(n^{-1} f \wedge \mathbf{1}) \le n\mathbf{1}_{A^c} = ng$. Since $g \in J \subset I$, we have $ng \in I$, whence $f_n \in I$. Now $(f_n)_n$ is increasing and converges pointwise, hence in norm to its supremum f. Since I is closed, $f \in I$ and this concludes the proof. □

Remarks 7.7. a) In most of the results of this section we required $p < \infty$, for good reasons. The space $L^\infty(\Omega, \Sigma, \mu)$ is a Banach lattice, but its norm is not order continuous. If the measure is finite, $L^\infty(\Omega, \Sigma, \mu)$ is still order complete (Exercise 7.10), but this is not true for general measure spaces. Moreover, if L^∞ is not finite dimensional, then there are always closed lattice ideals not of the form I_A.

b) If (Ω, Σ, μ) is a finite measure space, then the measure algebra $\Sigma/\!\sim$ is a lattice (with respect to the obvious order) and

$$\Sigma/\!\sim \longrightarrow \mathrm{L}^\infty, \qquad [A] \longmapsto [\mathbf{1}_A] \qquad (A \in \Sigma)$$

is an injective lattice homomorphism. The order-completeness of $\mathrm{L}^\infty(\Omega, \Sigma, \mu)$ (see a)) implies that $\Sigma/\!\sim$ is a complete lattice.

c) The space $\mathrm{L}^\infty(\Omega, \Sigma, \mu)$ is also a commutative, unital C^*-algebra. Moreover if $f \in \mathrm{L}^\infty$ and $g \in \mathrm{L}^p$, then $fg \in \mathrm{L}^p$ again, and hence $\mathrm{L}^p(\Omega, \Sigma, \mu)$ is a **module** over the algebra $\mathrm{L}^\infty(\Omega, \Sigma, \mu)$.

d) If (Ω, Σ, μ) is a finite measure space, we have

$$\mathrm{L}^\infty(\Omega, \Sigma, \mu) \subset \mathrm{L}^p(\Omega, \Sigma, \mu) \subset \mathrm{L}^1(\Omega, \Sigma, \mu) \quad (1 \le p \le \infty).$$

Particularly important in this scale will be the *Hilbert space* $\mathrm{L}^2(\Omega, \Sigma, \mu)$.

e) Finally we remark that the spaces $\mathrm{C}(K)$ from Lecture 4 are Banach lattices as well, see Exercise 7.8. One can show that the closed lattice ideals of $\mathrm{C}(K)$ coincide with the closed algebra ideals.

7.2 The Induced Operator and Ergodicity

We now study MDSs $(\Omega, \Sigma, \mu; \varphi)$ and the induced operator $T = T_\varphi$ on $\mathrm{L}^p(\Omega, \Sigma, \mu)$. We know that, for every $1 \le p \le \infty$,

1) T is an isometry on $\mathrm{L}^p(\Omega, \Sigma, \mu)$;
2) T is a **Banach lattice homomorphism** on $\mathrm{L}^p(\Omega, \Sigma, \mu)$; i.e., linear and $|Tf| = T|f|$ for every $f \in \mathrm{L}^p(\Omega, \Sigma, \mu)$; Note that the induced operator preserves $\vee$ and $\wedge$ on $\mathrm{L}^p(\Omega, \Sigma, \mu; \mathbb{R})$, hence is a lattice homomorphism (see page 61).
3) $T(fg) = Tf \cdot Tg$ for all $f \in \mathrm{L}^p(\Omega, \Sigma, \mu)$, $g \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)$;
4) T is unital C^*-algebra homomorphism on $\mathrm{L}^\infty(\Omega, \Sigma, \mu)$.

Any Banach lattice homomorphism $T : \mathrm{L}^p(\Omega, \Sigma, \mu) \longrightarrow \mathrm{L}^p(\Omega, \Sigma, \mu)$ is **positive**, i.e., $f \le g$ implies $Tf \le Tg$. Furthermore it is a homomorphism of the lattice $\mathrm{L}^p(\Omega; \mathbb{R})$. A positive operator T on $\mathrm{L}^p(\Omega, \Sigma, \mu)$ is called **irreducible** if the only T-invariant lattice ideals are $\{0\}$ and $\mathrm{L}^p(\Omega, \Sigma, \mu)$ itself.

The following result shows that the ergodicity of an MDS is characterised by the irreducibility of the induced operator. This is reminiscent (cf. Corollary 4.13) but also contrary to the TDS case, where minimality could not be characterised by the one-dimensionality of the fixed space of the induced operators.

Proposition 7.8. *Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS and consider the induced operator $T := T_\varphi$. The following statements are equivalent.*

(i) $(\Omega, \Sigma, \mu; \varphi)$ *is ergodic.*

(ii) *The fixed space*

$$\mathrm{Fix}(T) := \{f \in \mathrm{L}^p(\Omega,\Sigma,\mu) : Tf = f\},$$

is one-dimensional for some/all $1 \le p \le \infty$.

(iii) T *is irreducible on* $\mathrm{L}^p(\Omega,\Sigma,\mu)$, $1 \le p < \infty$.

Note that the constant functions are always contained in $\mathrm{Fix}(T) = \ker(\mathrm{I}-T)$. Hence (ii) is the same as to say that 1 is a simple eigenvalue of T.

Proof. We shall see that the proof does not depend on the choice of p. Notice also that by Lemma 6.9 A is φ-invariant if and only if $\varphi^*[A] = [A]$ if and only if $\mathbf{1}_A$ is a fixed point of T.

(i) $\Longrightarrow$ (ii): For any $f \in \mathrm{Fix}(T)$ and any $c \in \mathbb{R}$ the set $[f > c]$ is φ-invariant, and hence trivial (up to a null-set). Let

$$c_0 := \sup\{c \in \mathbb{R} : \mu[f > c] = 1\}.$$

Then for $c < c_0$ we have $\mu[f \le c] = 0$, and therefore $\mu[f < c_0] = 0$. For $c > c_0$ we have $\mu[f > c] \ne 1$, hence $\mu[f > c] = 0$, and therefore $\mu[f > c_0] = 0$, too. This implies $f = c_0$ almost everywhere.

(ii) $\Longrightarrow$ (i): If $A \in \Sigma/\sim$ satisfies $\varphi^*A = A$ and $A \ne \emptyset, A \ne \Omega$, then $\mathbf{1}_A$ is a non-constant fixed point of T, hence $\dim \mathrm{Fix}(T) \ge 2$.

(i) $\Longleftrightarrow$ (iii): This follows from Theorem 7.6 and from the fact that $A \in \Sigma$ is φ-invariant if and only if I_A is invariant under T. To see this, notice that $T(|f| \wedge \mathbf{1}_A) = T|f| \wedge T\mathbf{1}_A$ and $|Tf| = T|f|$. Hence $f \in I_A$ implies $Tf \in I_A$, whenever $\mathbf{1}_A = T\mathbf{1}_A$, i.e., when A is invariant. Now suppose that I_A is T-invariant. As $\mathbf{1}_{A^c} \in I_A$, we have $T\mathbf{1}_{A^c} \in I_A$, whence $T\mathbf{1}_{A^c} \le \mathbf{1}_{A^c}$. But then

$$T\mathbf{1}_A = T(\mathbf{1} - \mathbf{1}_{A^c}) = \mathbf{1} - T\mathbf{1}_{A^c} \ge \mathbf{1} - \mathbf{1}_{A^c} = \mathbf{1}_A,$$

thus $A \subset \varphi^*A$, i.e., A is invariant. □

We apply Proposition 7.8 to the rotations on the torus (see Lecture 5.3).

Proposition 7.9. *For* $a \in \mathbb{T}$ *the MDS* $(\mathbb{T},\mathfrak{B},\lambda;\varphi_a)$ *is ergodic if and only if* a *is not a root of unity.*

Proof. Let $T := T_{\varphi_a}$ be the induced operator on $\mathrm{L}^2(\mathbb{T},\mathfrak{B},\lambda)$, and suppose that $f \in \mathrm{Fix}(T)$. The functions $\chi_n : x \mapsto x^n$, $n \in \mathbb{Z}$ form a complete orthonormal system in $\mathrm{L}^2(\mathbb{T})$. Note that $T\chi_n = a^n\chi_n$, i.e., χ_n is an eigenvector of T with corresponding eigenvalue a^n. With appropriate Fourier-coefficients b_n we have

$$f = \sum_{n\in\mathbb{Z}} b_n\chi_n,$$

hence

$$Tf = \sum_{n\in\mathbb{Z}} b_n T\chi_n = \sum_{n\in\mathbb{Z}} a^n b_n\chi_n.$$

Since $Tf = f$ it follows, by the uniqueness of the Fourier-coefficients, that $b_n(a^n - 1) = 0$. This implies that either for all $n \in \mathbb{Z} \setminus \{0\}$ we have $b_n = 0$ hence f is constant, or there is $n \in \mathbb{Z} \setminus \{0\}$ with $a^n = 1$, hence a is a root of unity. This proves the claim. □

Remark 7.10. If φ_a is a rational rotation on the torus, then a non-constant fixed point of T_{φ_a} is easy to find: Suppose that $a^n = 1$ for some $n \in \mathbb{N}$, and divide $\mathbb{T}$ into n arcs: $I_j := \{z \in \mathbb{T} : \arg z \in [2\pi(j-1), 2\pi j/n)\}$, $j = 1, \dots, n$. Take any integrable function on I_1 and "copy it over" to the other segments. The thus arising function is a fixed point of T_{φ_a}.

Proposition 7.9 together with Example 2.19 says that a rotation on the torus is ergodic if and only if it is minimal. Exercise 7.2 generalises this to rotations on $\mathbb{T}^n$. Actually, the result is true for any rotation on a compact group as we shall prove it in a later lecture.

Similar to the case of minimal TDSs, the peripheral point spectrum of the induced operator has a nice structure.

Proposition 7.11. *Let $(\Omega, \Sigma, \mu; \varphi)$ be an ergodic MDS, and consider the induced operator $T := T_\varphi$ on $\mathrm{L}^p(\Omega, \Sigma, \mu)$ ($1 \le p \le \infty$). Then the peripheral point spectrum $\sigma_p(T) \cap \mathbb{T}$ is a subgroup of $\mathbb{T}$, and each eigenvalue is simple.*

We leave the proof as Exercise 7.3.

7.3 Isomorphism of Measure Preserving Systems

Let $(\Omega, \Sigma, \mu; \varphi)$ and $(\Omega', \Sigma', \mu'; \psi)$ be two MDSs and suppose that there is a measurable, measure-preserving map $\theta : \Omega \longrightarrow \Omega'$ such that the diagram below commutes.

$$\begin{tikzcd} \Omega \arrow[r, "\varphi"] \arrow[d, "\theta"'] & \Omega \arrow[d, "\theta"] \\ \Omega' \arrow[r, "\psi"'] & \Omega' \end{tikzcd}$$

Then θ is called a **point homomorphism** or a **factor map**. If $\theta : \Omega \longrightarrow \Omega'$ is a factor map and there exists a factor map $\tilde{\theta} : \Omega' \longrightarrow \Omega$ such that $\theta \circ \tilde{\theta} = \mathrm{Id}_{\Omega'}$ and $\tilde{\theta} \circ \theta = \mathrm{Id}_\Omega$ almost everywhere, then θ is called a **point isomorphism**.

A factor map θ induces an injective homomorphism

$$\theta^* : \Sigma'/\!\sim \longrightarrow \Sigma/\!\sim, \qquad [A] \longmapsto \big[\theta^{-1}(A)\big]$$

of the corresponding *measure algebras*. It is easy to see that $\theta^* \circ \psi^* = \varphi^* \circ \theta^*$. If θ is a point isomorphism, then θ^* is bijective, hence is an isomorphism in the sense of the following definition.

Definition 7.12. An (algebra) **isomorphism** of two MDSs $(\Omega,\Sigma,\mu;\varphi)$ and $(\Omega',\Sigma',\mu';\psi)$ is a lattice isomorphism $\Theta : \Sigma/{\sim} \longrightarrow \Sigma'/{\sim}$ of the corresponding measure-algebras, such that $\mu'(\Theta A) = \mu(A)$ for all $A \in \Sigma/{\sim}$ and the diagram

$$\begin{array}{ccc} \Sigma/{\sim} & \xrightarrow{\varphi^*} & \Sigma/{\sim} \\ \Theta \big\downarrow & & \big\downarrow \Theta \\ \Sigma'/{\sim} & \xrightarrow[\psi^*]{} & \Sigma'/{\sim} \end{array}$$

commutes. Two MDSs are (algebra) **isomorphic** if there exists an (algebra) isomorphism between them.

If $\theta : \Omega \longrightarrow \Omega'$ is a point isomorphism, then $\Theta := \theta^{*-1}$ is an algebra isomorphism. The following example shows that not every algebra isomorphism is induced by a point isomorphism: Take $\Omega = \{0\}$, $\Sigma = \mathscr{P}(\Omega)$, $\mu(\Omega) = 1$, $\varphi = \mathrm{Id}$ and $(\Omega,\Sigma',\mu';\psi)$ with $\Omega' = \{0,1\}$, $\Sigma' = \{\emptyset,\Omega'\}$, $\mu'(\Omega') = 1$, $\psi = \mathrm{Id}$. The two MDSs are not point isomorphic but it is easy to see that they are (algebra) isomorphic.

The next result indicates that the Banach lattice structure is adequate for the study of measure preserving systems.

Proposition 7.13. *Two MDSs $(\Omega,\Sigma,\mu;\varphi)$ and $(\Omega',\Sigma',\mu';\psi)$ are (algebra) isomorphic if and only if there exists an isometric Banach lattice isomorphism ι : $\mathrm{L}^1(\Omega,\Sigma,\mu) \longrightarrow \mathrm{L}^1(\Omega',\Sigma',\mu')$ with $\iota\mathbf{1}_\Omega = \mathbf{1}_{\Omega'}$ such that the diagram*

$$\begin{array}{ccc} \mathrm{L}^1(\Omega,\Sigma,\mu) & \xrightarrow{T_\varphi} & \mathrm{L}^1(\Omega,\Sigma,\mu) \\ \iota \big\downarrow & & \big\downarrow \iota \\ \mathrm{L}^1(\Omega',\Sigma',\mu') & \xrightarrow[T_\psi]{} & \mathrm{L}^1(\Omega',\Sigma',\mu') \end{array} \tag{7.4}$$

commutes.

Proof. Note first that a function $f \in \mathrm{L}^1$ is a characteristic function if and only if $f \wedge (\mathbf{1} - f) = 0$. Therefore, an isometric lattice isomorphism ι maps the characteristic functions on Ω onto the characteristic functions on Ω', and thereby induces a map

$$\Theta : \Sigma/{\sim} \longrightarrow \Sigma'/{\sim},$$

by setting $\Theta A := B$ if $\iota\mathbf{1}_A = \mathbf{1}_B$ ($A \in \Sigma/{\sim}$, $B \in \Sigma'/{\sim}$). It is clear that Θ is an lattice isomorphism preserving the measure.

Conversely, let $\Theta : \Sigma/{\sim} \longrightarrow \Sigma'/{\sim}$ be a measure-preserving lattice isomorphism. Then we can define $\iota\mathbf{1}_A := \mathbf{1}_{\Theta A}$. Clearly, we have

$$\iota(\mathbf{1}_A \wedge \mathbf{1}_B) = \iota\mathbf{1}_{A\cap B} = \mathbf{1}_{\Theta(A\cap B)} = \mathbf{1}_{\Theta A\cap\Theta B} = \mathbf{1}_{\Theta A} \wedge \mathbf{1}_{\Theta B},$$

and similarly for $\vee$; moreover $\iota \mathbf{1}_\Omega = \mathbf{1}_{\Omega'}$ holds. Observe that $\|\mathbf{1}_A\|_1 = \mu(A) = \mu(\Theta A) = \|\mathbf{1}_{\Theta A}\|_1 = \|\iota \mathbf{1}_A\|_1$. Therefore ι extends to an $\|\cdot\|_1$-isometric isomorphism first on the linear hull of the characteristic functions and then on the whole of $\mathrm{L}^1(\Omega,\Sigma,\mu)$. Since the identity $|\iota f| = \iota\,|f|$ is true for step functions, it must be true for all $f \in \mathrm{L}^1(\Omega,\Sigma,\mu)$. This shows that ι is a Banach lattice isomorphism.

To conclude the proof, note that

$$\iota T_\varphi(\mathbf{1}_A) = \iota \mathbf{1}_{\varphi^* A} = \mathbf{1}_{\Theta\varphi^* A} \quad \text{and} \quad T_\psi \iota(\mathbf{1}_A) = T_\psi \mathbf{1}_{\Theta A} = \mathbf{1}_{\psi^*\Theta A}$$

for every $A \in \Sigma$. Hence $\Theta \circ \varphi^* = \psi^* \circ \Theta$ if and only if $\iota \circ T_\varphi = T_\psi \circ \iota$, again by denseness of the step functions. □

Remark 7.14. The isometric lattice isomorphism $\iota : \mathrm{L}^1(\Omega,\Sigma,\mu) \longrightarrow \mathrm{L}^1(\Omega',\Sigma',\mu')$ in (7.4) may be restricted to the corresponding L^p-spaces, $1 \le p < \infty$. These restrictions are still isometric lattice isomorphisms for which the corresponding L^p-diagram commutes.

It is useful to introduce the following terminology. Two measure spaces (Ω,Σ,μ) and (Ω',Σ',μ') are called **point isomorphic**, or **algebra isomorphic** if $(\Omega,\Sigma,\mu;\mathrm{Id}_\Omega)$ and $(\Omega',\Sigma',\mu';\mathrm{Id}_{\Omega'})$ are point isomorphic, respectively algebra isomorphic. In certain situations, however, the distinction between "algebra isomorphic" and "point isomorphic" is unnecessary. J. von Neumann proved in [von Neumann (1932a)] that two compact metric (Borel) probability spaces are algebra isomorphic if and only if they are point isomorphic (for a proof and further reading we refer to [Royden (1963), Sec. 14.5] or Bogachev [Bogachev (2007), Ch. 9]). From this one can deduce the following classical result.

Theorem 7.15 (von Neumann, 1932). *Two MDSs on compact metric probability spaces are isomorphic if and only if they are point isomorphic.*

Exercises

1. Show that for an MDS $(\Omega,\Sigma,\mu;\varphi)$ the induced operator is an isometry on $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$.

2. Show that the rotation on the torus $\mathbb{T}^n$ is ergodic if and only if it is minimal. (Hint: Copy the proof of Proposition 7.9 using n-dimensional Fourier coefficients.)

3. Prove Proposition 7.11.

4. Show that ergodicity and invertibility of MDSs are (algebra) isomorphism invariants.

5. Let $(\Omega,\Sigma,\mu;\varphi)$ and $(\Omega',\Sigma',\mu';\psi)$ be two MDSs. Show that if ι is a Banach lattice isomorphism of the corresponding L^1-spaces, the respective measure-algebra isomorphism Θ is a σ-algebra isomorphism, i.e., preserves countable disjoint unions.

6.

a) Consider the Bernoulli shift $B(\frac{1}{2},\frac{1}{2}) = (\mathscr{W}_2^+,\Sigma,\mu;\tau)$ and endow $[0,1]$ with the Lebesgue measure. Show that the measure spaces $(\mathscr{W}_2,\Sigma,\mu)$ and $([0,1],\Lambda,\lambda)$, Λ the Lebesgue σ-algebra, are isomorphic. Prove that $B(\frac{1}{2},\frac{1}{2})$ and $([0,1],\Lambda,\lambda;\varphi)$, φ the doubling map, are isomorphic (see Example 5.1.2).

b) Prove the analogous statements for the Baker's Transformation (see Example 5.1.1) and the two-sided Bernoulli-shift $B(\frac{1}{2},\frac{1}{2}) = (\mathscr{W}_2^+,\Sigma,\mu;\tau)$.

(Hint: write the numbers in $[0,1]$ in binary form.)

7. Give an example of two non-isomorphic MDSs $(\Omega,\Sigma,\mu;\varphi)$ and $(\Omega',\Sigma',\mu';\psi)$ for which there is a Hilbert space isomorphism $\iota : L^2(\Omega,\Sigma,\mu) \longrightarrow L^2(\Omega',\Sigma',\mu')$ with the corresponding L^2-diagram (7.4) being commutative, i.e., with $T_\psi \circ \iota = \iota \circ T_\varphi$. In this case the two MDSs are called **spectrally isomorphic**. Prove that if two MDSs are spectrally isomorphic under ι such that ι and ι^{-1} map bounded functions into bounded functions and both are multiplicative on bounded functions, then the two MDSs are actually isomorphic. (Hint: conclude from multiplicativity that characteristic functions are mapped to characteristic functions.)

$*$ **8.** Prove that $C(K;\mathbb{R})$, K compact, is a Banach lattice.

$*$ **9.** Let (Ω,Σ,μ) be a measure space and let $1 \le p \le \infty$. Show that

$$|f| = \sup\{\operatorname{Re}(cf) : c \in \mathbb{T}\}$$

the supremum being taken in the order of $L^p(\Omega,\Sigma,\mu;\mathbb{R})$.

$*$ **10.** Let (Ω,Σ,μ) be a finite measure space. Show that the lattice

$$L^0(\Omega,\Sigma,\mu;[0,1])$$

is complete. (Hint: use Corollary 7.3.) Conclude (by setting up suitable "order isomorphisms") that the lattices

$$L^0(\Omega,\Sigma,\mu;\overline{\mathbb{R}}) \quad \text{and} \quad L^0(\Omega,\Sigma,\mu;[0,\infty])$$

are complete. Finally, prove that $L^\infty(\Omega,\Sigma,\mu;\mathbb{R})$ is an order complete Banach lattice, but its norm is not order continuous in general.

Lecture 8
The Mean Ergodic Theorem I

> One of the endlessly alluring aspects of mathematics is that its thorniest paradoxes have a way of blooming into beautiful theories.
>
> Philip J. Davis[1]

As was said at the beginning of Lecture 5, the reason for introducing measure-preserving dynamical systems is the intuition of a statistical equilibrium emergent from (very rapid) deterministic interactions of a multitude of particles. A measurement of the system can be considered as a random experiment, where repeated measurements appear to be independent since the time scale of measurements is far larger than the time scale of the internal dynamics.

In such a perspective, one expects that the (arithmetic) averages over the outcomes of these measurements ("time averages") should converge — in some sense or another — to a sort of "expected value". In mathematical terms, given an MDS $(\Omega, \Sigma, \mu; \varphi)$ and an "observable" $f : \Omega \longrightarrow \mathbb{R}$, the time averages take the form

$$\mathrm{A}_n f(\omega) := \frac{1}{n} \big(f(\omega) + f(\varphi(\omega)) + \cdots + f(\varphi^{n-1}(\omega)) \big) \tag{8.1}$$

if $\omega \in \Omega$ is the initial state of the system; and the expected value is the "space mean"

$$\int_\Omega f \,\mathrm{d}\mu$$

of f. In his original approach, Boltzmann assumed the so-called *"Ergodenhypothese"*, which allowed him to prove the convergence $\mathrm{A}_n f(\omega) \to \int_\Omega f \,\mathrm{d}\mu$ ("time mean equals space mean", cf. Lecture 1). However, the Ehrenfests in [Ehrenfest (1912)] doubted that this "Ergodenhypothese" is ever satisfied, a conjecture that was confirmed independently by Rosenthal and Plancherel only a few years later (see [Brush (1971)]). After some 20 more years, John von Neumann and George D. Birkhoff made a major step forward by separating the question of convergence of the averages $\mathrm{A}_n f$ from the question whether the limit is the space mean of

[1] Scientific American, 211, (Sept. 1964), pp. 51–59.

f or not. Their results — the *"Mean Ergodic Theorem"* and the *"Individual Ergodic Theorem"* — roughly state that under reasonable conditions on f the time averages always converge in some sense, while the limit is the expected "space mean" if and only if the system is ergodic (in our terminology). These theorems gave birth to Ergodic Theory as a mathematical discipline.

The present and the following two lectures are devoted to these fundamental results, starting with von Neumann's theorem. The linear operators induced by dynamical systems and studied in the previous lectures will be the main protagonists now.

8.1 Von Neumann's Mean Ergodic Theorem

Let $(\Omega,\Sigma,\mu;\varphi)$ be an MDS and let $T = T_\varphi$ be the induced operator. Note that the time mean of a function f under the first n iterates of T (8.1) can be written as

$$\mathrm{A}_n f = \frac{1}{n}\left(f + f\circ\varphi + \cdots + f\circ\varphi^{n-1}\right) = \frac{1}{n}\sum\nolimits_{j=0}^{n-1} T^j f.$$

Von Neumann's theorem deals with the averages $\mathrm{A}_n f$ for f from the Hilbert space $\mathrm{L}^2(\Omega,\Sigma,\mu)$.

Theorem 8.1 (von Neumann, 1931). *Let $(\Omega,\Sigma,\mu;\varphi)$ be an MDS and consider the induced operator $T := T_\varphi$. For each $f \in \mathrm{L}^2(\Omega,\Sigma,\mu)$ the limit*

$$\lim_{n\to\infty} \mathrm{A}_n f = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} T^j f$$

exists in the L^2*-sense, and is a fixed point of* T.

Proof. Let us abbreviate $H := \mathrm{L}^2(\Omega,\Sigma,\mu)$ and write $\left(\cdot\,|\,\cdot\right)$ for the scalar product in H. We consider the induced operator T as an operator on H and write $\mathrm{Fix}(T) = \{f \in H : Tf = f\}$ (cf. Proposition 7.8, $p = 2$). If $f \in \mathrm{Fix}(T)$ then

$$\mathrm{A}_n f = \frac{1}{n}\sum_{j=0}^{n-1} T^j f = \frac{1}{n}\sum_{j=0}^{n-1} f = f$$

for each $n \in \mathbb{N}$. Hence the sequence $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ (trivially) converges to f, which is a fixed a point of T. On the other hand, note the identity

$$(\mathrm{I}-T)\mathrm{A}_n f = \mathrm{A}_n(\mathrm{I}-T)f = \frac{1}{n}(f - T^n f), \tag{8.2}$$

valid for all $f \in H$. The operator T is a contraction (even an isometry) and hence $(1/n)(f - T^n f) \to 0$. This shows that

$$\mathrm{A}_n f \to 0 \quad \text{for all} \quad f \in \operatorname{ran}(\mathrm{I} - T).$$

Since $\|\mathrm{A}_n\| \le \frac{1}{n}\sum_{j=0}^{n-1}\left\|T^j\right\| \le 1$ for every n, we conclude by a standard argument from functional analysis (see Proposition C.16) that

$$\mathrm{A}_n f \to 0 \qquad \text{for all} \quad f \in \overline{\operatorname{ran}}(\mathrm{I} - T).$$

In the final step we show that

$$H = \operatorname{Fix}(T) \oplus \overline{\operatorname{ran}}(\mathrm{I} - T) \tag{8.3}$$

is an orthogonal decomposition. Having established this it follows that

$$\lim_{n\to\infty} \mathrm{A}_n f = Pf,$$

where $P : H \longrightarrow \operatorname{Fix}(T)$ is the orthogonal projection onto $\operatorname{Fix}(T)$. To prove (8.3), note first that this sum is direct since for $f \in \operatorname{Fix}(T) \cap \overline{\operatorname{ran}}(\mathrm{I} - T)$ we have $f = \mathrm{A}_n f \to 0$ as $n \to \infty$. Take now $f \in H$ with $0 \ne f \perp \operatorname{ran}(\mathrm{I} - T)$. Then $\big(f \,|\, f - Tf\big) = 0$ and hence $\big(f \,|\, Tf\big) = \big(f \,|\, f\big) = \|f\|^2$. This implies that

$$\|Tf - f\|^2 = \|Tf\|^2 - 2\operatorname{Re}\big(f \,|\, Tf\big) + \|f\|^2 = \|Tf\|^2 - \|f\|^2 \le 0$$

since T is a contraction. Consequently, $f = Tf$, i.e., $f \in \operatorname{Fix}(T)$. Hence we have proved that

$$\operatorname{ran}(\mathrm{I} - T)^\perp \subset \operatorname{Fix}(T),$$

but since $\overline{\operatorname{ran}}(\mathrm{I} - T) \cap \operatorname{Fix}(T) = \{0\}$, we obtain $\operatorname{ran}(\mathrm{I} - T)^\perp = \operatorname{Fix}(T)$ as claimed. □

The proof above yields the additional information that

$$Pf := \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} T^j f \qquad (f \in \mathrm{L}^2(\Omega, \Sigma, \mu))$$

is the orthogonal projection of f onto the fixed space

$$\operatorname{Fix}(P) = \operatorname{ran}(P) = \operatorname{Fix}(T).$$

Using more properties of T and $\operatorname{Fix}(T)$ we obtain even more information about P.

Corollary 8.2. *With hypotheses and notation as above, P is a positive projection onto the sublattice* $\operatorname{Fix}(T)$*, satisfying*

$$\int_\Omega (Pf)\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu \qquad (f \in \mathrm{L}^2(\Omega, \Sigma, \mu)) \tag{8.4}$$

and

$$P(f \cdot g) = f \cdot Pg \qquad (f \in \mathrm{L}^\infty(\Omega, \Sigma, \mu) \cap \operatorname{Fix}(P),\, g \in \mathrm{L}^2(\Omega, \Sigma, \mu)). \tag{8.5}$$

Proof. Since T is positive, so are A_n and hence $P = \lim_n \mathrm{A}_n$. Since $|Tf| = T|f|$ for each f, $\mathrm{Fix}(T)$ is a sublattice of $\mathrm{L}^2(\Omega,\Sigma,\mu)$. To prove (8.4), note that this is true for T in place of P since μ is φ-invariant. Hence it must be true for A_n in place of P and thus also for P. If $f \in \mathrm{Fix}(T) \cap \mathrm{L}^\infty$ and $g \in \mathrm{L}^2$, then $T(f \cdot g) = (Tf)\cdot(Tg) = f \cdot Tg$. Iterating this leads to $T^j(f \cdot g) = f \cdot T^j g$ for every $j \geq 0$ and hence $\mathrm{A}_n(f\cdot g) = f \cdot \mathrm{A}_n g$ for all $n \in \mathbb{N}$. Letting $n \to \infty$ yields (8.5). □

Remark 8.3. 1) Here is a different view on the last part of Corollary 8.2: The space $\mathscr{A} := \mathrm{L}^\infty(\Omega,\Sigma,\mu) \cap \mathrm{Fix}(T)$ is a closed C^*-subalgebra of $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$. Then $H = \mathrm{L}^2(\Omega,\Sigma,\mu)$ is a module over $\mathscr{A}$. A bounded operator S on H is an **$\mathscr{A}$-module homomorphism** if and only if $S(f \cdot g) = f \cdot Sg$ for all $f \in H, g \in \mathscr{A}$. The set of all bounded $\mathscr{A}$-module homomorphisms $\mathscr{L}_{\mathscr{A}}(H)$ is a subalgebra of $\mathscr{L}(H)$, closed with respect to the strong operator topology. So Corollary 8.2 tells $P \in \mathscr{L}_{\mathscr{A}}(H)$.

2) Here comes a probabilistic view on the projection P. Define

$$\Sigma_\varphi := \mathrm{Fix}(\varphi^*) := \{A \in \Sigma : \varphi^* A = A\} = \{A \in \Sigma : \mathbf{1}_A \in \mathrm{Fix}(T)\}.$$

This is obviously a sub-σ-algebra of Σ, called the **φ-invariant σ-algebra**. We claim that

$$\mathrm{Fix}(T) = \mathrm{L}^2(\Omega,\Sigma_\varphi,\mu).$$

Indeed, since step functions are dense in L^2, every $f \in \mathrm{L}^2(\Omega,\Sigma_\varphi,\mu)$ is contained in $\mathrm{Fix}(T)$. On the other hand, if $f \in \mathrm{Fix}(T)$, then for every Borel set $B \subset \mathbb{C}$ one has

$$\varphi^*[f \in B] = [\varphi \in [f \in B]] = [f \circ \varphi \in B] = \left[T_\varphi f \in B\right] = [f \in B],$$

which shows that $[f \in B] \in \Sigma_\varphi$. Hence f is Σ_φ-measurable. Now, by taking $A \in \Sigma_\varphi$ and $f := \mathbf{1}_A$ in (8.4) we obtain using (8.5)

$$\int_\Omega \mathbf{1}_A Pf \,\mathrm{d}\mu = \int_\Omega P(\mathbf{1}_A \cdot f)\,\mathrm{d}\mu = \int_\Omega \mathbf{1}_A \cdot f \,\mathrm{d}\mu.$$

This shows that $Pf = \mathbb{E}(f \,|\, \Sigma_\varphi)$ is the *conditional expectation* of f with respect to the φ-invariant σ-algebra Σ_φ.

As already mentioned, convergence to the "expected value" is characterised by ergodicity.

Corollary 8.4. *For an MDS $(\Omega,\Sigma,\mu;\varphi)$ the following assertions are equivalent.*

(i) *The MDS $(\Omega,\Sigma,\mu;\varphi)$ is ergodic.*

(ii) $\lim_{n\to\infty} \mathrm{A}_n f = \left(\int_\Omega f \,\mathrm{d}\mu\right) \cdot \mathbf{1}$ *for each* $f \in \mathrm{L}^2(\Omega,\Sigma,\mu)$.

Proof. By von Neumann's theorem, (ii) is equivalent to $\dim \mathrm{Fix}(T) = 1$ (as constants always belong to $\mathrm{Fix}(T)$) and this characterises (i) by Proposition 7.8.

Remark 8.5. Theorem 8.1 is called “Mean Ergodic Theorem” because the convergence is understood with respect to the L^2-norm, that is in “square mean”. Corollary 8.4 states thus that “time mean in mean” equals “space mean” for an ergodic MDS. The actual Ergodic Hypothesis “time mean equals space mean” from Lecture 1 aims at pointwise (a.e.) convergence. Note that this cannot be inferred from von Neumann’s theorem, since L^2-convergence does not imply pointwise convergence (a.e.). We shall return to this problem in Lecture 10 where we discuss Birkhoff’s “Individual Ergodic Theorem”.

8.2 Mean Ergodic Operators on Banach Spaces

The above proof of von Neumann’s theorem does not use the characteristic properties of the induced operator $T = T_\varphi$. Indeed, a closer look reveals that only the contractivity of T was actually used. So with the same proof one obtains that for every contraction T on a Hilbert space H the **Cesàro averages**

$$\mathrm{A}_n := \mathrm{A}_n[T] := \frac{1}{n}\sum_{j=0}^{n-1} T^j \tag{8.6}$$

converge in the strong operator topology to a (orthogonal) projection onto the *fixed space* $\mathrm{Fix}(T) = \ker(\mathrm{I} - T)$.

We shall now leave MDSs for a while to study this remarkable property in the context of a bounded linear operator T on an arbitrary Banach space X.

Definition 8.6. A bounded linear operator $T \in \mathscr{L}(X)$ on Banach space X is called **mean ergodic** if the sequence of its Cesàro averages $\mathrm{A}_n = \mathrm{A}_n[T]$ (defined as in (8.6)) converges in the strong operator topology, i.e., if

$$\lim_{n\to\infty} \mathrm{A}_n x = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} T^j x$$

exists for each $x \in X$.

The simple identity

$$(\mathrm{I} - T)\mathrm{A}_n = \mathrm{A}_n(\mathrm{I} - T) = \frac{1}{n}(\mathrm{I} - T^n) \tag{8.7}$$

was already used in the proof of von Neummann’s theorem (see (8.2)) and leads to a surprising analytic property. For simplicity we shall confine ourselves to the case that the operator T is **power bounded**, i.e., $\sup_{n\geq 0} \|T^n\| < \infty$.

Lemma 8.7. *Let X be a Banach space, let $T \in \mathscr{L}(X)$ be a power bounded operator, and let $Z \subset X'$ be a subspace of X' such that Z separates the points of X and $T'Z \subset Z$.*

If $x \in X$ and y is a $\sigma(X,Z)$-cluster point[2] *of the sequence $(\mathrm{A}_n x)_{n\in\mathbb{N}}$, then $Ty = y$, i.e., $y \in \mathrm{Fix}(T)$.*

Proof. Since $\mathrm{A}_n T = T\mathrm{A}_n$ we can write

$$y - Ty = (\mathrm{I}-T)(y-\mathrm{A}_n x) + \mathrm{A}_n(\mathrm{I}-T)x = (\mathrm{I}-T)(y-\mathrm{A}_n x) + \frac{1}{n}(\mathrm{I}-T^n)x$$

for every $n \in \mathbb{N}$. Since T is power bounded, $(1/n)(x - T^n x) \to 0$ in norm, and thus in the $\sigma(X,Z)$-topology. Fix $x' \in Z$ and let $\varepsilon > 0$. We choose N such that

$$\left|(1/n)\left\langle x - T^n x, x'\right\rangle\right| < \varepsilon/2$$

for all $n \geq N$. Since y is a $\sigma(X,Z)$-cluster point of $(\mathrm{A}_n x)_n$ and $x' - T'x' \in Z$, we find some $n \geq N$ such that

$$\left|\left\langle(\mathrm{I}-T)(y-\mathrm{A}_n x), x'\right\rangle\right| = \left|\left\langle y - \mathrm{A}_n x, x' - T'x'\right\rangle\right| < \varepsilon/2.$$

Then clearly $|\langle y - Ty, x'\rangle| < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, $\langle y - Ty, x'\rangle = 0$, and since Z separates the points of X, $y = Ty$. □

We now use Lemma 8.7 to characterise power bounded mean ergodic operators by quite different properties.

Theorem 8.8 (Mean Ergodic Operators). *Let T be a power bounded operator on some Banach space X. Then the following assertions are equivalent.*

(i) *T is mean ergodic.*

(ii) *T is weakly mean ergodic, i.e., weak-$\lim_{n\to\infty} \mathrm{A}_n x$ exists for each $x \in X$.*

(iii) *The sequence $(\mathrm{A}_n x)_{n\in\mathbb{N}}$ has a weak cluster point for each $x \in X$.*

(iv) *$\overline{\mathrm{conv}}\{T^j x : j \in \mathbb{N}_0\} \cap \mathrm{Fix}(T) \neq \emptyset$ for each $x \in X$.*

(v) *$\mathrm{Fix}(T)$ separates $\mathrm{Fix}(T')$, i.e., for $0 \neq x' \in \mathrm{Fix}(T')$ there exists $x \in \mathrm{Fix}(T)$ such that $\langle x, x'\rangle \neq 0$.*

(vi) *$X = \mathrm{Fix}(T) \oplus \overline{\mathrm{ran}}(I-T)$.*

Proof. The implications (i)$\Longrightarrow$(ii)$\Longrightarrow$(iii) are trivial.

(iii)$\Longrightarrow$(iv): Let $x \in X$ and let y be a weak cluster point of the sequence $(\mathrm{A}_n x)_{n\in\mathbb{N}}$. Then $y \in \mathrm{Fix}(T)$, by Lemma 8.7 (with $Z = X'$). Obviously $\mathrm{A}_n x \in \mathrm{conv}\{T^j x : j \in \mathbb{N}_0\}$ for every $n \in \mathbb{N}$. Since $\overline{\mathrm{conv}}\{T^j x : j \in \mathbb{N}_0\}$ is weakly closed by Mazur's Theorem C.7, $y \in \overline{\mathrm{conv}}\{T^j x : j \in \mathbb{N}_0\}$.

(iv)$\Longrightarrow$(v): For $0 \neq x' \in \mathrm{Fix}(T')$ take $0 \neq x \in X$ such that $\langle x, x'\rangle = 1$. Since $T'x' = x'$,

$$\left\langle \sum_{j=0}^{n} t_j T^j x, x' \right\rangle = \left\langle x, \sum_{j=0}^{n} t_j T'^j x' \right\rangle = \langle x, x'\rangle = 1$$

[2] Recall that y is a cluster point of a sequence $(a_n)_{n\in\mathbb{N}}$ if $y \in \overline{\{a_k : k \geq n\}}$ for all $n \in \mathbb{N}$, i.e. if every neighbourhood of y contains a_n for infinitely many $n \in \mathbb{N}$. See also Appendix A.2.

for every choice of $n \in \mathbb{N}$, $0 \le t_j \le 1$, $\sum_{j=0}^{n} t_j = 1$. Therefore $\langle y, x' \rangle = 1$ for each $y \in \overline{\mathrm{conv}}\{T^j x : j \in \mathbb{N}_0\}$. By (iv) there is such a y that is also a T-fixed point.

(v)$\Longrightarrow$(vi): Consider the space $Y := \mathrm{Fix}(T) + \overline{\mathrm{ran}}(I - T)$. By Exercise 8.1, the sum is direct and Y is closed in X. If $x' \in X'$ vanishes on Y, then x' vanishes in particular on $\mathrm{ran}(\mathrm{I} - T)$ and this just means that $x' \in \mathrm{Fix}(T')$ (see Exercise 8.2). Moreover, x' vanishes on $\mathrm{Fix}(T)$, which — by assumption — separates $\mathrm{Fix}(T')$. This forces $x' = 0$. It follows from the Hahn–Banach theorem that Y is dense in X, but as Y is closed, we arrive at $Y = X$.

(vi)$\Longrightarrow$(i): As in the proof of von Neumann's theorem, $\mathrm{A}_n x = x$ for $x \in \mathrm{Fix}(T), n \in \mathbb{N}$, and $\mathrm{A}_n x \to 0$ for $x \in \overline{\mathrm{ran}}(\mathrm{I} - T)$. Since $X = \mathrm{Fix}(T) \oplus \overline{\mathrm{ran}}(\mathrm{I} - T)$ by hypothesis, T is mean ergodic. □

Remarks 8.9. a) A closer examination of the proof shows that Lemma 8.7 and Theorem 8.8 still hold if $(1/n)T^n \to 0$ strongly on X and T is merely **Cesàro bounded**, i.e., $\sup_n \|\mathrm{A}_n\| < \infty$.

On the other hand, *if* an operator T is mean ergodic, then it is Cesàro bounded and $(1/n)T^n \to 0$ strongly. The first is a consequence of the Principle of Uniform Boundedness (Theorem C.1), the second follows from the identity

$$\frac{n+1}{n}\mathrm{A}_{n+1} = \frac{1}{n}T^n + \mathrm{A}_n \qquad (n \in \mathbb{N}).$$

b) If an operator T is mean ergodic, then we can define

$$Px := P_T x := \lim_n \mathrm{A}_n x \qquad (x \in X).$$

By a) and Theorem 8.8, P is a projection onto $\mathrm{ran}(P) = \mathrm{Fix}(T)$ with kernel $\ker(P) = \overline{\mathrm{ran}}(\mathrm{I} - T)$. This projection is called the **mean ergodic projection** associated with P.

c) The condition (v) in Theorem 8.8 can also be expressed as $\mathrm{Fix}(T') \cap \mathrm{Fix}(T)^{\perp} = \{0\}$. Here, $Y^{\perp}$ denotes the **annihilator** of a set $Y \subset X$ defined as

$$Y^{\perp} := \{x' \in X' : \langle x, x' \rangle = 0 \text{ for all } x \in Y\}.$$

d) One can replace condition (iii) in Theorem 8.8 by the weaker condition

(iii′) *The sequence* $(\mathrm{A}_n x)_{n \in \mathbb{N}}$ *has a weak cluster point for every* $x \in D$*, where* D *is a dense subset of* X.

Indeed, (iii') implies (iv) for all $x \in D$ which again implies (v). (In the proof of (iv)$\Rightarrow$(v) just take $x \in D$).

e) If $T \in \mathscr{L}(X)$ is mean ergodic, its adjoint T' need not to be mean ergodic (see Exercise 8.8). However, the Cesàro-means $\mathrm{A}_n[T']$ converge in the weak*-topology to the projection $P_{T'} := P_T'$. This is Exercise 8.4.

8.3 Examples

As a first example, we restate the conclusion from (the proof of) von Neumann's theorem.

Example 8.10 (Hilbert Space Contractions). *Every contraction on a Hilbert space is mean ergodic and the associated mean ergodic projection is orthogonal.*

In reflexive spaces, norm bounded subsets are relatively weakly compact. This leads to the following large class of mean ergodic operators.

Example 8.11 (Reflexive Spaces). *Every power bounded operator on a reflexive Banach space is mean ergodic. In particular, every power bounded operator on a space* $\mathrm{L}^p(\Omega,\Sigma,\mu)$*,* $1 < p < \infty$*, is mean ergodic.*

Proof. By power boundedness of T the set $\{\mathrm{A}_n x : n \in \mathbb{N}\}$ is norm-bounded for each $x \in X$, hence relatively weakly compact, since X is reflexive. Therefore the sequence $(\mathrm{A}_n x)_n$ has a weak cluster point. □

Example 8.12 (Complete Contractions). Let (Ω,Σ,μ) be a finite measure space. An operator T on $\mathrm{L}^1(\Omega,\Sigma,\mu)$ is called a **Dunford–Schwartz operator** or a **complete contraction**, if

$$\|Tf\|_1 \le \|f\|_1 \quad \text{and} \quad \|Tg\|_\infty \le \|g\|_\infty \qquad (f \in \mathrm{L}^1,\, g \in \mathrm{L}^\infty).$$

If T *is a Dunford–Schwartz operator, then* T *is mean ergodic.*

Proof. Let $\mathrm{B} := \{f \in \mathrm{L}^\infty(\Omega,\Sigma,\mu) : \|f\|_\infty \le 1\}$. View L^∞ as the dual of L^1 and equip it with the $\sigma(\mathrm{L}^\infty,\mathrm{L}^1)$-topology. By the Banach–Alaoglu theorem, B is weak*-compact. Since the embedding $(\mathrm{L}^\infty,\sigma(\mathrm{L}^\infty,\mathrm{L}^1)) \subset (\mathrm{L}^1,\sigma(\mathrm{L}^1,\mathrm{L}^\infty))$ is (obviously) continuous, B is weakly compact in L^1. Since B is invariant under the Cesàro averages A_n, the sequence $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ has a weak cluster point for each $f \in \mathrm{B}$. Thus (iii′) from Remark 8.9.d) is satisfied with $D := \mathrm{L}^\infty = \bigcup_{c>0} c\mathrm{B}$. □

Example 8.13 (Rotations on the Torus). *Let* $a \in \mathbb{T}$ *and let* φ_a *be the rotation by* a *on* $\mathbb{T}$*. Then the induced operator* $T = T_{\varphi_a}$ *on* $\mathrm{C}(\mathbb{T})$ *is mean ergodic.*

Proof. Write $\mathrm{A}_n = \mathrm{A}_n[T]$ for the Cesàro averages of T. The linear hull of the functions $\chi_n : z \longmapsto z^n$, $n \in \mathbb{Z}$, is a dense subalgebra of $\mathrm{C}(\mathbb{T})$, by the Stone–Weierstrass Theorem 4.3. Since T is power-bounded and therefore Cesàro-bounded, it suffices to show that $\mathrm{A}_n\chi_m$ converges for every $m \in \mathbb{Z}$. Note that $T\chi_m = a^m\chi_m$ for $m \in \mathbb{N}$, hence if $a^m = 1$, then $\chi_m \in \mathrm{Fix}(T)$ and there is nothing to show. So suppose that $a^m \ne 1$. Then

$$\mathrm{A}_n\chi_m = \frac{1}{n}\sum_{j=0}^{n-1} T^j\chi_m = \left(\frac{1}{n}\sum_{j=0}^{n-1} a^{mj}\right)\chi_m = \frac{1}{n}\frac{1-a^{mn}}{1-a^m}\chi_m \to 0$$

as $n \to \infty$. (Compare with the final problem in Lecture 1.) □

The above is a special case of the more general fact, the mean ergodicity of the induced operator of group rotations, which we shall prove in a later lecture.

Example 8.14 (Group Rotations). Let G be a compact group, $g \in G$ and consider the rotation TDS $(G;\varphi_g)$ (see Example 2.8). The induced operator T_{φ_g} is mean ergodic on $\mathrm{C}(G)$.

Example 8.15 (Shift). For a scalar sequence $x = (x_n)_{n\in\mathbb{N}_0}$ the **left shift** L and the **right shift** R are defined by

$$L(x_0,x_1,\dots) := (x_1,x_2,\dots) \quad \text{and} \quad R(x_0,x_1,\dots) := (0,x_0,x_1,x_2,\dots).$$

Clearly, a fixed vector for L must be constant, while 0 is the only fixed vector of R.

a) *The left and the right shift are mean ergodic on each $X = \ell^p$, $1 < p < \infty$.*
b) *The left shift on $X = \ell^1$ is mean ergodic, while the right shift is not.*
c) *The left and right shift are mean ergodic on $X = \mathrm{c}_0$, the space of null sequences.*
d) *The left shift* is *while the right shift* is not *mean ergodic on $X = \mathrm{c}$, the space of convergent sequences.*
e) *Neither the left nor the right shift is mean ergodic on $X = \ell^\infty$.*

The proof of these statements is left as Exercise 8.7.

Supplement: Powers and Convex Combinations of Mean Ergodic Operators

Let us illustrate how the various conditions in Theorem 8.8 can be used to check mean ergodicity, and thus to carry out certain constructions for mean ergodic operators.

Theorem 8.16. *Let X be a Banach space and let $S \in \mathscr{L}(X)$ be a power bounded mean ergodic operator with bounded powers. Let T be a k^{th}-root of S, i.e., $T^k = S$ for some $k \in \mathbb{N}$. Then T is also mean ergodic.*

Proof. Denote by P_S the mean ergodic projection of S. Define $P := \bigl(\frac{1}{k}\sum_{j=0}^{k-1} T^j\bigr)P_S$ and observe that $Px \in \overline{\mathrm{conv}}\{T^j x : j \in \mathbb{N}_0\}$ for all $x \in X$. Since T commutes with S, it commutes also with P_S. We now obtain

$$PT = TP = \Bigl(\frac{1}{k}\sum_{j=0}^{k-1} T^{j+1}\Bigr)P_S = P,$$

since $T^k P_S = SP_S = P_S$, and $Px \in \overline{\mathrm{conv}}\{T^j x \,:\, j \in \mathbb{N}_0\} \cap \mathrm{Fix}(T)$ by the above. So Theorem 8.8 (iv) implies that T is mean ergodic. It follows also that P is the corresponding mean ergodic projection. $\square$

On the contrary, it is possible that no power of a mean ergodic operator is mean ergodic.

Example 8.17. Take c, the space of (complexe) convergent sequences, and the multiplication operator
$$M : \mathrm{c} \longrightarrow \mathrm{c}, \quad (x_n)_n \longmapsto (\alpha_n x_n)_n$$
with some sequence $\{\alpha_n\}_{n=1}^\infty$ with $1 \neq \alpha_n \longrightarrow 1$. Now $\mathrm{Fix}(M) = \{0\}$, whereas $\mathrm{Fix}(M')$ contains x' defined by $x'((x_n)_{n=1}^\infty) := \lim_{n\to\infty} x_n$. By Theorem 8.8 (v) we conclude that M is not mean ergodic. Consider now a k^{th} root of unity $1 \neq a \in \mathbb{T}$ and define $T_k := aM$. Then it is easy to see that $\mathrm{Fix}(T_k') = \{0\}$, and hence, again by Theorem 8.8 (v), T_k is mean ergodic. It follows from the above that T_k^k is not mean ergodic. Now one can use a direct product construction to obtain a Banach space X and a mean ergodic operator $T \in \mathscr{L}(X)$ with no power T^k, $k \geq 2$, being mean ergodic (Exercise 8.10).

Convex Combinations of Mean Ergodic Operators

Other examples of "new" mean ergodic operators can be obtained by convex combinations of mean ergodic operators. Our first lemma, a nice application of the Kreĭn–Milman Theorem, is due to Kakutani.

Lemma 8.18. *Let X be a Banach space. Then the identity operator* I *is an extreme point of the closed unit ball in $\mathscr{L}(X)$.*

Proof. Take $T \in \mathscr{L}(X)$ such that $\|\mathrm{I}+T\| \leq 1$ and $\|\mathrm{I}-T\| \leq 1$. Then the same is true for the adjoints: $\|\mathrm{I}'+T'\| \leq 1$ and $\|\mathrm{I}'-T'\| \leq 1$. For $x' \in E'$ define $x_1' := (\mathrm{I}'+T')x'$, $x_2' := (\mathrm{I}'-T')x'$ and conclude that $x' = \frac{1}{2}(x_1'+x_2')$ and $\|x_1'\|, \|x_2'\| \leq \|x'\|$. Suppose now that x' is an extreme point of the unit ball in X'. Then we obtain $x' = x_1' = x_2'$ and hence $T'x' = 0$. The Banach–Alaoglu Theorem C.4 and the Kreĭn–Milman Theorem C.13 implies $T' = 0$, and hence $T = 0$.

Now assume that $\mathrm{I} = \frac{1}{2}(R+S)$ for contractions $R, S \in \mathscr{L}(X)$, and define $T := \mathrm{I} - R$. This implies $\mathrm{I} - T = R$ and $\mathrm{I} + T = 2\mathrm{I} - R = S$, so $\|\mathrm{I}-T\| \leq 1$, $\|\mathrm{I}+T\| \leq 1$. By the above considerations it follows that $T = 0$, i.e., $\mathrm{I} = R = S$. □

Lemma 8.19. *Let R, S be two power bounded, commuting operators on a Banach space X. Consider*
$$T := tR + (1-t)S$$
for some $0 < t < 1$. Then for the fixed spaces $\mathrm{Fix}(T)$, $\mathrm{Fix}(R)$ and $\mathrm{Fix}(S)$ we have
$$\mathrm{Fix}(T) = \mathrm{Fix}(R) \cap \mathrm{Fix}(S).$$

Proof. Only the inclusion $\mathrm{Fix}(T) \subset \mathrm{Fix}(R) \cap \mathrm{Fix}(S)$ is not obvious. Endow X with an equivalent norm $\|x\|_1 := \sup\{\|R^n S^m x\| : n, m \in \mathbb{N}_0\}$, $x \in X$, and observe that R and S become now contractive. From the definition of T we obtain

$$\mathrm{I}_{\mathrm{Fix}(T)} = T|_{\mathrm{Fix}(T)} = tR|_{\mathrm{Fix}(T)} + (1-t)S|_{\mathrm{Fix}(T)}$$

and $R|_{\mathrm{Fix}(T)}, S|_{\mathrm{Fix}(T)} \in \mathscr{L}(\mathrm{Fix}(T))$, since R and S commute. Lemma 8.18 implies $R|_{\mathrm{Fix}(T)} = S|_{\mathrm{Fix}(T)} = \mathrm{I}_{\mathrm{Fix}(T)}$, i.e., $\mathrm{Fix}(T) \subset \mathrm{Fix}(R) \cap \mathrm{Fix}(S)$. $\square$

Now we can prove the main result of this section.

Theorem 8.20. *Let X be a Banach space and R,S two power bounded, mean ergodic, commuting operators on X. Then for any $0 \le t \le 1$ also the convex combination $T := tR + (1-t)S$ is mean ergodic.*

Proof. Let $0 < t < 1$. By Lemma 8.19 we have $\mathrm{Fix}(T) = \mathrm{Fix}(R) \cap \mathrm{Fix}(S)$ and $\mathrm{Fix}(T') = \mathrm{Fix}(R') \cap \mathrm{Fix}(S')$. By Theorem 8.8 (v) it suffices to show that $\mathrm{Fix}(R) \cap \mathrm{Fix}(S)$ separates $\mathrm{Fix}(R') \cap \mathrm{Fix}(S')$. To this end pick $0 \neq x' \in \mathrm{Fix}(R') \cap \mathrm{Fix}(S')$. Then there is $x \in \mathrm{Fix}(R)$ with $\langle x, x'\rangle \neq 0$. Since $S\,\mathrm{Fix}(R) \subseteq \mathrm{Fix}(R)$ we have $P_S x \in \mathrm{Fix}(R) \cap \mathrm{Fix}(S)$ where P_S denotes the mean ergodic projection corresponding to S. Consequently, we have (see also Remark 8.9.e)))

$$\langle P_S x, x'\rangle = \langle x, P_S' x'\rangle = \langle x, P_{S'} x'\rangle = \langle x, x'\rangle \neq 0. \qquad \square$$

The following are immediate consequences of the above.

Corollary 8.21. *Let T be a convex combination of two power bounded, mean ergodic, commuting operators R and $S \in \mathscr{L}(X)$. Denote by P_R, resp. P_S the mean ergodic projections. Then the projection P_T corresponding to T is obtained as*

$$P_T = P_R P_S = P_S P_R = \lim_{n\to\infty} (R_n S_n).$$

Corollary 8.22. *Let $\{R_j\}_{j=1}^m$ be a family of power bounded, mean ergodic, commuting operators. Then every convex combination $T := \sum_{j=1}^m t_j R_j$ is mean ergodic.*

Exercises

1. Let T be a bounded linear operator on a Banach space X and let A_n be its n^{th} Cesàro average, $n \in \mathbb{N}$. Suppose that

$$\sup_n \|\mathrm{A}_n\| < \infty \quad \text{and} \quad \frac{1}{n} T^n x \to 0 \qquad (x \in X).$$

Show that

$$Y := \{x \in X : \; Px := \lim_{n\to\infty} \mathrm{A}_n x \text{ exists}\}$$

is a closed subspace of X and that $P \in \mathscr{L}(Y)$. Show further that $\operatorname{ran} P = \operatorname{Fix} T$, $\ker P = \overline{\operatorname{ran}}(I-T)$, P is a projection and thereby $Y = \operatorname{Fix} T \oplus \overline{\operatorname{ran}}(I-T)$.

2. Let X be a Banach space and $T \in \mathscr{L}(X)$. Show that

$$\operatorname{ran}(\mathrm{I}-T)^{\perp} = \operatorname{Fix}(T') \quad \text{and} \quad \operatorname{ran}(\mathrm{I}-T')^{\top} = \operatorname{Fix}(T).$$

Here, $Y'^{\top} := \{x \in X : \langle x, x' \rangle = 0 \text{ for all } x' \in Y'\}$ for $Y' \subset X'$.

3. Under the conditions of Exercise 8.1, show that $\operatorname{Fix}(T')$ always separates the points of $\operatorname{Fix}(T)$. (Hint: Take $0 \neq x \in \operatorname{Fix}(T)$ and consider the set $K := \{x' \in X' : \|x'\| \leq 1, \langle x, x' \rangle = \|x\|\}$. Then K is not empty by the Hahn–Banach theorem, $\sigma(X', X)$-closed and norm-bounded. Use Lemma 8.7 to show that $\operatorname{Fix}(T') \cap K \neq \emptyset$).

4. Let T be a mean ergodic operator on a Banach space X with associated projection $P : X \longrightarrow \operatorname{Fix}(T)$. Show that $\mathrm{A}'_n \to P'$ in the weak*-topology and that P' is a projection with $\operatorname{ran}(P') = \operatorname{Fix}(T')$.

5. Let (Ω, Σ, μ) be a finite measure space and let T be a *positive* operator on $\mathrm{L}^1(\Omega, \Sigma, \mu)$. Show that the following assertions are equivalent.

(i) T is a complete contraction, i.e., $\|Tf\|_\infty \leq \|f\|_\infty$ and $\|Tf\|_1 \leq \|f\|_1$ for all $f \in \mathrm{L}^\infty$.

(ii) $T\mathbf{1} \leq \mathbf{1}$ and $T'\mathbf{1} \leq \mathbf{1}$.

6. Show that the following operators are *not* mean ergodic:

a) $X = \mathrm{C}[0,1]$ and $(Tf)(x) = f(x^2)$, $x \in [0,1]$. (Hint: determine $\operatorname{Fix}(T)$ and $\operatorname{Fix}(T')$, cf. Lecture 3.)

b) $X = \mathrm{C}[0,1]$ and $(Tf)(x) = xf(x)$, $x \in [0,1]$. (Hint: look at $\mathrm{A}_n\mathbf{1}$, $n \in \mathbb{N}$.)

7. Prove the assertions a)–e) in Example 8.15.

8. Let X be a Banach space and $T \in \mathscr{L}(X)$ an operator whose adjoint T' is mean ergodic on X'. Show that T is mean ergodic. Give a counterexample for the converse implication.

9. Prove the assertions below. If not otherwise specified, X denotes a general Banach space.

a) A linear operator T on $X = \mathbb{C}$ is mean ergodic if and only if $\|T\| \leq 1$.

b) Let $T \in \mathscr{L}(X)$ be a *periodic operator*, i.e., $T^{n_0} = \mathrm{I}$ for some $n_0 \in \mathbb{N}$. Then T is mean ergodic with projection

$$P = \frac{1}{n_0} \sum_{i=0}^{n_0-1} T^i.$$

c) Assume that $T \in \mathscr{L}(X)$ has spectral radius $r(T) < 1$. Then T is mean ergodic with projection $P = 0$.

d) For the spectral radius of a mean ergodic operator T we have $r(T) \leq 1$.

e) Let $X = \mathbb{R}^k$ and T be a mean ergodic operator. Prove that it is power bounded.

10. Work out the details of Example 8.17.

Lecture 9
The Mean Ergodic Theorem II

The pendulum of mathematics swings back and forth towards abstraction and away from it with a timing that remains to be estimated.

Gian-Carlo Rota[1]

After the excursion into general Banach spaces in the previous lecture we now return to dynamical systems and apply the concept of mean ergodicity to the induced operators.

9.1 More on Ergodicity

Recall that *ergodicity* is a property of an MDS $(\Omega, \Sigma, \mu; \varphi)$ that can be characterised by

$$\mathrm{Fix}(T) = \langle \mathbf{1} \rangle$$

for the induced operator $T = T_\varphi$ in $\mathrm{L}^p(\Omega, \Sigma, \mu)$, $1 \leq p \leq \infty$ (Proposition 7.8). We have seen examples of MDSs which are ergodic (Bernoulli shift, irrational (minimal) rotations on the torus) and non-ergodic (rational (not minimal) rotations on the torus). On the other hand, *mean ergodicity* is a property of the induced operator T and it is always satisfied when T is considered on $\mathrm{L}^p(\Omega, \Sigma, \mu)$, $1 \leq p < \infty$ (Examples 8.11 and 8.12). This means that its Cesàro averages $\mathrm{A}_n := \frac{1}{n} \sum_{j=0}^{n-1} T^j$ converge to a projection onto $\mathrm{Fix}(T)$. Therefore, the MDS is ergodic if and only if the corresponding mean ergodic projection $P = \lim_{n\to\infty} \mathrm{A}_n$ is one-dimensional. In the following proposition this is expressed in various ways.

Proposition 9.1. *Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS and let $1 \leq p < \infty$. Then the induced operator $T = T_\varphi$ on $X = \mathrm{L}^p(\Omega, \Sigma, \mu)$ is mean ergodic and the following assertions are equivalent.*

[1] Indiscrete Thoughts, Birkhäuser Verlag Boston, 1997, p. 30.

(i) *The MDS* $(\Omega,\Sigma,\mu;\varphi)$ *is ergodic.*

(ii) $\displaystyle\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}T^j f = \left(\int_\Omega f\,\mathrm{d}\mu\right)\mathbf{1}$ *for every* $f\in \mathrm{L}^p(\Omega,\Sigma,\mu)$.

(iii) $\displaystyle\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\int_\Omega (f\circ\varphi^j)\cdot g\,\mathrm{d}\mu = \left(\int_\Omega f\,\mathrm{d}\mu\right)\left(\int_\Omega g\,\mathrm{d}\mu\right)$
for all $f\in \mathrm{L}^p(\Omega,\Sigma,\mu)$, $g\in \mathrm{L}^q(\Omega,\Sigma,\mu)$ *with* $\frac{1}{p}+\frac{1}{q}=1$.

(iv) $\displaystyle\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\mu(A\cap\varphi^{*j}(B)) = \mu(A)\mu(B)$ *for all* $A,B\in\Sigma$.

(v) $\displaystyle\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\mu(A\cap\varphi^{*j}(A)) = \mu(A)^2$ *for all* $A\in\Sigma$.

Proof. As noted above, the induced operator $T=T_\varphi$ is mean ergodic. Moreover, since φ is measure preserving, the corresponding mean ergodic projection P satisfies

$$\int_\Omega f\,\mathrm{d}\mu = \int_\Omega Tf\,\mathrm{d}\mu = \int_\Omega \mathrm{A}_n f\,\mathrm{d}\mu = \int_\Omega Pf\,\mathrm{d}\mu \quad \text{for every} \quad f\in \mathrm{L}^p(\Omega,\Sigma,\mu),$$

and $(\Omega,\Sigma,\mu;\varphi)$ is ergodic if and only if

$$Pf = \left(\int_\Omega f\mathrm{d}\mu\right)\mathbf{1} \quad \text{for all} \quad f\in \mathrm{L}^p(\Omega,\Sigma,\mu).$$

The above equivalences are more or less explicit reformulations of this insight.

(i)⇒(ii) follows from $\mathrm{Fix}(T)=\langle\mathbf{1}\rangle$ for ergodic systems, and (ii)⇒(iii) is straightforward by the duality of L^p-spaces. To obtain (iii)⇒(iv) just take $f:=\mathbf{1}_A$ and $g:=\mathbf{1}_B$. The implication (iv)⇒(v) is trivial.

To prove the implication (v)⇒(i) assume that A is φ-invariant, i.e., $\varphi^*(A)=A$ in the measure algebra. Then we have

$$\frac{1}{n}\sum_{j=0}^{n-1}\mu(A\cap\varphi^{*j}(A)) = \frac{1}{n}\sum_{j=0}^{n-1}\mu(A) = \mu(A)$$

and hence $\mu(A)^2=\mu(A)$ by (v). Therefore $\mu(A)\in\{0,1\}$, and (i) is proved. □

Remarks 9.2. a) Assertion (ii) in probabilistic language is a kind of "weak law of large numbers". To make this statement precise fix $f\in\mathrm{L}^p$ and let $X_j := f\circ\varphi^j$, $j\in\mathbb{N}_0$. Then the X_j are identically distributed random variables on the probability space (Ω,Σ,μ) with common expectation $\mathbb{E}(X_j)=\int_\Omega f\,\mathrm{d}\mu =: c$ for all j. Property (ii) just says that

$$\frac{X_0+\cdots+X_{n-1}}{n}\to c \quad \text{as} \quad n\to\infty$$

in L^p-norm. Note that this is slightly stronger than the "classical" weak law of large numbers that asserts convergence in measure (probability) only (see [Billingsley (1979)]). We shall come back to this in the next lecture.

b) Suppose that $\mu(A) > 0$. Then dividing by $\mu(A)$ in (iv) yields

$$\frac{1}{n}\sum_{j=0}^{n-1} \mu_A(\varphi^j \in B) \to \mu(B) \qquad (n \to \infty)$$

for every $B \in \Sigma$, where μ_A is the conditional probability given A (cf. Section 6.2). This shows that in an ergodic system the original measure μ is completely determined by μ_A.

c) By carrying out a standard density argument, one can replace "for all $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$" in assertion (ii) by "for all f in a dense subset of $\mathrm{L}^p(\Omega,\Sigma,\mu)$". The same holds for f and g in assertion (iii). Similarly, it suffices to take in (iv) and (v) sets from a dense subalgebra of Σ (more precisely of $\Sigma/\!\sim$; see Lecture 6.1 and in particular the proof of Proposition 6.12).

d) A stronger convergence property for the expression $\mu(A \cap \varphi^{*j}(A))$ appeared for the Bernoulli shifts in Proposition 6.12. In a later lecture, this will lead to the concept of *mixing*, a property stronger than ergodicity.

Recall that, by Proposition 6.12, Bernoulli shifts are always ergodic. We are now able to characterise ergodicity of Markov shifts, announced in Proposition 6.13.

Ergodicity of (irreducible) Markov shifts

Let $L = \{0,\dots,k-1\}$ and let $S = (s_{ij})_{0\le i,j<k}$ be a row-stochastic matrix. In Section 5.1 it was claimed that there is a probability (row) vector p such that $pS = p$. With the Mean Ergodic Theorem at hand this is now easy to establish. Namely, the operator $S : x \longmapsto xS$ is a contraction on $\mathbb{C}^k$ (considered as row vectors) with respect to the 1-norm and hence power-bounded (cf. Example 2.3). In Exercise 8.1 it was shown that the sum $\ker(\mathrm{I}-S)+\operatorname{ran}(\mathrm{I}-S)$ is direct, hence the dimension formula from elementary linear algebra yields that $\mathbb{C}^k = \ker(\mathrm{I}-S)\oplus\operatorname{ran}(\mathrm{I}-S)$. By (iv) of Theorem 8.8, S is mean ergodic[2], that is,

$$Q := \lim_n \frac{1}{n}\sum_{j=0}^{n-1} S^j$$

exists. Since S is row-stochastic, every power S^j is row-stochastic as well, whence also Q is row-stochastic. Since $QS = Q$, every row of Q is a left fixed probability vector of S.

[2] Alternatively, one can use Theorem 8.8 (iii) and the relative compactness of bounded sets in finite-dimensional spaces.

Let us take such a fixed probability vector p and consider the corresponding Markov shift $(\mathscr{W}_k^+, \Sigma, \mu(S,p); \tau)$ on $\mathscr{W}_k^+ = L^{\mathbb{N}_0}$. Our aim is to prove that this MDS is ergodic if and only if S is **irreducible**. Irreducibility of the matrix S means that for every pair of indices $(i,j) \in L \times L$ there is an $r \in \mathbb{N}_0$ such that the entry $s_{ij}^r := [S^r]_{ij}$ of the power S^r is strictly positive (cf. also the Proposition succeeding Example 2.22). Equivalently, there is $n \in \mathbb{N}$ such that $R := (\mathrm{I} + S)^n$ is **strictly positive**, i.e., has only strictly positive entries[3].

Lemma 9.3. *The matrix S is irreducible if and only if Q (defined above) is strictly positive. In this case every row of Q is equal to p.*

Proof. Suppose that S is irreducible and let R be defined as above. Since $QS = Q$, we have $QR = Q(\mathrm{I}+S)^n = 2^n Q$ and QR is strictly positive, since so is R. Hence also Q is strictly positive. Conversely, if Q is strictly positive, then for large n the Cesàro mean $\mathrm{A}_n[S]$ must be strictly positive, and hence S is irreducible.

It remains to show that Q has rank 1. If y is a *column* of Q, then $Qy = y$ (Q is a projection onto its column-space). Let $\varepsilon := \min_j y_j$ and $x := y - \varepsilon \mathbf{1}$. Then $x \geq 0$ and $Qx = x$, but since Q is strictly positive and x is positive, either $x = 0$ or x is also strictly positive. The latter is impossible by construction of x, and hence $x = 0$. This means that all entries of y are equal. Since y was an arbitrary column of Q, the claim is proved. □

By the lemma, an irreducible row-stochastic matrix S has a *unique* fixed probability vector p, which is moreover strictly positive. We are now in the position to prove the following characterisation.

Theorem 9.4. *Let S be a row-stochastic $k \times k$-matrix with fixed probability vector p. Then p is strictly positive and the Markov shift $(\mathscr{W}_k^+, \Sigma, \mu(S,p), \tau)$ is ergodic if and only if S is irreducible.*

Proof. Let $i_0, \ldots, i_l \in L$ and $j_0, \ldots, j_r \in L$. Then for $n \geq 1$ we have

$$\mu\big(\{i_0\} \times \cdots \times \{i_l\} \times L^{n-1} \times \{j_0\} \times \cdots \times \{j_r\} \times \prod L\big)$$
$$= p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} s_{i_l j_0}^n s_{j_0 j_1} \ldots s_{j_{r-1} j_r}$$

as a short calculation using (5.2) and involving the relevant indices reveals. If we consider the cylinder sets

$$E := \{i_0\} \times \cdots \times \{i_l\} \times \prod L, \qquad F := \{j_0\} \times \cdots \times \{j_r\} \times \prod L,$$

then if $n > l$

$$[\,\tau^n \in F\,] \cap E = \{i_0\} \times \cdots \times \{i_l\} \times L^{n-l-1} \times \{j_0\} \times \cdots \times \{j_r\} \times \prod L,$$

and hence

[3] It can be shown that this notion of irreducibility is equivalent to the one introduced for positive operators on L^p-spaces in Section 7.2, see [Schaefer (1974), p.20].

$$\mu([\tau^n \in F] \cap E) = p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} s^{n-l}_{i_l j_0} s_{j_0 j_1} \ldots s_{j_{r-1} j_r}.$$

By taking Cesàro averages of this expression and by splitting off the first summands we see that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu([\tau^n \in F] \cap E) = p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} q_{i_l j_0} s_{j_0 j_1} \ldots s_{j_{r-1} j_r}.$$

If S is irreducible, we know from Lemma 9.3 that $q_{i_l j_0} = p_{j_0}$, and hence

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu([\tau^n \in F] \cap E) = \mu(E)\mu(F).$$

Since cylinder sets form a dense subalgebra of Σ, by Remark 9.2.c) we conclude that (iv) of Proposition 9.1 holds, i.e., the MDS is ergodic.

Conversely, suppose that p is strictly positive and the MDS is ergodic and fix $i, j \in L$. Specialising $E = \{i\} \times \prod L$ and $F = \{j\} \times \prod L$ above, by Proposition 9.1 we obtain

$$p_i p_j = \mu(E)\mu(F) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu([\tau^n \in F] \cap E) = p_i q_{ij}.$$

Since $p_i > 0$ by assumption, $q_{ij} = p_j > 0$. Hence Q is strictly positive, and this implies that S is irreducible (again by Lemma 9.3) $\square$

9.2 From Topological to Measure-Preserving Systems

We now consider a TDS $(K;\varphi)$ and its induced operator $T = T_\varphi$ on $\mathrm{C}(K)$. In contrast to the L^p-case, it is much more difficult to establish weak compactness of subsets of $\mathrm{C}(K)$. Consequently, property (iii) of Theorem 8.8 plays a minor role here. Using instead condition (iv)

$$\text{“}\mathrm{Fix}(T) \quad \text{separates} \quad \mathrm{Fix}(T')\text{”}$$

of that theorem, we have already seen simple examples of TDSs where the induced operator is not mean ergodic (see Exercise 8.6.a)). Obviously, mean ergodicity of T is connected with $\mathrm{Fix}(T')$ being “not too large” (as compared to $\mathrm{Fix}(T)$). Now, if we identify $\mathrm{C}(K)' = \mathrm{M}(K)$ by virtue of the Riesz Representation Theorem 5.3, we see that a complex Baire measure $\mu \in \mathrm{M}(K)$ is a fixed point of T' if and only if it is φ-invariant. Mean ergodicity of T therefore becomes more likely if there are only few φ-invariant measures. To discuss the existence of such φ-invariant measures let us denote by

$$\mathrm{M}^1(K) := \big\{\mu \in \mathrm{M}(K) : \mu \geq 0,\ \langle \mu, \mathbf{1} \rangle = 1\big\}, \qquad \mathrm{M}^1_\varphi(K) := \mathrm{M}^1(K) \cap \mathrm{Fix}(T'_\varphi),$$

the set of all, and of all φ-invariant probability measures, respectively. The Krylov–Bogoljubov Theorem asserts that $\mathrm{M}^1_\varphi(K) \neq \emptyset$. We now have enough tools to give the proof of this important theorem.

Theorem 9.5 (Krylov–Bogoljubov). *For every TDS $(K;\varphi)$ there exists at least one φ-invariant probability measure on K. More precisely, for every $0 \neq f \in \mathrm{Fix}(T)$ in $\mathrm{C}(K)$ there exists a φ-invariant probability measure μ on K such that $\langle f, \mu \rangle \neq 0$.*

Proof. Since a measure $\mu \in \mathrm{M}(K)$ is positive if and only if $\langle f, \mu \rangle \geq 0$ for all $f \geq 0$, the set $\mathrm{M}^1(K)$ of probability measures is weakly* closed. However, by the Banach–Alaoglu theorem, the unit ball of $\mathrm{M}(K)$ is weakly* compact and hence $\mathrm{M}^1(K)$ is also weakly* compact. Given $0 \neq f \in \mathrm{Fix}(T)$ there is a $\nu \in \mathrm{M}^1(K)$ such that $\langle f, \nu \rangle \neq 0$ (take a Dirac measure). Since $\mathrm{M}^1(K)$ is convex and T'-invariant, all Cesàro averages

$$\mathrm{A}_n \nu = \frac{1}{n} \sum_{j=0}^{n-1} T'^j \nu$$

are contained in $\mathrm{M}^1(K)$, hence have a weak* cluster point μ. By Lemma 8.7 with $X = \mathrm{M}(K)$ and $Z = \mathrm{C}(K)$ we conclude that $T'\mu = \mu$, i.e., $\mu \in \mathrm{Fix}(T')$. Furthermore, since $Tf = f$,

$$0 \neq \langle f, \nu \rangle = \langle \frac{1}{n} \sum_{j=0}^{n-1} T^j f, \nu \rangle = \langle f, \frac{1}{n} \sum_{j=0}^{n-1} T'^j \nu \rangle \qquad (n \geq 1),$$

hence we obtain $0 \neq \langle f, \mu \rangle$. $\square$

Remarks 9.6. 1) As has been mentioned, $\mu \in \mathrm{M}(K)$ is φ-invariant if and only if $T'\mu = \mu$, i.e., $\mu \in \mathrm{Fix}(T')$. Hence Theorem 9.5 implies in particular that

$$\mathrm{Fix}(T') \quad \text{separates} \quad \mathrm{Fix}(T).$$

Remarkably enough, this holds for any contraction on a general Banach space (Exercise 8.3).

2) Theorem 9.5 has a generalisation to so-called **Markov operators**, i.e., operators T on $\mathrm{C}(K)$ such that $T \geq 0$ and $T\mathbf{1} = \mathbf{1}$ (see Exercise 9.2).

3) If in the above proof we take a cluster point of the sequence $\mathrm{A}_{n_k}\nu$ along an *arbitrary subsequence* $(n_k)_k \subset \mathbb{N}$, the construction still produces an invariant measure. This small trick will be exploited in a later lecture to produce invariant measure with particular properties.

Let $(K;\varphi)$ be a TDS. The Krylov–Bogoljubov theorem asserts that the set

$$\mathrm{M}^1_\varphi(K) = \left\{ \mu \in \mathrm{M}^1(K) : T'\mu = \mu \right\}$$

of φ-invariant probability measures is not empty, in fact it separates $\mathrm{Fix}(T_\varphi)$. Clearly the set $\mathrm{M}^1_\varphi(K)$ is convex and weakly* compact. The extreme points of this set (cf. Appendix C.6) are quite special, as the following lemma shows.

Lemma 9.7. *Let $(K;\varphi)$ be a TDS. Then $\mu \in \mathrm{M}^1_\varphi(K)$ is an extreme point of $\mathrm{M}^1_\varphi(K)$ if and only if the MDS $(K,\mathfrak{B}_0(K),\mu;\varphi)$ is ergodic.*

A φ-invariant probability measure $\mu \in \mathrm{M}^1_\varphi(K)$ is called **ergodic** if the corresponding MDS $(K,\mathfrak{B}_0(K),\mu;\varphi)$ is ergodic. By the lemma, μ is ergodic if and only if it is an extreme point of $\mathrm{M}^1_\varphi(K)$. By the Kreĭn–Milman theorem (Theorem C.13) such extreme points do always exist. Even more, we have

$$\mathrm{M}^1_\varphi(K) = \overline{\mathrm{conv}}\{\mu : \mu \text{ is a } \varphi\text{-invariant ergodic probability measure on } K\},$$

where the closure is to be taken in the weak* topology. Employing the so-called “Choquet theory” one can go even further and represent an arbitrary φ-invariant measure as a “barycenter” of ergodic measures, see [Phelps (1966)].

Proof. Assume that φ is not ergodic. Then there exists a set A with $0 < \mu(A) < 1$ such that $\varphi^* A = A$. The probability measure μ_A, defined by

$$\mu_A(B) := \mu(A \cap B)/\mu(A) \qquad (B \in \mathfrak{B}_0(K)),$$

is clearly φ-invariant, i.e., $\mu_A \in \mathrm{M}^1_\varphi(K)$. Analogously, $\mu_{A^c} \in \mathrm{M}^1_\varphi(K)$. But

$$\mu = \mu(A)\mu_A + (1-\mu(A))\mu_{A^c}$$

is a representation of μ as a non-trivial convex combination, and hence μ is not an extreme point of $\mathrm{M}^1_\varphi(K)$.

Conversely, suppose that μ is ergodic and that $\mu = (\mu_1 + \mu_2)/2$ for some $\mu_1, \mu_2 \in \mathrm{M}^1_\varphi(K)$. This implies in particular that $\mu_1 \le 2\mu$, hence

$$|\langle f, \mu_1 \rangle| \le 2\,\langle |f|, \mu \rangle = 2\,\|f\|_{\mathrm{L}^1(\mu)} \qquad (f \in \mathrm{C}(K)).$$

Since $\mathrm{C}(K)$ is dense in $\mathrm{L}^1(K,\mathfrak{B}_0,\mu)$, $\mu_1 \in \mathrm{L}^1(K,\mathfrak{B}_0,\mu)'$. Let us consider the induced operator T_φ on $\mathrm{L}^1(K,\mathfrak{B}_0,\mu)$. Since μ is ergodic, $\mathrm{Fix}(T_\varphi)$ is one-dimensional. By Example 8.12 T_φ is mean ergodic, and by Theorem 8.8(v) $\mathrm{Fix}(T_\varphi)$ separates $\mathrm{Fix}(T'_\varphi)$. Hence the latter is one-dimensional too, and since $\mu, \mu_1 \in \mathrm{Fix}(T'_\varphi)$ it follows that $\mu_1 = \mu$. Consequently, μ is an extreme point of $\mathrm{M}^1_\varphi(K)$. □

By Lemma 9.7, if there are two different φ-invariant probability measures then there is also a non-ergodic one. Conversely, all φ-invariant measures are ergodic if and only if there is exactly one φ-invariant probability measure. TDSs with this property are called **uniquely ergodic**. To analyse this notion further, we need the following useful information.

Lemma 9.8. *Let $(K;\varphi)$ be a TDS and let $T = T_\varphi$ be the induced operator on $\mathrm{C}(K)$. Then $\mathrm{Fix}(T')$ is a lattice, i.e., if $\nu \in \mathrm{Fix}(T')$ then also $|\nu| \in \mathrm{Fix}(T')$. Consequently, $\mathrm{M}^1_\varphi(K)$ linearly spans $\mathrm{Fix}(T')$.*

Proof. Take $\nu \in \mathrm{Fix}(T')$. By Exercise 9.8 we obtain $|\nu| = |T'\nu| \le T'|\nu|$ and

$$\langle \mathbf{1}, |\nu| \rangle \leq \langle \mathbf{1}, T'|\nu| \rangle = \langle \mathbf{1}, |\nu| \rangle.$$

This implies that $\langle \mathbf{1}, T'|\nu| - |\nu| \rangle = 0$, hence $|\nu| \in \mathrm{Fix}(T')$. To prove the second statement, note that $\nu \in \mathrm{Fix}(T')$ if and only if $\mathrm{Re}\,\nu, \mathrm{Im}\,\nu \in \mathrm{Fix}(T')$. But if ν is a real (=signed) measure in $\mathrm{Fix}(T')$, then $\nu = \nu^+ - \nu^-$ with $\nu^+ = (1/2)(|\nu| + \nu)$ and $\nu^- = (1/2)(|\nu| - \nu)$ both in M^+_φ (see Lecture 7.1). □

As a corollary we can characterise unique ergodicity.

Corollary 9.9. *Let $(K;\varphi)$ be a TDS with induced operator T on $\mathrm{C}(K)$. The following assertions are equivalent.*

(i) *$(K;\varphi)$ is uniquely ergodic, i.e., $\mathrm{M}^1_\varphi(K)$ is a singleton.*

(ii) $\dim \mathrm{Fix}(T') = 1$.

(iii) *T is mean ergodic and* $\dim \mathrm{Fix}(T) = 1$.

Proof. The equivalence of (i) and (ii) follows from Lemma 9.8 and the fact that $\mathrm{M}^1_\varphi(K)$ is never empty, by the Krylov–Bogoljubov Theorem 9.5. If (ii) is satisfied, then by Remark 9.2.a), we have that $\dim \mathrm{Fix}(T) = \dim \mathrm{Fix}(T') = 1$ and $\mathrm{Fix}(T)$ clearly separates $\mathrm{Fix}(T')$. Whence T is mean ergodic by Theorem 8.8. Conversely, suppose that (iii) holds. Then, by Theorem 8.8, we must have $\dim \mathrm{Fix}(T') = 1$ as well, hence (ii) follows. □

A further strengthening yields the following analogue of Proposition 7.8.

Corollary 9.10. *Let $(K;\varphi)$ be a TDS with induced operator $T = T_\varphi$ on $\mathrm{C}(K)$. The following properties are equivalent.*

(i) *$(K;\varphi)$ is minimal and T is mean ergodic.*

(ii) *$(K;\varphi)$ is uniquely ergodic, and the unique φ-invariant probability measure μ is strictly positive, i.e., $0 \leq f \in \mathrm{C}(K)$ and $\langle f, \mu \rangle = 0$ implies $f = 0$.*

A TDS having equivalent properties of Corollary 9.10 is called **strictly ergodic**.

Proof. (i)⇒(ii) Since minimality implies (forward) transitivity, $\mathrm{Fix}(T) = \langle \mathbf{1} \rangle$ follows by Lemma 4.14. By Corollary 9.9, the TDS $(K;\varphi)$ is uniquely ergodic. Let μ be the unique φ-invariant probability measure and define

$$J := \bigl\{ f \in \mathrm{C}(K) : \langle |f|, \mu \rangle = 0 \bigr\}.$$

Then J is a closed, T-invariant (algebra) ideal. By Corollary 4.13 this ideal must be equal to 0, hence μ is strictly positive.

(ii)⇒(i) Since $(K;\varphi)$ is uniquely ergodic, Corollary 9.9 implies that T is mean ergodic with one-dimensional mean ergodic projection

$$Pf = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} T^j f = \left(\int_\Omega f \,\mathrm{d}\mu \right) \cdot \mathbf{1} \qquad (f \in \mathrm{C}(K)).$$

Suppose that J is a closed, T-invariant, proper (algebra) ideal of $\mathrm{C}(K)$ and take $f \in J$. Then $g := |f| \in J$, since by Theorem 4.4 $J = I_A$ for some closed subset $A \subset K$. Since J is T-invariant and closed, $\langle g, \mu \rangle \mathbf{1} = Pg \in J$ as well. Since J is proper, $\langle g, \mu \rangle = 0$. Since μ is strictly positive, $g = 0$, i.e., $f = 0$. Thus, $J = 0$, and by Corollary 4.13 $(K;\varphi)$ must be minimal. □

Minimality is *not* characterised by unique ergodicity. An easy example of a uniquely ergodic system that is not minimal (not even transitive) is given in Exercise 9.4. Minimal systems that are not uniquely ergodic are much harder to obtain, see [Furstenberg (1961)] and [Raimi (1964)]. The following result shows that at least for a group rotation minimality and unique ergodicity are equivalent. Also, minimal group rotations are ergodic, a fact that has been announced in Proposition 6.13.

Theorem 9.11. *Let G be a compact group, with Haar probability measure μ. For $a \in G$ let $\varphi_a : g \longmapsto ag$ be the left rotation with a, and let $T = T_{\varphi_a}$ be the induced operator on $\mathrm{C}(G)$. Then the following assertions are equivalent:*

(i) $(G;\varphi_a)$ *is minimal.*

(ii) $\dim \mathrm{Fix}(T) = 1$.

(iii) $(G;\varphi_a)$ *is uniquely ergodic.*

(iv) $(G;\varphi_a)$ *is strictly ergodic.*

(v) $(G, \mathfrak{B}_0(G), \mu; \varphi_a)$ *is ergodic.*

Proof. (i)⇒(ii) follows from Lemma 4.14, (ii)⇒(iii) follows from Corollary 9.9 since T is mean ergodic by Example 8.14. (iii)⇒(iv) follows since the Haar measure μ is a strictly positive φ_a-invariant probability measure, and (iv)⇒(i) follows from Corollary 9.10.

If (v) holds then every fixed L^1-function is constant, so in particular this holds for every continuous function. (Note that the embedding $\mathrm{C}(G) \subset \mathrm{L}^\infty(G, \mathfrak{B}_0, \mu)$ is injective since the Haar measure μ is strictly positive.) Conversely, if (i)-(iv) hold then the Haar measure is the unique φ_a-invariant probability measure, hence by Lemma 9.7 gives rise to an ergodic MDS. □

Note that our proof rests on Example 8.14, a fact whose proof we still have to postpone.

9.3 Application: Equidistribution

Take a minimal rotation $\varphi_a : \mathbb{T} \longrightarrow \mathbb{T}$ on the torus and consider the induced operator $T = T_{\varphi_a}$ on the following function spaces:

a) $\mathrm{R}(\mathbb{T}) := \{f : f \text{ is bounded, Riemann integrable}\}$,

b) $\mathrm{B}(\mathbb{T}) := \{f : f \text{ is bounded, Borel measurable}\}$,

c) $\mathrm{L}^\infty(\mathbb{T}, \mathfrak{B}, \mathrm{d}z)$, each endowed with the supremum norm $\|\cdot\|_\infty$. On which of these spaces is T mean ergodic?

Proposition 9.12. *Let $a \in \mathbb{T}$ with $a^n \neq 1$ for all $n \in \mathbb{N}$. Then the corresponding rotation operator T_{φ_a} is mean ergodic on $\mathrm{R}(\mathbb{T})$ with one-dimensional fixed space, and the mean ergodic projection is given by $Pf = \int_{\mathbb{T}} f\,\mathrm{d}z \cdot \mathbf{1}$. However, it is mean-ergodic neither on $\mathrm{B}(\mathbb{T})$ nor on $\mathrm{L}^\infty(\mathbb{T},\mathfrak{B},\mathrm{d}z)$.*

Proof. Recall first that a real-valued function f is bounded and Riemann integrable if and only if for every $\varepsilon > 0$ there exist step functions g_ε, h_ε such that $g_\varepsilon \leq f \leq h_\varepsilon$ and $\int_{\mathbb{T}} (h_\varepsilon - g_\varepsilon)\,\mathrm{d}z \leq \varepsilon$.

We start by showing the convergence (in supremum norm) of the Cesàro averages $\mathrm{A}_n = \frac{1}{n}\sum_{j=0}^{n-1} T^j$ on a characteristic function χ of a "segment" of $\mathbb{T}$. For each such function and each $\varepsilon > 0$ there exist *continuous* functions g_ε, h_ε satisfying

$$g_\varepsilon \leq \chi \leq h_\varepsilon \quad \text{and} \quad \int_{\mathbb{T}} (h_\varepsilon - g_\varepsilon)\,\mathrm{d}z \leq \varepsilon.$$

By Example 8.13, $\mathrm{A}_n g_\varepsilon$ converges uniformly to $\left(\int_{\mathbb{T}} g_\varepsilon\,\mathrm{d}z\right)\cdot\mathbf{1}$, while $\mathrm{A}_n h_\varepsilon$ converges uniformly to $\left(\int_{\mathbb{T}} h_\varepsilon\,\mathrm{d}z\right)\cdot\mathbf{1}$. Since $\mathrm{A}_n g_\varepsilon \leq \mathrm{A}_n \chi \leq \mathrm{A}_n h_\varepsilon$ for all $n \in \mathbb{N}$, we obtain

$$\left(\int_{\mathbb{T}} g_\varepsilon\,\mathrm{d}z - \varepsilon\right)\cdot\mathbf{1} \leq \mathrm{A}_n\chi \leq \left(\int_{\mathbb{T}} h_\varepsilon\,\mathrm{d}z + \varepsilon\right)\cdot\mathbf{1}$$

for sufficiently large n. So we have for such n that

$$\left\|\mathrm{A}_n\chi - \int_\Omega \chi\,\mathrm{d}z\cdot\mathbf{1}\right\|_\infty \leq 2\varepsilon$$

by the choice of g_ε and h_ε. Therefore

$$\|\cdot\|_\infty - \lim_{n\to\infty} \mathrm{A}_n\chi = \left(\int_{\mathbb{T}} \chi\,\mathrm{d}z\right)\cdot\mathbf{1}.$$

Clearly the convergence on these characteristic functions extends to all step functions. Take now f real-valued, bounded and Riemann integrable. As recalled above, for $\varepsilon > 0$ we find step functions g_ε, h_ε satisfying $g_\varepsilon \leq f \leq h_\varepsilon$ and $\int_{\mathbb{T}}(h_\varepsilon - g_\varepsilon)\,\mathrm{d}z \leq \varepsilon$. Arguing similarly as above, we can prove

$$\|\cdot\|_\infty - \lim_{n\to\infty} \mathrm{A}_n f = \left(\int_{\mathbb{T}} f\,\mathrm{d}z\right)\cdot\mathbf{1}.$$

The proof of the remaining statements is left as Exercise 9.7. □

The mean ergodicity of T on $\mathrm{R}(\mathbb{T})$ has a nice application to the following property of sequences in $[0,1]$: A sequence $(\alpha_n)_{n=0}^\infty \subset [0,1]$ is called **equidistributed** (in [0,1]) if

$$\lim_{n\to\infty} \frac{\operatorname{card}\{j : 0 \leq j < n,\ \alpha_j \in [a,b]\}}{n} = b - a$$

for every $[a,b] \subset [0,1]$. In other words, the relative frequency of the first n elements of the sequence falling into an arbitrary fixed interval $[a,b]$ converges to the length of that interval (independently of its location).

The following classical theorem from [Weyl (1916)] gives an important example of an equidistributed sequence and is a quantitative version of Kronecker's theorem in Example 2.19.

Theorem 9.13 (Weyl). *Let $\alpha \in [0,1] \setminus \mathbb{Q}$. Then the sequence $(n\alpha \pmod 1)_{n=0}^{\infty}$ is equidistributed in $[0,1]$.*

Proof. Denote $\alpha_n := n\alpha \pmod 1$. We have seen in Exercise 2.2 that the translation by $\alpha \pmod 1$ is isomorphic to the rotation on $\mathbb{T}$. Therefore the induced operator

$$(Tf)(x) = f(\alpha + x), \quad x \in [0,1] \pmod 1,$$

is mean ergodic on the Riemann integrable functions f on $[0,1]$ by Proposition 9.12. Furthermore, the Cesàro averages

$$\mathrm{A}_n f(0) = \frac{1}{n}\sum_{j=0}^{n-1} T^j f(0) = \frac{1}{n}\sum_{j=0}^{n-1} f(\alpha_j)$$

converge to $\int_0^1 f(s)\,\mathrm{d}s$. By taking $f = \mathbf{1}_{[a,b]}$ we obtain the assertion. □

In the proof it was decisive that we have

$$\frac{1}{n}\sum_{j=0}^{n-1} f(\alpha_j) \to \int_0^1 f(s)\mathrm{d}s \quad (n \to \infty).$$

This is actually a property of mod 1 equidistributed sequences. The proof of the following characterisation is left as Exercise 9.9.

Proposition 9.14. *A sequence $(\alpha_n)_{n=0}^{\infty} \subset [0,1]$ is equidistributed if and only if*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(\alpha_j) = \int_0^1 f(s)\,\mathrm{d}s$$

holds for every bounded and Riemann-integrable (equivalently: for every continuous) function $f : [0,1] \longrightarrow \mathbb{C}$.

This result applied to exponential functions $s \longmapsto \mathrm{e}^{2\pi i n s}$ was the basis of Weyl's original proof. More on this circle of ideas can be found in [Hlawka (1979)] or [Kuipers and Niederreiter (1974)].

Exercises

1.[Mean Ergodicity on $\mathrm{C}(K)$] Let $(K;\varphi)$ be a TDS with an invariant probability measure μ. Suppose that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f(\varphi^j(x)) = \int_K f\,\mathrm{d}\mu$$

for every $f\in\mathrm{C}(K)$ and every $x\in K$. Show that the induced operator T_φ on $\mathrm{C}(K)$ is mean ergodic and that $(K,\mathfrak{B}_0(K),\mu;\varphi)$ is ergodic. (Hint: Use the Dominated Convergence Theorem to prove *weak mean ergodicity* and then use part (iii) of Theorem 8.8.)

2.[Markov Operators] Let K be a compact topological space. A bounded linear operator $T:\mathrm{C}(K)\longrightarrow\mathrm{C}(K)$ is called a **Markov operator** if it is *positive* (i.e., $f\geq 0$ implies $Tf\geq 0$) and satisfies $T\mathbf{1}=\mathbf{1}$.

Let T be a Markov operator on $\mathrm{C}(K)$. Show that $\|T\|=1$ and that there exists a probability measure $\mu\in\mathrm{M}^1(K)$ such that

$$\int_K Tf\,\mathrm{d}\mu = \int_K f\,\mathrm{d}\mu \quad \text{for all} \quad f\in\mathrm{C}(K).$$

(Hint: Imitate the proof of the Krylov–Bogoljubov theorem.)

3. A linear functional ψ on ℓ^∞ with the following properties is called a **Banach limit**.

(i) ψ is positive, i.e., $(x_n)_{n\in\mathbb{N}_0}\in\ell^\infty$, $x_n\geq 0$ imply $\psi(x_n)_{n\in\mathbb{N}_0}\geq 0$.

(ii) ψ is translation invariant, i.e., $\psi(x_n)_{n\in\mathbb{N}_0}=\psi(x_{n+1})_{n\in\mathbb{N}}$.

(iii) $\psi(\mathbf{1})=1$, where $\mathbf{1}$ denotes the constant 1 sequence.

Prove the following assertions.

a) A Banach limit ψ is continuous, i.e., $\psi\in(\ell^\infty)'$.

b) If $(x_n)_{n\in\mathbb{N}_0}\in\ell^\infty$ is periodic with period $k\geq 1$, then for a Banach limit ψ

$$\psi(x_n)=\frac{1}{k}\sum_{j=0}^{k-1}x_j.$$

c) If $(x_n)\in\mathrm{c}$ and ψ is a Banach limit, then $\psi(x_n)=\lim_{n\to\infty}x_n$.

d) If $(x_n)\in\ell^\infty$ is Cesàro convergent (i.e., $c_n:=\frac{1}{n}\sum_{j=0}^{n-1}x_j$ converges), then for a Banach limit ψ

$$\psi(x_n)=\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}x_j.$$

e) There exist Banach limits. (Hint: Imitate the proof of the Krylov–Bogoljubov theorem.)

f) For any $(x_n)_{n\in\mathbb{N}} \in \ell^\infty(\mathbb{R})$ and $\alpha \in [\liminf_{n\to\infty} x_n, \limsup_{n\to\infty} x_n]$ there is a Banach limit ψ with $\psi(x_n) = \alpha$. (Hint: refine the construction in e), cf. also Remark 9.6.3).

4. Let $K := [0,1]$ and $\varphi(x) = 0$ for all $x \in [0,1]$. Show that $(K;\varphi)$ is uniquely ergodic, but not minimal.

5. Take $K = \{z \in \mathbb{C} : |z| \leq 1\}$ and $\varphi_a(z) = az$ for some $a \in \mathbb{T}$ satisfying $a^n \neq 1$ for all $n \in \mathbb{N}$. Find the ergodic measures for φ_a.

6. For the doubling map on $K = [0,1)$ (cf. Exercise 2.4) and for the tent map on $[0,1]$ (cf. Exercise 2.5) answer the following questions:

1) Is the TDS uniquely ergodic?
2) Is the induced operator on $\mathrm{C}(K)$ mean ergodic?

$*$ **7.** Let $a \in \mathbb{T}$ with $a^n \neq 1$ for all $n \in \mathbb{N}$. Show that the operator $T = T_{\varphi_a}$ induced by the rotation φ_a is not mean ergodic on $\mathrm{B}(\mathbb{T})$ and on $\mathrm{L}^\infty(\mathbb{T},\mathfrak{B},\mathrm{d}z)$. (Hint: For the first part, take a 0-1-sequence $\{c_n\}_{n=1}^\infty$ which is not Cesàro convergent and consider the characteristic function of the set $\{a^n : c_n = 1\}$. For the second part, consider $M := \{a^n : n \in \mathbb{N}\}$ and

$$I := \big\{[f] \in \mathrm{L}^\infty(\mathbb{T},\mathfrak{B},\mathrm{d}z) : f \in \mathrm{B}(\mathbb{T}) \text{ vanishes on a neighbourhood of } M\big\}.$$

Show that I is an ideal of $\mathrm{L}^\infty(\mathbb{T},\mathfrak{B},\mathrm{d}z)$, with $\mathbf{1} \notin J := \overline{I}$. Conclude that there is $\nu \in \mathrm{L}^\infty(\mathbb{T},\mathfrak{B},\mathrm{d}z)'$ which vanishes on J and satisfies $\langle \mathbf{1}, \nu\rangle = 1$. Use this property for $T'^n\nu$ and $\mathrm{A}_n'\nu$ and then a weak*-compactness argument.)

$*$ **8.** Let $(K;\varphi)$ be a TDS, with induced operator $T = T_\varphi$ on $\mathrm{C}(K)$. Show that $T'\mu = \varphi_*\mu$ is the push-forward measure, for every $\mu \in \mathrm{M}(K)$. Then show that $|T'\mu| \leq T'|\mu|$ for every $\mu \in \mathrm{M}(K)$, e.g., by using the definition of the modulus of a measure in Appendix B.12. (Note that the “inf” there has to be corrected to a “sup”.)

9. Prove Proposition 9.14.

Lecture 10
The Pointwise Ergodic Theorem

I say that almost everywhere there is beauty enough to fill a person's life if one would only be sensitive to it.

Elizabeth J. Coatsworth[1]

While von Neumann's mean ergodic theorem is powerful and far reaching, it does not actually solve the original problem of establishing that "time mean equals space mean" for a given MDS $(\Omega, \Sigma, \mu; \varphi)$. For this we need the *pointwise* limit

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

for states $x \in \Omega$ and observables $f : \Omega \longrightarrow \mathbb{R}$. Of course, by Proposition 9.1 this limit equals the "space mean" $\int_\Omega f \, d\mu$ for each observable f only if the system is ergodic. Moreover, due to the presence of null-sets we cannot expect the convergence to hold for *all* points $x \in \Omega$. Hence we should ask merely for convergence *almost everywhere*.

Shortly after and inspired by von Neumann's result, G.D. Birkhoff in [Birkhoff (1931)] proved the desired result.

Theorem 10.1 (Birkhoff). *Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS and let $f \in \mathrm{L}^1(\Omega, \Sigma, \mu)$. Then the limit*

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

exists for almost every $x \in \Omega$.

Birkhoff's theorem is also called the *"Pointwise (or: Individual) Ergodic Theorem"*. Using the induced operator $T = T_\varphi$ and its Cesàro averages $\mathrm{A}_n = \mathrm{A}_n[T]$ we obtain by Birkhoff's theorem that for every $f \in \mathrm{L}^1(\Omega, \Sigma, \mu)$, the sequence $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ converges pointwise μ-almost everywhere. Since we already know that T is mean

[1] Personal Geography: Almost an Autobiography, S. Greene Press, 1976, page 110.

ergodic, the Cesàro averages $\mathrm{A}_n f$ converge in L^1-norm. Hence Birkhoff's theorem in combination with Proposition 9.1.(ii) implies the following characterisation of ergodic MDSs.

Corollary 10.2. *An MDS $(\Omega, \Sigma, \mu; \varphi)$ is ergodic if and only if for every ("observable") $f \in \mathrm{L}^1(\Omega, \Sigma, \mu)$ one has "time mean equals space mean"*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\varphi^j(x)) = \int_\Omega f \,\mathrm{d}\mu$$

for μ-almost every ("state") $x \in \Omega$.

As in the case of von Neumann's theorem, Birkhoff's result is operator-theoretic in nature. We therefore proceed as in Lecture 8 and use an abstract approach.

10.1 Pointwise Ergodic Operators

Definition 10.3. Let (Ω, Σ, μ) be a measure space and let $1 \le p \le \infty$. A bounded linear operator T on $\mathrm{L}^p(\Omega, \Sigma, \mu)$ is called **pointwise ergodic** if the limit

$$\lim_{n\to\infty} \mathrm{A}_n f = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} T^j f$$

exists μ-almost everywhere for every $f \in \mathrm{L}^p(\Omega, \Sigma, \mu)$.

Birkhoff's theorem says that induced operators arising from MDSs are pointwise ergodic. We shall obtain Birkhoff's theorem from a more general result which goes back to [Hopf (1954)] and [Dunford and Schwartz (1958)]. Recall that an operator T on $\mathrm{L}^1(\Omega, \Sigma, \mu)$, (Ω, Σ, μ) a measure space, is called a Dunford–Schwartz operator (or a complete contraction) if one has

$$\|Tf\|_1 \le \|f\|_1 \quad \text{and} \quad \|Tf\|_\infty \le \|f\|_\infty$$

for all $f \in \mathrm{L}^\infty \cap \mathrm{L}^1$ (cf. Example 8.12). The Hopf–Dunford–Schwartz result then reads as follows.

Theorem 10.4 (Pointwise Ergodic Theorem). *Let (Ω, Σ, μ) be a measure space and let T be a positive Dunford–Schwartz operator. Then T is pointwise ergodic.*

Before we turn to the proof of Theorem 10.4 let us discuss some further results.

Dunford and Schwartz [Dunford and Schwartz (1958)] have shown that one can omit the condition of positivity from Theorem 10.4. This is due to the fact that for a Dunford– Schwartz operator T there always exists a *positive* Dunford-Schwartz operator S such that $|Tf| \le S|f|$ for all $f \in \mathrm{L}^1$. On the other hand, a general positive contraction on $\mathrm{L}^1(\Omega, \Sigma, \mu)$ need not be pointwise ergodic. A first example was given in [Chacon (1964)]. Shortly after, A. Ionescu Tulcea proved even that the class of positive isometric isomorphisms on $\mathrm{L}^1(0,1)$ which are not pointwise ergodic is "rich" in the sense of category [Ionescu Tulcea (1965)].

One can weaken the condition on L^∞-contractivity, though. E.g., Hopf has shown that in place of $T\mathbf{1} \le \mathbf{1}$ (cf. Exercise 10.1) one may suppose that there is a strictly positive function f such that $Tf \le f$ [Krengel (1985), Theorem 3.3.5]. For general positive L^1-contractions there is the following result from [Krengel (1985), Theorem 4.9].

Theorem 10.5 (Stochastic Ergodic Theorem). *Let (Ω,Σ,μ) be a finite measure space and let T be a positive contraction on $\mathrm{L}^1(\Omega,\Sigma,\mu)$. Then the Cesàro averages $(\mathrm{A}_n[T]f)_n$ converge in measure for every $f \in \mathrm{L}^1(\Omega,\Sigma,\mu)$.*

If we pass to L^p spaces with $1 < p < \infty$, the situation improves. Namely, building on [Ionescu Tulcea (1964)] Akcoglu established in [Akcoglu (1975)] the following celebrated result.

Theorem 10.6 (Akcoglu's Ergodic Theorem). *Let (Ω,Σ,μ) be a measure space and let T be a positive contraction on $\mathrm{L}^p(\Omega,\Sigma,\mu)$ for some $1 < p < \infty$. Then T is pointwise ergodic.*

For $p = 2$ and T self-adjoint, this is due to Stein and has an elementary proof, see [Stein (1961b)]. In the general case the proof is quite involved and beyond the scope of these lectures, see [Krengel (1985), Section 5.2] or [Kern et al. (1977)] and [Nagel and Palm (1982)]. Burkholder [Burkholder (1962)] has shown that if $p = 2$, the condition of positivity cannot be omitted. (The question whether this is true also for $p \neq 2$ is still open.)

Let us return to Theorem 10.4 and its proof. For simplicity, we shall deal with finite measure spaces only, so let T be a Dunford–Schwartz operator on a finite measure space (Ω,Σ,μ). By Example 8.12 we already know that T is mean ergodic, whence

$$\mathrm{L}^1(\Omega,\Sigma,\mu) = \mathrm{Fix}(T) \oplus \overline{\mathrm{ran}}(\mathrm{I}-T).$$

Since $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$ is dense in $\mathrm{L}^1(\Omega,\Sigma,\mu)$, the space

$$F := \mathrm{Fix}(T) \oplus (\mathrm{I}-T)\mathrm{L}^\infty(\Omega,\Sigma,\mu)$$

is dense in $\mathrm{L}^1(\Omega,\Sigma,\mu)$. As before, we write $\mathrm{A}_n := \mathrm{A}_n[T]$, $n \in \mathbb{N}$.

Lemma 10.7. *In the situation described above, the sequence $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ is $\|\cdot\|_\infty$-convergent for every $f \in F$.*

Proof. Write $f = g + (\mathrm{I}-T)h$ with $g \in \mathrm{Fix}(T)$ and $h \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$. Then as already seen many times, $\mathrm{A}_n f = g + (1/n)(h - T^n h)$, and since $\sup_n \|T^n h\|_\infty \le \|h\|_\infty$, it follows that $\|\mathrm{A}_n f - g\|_\infty \to 0$ for $n \to \infty$. □

By this lemma we have a.e.-convergence of $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ for every f from the dense subspace F of $\mathrm{L}^1(\Omega,\Sigma,\mu)$. What we need is a tool that allows us to pass from F to its closure. This is considerably more difficult than in the case of norm convergence, and will be treated in the next section.

10.2 Banach's Principle and Maximal Inequalities

Let (Ω,Σ,μ) be a measure space, $1 \le p \le \infty$, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of bounded linear operators on $X := \mathrm{L}^p(\Omega,\Sigma,\mu)$. The statement "$\lim_n T_n f$ exists pointwise almost everywhere" can be reformulated as

$$\limsup_{k,l\to\infty} |T_k f - T_l f| = \inf_{n\in\mathbb{N}} \sup_{k,l\ge n} |T_k f - T_l f| = 0, \tag{10.1}$$

where suprema and infima are taken in the complete lattice $\mathrm{L}^0 := \mathrm{L}^0(\Omega,\Sigma,\mu;\overline{\mathbb{R}})$ (see Section 7.1).

If (10.1) is already established for f from some dense subspace F of X, one would like to infer that it holds for every $f \in X$. To this purpose we consider the associated **maximal operator** $T^* : X \longrightarrow \mathrm{L}^0$ defined by

$$T^* f := \sup_{n\in\mathbb{N}} |T_n f| \qquad (f \in X).$$

Note that $T^* f \ge 0$ and $T^*(\alpha f) = |\alpha|\, T^* f$ for every $f \in X$ and $\alpha \in \mathbb{C}$. However, the operator T^* is not additive but *subadditive*, i.e., $T^*(f+g) \le T^* f + T^* g$ for all $f, g \in X$.

Definition 10.8. We say that the sequence of operators $(T_n)_n$ satisfies an (abstract) **maximal inequality** if there is a function $c : (0,\infty) \to [0,\infty)$ with $\lim_{\lambda\to\infty} c(\lambda) = 0$ such that

$$\mu[\, T^* f > \lambda \,] \le c(\lambda) \qquad (\lambda > 0, f \in X, \|f\| \le 1).$$

The following result shows that an abstract maximal inequality is exactly what we need.

Proposition 10.9 (Banach's Principle). *Let (Ω,Σ,μ) be a measure space, $1 \le p < \infty$, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of bounded linear operators on $X = \mathrm{L}^p(\Omega,\Sigma,\mu)$. If the associated maximal operator T^* satisfies a maximal inequality, then the set*

$$C := \big\{ f \in X : (T_n f)_{n\in\mathbb{N}} \text{ is a.e.-convergent} \big\}$$

is a closed subspace of X.

Proof. Since the operators T_n are linear, C is a subspace of X. To see that it is closed, let $f \in X$ and $g \in C$. For any natural numbers k, l we have

$$|T_k f - T_l f| \le |T_k(f-g)| + |T_k g - T_l g| + |T_l(g-f)| \le 2T^*(f-g) + |T_k g - T_l g|\,.$$

Taking the limsup in L^0 with $k, l \to \infty$ (see (10.1)) one obtains

$$h := \limsup_{k,l\to\infty} |T_k f - T_l f| \le 2T^*(f-g) + \limsup_{k,l\to\infty} |T_k g - T_l g| = 2T^*(f-g),$$

since $g \in C$. For $\lambda > 0$ we thus have $[\, h > 2\lambda \,] \subset [\, T^*(f-g) > \lambda \,]$ and hence

$$\mu[\,h > 2\lambda\,] \le \mu[\,T^*(f-g) > \lambda\,] \le c\left(\frac{\lambda}{\|f-g\|}\right).$$

If $f \in \overline{C}$ we can make $\|f-g\|$ arbitrarily small, and since $\lim_{t\to\infty} c(t) = 0$, we obtain $\mu[\,h > 2\lambda\,] = 0$. Since $\lambda > 0$ was arbitrary, it follows that $h = 0$. This shows that $f \in C$, whence C is closed. □

To conclude the proof of the Pointwise Ergodic Theorem we need a maximal inequality for the maximal operator A^* associated with the sequence $(\mathrm{A}_n)_{n\in\mathbb{N}}$ of Cesàro averages. The key here is the following lemma by E. Hopf.

Lemma 10.10 (Hopf). *For a positive contraction T on $\mathrm{L}^1(\Omega,\Sigma,\mu)$ define*

$$S_n f := \sum\nolimits_{j=0}^{n-1} T^j f \quad \textit{and} \quad M_n f := \max_{1\le k\le n} S_k f \qquad (f \in \mathrm{L}^1(\Omega,\Sigma,\mu;\mathbb{R}))$$

for $n \in \mathbb{N}$. Then

$$\int_{[M_n f \ge 0]} f \,\mathrm{d}\mu \ge 0$$

for every $f \in \mathrm{L}^1(\Omega,\Sigma,\mu;\mathbb{R})$.

Proof. For $2 \le k \le n$ one clearly has $S_{k-1} f \le M_n f \le (M_n f)^+$, and therefore

$$S_k f = f + T S_{k-1} f \le f + T(M_n f)^+ \qquad (2 \le k \le n),$$

since T is positive. But $S_1 f = f \le f + T(M_n f)^+$ is obvious, and hence we obtain

$$M_n f = \max_{1\le k\le n} S_k f \le f + T(M_n f)^+.$$

Integration yields

$$\begin{aligned}\int_\Omega (M_n f)^+ \mathrm{d}\mu &= \int_{[M_n f\ge 0]} M_n f \,\mathrm{d}\mu \le \int_{[M_n f\ge 0]} f\,\mathrm{d}\mu + \int_{[M_n f\ge 0]} T(M_n f)^+\,\mathrm{d}\mu \\ &\le \int_{[M_n f\ge 0]} f\,\mathrm{d}\mu + \int_\Omega T(M_n f)^+\,\mathrm{d}\mu \le \int_{[M_n f\ge 0]} f\,\mathrm{d}\mu + \int_\Omega (M_n f)^+\,\mathrm{d}\mu\end{aligned}$$

because T is positive and contractive. From this the assertion follows. □

As a consequence of Hopf's lemma we obtain what is called a *maximal theorem*.

Theorem 10.11 (Maximal Ergodic Theorem). *Let (Ω,Σ,μ) be a finite measure space, and let T be a positive Dunford–Schwartz operator. Then, with $\mathrm{A}_n := \mathrm{A}_n[T]$ and*

$$\mathrm{A}_n^* f := \max_{1\le k\le n} \mathrm{A}_k f \qquad (n \in \mathbb{N}, f \in \mathrm{L}^1(\Omega,\Sigma,\mu;\mathbb{R}))$$

one has

$$\lambda\,\mu[\,\mathrm{A}_n^* f > \lambda\,] \le \int_{[\mathrm{A}_n^* > \lambda\,]} f\,\mathrm{d}\mu \le \|f\|_1$$

for every $n \in \mathbb{N}$, $\lambda > 0$ *and* $f \in \mathrm{L}^1(\Omega, \Sigma, \mu; \mathbb{R})$.

Proof. We use the notation from Hopf's lemma and apply it to the function $g := f - \lambda \mathbf{1}$ obtaining

$$\int_{[M_n g \geq 0]} g \, \mathrm{d}\mu \geq 0.$$

Since $[M_n g = 0] \subset [g \leq 0]$ we even have

$$\int_{[M_n g > 0]} (f - \lambda \mathbf{1}) \, \mathrm{d}\mu = \int_{[M_n g > 0]} g \, \mathrm{d}\mu \geq 0. \tag{10.2}$$

Fix $1 \leq k \leq n$. Then $T^j g = T^j f - \lambda T^j \mathbf{1} \geq T^j f - \lambda \mathbf{1}$ for all $0 \leq j \leq k$. Hence for the Cesàro averages we have $\mathrm{A}_k g \geq \mathrm{A}_k f - \lambda \mathbf{1}$. Taking maxima over $k = 1, \dots, n$ yields $\mathrm{A}_n^* g \geq \mathrm{A}_n^* f - \lambda \mathbf{1}$. This implies

$$[\, f > \lambda \,] \subset [\, \mathrm{A}_n^* f > \lambda \,] \subset [\, \mathrm{A}_n^* g > 0 \,] = [\, M_n g > 0 \,].$$

Hence in

$$\int_{[M_n g > 0]} (f - \lambda \mathbf{1}) \, \mathrm{d}\mu = \int_{[\mathrm{A}_n^* f > \lambda]} (f - \lambda \mathbf{1}) \, \mathrm{d}\mu + \int_{[M_n g > 0] \setminus [\mathrm{A}_n^* f > \lambda]} (f - \lambda \mathbf{1}) \, \mathrm{d}\mu$$

the second summand on the right-hand side is non-positive. This together with (10.2) concludes the proof. □

We are now able to prove Theorem 10.4.

Proof of Theorem 10.4. Since T is positive, each A_n is also positive. Hence, for $f \in \mathrm{L}^1(\Omega, \Sigma, \mu)$ we have $\max_{1 \leq k \leq n} |\mathrm{A}_k f| \leq \mathrm{A}_n^* |f|$ and thus

$$\mu \left[\max_{1 \leq k \leq n} |\mathrm{A}_k f| > \lambda \right] \leq \mu [\, \mathrm{A}_n^* |f| > \lambda \,] \leq \frac{\|f\|_1}{\lambda}$$

by the Maximal Ergodic Theorem 10.11. If we let $n \to \infty$, then $\max_{1 \leq k \leq n} |\mathrm{A}_k f| \nearrow \sup_{k \in \mathbb{N}} |\mathrm{A}_k f| = \mathrm{A}^* f$. Hence $[\max_{1 \leq k \leq n} |\mathrm{A}_k f| > \lambda \,] \nearrow [\, \mathrm{A}^* f > \lambda \,]$ and we obtain

$$\mu [\, \mathrm{A}^* f > \lambda \,] \leq \frac{\|f\|_1}{\lambda} \qquad (\lambda > 0),$$

i.e., the sequence $(\mathrm{A}_n[T])_{n \in \mathbb{N}}$ satisfies a maximal inequality. Combining Banach's principle with Lemma 10.7 concludes the proof of Theorem 10.4. □

Remark 10.12. Hopf's lemma and the Maximal Ergodic Theorem are from [Hopf (1954)] generalising results from [Yosida and Kakutani (1939)]. A slightly weaker form was already obtained by [Wiener (1939)]. Our proof of Hopf's lemma is due to [Garsia (1965)].

The role of maximal inequalities for almost everywhere convergence results is known at least since [Kolmogoroff (1925)] and is demonstrated impressively in [Stein (1993)]. Employing a Baire category argument one can show that an abstract maximal inequality is indeed *necessary* for pointwise convergence results of quite a general type [Krengel (1985), Ch. 1, Theorem 7.2]; a thorough study of this connection has been carried out in [Stein (1961a)].

Finally, we recommend [Garsia (1970)] for more results on almost everywhere convergence.

10.3 Applications

Weyl's theorem revisited

Let $\alpha \in [0,1] \setminus \mathbb{Q}$. In Lecture 9 we proved *Weyl's theorem* stating that the sequence $(n\alpha \pmod 1)_{n\in\mathbb{N}}$ is equidistributed in $[0,1]$. The mean ergodicity of the involved operator with respect to the sup-norm implies the following stronger statement.

Corollary 10.13. *If $\alpha \in [0,1] \setminus \mathbb{Q}$, then for every interval $B = [a,b] \subset [0,1]$*

$$\frac{1}{n}\,\mathrm{card}\big\{ j \in [0,n) \cap \mathbb{N}_0 \,:\, x + j\alpha \pmod 1 \in B \big\} \;\to\; b-a$$

uniformly in $x \in [0,1]$.

The Pointwise Ergodic Theorem accounts for the analogous statement allowing for general Borel set B in place of just intervals. However, one has to pay the price of loosing convergence *everywhere*. Denoting by λ the Lebesgue measure on $\mathbb{R}$ we obtain the following result.

Corollary 10.14. *If $\alpha \in [0,1] \setminus \mathbb{Q}$, then for every Borel set $B \subset [0,1]$*

$$\frac{1}{n}\,\mathrm{card}\big\{ j \in [0,n) \cap \mathbb{N}_0 \,:\, x + j\alpha \pmod 1 \in B \big\} \;\to\; \lambda(B)$$

for almost every $x \in [0,1]$.

Proof. This is Exercise 10.3.

Borel's theorem on (simply) normal numbers

A number $x \in [0,1]$ is called **simply normal** (in base 10) if in its decimal expansion

$$x = 0,x_1x_2x_3\ldots \qquad x_j \in \{0,\ldots,9\}, \quad j \in \mathbb{N},$$

each digit appears asymptotically with frequency $1/10$. The following goes back to [Borel (1909)].

Theorem 10.15 (Borel). *Almost every number $x \in [0,1]$ is simply normal.*

Proof. First of all note that the set of numbers with non-unique decimal expansion is countable. Let $\varphi(x) = 10x \pmod 1$ for $x \in [0,1)$. As in Exercise 7.6, the MDS $([0,1],\mathfrak{B},\lambda;\varphi)$ is isomorphic to the Bernoulli shift (Section 5.1.5)

$$B(1/10,\dots,1/10) = (\mathscr{W}_{10}^{+},\Sigma,\mu;\tau).$$

The isomorphism is induced by the point isomorphism (modulo null sets)

$$\mathscr{W}_{10}^{+} \longrightarrow [0,1], \qquad (x_1,x_2,\dots) \longmapsto 0,x_1x_2\dots.$$

Since the Bernoulli shift is ergodic (Proposition 6.12), the MDS $([0,1],\mathfrak{B},\lambda;\varphi)$ is ergodic, too. Fix a digit $k \in \{0,\dots,9\}$ and consider

$$A := \{x \in [0,1) : x_1 = k\} = [k/10,(k+1)/10),$$

so $\lambda(A) = 1/10$. Then

$$\frac{\operatorname{card}\{1 \le j \le n : x_j = k\}}{n} = \frac{1}{n}\sum_{j=0}^{n-1} \mathbf{1}_A(\varphi^j(x)) \to \int_{[0,1]} \mathbf{1}_A \,\mathrm{d}\lambda = \lambda(A) = 1/10$$

for almost every $x \in [0,1]$ by Corollary 10.2. □

We refer to Exercise 10.6 for definition and analogous property of *normal* numbers.

The strong law of large numbers

In Lecture 9 we briefly pointed at a connection between the (Mean) Ergodic Theorem and the (weak) law of large numbers. In this section we shall make this more precise and prove the following classical result, going back to [Kolmogoroff (1933)]. We freely use the terminology common in probability theory, cf. [Billingsley (1979)].

Theorem 10.16 (Kolmogorov). *Let $(\Omega,\Sigma,\mathrm{P})$ be a probability space, and let $(X_n)_{n\in\mathbb{N}} \subset \mathrm{L}^1(\Omega,\Sigma,\mathrm{P})$ be a sequence of independent and identically distributed real random variables. Then*

$$\lim_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n) \;=\; \mathbb{E}(X_1) \qquad \mathrm{P}\text{-}\textit{almost surely.}$$

Proof. Since the X_j are identically distributed, $\mu := \mathrm{P}_{X_j}$ (the distribution of X_j) is a Borel probability measure on $\mathbb{R}$, independent of j, and

$$\mathbb{E}(X_1) = \int_{\mathbb{R}} t \,\mathrm{d}\mu(t)$$

is the common expectation. Define the product space

$$(\Omega', \Sigma', \mu') := \left(\mathbb{R}^{\mathbb{N}}, \bigotimes\nolimits_{\mathbb{N}} \mathfrak{B}(\mathbb{R}), \bigotimes\nolimits_{\mathbb{N}} \mu\right).$$

As mentioned in Lecture 5, the left shift τ is a measurable transformation of (Ω', Σ') and μ' is τ-invariant. The MDS $(\Omega', \Sigma', \mu'; \tau)$ is ergodic; this can be shown in exactly the same way as it was done for the Bernoulli shift (Proposition 6.12).

For $n \in \mathbb{N}$ let $Y_n : \Omega' \longrightarrow \mathbb{R}$ be the n^{th} projection and write $g := Y_1$. Then $Y_{j+1} = g \circ \tau^j$ for every $j \geq 0$, hence Corollary 10.2 yields that

$$\lim_{n\to\infty} \frac{1}{n}(Y_1 + \cdots + Y_n) \;=\; \lim_{n\to\infty} \frac{1}{n} \sum\nolimits_{j=0}^{n-1} (g \circ \tau^j) \;=\; \int_{\Omega'} g \, \mathrm{d}\mu'$$

pointwise μ'-almost everywhere. Note that $g_*\mu' = \mu$ and hence

$$\int_{\Omega'} g \, \mathrm{d}\mu' = \int_{\mathbb{R}} t \, \mathrm{d}\mu(t).$$

It remains to show that the $(Y_n)_{n\in\mathbb{N}}$ are in a certain sense "the same" as the originally given $(X_n)_{n\in\mathbb{N}}$. This is done by devising an injective lattice homomorphism

$$\Phi : \mathrm{L}^0(\Omega', \Sigma', \mu'; \overline{\mathbb{R}}) \longrightarrow \mathrm{L}^0(\Omega, \Sigma, \mathrm{P}; \overline{\mathbb{R}})$$

which carries Y_n to $\Phi(Y_n) = X_n$ for every $n \in \mathbb{N}$. Define

$$\varphi : \Omega \longrightarrow \Omega', \quad \varphi(\omega) := (X_n(\omega))_{n\in\mathbb{N}} \qquad (\omega \in \Omega).$$

Since $(X_n)_{n\in\mathbb{N}}$ is an independent sequence, the push-forward measure satisfies $\varphi_*\mathrm{P} = \mu'$. Let $\Phi = T_\varphi : f \longmapsto f \circ \varphi$ be the operator induced by φ, considered on $\mathrm{L}^0(\Omega', \Sigma', \mu'; \overline{\mathbb{R}})$. The operator , Φ is well-defined since φ is measure-preserving.

By construction, $\Phi Y_n = Y_n \circ \varphi = X_n$ for each $n \in \mathbb{N}$. Moreover, Φ is clearly a homomorphism of lattices (see Lecture 7) satisfying

$$\sup_{n\in\mathbb{N}} \Phi\big(f_n\big) = \Phi\big(\sup_{n\in\mathbb{N}} f_n\big) \quad \text{and} \quad \inf_{n\in\mathbb{N}} \Phi\big(f_n\big) = \Phi\big(\inf_{n\in\mathbb{N}} f_n\big)$$

for every sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{L}^0(\Omega', \Sigma', \mu'; \overline{\mathbb{R}})$. Since the almost everywhere convergence of a sequence can be described in purely lattice-theoretic terms involving only countable suprema and infima (cf. also (10.1)), one has

$$\lim_{n\to\infty} f_n = f \quad \mu'\text{-almost everywhere} \quad \Longrightarrow \quad \lim_{n\to\infty} \Phi(f_n) = \Phi(f) \quad \mathrm{P}\text{-almost surely}.$$

This for $f_n := (1/n)(Y_1 + \cdots + Y_n)$ and $f := \mathbb{E}(X_1)\mathbf{1}$ concludes the proof. □

Remark 10.17. By virtue of the same product construction one can show that the Mean Ergodic Theorem implies a general *weak* law of large numbers.

Skew Shifts

Let $\alpha \in [0,1]$ be an irrational number. By Example 2.19 and Theorem 9.11, the shift $\varphi_\alpha : x \longmapsto x+\alpha \pmod 1$ is a uniquely ergodic TDS. We consider here its group extension along the function $\Phi : [0,1) \longrightarrow [0,1)$, $\Phi(x) = x$, i.e.,

$$K := [0,1]^2 \pmod 1 \quad \text{and} \quad \psi(x,y) = (\varphi_\alpha(x), \Phi(x)+y) = (x+\alpha, x+y).$$

Then $(K;\psi)$ is a TDS, see Lecture 2, Example 2.10.3.

Our aim is to show that this TDS, called the **skew shift**, is uniquely ergodic. The Lebesgue measure λ on K is ψ-invariant, so it remains to prove that this measure is actually the only ψ-invariant probability measure on K. The next is an auxiliary result.

Lemma 10.18. *Let $\alpha \in [0,1]$ be an irrational number. Then the skew shift $(K, \mathfrak{B}(K), \lambda; \psi)$ is ergodic.*

Proof. To prove the assertion we use Proposition 7.8 and show that the fixed space of the induced operator $T = T_\psi$ on $\mathrm{L}^2(K,\mathfrak{B},\lambda)$ consists of the constant functions only. For $f,g \in \mathrm{L}^2[0,1]$ the function $f \otimes g : (x,y) \longmapsto f(x)g(y)$ belongs to $\mathrm{L}^2(K,\mathfrak{B},\lambda)$. Now every $F \in \mathrm{L}^2(K,\mathfrak{B},\lambda)$ can be written uniquely as an L^2-sum

$$F = \sum_{n\in\mathbb{Z}} f_n \otimes e_n$$

where $(f_n)_{n\in\mathbb{Z}} \subset \mathrm{L}^2[0,1]$ and $(e_n)_{n\in\mathbb{Z}}$ is the orthonormal basis

$$e_n(y) := \mathrm{e}^{2\pi i n y}, \qquad (y \in [0,1],\ n \in \mathbb{Z})$$

of $\mathrm{L}^2[0,1]$. Applying T to F yields

$$TF = \sum_{n\in\mathbb{Z}} (e_n T_{\varphi_\alpha} f_n) \otimes e_n,$$

hence the condition $F \in \mathrm{Fix}(T)$ is equivalent to

$$T_{\varphi_\alpha} f_n = e_{-n} f_n \quad \text{for every } n \in \mathbb{Z}.$$

Fix $n \in \mathbb{Z}$ with $f_n \neq 0$. We have to show that $n = 0$. Since

$$T_{\varphi_\alpha} |f_n| = \left|T_{\varphi_\alpha} f_n\right| = |e_{-n} f_n| = |f_n|$$

and α is irrational, $|f_n|$ is a constant (non-zero) function. Now take $\theta \in [0,1)$ rationally independent from α. Consider the operator $T_{\varphi_\theta} \in \mathscr{L}(\mathrm{L}^2[0,1]\,)$ induced by the shift φ_θ on $[0,1]$, and set $g_n := f_n / T_{\varphi_\theta} f_n$. Then we have

$$T_{\varphi_\alpha} g_n = T_{\varphi_\alpha} \frac{f_n}{T_{\varphi_\theta} f_n} = \frac{T_{\varphi_\alpha} f_n}{T_{\varphi_\theta} T_{\varphi_\alpha} f_n} = \frac{e_{-n} f_n}{e_{-n}(\theta) e_{-n} T_{\varphi_\theta} f_n} = e_n(\theta) g_n.$$

By Exercise 10.4, $e_n(\theta) = e_m(\alpha)$ for some $m \in \mathbb{Z}$, but since θ is rationally independent from α, it follows that $n = 0$ and thus $F = f_0 \otimes e_0$. Since $T_{\varphi_\alpha} f_0 = f_0$ and φ_α is ergodic, we have that f_0 and hence F is constant. □

We are now prepared for the main result.

Proposition 10.19. *The skew shift* $(K;\psi)$ *is uniquely ergodic.*

Proof. We use the concept of *generic points*, introduced in Exercise 10.5. By Lemma 10.18 the measure λ is ergodic for the skew shift ψ. Since $\mathrm{C}(K)$ is separable, Exercise 10.5.b) yields that λ-almost every point $x \in K$ is generic for λ, i.e.,

$$\frac{1}{n}\sum_{j=0}^{n-1} f(\psi^j(x,y)) \to \int_K f \,\mathrm{d}\lambda \qquad (n \to \infty)$$

for all $f \in \mathrm{C}(K)$, for λ-almost all $(x,y) \in K$. If (x,y) is generic for λ, then $(x, y+\theta \pmod 1)$ is generic for λ, too, for every $\theta \in [0,1)$. This holds since the mapping

$$\rho_\theta : (x,y) \longmapsto (x, y+\theta \pmod 1)$$

commutes with ψ and preserves the measure λ. Hence the set of points which are *not* λ-generic is contained in a set of the form $A \times [0,1]$ for some Borel set $A \subset [0,1]$ of measure zero.

Now let $\mu \in \mathrm{M}^1(K)$ be an arbitrary ψ-invariant probability measure on K. Denoting by $\pi : [0,1]^2 \longrightarrow [0,1]$ the projection to the first coordinate, we obtain that $\pi_*\lambda$ and $\pi_*\mu$ are both φ_α-invariant Borel probability measures on $[0,1]$. Since α is irrational, the shift $([0,1];\varphi_\alpha)$ is uniquely ergodic (Theorem 9.11), hence $\pi_*\lambda = \pi_*\mu$. In particular, $\mu(A\times[0,1]) = \lambda(A \times [0,1]) = 0$. For fixed $f \in \mathrm{C}(K)$ we hence obtain that

$$\frac{1}{n}\sum_{j=0}^{n-1} f(\psi^j(x,y)) \to \int_K f \,\mathrm{d}\lambda \qquad (n \to \infty)$$

for μ-almost all $(x,y) \in K$. Integrating with respect to μ, the Dominated Convergence Theorem yields

$$\int_K f \,\mathrm{d}\mu = \int_K \frac{1}{n}\sum_{j=0}^{n-1} f \circ \psi^j \,\mathrm{d}\mu \;\to\; \int_K f \,\mathrm{d}\lambda.$$

Hence $\langle f, \mu \rangle = \langle f, \lambda \rangle$ for all $f \in \mathrm{C}(K)$, and thus $\mu = \lambda$. □

Since λ is strictly positive, the TDS $(K;\psi)$ is even strictly ergodic. So by Corollary 9.10 we obtain that this TDS is minimal and its induced operator is mean ergodic. This leads to the following number-theoretic application.

Proposition 10.20. *Let* $\alpha \in [0,1] \setminus \mathbb{Q}$. *Then the sequence* $(n^2\alpha \pmod 1)_{n=0}^\infty$ *is equidistributed in* $[0,1]$, *i.e.,*

$$\lim_{n\to\infty} \frac{\operatorname{card}\{\, j : 0 \le j < n,\ n^2\alpha \in [a,b] \pmod 1 \,\}}{n} = b - a$$

holds for any $0 \le a < b \le 1$.

Proof. Consider the TDS $(K;\psi_{2\alpha})$, whose induced operator $T_{\varphi_{2\alpha}}$ is now known to be mean ergodic on $\mathrm{C}(K)$. In particular, the Cesàro means A_n of the induced operator converge to the mean ergodic projection P in the supremum norm, hence pointwise (cf. also Exercise 9.1). We have seen in Corollary 9.10 that $Pf = \int_K f \, \mathrm{d}\lambda$ for any $f \in \mathrm{C}(K)$. Let now $g \in \mathrm{C}[0,1]$ be arbitrary and set $f(x,y) := g(y)$. It is easy to see that one has $\psi_{2\alpha}^n(\alpha,0) = ((2n+1)\alpha, n^2\alpha)$. So the convergence

$$\frac{1}{n}\sum_{j=0}^{n-1} g(n^2\alpha \pmod 1) = \mathrm{A}_n f(\alpha,0) \to \int_K f \, \mathrm{d}\lambda = \int_0^1 g(s)\, \mathrm{d}s$$

and Proposition 9.14 conclude the proof. □

Final Remark: Birkhoff vs. von Neumann

From the point of view of statistical mechanics, Birkhoff's theorem seems to outrun von Neumann's. By virtue of the Dominated Convergence Theorem and the density of L^∞ in L^2, the latter is even a corollary of the former (cf. the presentation in [Walters (1982), Corollary 1.14.1]). Reed and Simon take a moderate viewpoint when they write in [Reed and Simon (1972), p.60] (annotation in square brackets by the authors).

> This [i.e., the pointwise ergodic] theorem is closer to what one wants to justify [in?] statistical mechanics than the von Neumann theorem, and it is fashionable to say that the von Neumann theorem is unsuitable for statistical mechanics. We feel that this is an exaggeration. If we had only the von Neumann theorem we could probably live with it quite well. Typically, initial conditions are not precisely measurable anyway, so that one could well associate initial states with measures $f\mathrm{d}\mu$ where $\int f \, \mathrm{d}\mu = 1$, in which case the von Neumann theorem suffices. However, the Birkhoff theorem does hold and is clearly a result that we are happier to use in justifying the statement that phase-space averages and time averages are equal.

However, von Neumann's theorem inspired the operator-theoretic concept of mean ergodicity and an enormous amount of research in the field of asymptotics of discrete (and continuous) operator semigroups with tantamount applications to various other fields. Certainly it would be too much to say that Birkhoff's theorem is overrated, but von Neumann's theorem should not be underestimated either.

Exercises

1.

a) Let (Ω, Σ, μ) be a finite measure space and let T be a positive operator on $\mathrm{L}^1(\Omega, \Sigma, \mu)$. Show that T is a Dunford–Schwartz operator if and only if $T'\mathbf{1} \le \mathbf{1}$ and $T\mathbf{1} \le \mathbf{1}$. siehe Exercise 8.5!!!

b) Let K be a compact topological space, let T be a Markov operator on $\mathrm{C}(K)$, and let $\mu \in \mathrm{M}^1(K)$ be such that $T'\mu \le \mu$. Show that T extends uniquely to a positive Dunford–Schwartz operator on $\mathrm{L}^1(K, \mathfrak{B}_0(K), \mu)$.

2. Let (Ω, Σ, μ) be a measure space, $1 \le p < \infty$, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of bounded linear operators on $X = \mathrm{L}^p(\Omega, \Sigma, \mu)$. Moreover, let T be a bounded linear operator on X. If the associated maximal operator T^* satisfies a maximal inequality, then the set

$$C := \left\{ f \in X \,:\, T_n f \to T f \text{ almost everywhere} \right\}$$

is a closed subspace of X.

3. Prove Corollaries 10.13 and 10.14.

4. Let $\alpha \in [0,1)$ be an irrational number and consider the shift $\varphi_\alpha : x \longmapsto x + \alpha \pmod 1$ on $[0,1]$. Show that for the point spectrum of the induced operator $T = T_{\varphi_\alpha}$ on $\mathrm{L}^2[0,1]$ we have $\sigma_p(T) = \{\mathrm{e}^{2\pi i m\alpha} : m \in \mathbb{Z}\}$. (Hint: Use a Fourier expansion as in the proof of Proposition 7.9.)

5. Let $(K;\varphi)$ be a TDS with induced operator $T = T_\varphi$ and associated Cesàro averages $\mathrm{A}_n = \mathrm{A}_n[T]$, $n \in \mathbb{N}$. Let μ be a φ-invariant probability measure on K. A point $x \in K$ is called **generic** for μ if

$$\lim_{n\to\infty} (\mathrm{A}_n f)(x) = \int_K f \,\mathrm{d}\mu \tag{10.3}$$

for *all* $f \in \mathrm{C}(K)$.

a) Show that $x \in K$ is generic for μ if (10.3) holds for each f from a dense subset of $\mathrm{C}(K)$.

b) Show that in the case that $\mathrm{C}(K)$ is separable and μ is ergodic, μ-almost every $x \in K$ is generic for μ. (Hint: apply Corollary 10.2 to every f from a countable dense set $D \subset \mathrm{C}(K)$.)

6. A number $x \in [0,1]$ is called **normal** (in base 10) if every finite combination (of length k) of the digits $\{0,1,\dots,9\}$ appears in the decimal expansion of x with asymptotic frequency 10^{-k}. Prove that almost all numbers in $[0,1]$ are normal.

$*$ **7.**[Dominated Ergodic Theorem] Let (Ω, Σ, μ) be a finite measure space and let $f \in \mathrm{L}^0_+(\Omega, \Sigma, \mu)$. Show that

$$\int_\Omega f \,\mathrm{d}\mu = \int_0^\infty \mu[\, f > t\,] \,\mathrm{d}t$$

and derive from this that $\|f\|_p^p = \int_0^\infty p t^{p-1} \mu[\,|f| > t\,] \,\mathrm{d}t$ for all $f \in \mathrm{L}^0(\Omega, \Sigma, \mu)$ for $1 \le p < \infty$.

Now let T be a Dunford–Schwartz operator on $\mathrm{L}^1(\Omega,\Sigma,\mu)$ with its Cesàro averages $(\mathrm{A}_n)_{n\in\mathbb{N}}$ and the associated maximal operator A^*. Use the Maximal Ergodic Theorem to show that

$$\|\mathrm{A}^* f\|_p \le \frac{p}{p-1}\,\|f\|_p \qquad (1 < p < \infty, \quad f \in \mathrm{L}^p(\Omega,\Sigma,\mu)).$$

$*$ **8.** Prove Theorem 10.4 for σ-finite measure spaces.

Lecture 11
Compact Semigroups and Groups

I hail a semigroup when I see one and I seem to see them everywhere! [1]

Einar Hille[2]

Compact groups have been introduced already in Lecture 2 as simple examples of TDSs. In Lecture 5 they reappeared, endowed with their Haar measure, as simple examples of MDSs. As we shall see later, these examples are by no means artificial: within a structure theory of general dynamical systems they form the fundamental building blocks.

Another reason is that groups (or at least semigroups) are present in the mere setting of general dynamical systems already. If $(K;\varphi)$ is an invertible TDS, then $\{\varphi^n : n \in \mathbb{Z}\}$ is an Abelian group of transformations of K and — passing to the induced operator — $\{T_\varphi^n : n \in \mathbb{Z}\}$ is a group of operators on $\mathrm{C}(K)$. Analogous remarks hold for MDSs, of course. If the dynamical system is not invertible, then one obtains *semi*groups instead of groups (see definition below), and indeed, many notions and results from Lecture 2 and Lecture 3 can be generalised to general (semi)group actions on compact spaces (see, e.g., [Gottschalk and Hedlund (1955)]).

In this lecture we shall present some facts about compact semigroups and groups. We concentrate on those that are relevant for a deeper analysis of dynamical systems, e.g., for the decomposition theorem to come in the following lecture. Our treatment is therefore far from being complete and the reader is referred to [Hofmann and Mostert (1966)], [Hewitt and Ross (1979)], [Rudin (1990)] and [Hofmann and Morris (2006)] for further information.

[1] Hille continues: "... Friends have observed, however, that there are mathematical objects which are not semigroups."

[2] Functional Analysis and Semigroups, AMS Coll. Publ. vol. 31, Providence R.I., 1948. — Foreword

11.1 Compact Semigroups

A **semigroup** is a non-empty set S with an operation $S \times S \longrightarrow S$, $(s,t) \longmapsto st := s \cdot t$ — usually called *multiplication* — which is associative, i.e., one has $r(st) = (rs)t$ for all $r,s,t \in S$. If the multiplication is commutative (i.e., $st = ts$ for all $s,t \in S$), then the semigroup is called **Abelian**. An element e of a semigroup S is called an **idempotent** if $e^2 = e$, and is called a **neutral element** if $se = es = s$ for all $s \in S$. It is easy to see that there may be many idempotents but at most one neutral element in a semigroup. The semigroup S is called a **group** if it has a neutral element e and for every $s \in S$ the equation $sx = xs = e$ has a solution x. It is easy to see that this solution is unique, called the **inverse** s^{-1} of s. (We assume the reader to be familiar with the fundamentals of group theory and refer, e.g., to [Lang (2005)].)

For subsets $A,B \subset S$ and elements $s \in S$ we write

$$sA = \{sa \,:\, a \in A\}, \quad As = \{as \,:\, a \in A\}, \quad AB := \{ab \,:\, a \in A, b \in B\}.$$

An **ideal** of S is a non-empty subset J such that $JS \cup SJ \subset J$. Obviously, in an Abelian semigroup this condition reduces to $JS \subset J$.

As we are doing analysis, semigroups are interesting to us only if endowed with a topology which is in some sense related to the algebraic structure. One has different possibilities here.

Definition 11.1. A semigroup endowed with a topology is called a **semitopological semigroup** if for every $s \in S$ the *right* and the *left multiplication* with s is continuous on S, i.e., if the mappings

$$\rho_s : t \longmapsto ts \quad \text{and} \quad \varphi_s : t \longmapsto st$$

are continuous for all $s \in S$. A semigroup S endowed with a topology is called a **topological semigroup** if the multiplication mapping $S \times S \longrightarrow S$ is continuous. A topological semigroup which is a group is called a **topological group** if also the inversion mapping $s \longmapsto s^{-1}$ is continuous.

One says that in a semitopological semigroup the multiplication is *separately* continuous, whereas in a topological semigroup it is *jointly* continuous. Of course there are examples of semitopological semigroups that are not topological semigroups, i.e., such that the multiplication is not jointly continuous (see Exercise 11.1). The example particularly interesting for us is $\mathscr{L}(X)$, X a Banach space, endowed with the weak operator topology. By Proposition C.19 and Example C.20 this is a semitopological semigroup, which is in general not a topological semigroup. hemce

A semitopological semigroup is called **compact** if its topology is compact. Compact semitopological *Abelian* semigroups have a remarkable property.

Theorem 11.2. *Every compact semitopological Abelian semigroup S contains a unique minimal (two-sided) ideal $K(S)$. Moreover, $K(S)$ is compact and a group.*

Proof. The set $\mathscr{M}$ of all closed (two-sided) ideals in S is non-empty, since $S \in \mathscr{M}$. Moreover, it has the finite intersection property: if $J_1, \ldots, J_n$ are closed ideals,

then
$$\emptyset \neq J_1 J_2 \ldots J_n \subset \bigcap_{k=1}^{n} J_k.$$

By compactness $K(S) := \bigcap \mathscr{M}$ is a non-empty compact set, therefore a closed ideal.

Now let J be *any* ideal of S, and let $s \in J$. Then sS is an ideal, and compact by separate continuity. Hence sS is closed and, by construction, $K(S) \subset sS \subset J$. Hence $K(S)$ is contained in every ideal of S, whence it is the unique minimal ideal of S.

To show that $K(S)$ is a group, note that $tK(S) = K(S)$ for every $t \in K(S)$. Indeed, $K(S) \subset tK(S)$ (since $tK(S)$ is an ideal) and $tK(S) \subset K(S)$ (since $K(S)$ is an ideal). Fix now $s \in K(S)$ and let $q \in K(S)$ be such that $sq = s$. For each $r \in K(S)$ there exists $r' \in K(S)$ such that $r's = r$. This implies that $rq = r'sq = r's = r$, i.e., q is a neutral element in $K(S)$. Again from $sK(S) = K(S)$ we infer the existence of $s' (= s^{-1})$ such that $ss' = q$. □

To understand minimal ideals in compact Abelian semitopological semigroups, we have to study therefore compact semitopological groups. Surprisingly, every such group is actually topological as the following celebrated result of Ellis shows (see [Ellis (1957)] or [Hindman and Strauss (1998), Section 2.5]).

Theorem 11.3 (Ellis). *Let G be a semitopological group whose topology is locally compact. Then G is a topological group.*

Combining Theorem 11.2 with Ellis' result, we obtain the following.

Corollary 11.4. *Let S be a compact semitopological Abelian semigroup. Then its minimal ideal $K(S)$ is a compact (topological) group.*

The proof of Theorem 11.3 is fairly involved. The major difficulty is deducing joint continuity from separate continuity. One possible strategy to achieve this is due to Namioka [Namioka (1974)].

1) Find *at least one point* $(g,h) \in G \times G$ at which multiplication is continuous.
2) Step 1 is accomplished by finding, as a matter of fact, *many* points where the multiplication is jointly continuous. More precisely, one shows that the set of points where the multiplication is *not* continuous is of *first category*.
3) By separate continuity of the multiplication and by using the group property conclude that the multiplication is continuous at *all points* $(g,h) \in G \times G$.

Once Step 1 (actually, Step 2) is established, Step 3) is trivial. For Step 2) there is a relatively short proof if G is compact and metrisable, while the general statement is in [Namioka (1974)].

Lemma 11.5. *Let K be a complete metric space and $f : K \times K \longrightarrow K$ a separately continuous function. Then the set of continuity of f is a dense G_δ set (hence its complement is of first category).*

Proof. Consider the function

$$D(x,y) := \inf\big\{\operatorname{diam} f(U) : (x,y) \in U \text{ open}\big\},$$

and set $U_n := [D < 1/n]$. It is a routine exercise to show that $U_n \subset K \times K$ is open and that $G := \bigcap_{n\in\mathbb{N}} U_n$ is actually the set of points where f is continuous. We prove below that U_n is dense in $K \times K$. Then by Theorem A.5.c) G is a dense G_δ set, hence its complement is of first category.

Let $m \in \mathbb{N}$. We show that U_m is dense in $K \times K$. Set $\varepsilon := \frac{1}{2m}$ and take $V \subset K$ open with $b \in V$. For $(x,b) \in K \times K$ set

$$A_{x,b,\varepsilon} := \big\{0 \le \delta \le 1 : d(b,y) \le \delta \implies d(f(x,b), f(x,y)) \le \varepsilon\big\}.$$

By the continuity in the second variable, we have that $A_{x,b,\varepsilon}$ is non-empty. We claim that the function $g_{b,\varepsilon} : K \longrightarrow \mathbb{R}_+$, $g_{b,\varepsilon}(x) := \sup A_{x,b,\varepsilon}$ is upper semicontinuous, i.e., for any $c \ge 0$ the set $\big[g_{b,\varepsilon} \ge c\big]$ is closed. If $c = 0$, this is trivial, so assume $c > 0$ and let $x_n \to x$ with $g_{b,\varepsilon}(x_n) \ge c$. Then there is $c - 1/n \le \delta_n \in A_{x_n,b,\varepsilon}$. Now let $\delta := \liminf_{n\to\infty} \delta_n \ge c$ and $0 < \delta' < \delta$. There is some $n_0 \in \mathbb{N}$ with $0 < \delta' < \delta_n$ for all $n \ge n_0$, so if $d(b,y) \le \delta' < \delta_n$, then $d(f(x_n,b), f(x_n,y)) \le \varepsilon$ for all $n \ge n_0$. Thus, by continuity of f in the first variable, $d(f(x,b), f(x,y)) \le \varepsilon$. This shows $\delta' \in A_{x,b,\varepsilon}$, so $\delta' \le \sup A_{x,b,\varepsilon}$ and by arbitrariness of δ' we obtain $c \le \delta \le g_{b,\varepsilon}(x)$. Now consider the closed sets $F_{b,\varepsilon,k} := \big[g_{b,\varepsilon} \ge 1/k\big]$, which by the above cover K, hence for the open set $V \subset K$ we have $V = \bigcup_{k\in\mathbb{N}} (F_{b,\varepsilon,k} \cap V)$. By the Baire Category Theorem A.4 there is $k \in \mathbb{N}$ such that $F_{b,\varepsilon,k} \cap V$ contains a ball $\mathrm{B}(a,r)$, $r > 0$. If we take $0 < r$ possibly smaller, we can achieve that $\mathrm{B}(a,r) \subset [d(f(a,b), f(\cdot,b)) < \varepsilon]$ (use continuity in the first variable). If $d(a,x) < r$ and $d(b,y) < 1/k$, then $d(f(a,b), f(x,b)) + d(f(x,b), f(x,y)) < 2\varepsilon$. This shows that with $U := \mathrm{B}(a,r) \times \mathrm{B}(b,1/k)$, $\varepsilon = \frac{1}{2m}$ we have $\operatorname{diam} f(U) < \frac{1}{m}$, hence $(a,b) \in U_m \cap (V \times V)$, establishing the desired density of U_m □

Proof of Ellis' Theorem (compact, metrisable case). Apply Lemma 11.5 to the multiplication to obtain joint continuity at one point, then Step 3) to obtain the joint continuity everywhere.

We prove that the mapping $g \mapsto g^{-1}$ is continuous on G. Let $g_n \to g$. It suffices to show that there is a subsequence of $(g_n^{-1})_{n\in\mathbb{N}}$ convergent to g^{-1}. By compactness, however, we find a subsequence $(g_{n_k}^{-1})_{k\in\mathbb{N}}$ convergent to some h as $k \to \infty$. By the joint continuity of the multiplication $1 = g_{n_k} g_{n_k}^{-1} \to gh$, whence $h = g^{-1}$. □

As we have seen, by passing to the minimal ideal we are naturally led from compact semitopological Abelian semigroups to compact Abelian *topological* groups. Their study is the subject of the next section.

11.2 Compact Groups

In the following, G always denotes a topological group with neutral element $1 = 1_G$. For a set $A \subset G$ we write $A^{-1} := \{a^{-1} : a \in A\}$ and call it **symmetric** if $A^{-1} = A$. Note that A is open if and only A^{-1} is open, since the inversion mapping is a homeomorphism of G. In particular, every open neighbourhood U of 1 contains a symmetric open neighbourhood of 1, namely $U \cap U^{-1}$.

Since the left and right multiplications by any given element $g \in G$ are homeomorphisms of G, the set of open neighbourhoods of the unit element 1 completely determines the topology of G. Indeed, U is an open neighbourhood of 1 if and only if Ug (or gU) is an open neighbourhood of $g \in G$.

The following are the basic properties of topological groups needed later.

Lemma 11.6. *For a topological group G the following statements hold.*

a) *If V is an open neighbourhood of* 1*, then there is* $1 \in W$ *open, symmetric set with* $WW \subset V$.
b) *If H is a topological group and $\varphi : G \longrightarrow H$ is a group homomorphism, then φ is continuous if and only if it is continuous at 1_G.*
c) *If H is a normal subgroup of G, then the factor group G/H is a topological group with respect to the quotient topology.*
d) *If H is a (normal) subgroup of G, then $\overline{H}$ is also a (normal) subgroup of G.*

Proof. To prove a) we use that the multiplication is (jointly) continuous to find $1 \in U$ open with $UU \subset V$. Then $W := U \cap U^{-1}$ has the desired property. For b), c) and d) see Exercise 11.2 and Exercise 11.3. □

A topological group is called a (locally) **compact group** if its topology is (locally) compact. Important (locally) compact groups are the additive groups $\mathbb{R}^d$ and the (multiplicative) tori $\mathbb{T}^d$, $d \in \mathbb{N}$. Furthermore, every group G can be made a locally compact group just by endowing it with the discrete topology. (To indicate this choice of topology we shall write G_{d}.)

Proposition 11.7. *Let G be a compact group, and let $f \in \mathrm{C}(G)$. Then f is* ***uniformly continuous****, i.e., for any $\varepsilon > 0$ there is a symmetric open neighbourhood U of* 1 *such that $gh^{-1} \in U$ implies $|f(g) - f(h)| < \varepsilon$.*

Proof. Fix $\varepsilon > 0$. For each $g \in G$ the set

$$V_g := \big\{z \in G : |f(zg) - f(g)| < \varepsilon/2\big\}$$

is an open neighbourhood of 1, by continuity of f and ρ_g. By Lemma 11.6.a) there is a symmetric open neighbourhood W_g of 1 such that $W_g W_g \subset V_g$. Now $G = \bigcup_{g \in G} W_g g$, and hence by compactness $G = \bigcup_{g \in F} W_g g$ for some finite set $F \subset G$. Define $U := \bigcap_{g \in F} W_g$, which is a symmetric open neighbourhood of 1. For $xy^{-1} \in U$ we find $g \in F$ such that $y \in W_g g \subset V_g g$. Hence $x \in Uy \subset W_g W_g g \subset V_g g$. Therefore $xg^{-1}, yg^{-1} \in V_g$, which implies that

$$|f(x) - f(y)| \le |f(x) - f(g)| + |f(y) - f(g)| < \varepsilon. \qquad \square$$

We have the following corollary concerning the operators T_{φ_g} and T_{ρ_g} induced on $\mathrm{C}(G)$ by left and right multiplications.

Corollary 11.8. *Let G be a compact group and let $f \in \mathrm{C}(G)$. Then the mappings*

$$G \longrightarrow \mathrm{C}(G), \quad g \longmapsto T_{\varphi_g} f \tag{11.1}$$
$$G \longrightarrow \mathrm{C}(G), \quad g \longmapsto T_{\rho_g} f \tag{11.2}$$

are continuous.

Proof. Let $\varepsilon > 0$. Choose U as in Proposition 11.7. If $gh^{-1} \in U$, then $(gx)(hx)^{-1} = gh^{-1} \in U$ for every $x \in G$. Hence

$$\left\| T_{\varphi_g} f - T_{\varphi_h} f \right\|_\infty = \sup_{x \in G} |f(gx) - f(hx)| \le \varepsilon.$$

This proves the first statement. To prove the second, apply the first to the function $f^\sim(x) := f(x^{-1})$ and note that $g \longmapsto g^{-1}$ is a homeomorphism of G. $\square$

As a consequence of Corollary 11.8 we can immediately deduce the following result already announced in Example 8.14.

Corollary 11.9. *Let G be a compact group and let $a \in G$. Then the operator $T := T_{\varphi_a}$ induced on $\mathrm{C}(G)$ by the group rotation $(G; \varphi_a)$ is mean ergodic.*

Proof. Let $f \in \mathrm{C}(G)$. Then by Corollary 11.8 the set $\{T_{\varphi_g} f : g \in G\} \subset \mathrm{C}(G)$ is compact. A fortiori, the set $\{T^n f : n \in \mathbb{N}_0\}$ is relatively compact. Consequently, its closed convex hull is compact, whence Theorem 8.8 (iii) concludes the proof. $\square$

The Haar-measure

As explained in Lecture 5, on any compact group there is a unique translation invariant (Baire) probability measure μ, called the Haar measure. The Haar measure is also inversion invariant. We often write simply $\mathrm{L}^p(G)$ in place of $\mathrm{L}^p(G, \mathfrak{B}_0(G), \mu)$ and write $\int_G f(x)\,\mathrm{d}x$ to indicate integration with respect to the Haar measure.

11.3 The Character Group

An important tool in the study of locally compact Abelian groups is the so-called character group. A **character** of a locally compact group G is a continuous homomorphism $\chi : G \longrightarrow \mathbb{T}$. The set

$$G^* := \left\{ \chi : \chi \text{ character of } G \right\}$$

is called the **character** or **dual group**. The dual group G^* is indeed an (Abelian) group with respect to pointwise products, the constant function $\mathbf{1}$ being its neutral element.

For many locally compact Abelian groups the dual group can be described explicitly. For example, the character group of $\mathbb{R}$ is given by

$$\mathbb{R}^* = \big\{t \longmapsto \mathrm{e}^{2\pi\mathrm{i}\alpha t} : \alpha \in \mathbb{R}\big\},$$

and the character group of $\mathbb{T} \simeq [0,1)$ by

$$\mathbb{T}^* = \big\{t \longmapsto \mathrm{e}^{2\pi\mathrm{i}nt} : n \in \mathbb{Z}\big\},$$

i.e., $\mathbb{R}^* \simeq \mathbb{R}$ and $\mathbb{T}^* \simeq \mathbb{Z}$ (see Examples 11.21 and Exercise 11.5; there are elementary proofs for these equalities, however, we prove the second one later by using tools from the general theory). In these examples, the character group seems to be quite "rich", and an important result says that this is actually a general fact.

Theorem 11.10. *Let G be a locally compact Abelian group. Then its dual group G^* separates the points of G, i.e., for every $1 \neq a \in G$ there is a character $\chi \in G^*$ such that $\chi(a) \neq 1$.*

The proof of this result, in the case G is compact or discrete, is in the supplement to this lecture. Here we look at some of its consequences. The first resembles a standard application of the Hahn–Banach theorem. The proof is Exercise 11.6.

Corollary 11.11. *Let G be a locally compact Abelian group and let H be a closed subgroup of G. If $g \notin H$, then there is a character $\chi \in G^*$ with $\chi(g) \neq 1$ and $\chi|_H = 1$.*

We now turn to compact groups, where $G^* \subset \mathrm{C}(G) \subset \mathrm{L}^2(G)$. The following is a preliminary result.

Proposition 11.12 (Orthogonality). *Let G be a compact Abelian group. Then G^* is an orthonormal set in $\mathrm{L}^2(G)$.*

Proof. Let $\chi_1, \chi_2 \in G^*$ be two characters and consider $\chi := \chi_1\overline{\chi_2}$. Then

$$\alpha := \big(\chi_1 \,|\, \chi_2\big) = \int_G \chi_1\overline{\chi_2}\,\mathrm{d}\mu = \int_G \chi(g)\,\mu(\mathrm{d}g) = \int_G \chi(gh)\,\mu(\mathrm{d}g) = \chi(h)\alpha$$

for every $h \in G$. Hence either $\chi = \mathbf{1}$ (in which case $\chi_1 = \chi_2$) or $\alpha = 0$. □

Using the orthogonality of the characters, we obtain the following.

Proposition 11.13. *Let G be a compact Abelian group and let $X \subset G^*$ be a subset separating the points of G. Then the subgroup $\langle X\rangle$ generated by X is the whole of G^* and $\operatorname{lin}\langle X\rangle$ is (sup-norm) dense in $\mathrm{C}(G)$.*

Proof. Consider $\mathscr{A} = \operatorname{lin}\langle X\rangle$ which is a conjugation invariant subalgebra(!) of $\mathrm{C}(G)$ separating the points of G. So by the Stone–Weierstraß Theorem 4.3 it is dense in $\mathrm{C}(G)$. If there is $\chi \in G^* \setminus \langle X\rangle$, take $f \in \mathscr{A}$ with $\|\chi - f\|_\infty < 1$. Then by Proposition 11.12 χ is orthogonal to $\mathscr{A}$ and we obtain the following contradiction

$$1 > \|f-\chi\|_{\mathrm{L}^2}^2 = \|f\|_{\mathrm{L}^2}^2 - \big(f \,|\, \chi\big) - \big(\chi \,|\, f\big) + \|\chi\|_{\mathrm{L}^2}^2 = 1 + \|f\|_{\mathrm{L}^2}^2 \geq 1. \qquad \square$$

Because of Theorem 11.10 we can specialise $X = G^*$ in Proposition 11.13 and obtain the following.

Corollary 11.14. *Let G be a compact Abelian group. Then the set* $\operatorname{lin} G^*$ *of "trigonometric polynomials" is dense in* $\mathrm{C}(G)$.

Since $\operatorname{lin} G^*$ is dense in $\mathrm{C}(G)$, it is dense in $\mathrm{L}^2(G)$, and we obtain our final result on G^*.

Corollary 11.15. *Let G be a compact Abelian group. Then G^* is an orthonormal basis of* $\mathrm{L}^2(G)$.

Topology on the Character Group

Let G be a locally compact Abelian group. In this section we describe a topology turning G^* into a locally compact group as well.

Consider the product space $\mathbb{T}^G$ of all functions from G to $\mathbb{T}$ and endow it with the product topology (Appendix A.5). Then by Tychonov's Theorem A.2, this is a compact space, and it is an easy exercise to show that it is actually a *compact group* with respect to the pointwise operations. The subspace topology on $G^* \subset \mathbb{T}^G$, called the **pointwise topology**, makes G^* a topological group. However, only in exceptional cases this topology is (locally) compact. Instead we consider another, generally finer, topology. This is the topology of uniform convergence on compact sets, called the **compact-open topology**, which again makes G^* a topological group. Unless otherwise specified, we always take this topology on the dual group. The important examples for us are the following.

Examples 11.16. 1) If the group G is compact, the compact-open topology is the same as the topology inherited from the sup-norm topology on $\mathrm{C}(G)$.

2) If the group G is discrete, then the compact-open topology is just the pointwise topology.

One has the following nice duality.

Proposition 11.17. *Let G be a locally compact Abelian group. If G is compact, then, G^* is discrete; and if G is discrete, then G^* is compact.*

Proof. Suppose that G is compact. Take $\chi \in G^*$, then $\chi(G) \subset \mathbb{T}$ is a compact subgroup of $\mathbb{T}$. If $\chi \in G^*$ satisfies

$$\|\mathbf{1} - \chi\|_\infty = \sup_{g \in G} |1 - \chi(g)| < \sqrt{3},$$

then actually $\chi(G) = \{1\}$, hence $\chi = \mathbf{1}$. So we obtain that $\{\mathbf{1}\}$ is an open neighbourhood of $\mathbf{1}$, thus G^* is discrete.

Suppose G to be discrete. By Example 11.16.2 the dual group G^* has the pointwise topology. For any $g \in G$ consider the projection $\pi_g : \mathbb{T}^G \longrightarrow \mathbb{T}$, $\pi_g(\psi) = \psi(g)$. Since any group homomorphism $\chi : G \longrightarrow \mathbb{T}$ is continuous, we have

$$G^* = \bigcap_{g,h \in G} \left[\pi_{gh} \pi_h^{-1} \pi_g^{-1} = 1 \right].$$

However, the functions $\pi_{gh}\pi_h^{-1}\pi_g^{-1}$ are continuous from $\mathbb{T}^G$ to $\mathbb{T}$ for any $g, h \in G$, so G^* is closed, hence compact in $\mathbb{T}^G$. □

Thus we see that if either G is compact or G is discrete, then G^* endowed with the compact-open topology will be a locally compact group. It is actually true that the compact-open topology on G^* is always locally compact, see [Hewitt and Ross (1979), §23].

The next is an auxiliary result, whose proof we leave as Exercise 11.9.

Proposition 11.18. *Let G and H be locally compact Abelian groups. Then the product group $G \times H$ with the product topology is a locally compact topological group. For any $\chi \in G^*$ and $\psi \in H^*$ we have $\chi \otimes \psi \in (G \times H)^*$, where $\chi \otimes \psi(x,y) := \chi(x)\psi(y)$. Moreover, the mapping*

$$\Psi : G^* \times H^* \longrightarrow (G \times H)^*, \quad (\chi, \psi) \longmapsto \chi \otimes \psi$$

is a topological isomorphism.

11.4 The Pontryagin Duality Theorem

For a locally compact Abelian group G consider the compact space $\mathbb{T}^{G^*}$ which endowed with the pointwise multiplication is a compact group. Define the mapping

$$\Phi : G \longrightarrow \mathbb{T}^{G^*}, \quad \Phi(g)\chi = \chi(g). \tag{11.3}$$

It is easy to see that Φ is a continuous homomorphism, and by Theorem 11.10 it is injective. It is Exercise 11.7 to show that $\Phi(g)$ is actually a character of the dual group G^* for any $g \in G$, and that the mapping

$$\Phi : G \longrightarrow G^{**} := (G^*)^*$$

is continuous. Now, the fundamental theorem of Pontryagin asserts that G is topologically isomorphic to its **bi-dual group** G^{**} under the mapping Φ.

Theorem 11.19 (Pontryagin Duality Theorem). *For G a locally compact Abelian group the mapping $\Phi : G \longrightarrow G^{**}$ defined in* (11.3) *is a topological isomorphism.*

We prove this theorem only for discrete or compact groups. By the above, it remains prove that Φ is surjective and its inverse in continuous.

Proof. First, suppose that G is discrete. Then, by Proposition 11.17, G^* is compact and the subgroup $\operatorname{ran}\Phi \subset G^{**}$ trivially separates the points of G^*. By Proposition 11.13 we have $\operatorname{ran}\Phi = G^{**}$. Since both G and G^{**} are discrete, the mapping Φ is a homeomorphism.

Let now G be compact. Then G^* is discrete and G^{**} is compact by Proposition 11.17. Since Φ is continuous and injective, $\operatorname{ran}\Phi$ is a compact hence closed subgroup of G^{**}. If $\operatorname{ran}\Phi \neq G^{**}$, then by Corollary 11.11 there is a character $\gamma \neq \mathbf{1}$ of G^{**} with $\gamma|_{\operatorname{ran}\Phi} = 1$. Now, as G^* is discrete, by the already proved first part we see that there is a $\chi \in G^*$ with $\gamma(\varphi) = \varphi(\chi)$ for all $\varphi \in G^{**}$. In particular, for $g \in G$ and $\varphi = \Phi(g)$ we have $\chi(g) = \Phi(g)\chi = \gamma(\Phi(g)) = 1$, a contradiction. Since G is compact and Φ continuous onto G^{**}, it is actually a homeomorphism (see Appendix A.8). □

An important message of Pontryagin's Theorem is that the dual group determines the group itself. This is stated as follows.

Corollary 11.20. *Two locally compact Abelian groups are topologically isomorphic if and only if their duals are topologically isomorphic.*

Examples 11.21. 1) The mapping $\Psi : \mathbb{Z} \longrightarrow \mathbb{T}^*$, $\Psi(n) = (z \longmapsto z^n)$ is a topological isomorphism. Indeed, it is obvious that $\Psi(n)$ is a character of $\mathbb{T}$ for any $n \in \mathbb{Z}$ and that Ψ is an injective homomorphism. Since $\operatorname{ran}\Psi$ separates $\mathbb{T}$, by Proposition 11.13 we obtain $\operatorname{ran}\Psi = \mathbb{T}^*$. Because both spaces have the discrete topology, Ψ is a homeomorphism.

2) By inductive application of Proposition 11.18 and by Example 1) we obtain that the characters of $\mathbb{T}^n$ have the form $z \longmapsto z_1^{k_1} z_2^{k_2} \cdots z_n^{k_n}$, $k_i \in \mathbb{Z}$, and that $(\mathbb{T}^n)^*$ and $\mathbb{Z}^n$ are topologically isomorphic.

3) The dual group $\mathbb{Z}^*$ is topologically isomorphic to $\mathbb{T}$. This follows from Example 1) and Pontryagin's Theorem 11.19.

4) The mapping $\Psi : \mathbb{R} \longrightarrow \mathbb{R}^*$, $\Psi(\alpha) = (t \longmapsto \mathrm{e}^{2\pi\mathrm{i}\alpha t})$ is a topological isomorphism. We leave the proof as Exercise 11.5.

Let us return now to the mapping Φ from G into $\mathbb{T}^{G^*}$ (see (11.3)). Since this latter space is compact, the closure $\mathrm{b}G$ of $\operatorname{ran}\Phi$ in $\mathbb{T}^{G^*}$ is compact. Then $\mathrm{b}G$ is a compact group (see Exercise 11.3.b) and G is densely and continuously embedded in $\mathrm{b}G$ by Φ; $\mathrm{b}G$ is called the **Bohr-compactification** of G. Note that if G is compact, then so is $\operatorname{ran}\Phi$, hence $\operatorname{ran}\Phi = \mathrm{b}G$, and G is topologically isomorphic to its Bohr-compactification $\mathrm{b}G$.

Next we determine the dual group of $\mathrm{b}G$.

Proposition 11.22. *Let G be a locally compact Abelian group and* $\mathrm{b}G$ *its Bohr-compactification. The following are true.*

a) *The dual group of* $\mathrm{b}G$ *is topologically isomorphic to* $(G^*)_{\mathrm{d}}$.
b) *The Bohr-compactification* $\mathrm{b}G$ *of G is topologically isomorphic to* $(G^*)_{\mathrm{d}}^*$.

Proof. a) Consider the projections $\pi_\chi : \mathbb{T}^{G^*} \longrightarrow \mathbb{T}$, $\pi_\chi(\gamma) = \gamma(\chi)$, which are obviously characters of $\mathbb{T}^{G^*}$ and hence their restrictions are characters of $\mathrm{b}G$ still denoted by π_χ. Let $\chi_i \in G^*$ and $n_i \in \mathbb{Z}$, $i = 1, \dots, k$. Then $\chi_1^{n_1}\chi_2^{n_2}\cdots\chi_k^{n_k} = \mathbf{1}$ if and only if $\Phi(g)(\chi_1^{n_1}\chi_2^{n_2}\cdots\chi_k^{n_k}) = 1$ for all $g \in G$ and hence (exploiting continuity and denseness) if and only if $\gamma(\chi_1^{n_1}\chi_2^{n_2}\cdots\chi_k^{n_k}) = 1$ for all $\gamma \in \mathrm{b}G$, i.e., $\pi_{\chi_1}^{n_1}\pi_{\chi_2}^{n_2}\cdots\pi_{\chi_k}^{n_k} = \mathbf{1}$. This shows that the mapping

$$\Psi : (G^*)_{\mathrm{d}} \longrightarrow (\mathrm{b}G)^*, \quad \chi_1^{n_1}\chi_2^{n_2}\cdots\chi_k^{n_k} \longmapsto \pi_{\chi_1}^{n_1}\pi_{\chi_2}^{n_2}\cdots\pi_{\chi_k}^{n_k}$$

is well-defined and injective. Of course it is a homeomorphism onto its image, because both space have the discrete topology (use Proposition 11.17).

To prove the surjectivity of Ψ we apply Proposition 11.13 to the compact group $\mathrm{b}G$ with the choice $X = \{\pi_\chi : \chi \in G^*\}$ (obviously separating), and obtain that $(\mathrm{b}G)^* = \langle \pi_\chi : \chi \in G^* \rangle$.

b) This follows from a) and Pontryagin's Theorem 11.19. □

11.5 Application: Ergodic Rotations and Kronecker's Theorem

We now use (parts of) the above duality theory to characterise dynamical properties of group rotations.

Proposition 11.23. *Let G be a compact Abelian group with Haar measure μ, let $a \in G$ and consider the MDS $(G, \mathfrak{B}_0, \mu; \varphi_a)$ and its induced operator $T := T_{\varphi_a}$ on $\mathrm{L}^2(G)$. Then the following assertions are true.*

a) $\sigma_p(T) = \{\chi(a) : \chi \in G^*\}$.
b) *The MDS is ergodic if and only if $\chi(a) = 1$ implies $\chi = \mathbf{1}$.*
c) *The MDS is ergodic if and only if* $\mathrm{orb}(1) = \{a^n : n \in \mathbb{Z}\}$ *is dense in G, i.e., if and only if the TDS $(G; \varphi_a)$ is minimal.*

Proof. It is clear that $T\chi = \chi(a)\chi$, so $\chi(a) \in \sigma_p(T)$ and the "only if" part of b) follow (use Proposition 7.8). Let now $\lambda \in \sigma_p(T)$ and $f \in \mathrm{L}^2(G)$ be such that $Tf = \lambda f$. By Corollary 11.14 we can write f as an L^2-sum

$$f = \sum_{\chi \in G^*} (f \,|\, \chi)\chi,$$

and hence $$Tf = \sum_{\chi \in G^*} \chi(a)(f \,|\, \chi)\chi = \sum_{\chi \in G^*} \lambda (f \,|\, \chi)\chi.$$

Hence $\chi(a)\big(f\,|\,\chi\big) = \lambda\big(f\,|\,\chi\big)$ for all $\chi \in G^*$, so either $f = 0$ or $\chi(a) = \lambda$ for some $\chi \in G^*$. Finally with $\lambda = 1$ the "if" part of b) follows (use Proposition 7.8). (Cf. also Proposition 7.9 and Exercise 7.2.)

c) Set $H = \overline{\mathrm{orb}}(1)$, which is a compact subgroup in G. For $\chi \in G^*$ the conditions $\chi(a) = 1$ and $\chi|_H = 1$ are equivalent. By Corollary 11.11 we see that ergodicity implies denseness of H, hence minimality (use Proposition 3.5). Conversely, if the group rotation is minimal, by the above equivalence and b) we obtain ergodicity. □

Now we are in the position to prove the hitherto postponed Theorem 2.21 of Kronecker. We present two proofs: one exploiting the fact that the characters form an orthonormal basis (cf. Exercise 7.2), and another making use of the Bohr-compactification.

Theorem 11.24. *Let $a = (a_1, \ldots, a_n) \in \mathbb{T}^n$ and consider the rotation $\varphi_a : \mathbb{T}^n \longrightarrow \mathbb{T}^n$. Then the TDS $(\mathbb{T}^n; \varphi_a)$ is topologically transitive (i.e., minimal) if and only if $a_1, a_2, \ldots, a_n$ are linearly independent in the $\mathbb{Z}$-module $\mathbb{T}$.*

Proof I. By using Example 11.21.2) the assertion follows from Proposition 11.23.b) and c). □

Proof II. If $a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n} = 1$ for $0 \neq (k_1, k_2, \ldots, k_n) \in \mathbb{Z}^n$, then the set

$$A := \big\{x \in \mathbb{T}^n \,:\, x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} = 1\big\}$$

is closed, φ_a-invariant and $a \in A$, so $A \neq \emptyset$. If $k_i \neq 0$, there are exactly k_i points in A with given coordinates $x_j \in \mathbb{T}$, $i \neq j$, so $A \neq \mathbb{T}^n$, so the TDS $(\mathbb{T}^n; \varphi_a)$ is not minimal.

For the converse implication suppose that $a_1, a_2, \ldots, a_n$ are linearly independent in the $\mathbb{Z}$-module $\mathbb{T}$, and recall from Example 11.21 that $\mathbb{Z}^*$ and $\mathbb{T}$ are topologically isomorphic, so without further reference we identify them.

Let $b := (b_1, \ldots, b_n) \in \mathbb{T}^n$ and $\varepsilon > 0$ be given. Denote by $G = \langle a_1, a_2, \ldots, a_n \rangle$. Since $\{a_1, \ldots, a_n\}$ is linearly independent in the $\mathbb{Z}$-module (Abelian group) $\mathbb{T}$, there exists a $\mathbb{Z}$-linear mapping (group homomorphism)

$$\psi : G \longrightarrow \mathbb{T} \quad \text{with } \psi(a_i) = b_i \text{ for } i = 1, \ldots, n.$$

By Proposition 11.25 there is a group homomorphism $\chi : \mathbb{T} \longrightarrow \mathbb{T}$ extending ψ. Obviously $\chi \in (\mathbb{T}_{\mathrm{d}})^*$, which by Proposition 11.22 is topologically isomorphic to $\mathrm{b}\mathbb{Z}$. By the definition of $\mathrm{b}\mathbb{Z}$ and the topology on it, we find $k \in \mathbb{Z}$ such that

$$|b_i - a_i^k| = \big|\chi(a_i) - \Phi(n)(a_i)\big| < \varepsilon$$

for $i = 1, \ldots, n$ (recall that $\Phi(n)(z) = z^n$ for $z \in \mathbb{T} \simeq \mathbb{Z}^*$). Therefore $\mathrm{orb}(1) = \big\{(a_1^k, a_2^k, \ldots, a_n^k) \,:\, k \in \mathbb{Z}\big\}$ is dense in $\mathbb{T}^n$. □

Supplement: The Character Group Separates the Points

In this section we present a proof of Theorem 11.10 in the special case that the group is discrete or compact.

The discrete case is of course pure algebra and actually the proof resembles very much to that of the Hahn–Banach theorem.

Proposition 11.25. *Let G be an Abelian group, $H \subset G$ a subgroup, and let $\psi : H \longrightarrow \mathbb{T}$ a homomorphism. Then there is a homomorphism $\chi : G \longrightarrow \mathbb{T}$ extending ψ.*

Proof. Consider the family

$$\mathscr{M} := \big\{(K,\varphi) \,:\, K \subset G \text{ subgroup}, \ \varphi : K \longrightarrow \mathbb{T} \text{ homomorphism}, \ \varphi|_H = \psi\big\}.$$

This set is partially ordered by the following relation: $(K_1,\varphi_1) \leq (K_2,\varphi_2)$ if and only if $K_1 \subset K_2$ and $\varphi_2|_{K_1} = \varphi_1$. By Zorn's Lemma, as the chain condition is obviously satisfied, there is a maximal element $(K,\varphi) \in \mathscr{M}$. We prove that $K = G$.

If $x \in G \setminus K$, then we construct an extension of φ to $K_1 := \langle K \cup \{x\}\rangle$ contradicting maximality. We have $K_1 = \{hx^n \,:\, n \in \mathbb{Z}, h \in K\}$. If no power x^n, $n \geq 2$ belongs to K, then the mapping $\varphi_1(x^i h) := \varphi(h)$, $i \in \mathbb{Z}$, $h \in K$ is a well-defined group homomorphism. If $x^n \in K$ for some $n \geq 2$, then take the smallest such n. Let $\alpha \in \mathbb{T}$ be some n^{th} root of $\varphi(x^n)$ and set $\varphi_1(x^i h) := \alpha^i \varphi(h)$, for $i \in \mathbb{Z}$, $h \in K$. This mapping is well-defined: if $x^i h = x^j h'$ and $i > j$, then $x^{i-j}h = h' \in K$ and $x^{i-j} \in K$, so $\varphi(h') = \varphi(x^{i-j}h) = \varphi(x^{i-j})\varphi(h)$. By assumption n divides $i - j$, so $\varphi(x^{i-j}h) = \alpha^{i-j}\varphi(h)$. Of course, $\varphi_1 : K_1 \longrightarrow \mathbb{T}$ is a homomorphism that extends φ (hence ψ), so $(K_1,\varphi_1) \in \mathscr{M}$ with $(K_1,\varphi_1) > (K,\varphi)$, a contradiction. □

As a corollary we obtain the version of Theorem 11.10 for the discrete case.

Proposition 11.26. *If G is an Abelian group the characters separate the points of G.*

Proof. Let $x \in G$, $x \neq 1$. If the there is $n \geq 2$ with $x^n = 1$, then take $1 \neq \alpha \in \mathbb{T}$ an n^{th} root of unity. If the order of x is infinite, then take any $1 \neq \alpha \in \mathbb{T}$. For $k \in \mathbb{Z}$ set $\psi(x^k) = \alpha^k$, which is a well-defined non-trivial character of $\langle x\rangle$. By Proposition 11.25 we can extend it to a character χ on G that separates 1 and x. □

For the proof in the compact case we shall employ the theory of commutative Banach algebras, briefly touched upon in Lecture 4. (With some more or less standard changes the proof works also for locally compact groups.)

For an compact Abelian group G we define a multiplication on $\mathrm{M}(G) = \mathrm{C}(G)'$ in the following way: if $\mu_1,\mu_2 \in \mathrm{M}(G)$, then let $\mu_1 * \mu_2$ be the linear functional on $\mathrm{C}(G)$ defined by

$$\langle f, \mu_1 * \mu_2\rangle = \int_G \int_G f(xy)\,\mu_1(\mathrm{d}x)\mu_2(\mathrm{d}y).$$

The measure $\mu_1 * \mu_2$ is called the **convolution** of the measures μ_1 and μ_2. Clearly,

$$* : \mathrm{M}(G) \times \mathrm{M}(G) \longrightarrow \mathrm{M}(G)$$

is bilinear and one has $\|\mu_1 * \mu_2\| \leq \|\mu_1\| \, \|\mu_2\|$. Moreover, Fubini's theorem shows that convolution is associative and commutative. The Dirac measure δ_1 at 1 acts as a neutral element: $\delta_1 * \nu = \nu * \delta_1 = \nu$ for every $\nu \in \mathrm{M}(G)$. In this way $\mathrm{M}(G)$ becomes a commutative unital Banach algebra.

The mapping

$$\iota : \mathrm{L}^1(G) \longrightarrow \mathrm{M}(G), \quad f \longmapsto f(x)\mathrm{d}x$$

is an isometry onto a closed subspace of $\mathrm{M}(G)$. (This is a general fact from integration theory. By the Radon–Nikodym theorem the range of ι consists precisely of those complex measures on G that are absolutely continuous with respect to the Haar measure μ.) Now for $f, g \in \mathrm{C}(G)$ we define their convolution $f * g$ by

$$(f * g)(x) := \int_G f(xy) g(y^{-1}) \, \mathrm{d}y = \int_G f(y) g(y^{-1}x) \, \mathrm{d}y \qquad (x \in G). \tag{11.4}$$

Lemma 11.27. *Let $f, g \in \mathrm{C}(G)$, then $f * g \in \mathrm{C}(G)$ again.*

Proof. For a function $g \in \mathrm{C}(G)$ and $x, y \in G$ we write $(\tau_y g)(x) := g(y^{-1}x)$. By Corollary 11.8, the mapping $y \longmapsto f(y)\tau_y g$ is continuous from G into $\mathrm{C}(G)$. So we can integrate this vector-valued function to obtain a function

$$h := \int_G f(y) \tau_y g \, \mathrm{d}y \in \mathrm{C}(G).$$

Since point evaluations are continuous linear functionals on $\mathrm{C}(G)$, we can evaluate under the integral sign, giving

$$h(x) = \int_G f(y) g(y^{-1}x) \, \mathrm{d}y = (f * g)(x)$$

for every $x \in G$. □

An easy application of the Fubini's theorem and translation invariance of the Haar measure shows that $\iota(f * g) = (\iota f) * (\iota g)$, whence there is no danger of identifying f with ιf and we drop the explicit reference to ι. In particular we obtain $\|f * g\|_1 \leq \|f\|_1 \, \|g\|_1$ for all $f, g \in \mathrm{C}(G)$. Since $\mathrm{C}(G)$ is dense in $\mathrm{L}^1(G)$, we see that $\mathrm{L}^1(G) * \mathrm{L}^1(G) \subset \mathrm{L}^1(G)$. Hence the space $\mathrm{L}^1(G)$ is a Banach algebra with respect to convolution, the so-called **group algebra** of G.

Consider the space

$$\mathscr{A} := \mathrm{L}^1(G) + \langle \delta_1 \rangle$$

contained in $\mathrm{M}(G)$. The considerations above show that $\mathscr{A}$ is a closed subalgebra of $\mathrm{M}(G)$ and thus a commutative Banach algebra with unit $\mathrm{e} = \delta_1$. We establish a correspondence between the characters of G and certain unital algebra homomorphisms $\mathscr{A} \longrightarrow \mathbb{C}$ (cf. Lecture 4). For $\chi \in G^*$ consider the mapping

$$[\chi] : \nu \longmapsto \langle \chi, \nu \rangle = \int_G \chi(x) \, \nu(\mathrm{d}x), \qquad \mathrm{M}(G) \longrightarrow \mathbb{C}.$$

An easy calculation shows that $[\chi]$ is a unital algebra homomorphism. Hence $\psi_\chi := [\chi]\big|_{\mathscr{A}}$ is an element of the Gelfand space $\Gamma(\mathscr{A})$ of $\mathscr{A}$ (see Lecture 4).

Proposition 11.28. *Let G be a compact Abelian group. Then the mapping $\psi : \chi \longmapsto \psi_\chi\big|_{\mathscr{A}}$ establishes a bijective correspondence between characters of G and those complex unital algebra homomorphisms on $\mathscr{A}$ that are non-zero on $\mathrm{L}^1(G)$. More precisely,*

$$G^* \longrightarrow \left\{\psi \in \Gamma(\mathscr{A}) : \psi\big|_{\mathrm{L}^1(G)} \neq 0\right\}, \qquad \chi \longmapsto \psi_\chi$$

is bijective.

Proof. We have seen above that indeed $\psi_\chi \in \Gamma(\mathscr{A})$. Since $\chi(1) = 1$ and χ is continuous, by Urysohn's lemma we can find a continuous function $f \in \mathrm{C}(G)$ such that $\psi_\chi(f) = \int_G \chi f \,\mathrm{d}\mu \neq 0$. To prove the surjectivity, let $\psi : \mathscr{A} \longrightarrow \mathbb{C}$ be any unital algebra homomorphism. If $\psi\big|_{\mathrm{L}^1(G)} \neq 0$, there is $h \in \mathrm{C}(G)$ such that $\psi(h) = 1$. Define

$$\chi(x) := \psi(\tau_x h) \qquad (x \in G)$$

with $(\tau_x f)(y) = f(x^{-1}y)$. Then χ is continuous and $\chi(1) = 1$. We claim that $\chi \in G^*$ and $\psi = \psi_\chi$. Now

$$\chi(x)\chi(y) = \psi(\tau_x h)\psi(\tau_y h) = \psi(\tau_x h * \tau_y h) = \psi(h * \tau_{xy} h) = \psi(h)\psi(\tau_{xy} h) = \chi(xy)$$

for any $x, y \in G$. (Here we used the identity $\tau_x(f * g) = (\tau_x f) * g = f * (\tau_x g)$, which is immediate from (11.4).)

Taking $y = x^{-1}$ in the above yields that χ never vanishes. Since $\psi \in \mathscr{A}'$ by Lemma 4.8, the function χ is clearly bounded, and thus it must map G into $\mathbb{T}$ (look at the sequence $(\chi(x^n))_{n\in\mathbb{Z}}$). Hence $\chi \in G^*$. Finally, take $f \in \mathrm{C}(G)$ and note that $(x \longmapsto f(x)\tau_x h) : G \longmapsto \mathrm{C}(G) \subset \mathrm{L}^1(G)$ is continuous, hence integrable. Since ψ is continuous we obtain

$$\begin{aligned}\psi_\chi(f) &= \int_G f(y)\chi(y)\,\mathrm{d}y = \int_G f(y)\psi(\tau_y h)\,\mathrm{d}y = \psi\left(\int_G f(y)(\tau_y h)\,\mathrm{d}y\right)\\ &= \psi(h * f) = \psi(h)\psi(f) = \psi(f),\end{aligned}$$

and since $\mathrm{C}(G)$ is dense in $\mathrm{L}^1(G)$, $\psi_\chi = \psi$ as claimed. □

The next step in our proof consists in *representing* $\mathscr{A}$ as a selfadjoint subalgebra of $\mathscr{L}(H)$, where H is the Hilbert space $\mathrm{L}^2(G)$. To this aim, define

$$(\nu * f)(x) := \int_G f(x^{-1}y)\,\nu(\mathrm{d}y) \qquad (x \in G)$$

for $\nu \in \mathrm{M}(G)$ and $f \in \mathrm{C}(G)$. Then $\nu * f \in \mathrm{C}(G)$ (similarly to Lemma 11.27) and Fubini's theorem shows that $\iota(\nu * f) = \nu * (\iota f)$, so the notation is consistent.

Lemma 11.29 (Young Inequality). *Let G be a compact Abelian group. Then*

$$\|\nu * f\|_2 \leq \|\nu\|_{\mathrm{M}} \|f\|_2$$

for all $\nu \in \mathrm{M}(G)$, $f \in \mathrm{C}(G)$.

Proof. For $f \in \mathrm{C}(G)$ fixed, define $F : G \longrightarrow \mathrm{C}(G) \subset \mathrm{L}^2(G)$ by $F(y)(x) := f(x^{-1}y)$. Then F is continuous and hence we can integrate it against ν, giving $\int_G F \,\mathrm{d}\nu = \nu * f$. The triangle inequality for (vector-valued) integrals hence yields

$$\begin{aligned} \|\nu * g\|_2 &= \left\| \int_G F(y)\, \nu(\mathrm{d}y) \right\|_2 \le \int_G \|F(y)\|_2 \, |\nu|\,(\mathrm{d}y) \\ &= \int_G \left(\int_G \left| f(x^{-1}y) \right|^2 \mathrm{d}x \right)^{1/2} |\nu|\,(\mathrm{d}y) \\ &= \int_G \left(\int_G |f(x)|^2 \,\mathrm{d}x \right)^{1/2} |\nu|\,(\mathrm{d}y) = \|\nu\|_{\mathrm{M}} \, \|f\|_2 \,. \end{aligned}$$

(Here we used that the Haar measure is also inversion invariant.) □

Consider the Hilbert space $H := \mathrm{L}^2(G)$. Since $\mathrm{C}(G)$ is dense in H, the lemma shows that for each $\nu \in \mathrm{M}(G)$ we can form the convolution operator

$$L_\nu : H \longrightarrow H, \qquad f \longmapsto \nu * f.$$

The associativity of convolution implies that $L : \nu \longmapsto L_\nu$ is an algebra homomorphism $\mathrm{M}(G) \longrightarrow \mathscr{L}(H)$, a so-called *representation*.

For a continuous function $f \in \mathrm{C}(G)$ we define

$$f^*(x) = \overline{f(x^{-1})} \qquad (x \in G).$$

For $f, g, h \in \mathrm{C}(G)$ the usual manipulations (Fubini's theorem + invariance) yield $\big(h \,|\, f^* * g\big) = \big(f * h \,|\, g\big)$, whence $L_{f^*} = (L_f)^*$ is the Hilbert space adjoint of L_f. Using common knowledge about Hilbert space operators we are now able to prove the following result.

Lemma 11.30. *In the situation above let* $0 \neq h \in \mathrm{C}(G)$ *with* $h = h^*$. *Then there is a character* $\chi \in G^*$ *such that* $\int_G h\chi \,\mathrm{d}x \neq 0$.

Proof. Since $h \neq 0$ We have $(h * h^*)(1) = \|h\|_2^2 > 0$ and since $h * h^*$ is continuous, it follows that $\|L_h(h^*)\|_2 = \|h * h^*\|_2 > 0$. This shows in particular that $L_h \neq 0$, i.e., $\|L_h\| \neq 0$. Since $h = h^*$, the operator L_h is self-adjoint, hence its operator norm must coincide with its spectral radius: $\|L_h\| = r(L_h)$. But $\nu \longmapsto L_\nu$ is an algebra homomorphism, so every invertible element of $\mathscr{A}$ is carried to an invertible element of $\mathscr{L}(H)$. This shows that

$$\big\{ \lambda \in \mathbb{C} : \lambda\delta_1 - h \text{ not invertible in } \mathscr{A} \big\} \supset \sigma(L_h).$$

As seen above, $\sigma(L_h) \neq \{0\}$, hence there is $0 \neq \lambda$ such that $\lambda\delta_1 - h$ is not invertible in $\mathscr{A}$. By a standard application of Zorn's lemma there is a maximal (algebra) ideal I of $\mathscr{A}$ containing $\lambda\delta_1 - h$. Lemma 4.8 shows that $I = \ker\psi$ for some $\psi \in \Gamma(\mathscr{A})$.

Then $0 = \psi(\lambda\delta_1 - h) = \lambda - \psi(h)$, i.e., $\psi(h) = \lambda \neq 0$. By Proposition 11.28 above, there is a character $\chi \in G^*$ such that $\psi = \psi_\chi$. This concludes the proof. □

Proof of Theorem 11.10. Suppose that $1 \neq a \in G$. Then $a^{-1} \neq 1$ as well. Let U be a symmetric neighbourhood of 1 such that $U \cap aU = \emptyset$. By Urysohn's lemma, there is a function $0 \leq f \in \mathrm{C}(G)$ with $\operatorname{supp}(f) \subset U$, $f(1) \neq 0$. By passing to $f \cdot f^*$ and scaling appropriately, we may suppose that $f = f^*$ and $\int_G f \,\mathrm{d}\mu = 1$. Now consider the function

$$h(x) := f(x) - \frac{f(ax) + f(a^{-1}x)}{2} \qquad (x \in G).$$

Then also $h^* = h$ and $h(1) = f(1) \neq 0$ (since $f(a) = f(a^{-1}) = 0$). In particular, $h \neq 0$. By Lemma 11.30 there is a character $\chi \in G^*$ such that $\langle \chi, h\rangle \neq 0$, i.e.,

$$\begin{aligned}\langle \chi, f\rangle &\neq \frac{1}{2}\int_G f(ax)\chi(x)\,\mathrm{d}x + \frac{1}{2}\int_G f(a^{-1}x)\chi(x)\,\mathrm{d}x \\ &= \frac{\chi(a^{-1})}{2}\int_G f(x)\chi(x)\,\mathrm{d}x + \frac{\chi(a)}{2}\int_G f(x)\chi(x)\,\mathrm{d}x = \operatorname{Re}\chi(a)\,\langle \chi, f\rangle\,.\end{aligned}$$

Hence $\chi(a) \neq 1$ as desired. □

Exercises

1. Consider $S := \mathbb{R} \cup \{\infty\}$, the one-point compactification of $\mathbb{R}$, and define thereon the addition $t + \infty := \infty + t := \infty$, $\infty + \infty := \infty$ for $t \in \mathbb{R}$. Show that S is a compact semitopological but not a topological semigroup. Determine the minimal ideal.

2. Prove assertions b), c) in Lemma 11.6. Prove also the following generalisation of Proposition 11.7: if G and H are topological group, G is compact and $f : G \longrightarrow H$ is continuous, it is even uniformly continuous in the sense that for any V open neighbourhood of 1_H there is an open neighbourhood U of 1_G such that $gh^{-1} \in U$ implies $f(g)f(h)^{-1} \in V$.

3.

a) Let S be a semitopological semigroup. A **subsemigroup** of S is a subset $H \subset S$ such that $HH \subset H$. Show that if H is a subsemigroup (ideal) then $\overline{H}$ is a subsemigroup (ideal), too. Show that if H is Abelian then also $\overline{H}$ is.

b) Show that if G is a topological group and H is a (normal) subgroup of G then $\overline{H}$ is a (normal) subgroup as well.

4. Let G be a topological group and $H \subset G$ an open topological subgroup. Show that H is closed. Show that the connected component G_0 of $1 \in G$ is a closed normal subgroup.

5. Show that the dual group of $\mathbb{R}$ is as claimed in Section 11.3.

6. Prove Corollary 11.11 for

a) G a compact or discrete Abelian topological group,

*b) G a general locally compact Abelian group.

(Hint: Prove that the factor group G/H is a locally compact (in particular Hausdorff) Abelian group and apply Theorem 11.10 to this group.)

7. For $\Phi : G \longrightarrow \mathbb{T}^{G^*}$ defined in (11.3) prove the following assertions.

a) Φ maps into G^{**}.

b) If G is a discrete or compact Abelian group, then $\Phi : G \longrightarrow G^{**}$ is continuous.

*c) For general locally compact Abelian groups $\Phi : G \longrightarrow G^{**}$ is continuous.

8. Determine the dual of the group $\mathbb{A}$ defined in Exercise 3.4. (Hint: Consider the TDS from Exercise 3.4. Show that $\varphi^{2^n}(0) \to 0$, $n \to \infty$ (0 denotes the unit in $\mathbb{A}$). For $\chi \in \mathbb{A}^*$, set $a = \chi(\mathbf{1})$ and prove that $a^{2^n} \to 1$ as $n \to \infty$.)

9. Prove Proposition 11.18.

$*$ **10.** Let K be a compact metric space, $f : K \times K \longrightarrow K$ be a separately continuous function, and let $b \in K$. Show that there is there is $F_b \subset K$ such that $K \setminus F_b$ is of first category and for each $a \in F_b$ the function f is continuous at (a,b).

Lecture 12
The Jacobs–de Leeuw–Glicksberg Decomposition

Das also war des Pudels Kern! [1]

Johann Wolfgang von Goethe[2]

In this chapter we shall apply the previous results to compact semigroups of bounded linear operators on a Banach space X. Let us write $\mathscr{L}_s(X)$ and $\mathscr{L}_w(X)$ to denote the space $\mathscr{L}(X)$ endowed with the strong and the weak operator topology, respectively. To simplify terminology, we shall speak of weakly/strongly closed, open, relatively compact, compact etc. sets when we intend the weak/strong operator topology.[3] Also, to render the notation more feasible, we shall write $\mathrm{cl}_w\mathscr{T}$, $\mathrm{cl}_s\mathscr{T}$ for the closure of a set $\mathscr{T} \subset \mathscr{L}(X)$ with respect to the weak/strong operator topology. For a subset $A \subset X$ we shall write $\mathrm{cl}_\sigma A$ to denote its closure in the $\sigma(X,X')$-topology.

Note that a subset $\mathscr{T} \subset \mathscr{L}(X)$ is a semigroup (with respect to operator multiplication) if and only if it is closed under multiplication. The operator multiplication is separately continuous with respect to both the strong and the weak operator topologies (cf. Appendix C.8). Hence — in the terminology of Lecture 11 — both $\mathscr{L}_s(X)$ and $\mathscr{L}_w(X)$ are semitopological semigroups with respect to the operator multiplication. We shall concentrate mainly on $\mathscr{L}_w(X)$.

To be more specific, for a linear operator $T \in \mathscr{L}(X)$

$$\mathscr{T}_T := \{T^n : n \in \mathbb{N}_0\} \quad \text{and} \quad \mathrm{conv}(\mathscr{T}_T) = \mathrm{conv}\{T^n : n \in \mathbb{N}_0\} \tag{12.1}$$

are both Abelian subsemigroups of $\mathscr{L}(X)$ (cf. Exercise 12.1). The semigroup $\mathscr{T}_T$ is called the semigroup **generated** by T. We are interested in the case when these semigroups are relatively weakly compact. The following useful characterisation is Proposition C.18 from Appendix C.

Lemma 12.1. *Let X be a Banach space, let $\mathscr{T} \subset \mathscr{L}(X)$, and let $D \subset X$ be a subset such that* $\mathrm{lin}\,D$ *is dense in X. Then the following assertions are equivalent.*

[1] This then the kernel of the brute! (Translation: Bayard Taylor)

[2] Faust, Part I.

[3] Do not confuse this with the corresponding notion for the weak topology of $\mathscr{L}(X)$ as a Banach space. This topology will not appear in this course.

(i) *$\mathscr{T}$ is relatively weakly compact, i.e, relatively compact in the weak operator topology.*

(ii) *$\mathscr{T}x = \{Tx : T \in \mathscr{T}\}$ is relatively weakly compact in X for each $x \in X$.*

(iii) *$\mathscr{T}$ is norm-bounded and $\mathscr{T}x = \{Tx : T \in \mathscr{T}\}$ is relatively weakly compact in X for each $x \in D$.*

For historical reasons a relatively weakly compact semigroup $\mathscr{T} \subset \mathscr{L}(X)$ is sometimes called **weakly almost periodic** (e.g. in [Krengel (1985)]). Analogously, relatively strongly compact semigroups are termed **(strongly) almost periodic**. In Exercise 12.2 the analogue of Lemma 12.1 for the strong operator topology is stated.

Remarks 12.2. 1) Let $T \in \mathscr{L}(X)$, X a Banach space. Then $\mathscr{T}_T$ is relatively weakly compact if and only if $\mathrm{conv}(\mathscr{T}_T)$ is relatively weakly compact. This follows from Lemma 12.1 together with the Kreĭn–Šmulian Theorem C.11. An analogous remark holds for the strong operator topology.

2) If T generates a relatively weakly compact semigroup, then T is mean ergodic. This follows from Lemma 12.1 and Theorem 8.8.

3) Let X be a reflexive Banach space. Then every norm-bounded subset of $\mathscr{L}(X)$ is relatively weakly compact by Lemma 12.1 and Theorem C.12. In particular, the unit ball of $\mathscr{L}(X)$ is relatively weakly compact, and if $T \in \mathscr{L}(X)$ is power bounded, then $\mathscr{T}_T$ is relatively weakly compact.

4) Let (Ω, Σ, μ) be a finite measure space. Then the set

$$\left\{T \in \mathscr{L}(\mathrm{L}^1(\Omega, \Sigma, \mu)) : T \text{ is a Dunford–Schwartz operator}\right\}$$

is relatively weakly compact. In particular, every Dunford–Schwartz operator generates a relatively weakly compact semigroup (cf. Example 8.12).

5) If $(\Omega, \Sigma, \mu; \varphi)$ is an MDS, then the induced operator $T = T_\varphi$ generates a relatively weakly compact semigroup on each of the spaces $\mathrm{L}^p(\Omega, \Sigma, \mu)$, $1 \le p < \infty$.

6) If S is a compact semitopological semigroup, then its action on $\mathrm{C}(S)$ by left rotations is relatively weakly compact. The proof is not hard but rests on an important result of Grothendieck, see the supplement to this lecture.

Let $\mathscr{T} \subset \mathscr{L}(X)$ be an Abelian relatively weakly compact semigroup of operators on a Banach space X, and let $\mathscr{S} := \mathrm{cl}_w \mathscr{T}$ be its weak (operator) closure. Then $\mathscr{S}$ is a compact semitopological Abelian semigroup, hence Theorem 11.2 applies and $\mathscr{S}$ has a minimal ideal $K(\mathscr{S})$, which is a group. Let Q be the neutral element of this group. Then $K(\mathscr{S}) = Q\mathscr{S}$ and $Q = Q^2$, i.e., Q is a projection. Therefore it induces a decomposition of X into a direct sum

$$X = X_\mathrm{r} \oplus X_\mathrm{s}$$

of $\mathscr{T}$-invariant closed subspaces $X_\mathrm{r} := \mathrm{ran}\, Q$ and $X_\mathrm{s} := \ker Q$, called the (abstract) **Jacobs–de Leeuw–Glicksberg decomposition** (JdLG-decomposition) of X corresponding to $\mathscr{T}$. It goes back to [Jacobs (1956)], [de Leeuw and Glicksberg (1959)] and [de Leeuw and Glicksberg (1961)].

Furthermore, one has the corresponding decomposition of every operator $T \in \mathscr{T}$ as $T = T_{\mathrm{r}} \oplus T_{\mathrm{s}}$, where

$$T_{\mathrm{r}} := T|_{X_{\mathrm{r}}}, \quad T_{\mathrm{s}} := T|_{X_{\mathrm{s}}}.$$

In order to study $\mathscr{T}$ on X it therefore suffices to study $\mathscr{T}_{\mathrm{r}} := \{T_{\mathrm{r}} : T \in \mathscr{T}\}$ and $\mathscr{T}_{\mathrm{s}} := \{T_{\mathrm{s}} : T \in \mathscr{T}\}$ on X_{r} and X_{s}, respectively.

The space X_{r} is called the **reversible part** of X and its elements the **reversible vectors**; the space X_{s} is called the **stable** part and its element the **almost weakly stable** vectors. This terminology stems from the following characterisation.

Lemma 12.3. *Let $X = X_{\mathrm{r}} \oplus X_{\mathrm{s}}$ be the JdLG-decomposition of a Banach space X with respect to a relatively weakly compact semigroup $\mathscr{T} \subset \mathscr{L}(X)$. Then*

$$\begin{aligned} X_{\mathrm{r}} &= \{x \in X : y \in \mathrm{cl}_\sigma(\mathscr{T}x) \Rightarrow x \in \mathrm{cl}_\sigma(\mathscr{T}y)\} \\ &= \{x \in X : y \in \mathscr{S}x \Rightarrow x \in \mathscr{S}y\}, \\ X_{\mathrm{s}} &= \{x \in X : 0 \in \mathrm{cl}_\sigma \mathscr{T}x\} = \{x \in X : 0 \in \mathscr{S}x\}, \end{aligned}$$

where $\mathscr{S} := \mathrm{cl}_w \mathscr{T}$ is the closure of $\mathscr{T}$ in $\mathscr{L}_w(X)$.

Hence a vector x is reversible if and only if whenever some vector y can be reached by the action of $\mathscr{S}$ starting from x, then one can return to x again by the action of $\mathscr{S}$. On the other hand, a vector is almost weakly stable if and only if 0 can be reached by the action of $\mathscr{S}$.

Proof. Since the mapping $T \longmapsto Tx$ is continuous from $\mathscr{L}_w(X)$ to $(X, \sigma(X, X'))$ and since $\mathscr{S} = \mathrm{cl}_w \mathscr{T}$ is compact, one has $\mathrm{cl}_\sigma(\mathscr{T}x) = (\mathrm{cl}_w \mathscr{T})x = \mathscr{S}x$.

Now, if $x \in X_{\mathrm{r}} = \operatorname{ran} Q$, then $\mathscr{S}x = \mathscr{S}Qx = K(\mathscr{S})x$. Hence if $y \in \mathscr{S}x$, then there is $R \in K(\mathscr{S})$ with $y = Rx$. Since $K(\mathscr{S})$ is a group, there is $S \in K(\mathscr{S})$ with $SR = Q$. Hence $x = Qx = SRx = Sy \in \mathscr{S}y$. Conversely, if $x \in X$ is such that there is $S \in \mathscr{S}$ with the property that $SQx = x$, then $x = SQx = QSx \in \operatorname{ran} Q$.

If $x \in X_{\mathrm{s}} = \ker Q$, then immediately $0 = Qx \in \mathscr{S}x$. On the other hand, if $0 \in \mathscr{S}x$, then there is $S \in \mathscr{S}$ such that $Sx = 0$. Since $QS \in K(\mathscr{S})$ and $K(\mathscr{S})$ is a group, there is $R \in K(\mathscr{S})$ such that $RQS = Q$. Hence $Qx = RQSx = RQ0 = 0$, whence $x \in X_{\mathrm{s}}$. □

Example 12.4. Let us look at the particular example $\mathscr{T} := \mathrm{conv}(\mathscr{T}_T)$ for some operator $T \in \mathscr{L}(X)$ generating a relatively weakly compact semigroup $\mathscr{T}_T$. By Remark 12.2.2, the operator T is mean ergodic, and one has the mean ergodic decomposition

$$X = \mathrm{Fix}(T) \oplus \overline{\mathrm{ran}(\mathrm{I} - T)}$$

with ergodic projection $P = P_T$. Since P is the strong limit of the Cesàro means $\mathrm{A}_n[T]$, it belongs to $\mathscr{S} := \mathrm{cl}_w \mathscr{T}$. But since $PT = TP = P$, one has $\mathscr{T}P = \{P\}$ and hence $\mathscr{S}P = \{P\}$. Hence $\{P\}$ is the minimal ideal of $\mathscr{S}$, and the JdLG-decomposition for $\mathrm{conv}(\mathscr{T}_T)$ coincides with the mean ergodic decomposition for T.

Example 12.4 shows that we do not obtain new information when applying the JdLG-decomposition to the semigroup $\mathrm{conv}(\mathscr{T}_T)$. We shall see next that this is different for the semigroup $\mathscr{T}_T$ itself.

As we have said above, to study an Abelian relatively weakly compact semigroup $\mathscr{T} \subset \mathscr{L}(X)$ it suffices to consider the restricted semigroups $\mathscr{T}_\mathrm{r}$ on X_r and $\mathscr{T}_\mathrm{s}$ on X_s. The following result states in particular that on the parts X_r and X_s the JdLG-decomposition is then trivial and a further reduction is not possible.

Proposition 12.5. *Let X be a Banach space, and let $\mathscr{T} \subset \mathscr{L}(X)$ be an Abelian relatively weakly compact semigroup with associated JdLG-decomposition $X = X_\mathrm{r} \oplus X_\mathrm{s}$. With the terminology introduced above, the following assertions hold.*

a) *The semigroup $\mathscr{T}_\mathrm{r} \subset \mathscr{L}(X_\mathrm{r})$ is relatively weakly compact, and its weak closure $\mathscr{S}_\mathrm{r} := \mathrm{cl}_w \mathscr{T}_\mathrm{r}$ is a weakly compact* group *of invertible operators on X_r. Moreover,*
$$\mathscr{S}_\mathrm{r} = \big\{ S|_{X_\mathrm{r}} : S \in \mathscr{S} \big\} = \big\{ S|_{X_\mathrm{r}} : S \in K(\mathscr{S}) \big\}$$
and the neutral element of $K(\mathscr{S}_\mathrm{r})$ is the identity operator on X_r.

b) *The semigroup $\mathscr{T}_\mathrm{s} \subset \mathscr{L}(X_\mathrm{s})$ is relatively weakly compact, and its weak closure $\mathscr{S}_\mathrm{s} := \mathrm{cl}_w \mathscr{T}_\mathrm{s}$ contains the zero operator. Moreover,*
$$\mathscr{S}_\mathrm{s} = \big\{ S|_{X_\mathrm{s}} : S \in \mathscr{S} \big\}$$
and $K(\mathscr{S}_\mathrm{s}) = \{0\}$.

Proof. The restriction mapping is continuous for the weak operator topologies, so $\mathscr{T}_\mathrm{r}$ is a relatively weakly compact semigroup on X_r with weak closure
$$\mathscr{S}_\mathrm{r} = \big\{ S|_{X_\mathrm{r}} : S \in \mathscr{S} \big\}.$$
However, since $Sx = SQx$ for all $x \in X_\mathrm{r}$ and $Q\mathscr{S} = K(\mathscr{S})$ is a group,
$$\mathscr{S}_\mathrm{r} = \big\{ S|_{X_\mathrm{r}} : S \in K(\mathscr{S}) \big\}$$
is a group with neutral element $Q|_{X_\mathrm{r}} = \mathrm{I}_{X_\mathrm{r}}$. The arguments for the stable part are similar. □

Example 12.6. Let H be a Hilbert space and let $T \in \mathscr{L}(H)$ be a contraction. Since the set of contractions is compact in $\mathscr{L}_w(H)$ (see Remark 12.2.3), the semigroup $\mathscr{T}_T$ generated by T is relatively weakly compact. The associated JdLG-decomposition $H = H_\mathrm{r} \oplus H_\mathrm{s}$ is orthogonal since the projection Q is a contraction, see Exercise 12.3. The group $\mathscr{S}_\mathrm{r}$ is a group of invertible contractions, i.e., of unitary operators. In particular, the restricted operator T_r on H_r is unitary.

The following example shows that the reversible part can be trivial even if $\mathscr{T}$ is itself a group of invertible operators.

Example 12.7. Let $H = \ell^2(\mathbb{Z})$ with T being the (right or left) shift. Then T is a unitary operator. In particular, $\mathscr{T} := \{T^n : n \in \mathbb{Z}\}$ is a relatively weakly compact *group* of operators. However, $T^n f \to 0$ weakly for every $f \in H$. Hence $H = H_s$ and the reversible part is trivial.

12.1 Compact Group Actions on Banach Spaces

In this section we study the case $X = X_r$. Let X be a Banach space and let $\mathscr{T} \subset \mathscr{L}(X)$ be an Abelian relatively weakly compact semigroup whose weak closure $G := \mathrm{cl}_w \mathscr{T} \subset \mathscr{L}_w(X)$ is a (compact, Abelian) group of invertible operators with the identity operator I being its neutral element. (This means $X_s = \{0\}$ in the JdLG-decomposition.) Let us call $0 \neq x \in X$ an **eigenvector** for $\mathscr{T}$ if the one-dimensional subspace $\mathbb{C}x$ is invariant under $\mathscr{T}$.

Lemma 12.8. *In the situation above, if x is an eigenvector of $\mathscr{T}$, then there is a unique character $\chi \in G^*$ such that*

$$Tx = \chi(T)x \qquad (T \in \mathscr{T}).$$

Proof. If $\mathbb{C}x$ is invariant under $\mathscr{T}$, it is also invariant under G. Hence for every $g \in G$ there is a unique $\chi(g) \in \mathbb{C}$ such that $gx = \chi(g)x$. Observe that $\chi : G \longrightarrow \mathbb{C}$ is continuous and multiplicative, and since G is compact, χ is a character, i.e., $\chi(G) \subset \mathbb{T}$. Uniqueness is clear since $\mathscr{T}$ is dense in G. □

Up to now it is not clear that such eigenvectors do exist. To construct them we form — for given $\chi \in G^*$ and $x \in X$ — the integral

$$P_\chi x := \int_G \overline{\chi(g)}\,(gx)\,\mathrm{d}g.$$

Since the mapping $g \longmapsto gx$ is only known to be weakly continuous, this integral is only defined in a weak sense, i.e., $P_\chi x$ is the unique element in the *bidual* X'' satisfying

$$\langle P_\chi x, x' \rangle = \int_G \overline{\chi(g)}\,\langle gx, x' \rangle\,\mathrm{d}g \qquad (x' \in X').$$

Let $K := \overline{\mathrm{conv}}\{(gx)\overline{\chi(g)} : g \in G\}$. By the Kreĭn–Šmulian theorem C.11, the set K is weakly compact. Since the canonical embedding $X \subset X''$ is weak-to-weak* continuous, K is also weak* closed in X'' (and of course convex). But then an easy Hahn–Banach type argument shows that $P_\chi x \in K$, hence $P_\chi x \in X$. (See also [Rudin (1991), Theorem 3.27] for more information on weak integration.)

Once we know that $P_\chi x \in X$, it follows from the invariance of the Haar measure and from the fact that χ is a character that

$$hP_\chi x = \int_G \overline{\chi(g)}\,[(hg)x]\,\mathrm{d}g = \chi(h)P_\chi x$$

for each $h \in G$. (Apply functionals to justify this computation rigorously.) So if $P_\chi x \neq 0$, then it is indeed an eigenvector of $\mathscr{T}$. (By Exercise 12.4, the operators P_χ

form a set of "orthogonal" projections.) We are now in the position to show that there are actually "many" eigenvectors for $\mathscr{T}$.

Proposition 12.9. *Let X be a Banach space and let $G \subset \mathscr{L}(X)$ be an Abelian group of invertible operators which is compact in the weak operator topology. Then the following assertions hold.*

a) *The space $\bigcup_{\chi \in G^*} \operatorname{ran} P_\chi$ is dense in X.*
b) *The strong and the weak operator topologies coincide on G.*
c) *The action $G \times X \longrightarrow X$, $(g,x) \longmapsto gx$ is jointly continuous, where X carries the norm topology and G carries the weak/strong operator topology.*

Proof. a) Take $x' \in X'$ vanishing on every element $P_\chi x$, $x \in X, \chi \in G^*$. This means that for each $x \in X$

$$0 = \langle P_\chi x, x' \rangle = \int_G \langle gx, x' \rangle \, \overline{\chi(g)} \, \mathrm{d}g$$

for every $\chi \in G^*$. Since the characters form an orthonormal basis of $\mathrm{L}^2(G)$, this implies that $\langle gx, x' \rangle = 0$ for all $g \in G$, in particular for $g = \mathrm{I}$. Hence $\langle x, x' \rangle = 0$ for all $x \in X$, i.e., $x' = 0$.

b) Let $Y := \{x \in X : g \longmapsto gy \text{ is strongly continuous}\}$. It suffices to show that $Y = X$. Since G is norm-bounded in $\mathscr{L}(X)$, Y is a closed subspace of X. Moreover, Y contains every element of the form $P_\chi x$, $x \in X, \chi \in G^*$, and hence, by a), $Y = X$.

c) This follows from b) and Proposition C.19. □

We sum up our considerations in the following theorem.

Theorem 12.10 (Jacobs–de Leeuw–Glicksberg). *Let X be a Banach space, and let $\mathscr{T} \subset \mathscr{L}(X)$ be an Abelian relatively weakly compact semigroup. Then one has*

$$X_\mathrm{r} = \overline{\mathrm{lin}}\{x \in X : \forall T \in \mathscr{T} \; \exists \lambda \in \mathbb{T} : Tx = \lambda x\}.$$

Proof. The inclusion $\subset$ follows immediately from the preceding considerations. Conversely, let $0 \neq x \in X$ be such that $Tx = \lambda(T)x$ for every $T \in \mathscr{T}$, and $\lambda(T) \in \mathbb{T}$. Let Q be — as above — the neutral element of $K(\mathscr{S})$, where $\mathscr{S}$ is the weak closure of $\mathscr{T}$. Then there must be $\lambda \in \mathbb{T}$ such that $Qx = \lambda x$. But since $Q^2 = Q$ and $x \neq 0$, $\lambda = 1$, whence $x \in X_\mathrm{r}$. □

Finally, we apply this to the semigroup $\mathscr{T}_T$ generated by a single operator $T \in \mathscr{L}(X)$.

Corollary 12.11. *Let $T \in \mathscr{L}(X)$ generate a relatively weakly compact semigroup $\mathscr{T}_T$. Then*

$$X_\mathrm{r} = \overline{\mathrm{lin}}\{x \in X : \exists \lambda \in \mathbb{T} : Tx = \lambda x\},$$

i.e., X_r is spanned by eigenvectors associated to unimodular eigenvalues of T.

Proof. This is an immediate consequence of Theorem 12.10 since if $Tx = \lambda x$, then $T^n x = \lambda^n x$ for all $n \in \mathbb{N}_0$. □

Remarks 12.12. 1) A power bounded operator T on a Banach space X is said to have **discrete spectrum** if X is spanned by the set of eigenvectors of T corresponding to unimodular eigenvalues. By Corollary 12.11 this holds if and only if $\mathrm{cl}_w \mathscr{T}_T$ is a compact group or, equivalently, if $X = X_\mathrm{r}$ in the corresponding JdLG-decomposition.

2) A topological group G is called **monothetic** if there is $a \in G$ such that $\{a^n : n \in \mathbb{N}_0\}$ is dense in G, see Exercise 12.5. If T has discrete spectrum, then $G := \mathrm{cl}_w \mathscr{T}_T$ is a monothetic group. For more information on compact monothetic groups see [Hewitt and Ross (1979), pp. 407–409] and [Hofmann and Morris (2006)].

12.2 Almost Weakly Stable Vectors

Let us now turn to the stable part associated with a relatively weakly compact semigroup $\mathscr{T}$ on a Banach space X. We confine our considerations to the case that $\mathscr{T}$ is generated by a single operator T, i.e.,

$$\mathscr{T} := \{\mathrm{I}, T, T^2, \dots\}.$$

By Lemma 12.3, $x \in X_\mathrm{s}$ if and only if $0 \in \mathrm{cl}_\sigma\{x, Tx, T^2x, \dots\}$, and we aim to characterise this property by various other asymptotic properties of the sequence $(T^n x)_{n\in\mathbb{N}}$. To formulate the main result, we need a new concept.

Definition 12.13. The **(asymptotic) density** of a set $A \subset \mathbb{N}_0$ is

$$d(A) := \lim_{n\to\infty} \frac{\mathrm{card}(A \cap [0,n))}{n}$$

if it exists. The asymptotic density of a sequence $(n_j)_j \subset \mathbb{N}_0$ is the asymptotic density of the set $\{n_j : j \in \mathbb{N}\}$. Notice that $d(A)$ is just the Cesàro limit of the characteristic sequence $(x_j)_{j\in\mathbb{N}_0}$ of A (if exists).

The density $d(A)$ is a number in $[0,1]$ and measures how large a set/sequence is in $\mathbb{N}_0$ asymptotically. Using this notion, we are now able to state the main result. It explains why the vectors $x \in X_\mathrm{s}$ are called "almost weakly stable".

Theorem 12.14. *Let T generate a relatively weakly compact semigroup $\mathscr{T}_T$ on a Banach space X. Then for $x \in X$ the following assertions are equivalent.*

(i) $x \in X_\mathrm{s}$*;*

(ii) $0 \in \mathrm{cl}_\sigma\{T^n x : n \in \mathbb{N}\}$*;*

(iii) *There exists a subsequence $(n_j)_{j\in\mathbb{N}} \subset \mathbb{N}$ such that $T^{n_j}x \to 0$ weakly as $j \to \infty$;*

(iv) *There exists a subsequence $(n_j)_{j\in\mathbb{N}} \subset \mathbb{N}$ with density 1 such that $T^{n_j}x \to 0$ weakly as $j \to \infty$;*

(v) $\displaystyle\sup_{x'\in X',\|x'\|\le 1} \frac{1}{n}\sum_{k=0}^{n-1}\left|\left\langle T^k x, x'\right\rangle\right| \underset{n\to\infty}{\longrightarrow} 0.$

The proof of this theorem is split into parts. To begin with, we note that the equivalence (i)⇔(ii) is Lemma 12.3, and that the chain of implications (iv)⇒(iii)⇒(ii) holds trivially.

Proof of (ii)⇒(iii): Define $Y := \overline{\operatorname{lin}}\{T^n x : n \ge 0\}$. Then Y is a closed separable subspace of X. Since Y is also weakly closed, the set $A := \operatorname{cl}_\sigma\{T^n x : n \ge 0\}$ (closure in X) is contained in Y and hence a weakly compact subset of Y. This shows that we may suppose without loss of generality that X is separable. By hypothesis,

$$0 \in \operatorname{cl}_\sigma\{T^n x : n \ge 0\} = \{x, Tx, \ldots, T^{k-1}x\} \cup \operatorname{cl}_\sigma\{T^n x : n \ge k\}$$

for every $k \in \mathbb{N}$. Since $T^k x = 0$ implies $T^m x = 0$ for all $m \ge k$, one has

$$0 \in \bigcap_{k\in\mathbb{N}} \operatorname{cl}_\sigma\{T^n x : n \ge k\},$$

that is, 0 is a weak cluster point of $(T^n x)_{n\in\mathbb{N}_0}$. Since X is separable, the weak topology on A is metrisable (see Proposition C.10). This establishes (iii). □

The implication (iii)⇒(v) is a special case of the following result from [Jones and Lin (1976)].

Proposition 12.15 (Jones–Lin). *Let T be a power bounded operator on a Banach space X, and let $x \in X$. If $\lim_j T^{n_j} x = 0$ weakly for some subsequence $(n_j)_j \subset \mathbb{N}$, then*

$$\sup_{x'\in X',\|x'\|\le 1} \frac{1}{n}\sum_{k=0}^{n-1}\left|\left\langle T^k x, x'\right\rangle\right| \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. First, we suppose that T is a contraction. The dual unit ball $\mathrm{B} := \{x' \in X' : \|x'\| \le 1\}$ is a compact space with respect to the weak*-topology. Since T is a contraction, T' leaves B invariant, and thus gives rise to a TDS $(\mathrm{B}; T')$ (see Example 2.4). We consider its induced operator S on $\mathrm{C}(\mathrm{B})$, i.e.,

$$(Sf)(x') := f(T'x') \qquad (x' \in \mathrm{B}).$$

Consider the function $f(x') := |\langle x, x'\rangle|$, $x' \in \mathrm{B}$. Then $f \in \mathrm{C}(\mathrm{B})$, and we claim that

$$f \in \overline{\operatorname{ran}}(\mathrm{I} - S).$$

To establish this we use the Hahn–Banach theorem. Take $\mu \in \mathrm{M}(\mathrm{B}) = \mathrm{C}(\mathrm{B})'$ which vanishes on $\operatorname{ran}(\mathrm{I} - S)$, i.e., $\langle g, \mu\rangle = \langle Sg, \mu\rangle$ for all $g \in \mathrm{C}(\mathrm{B})$. Then in particular

$$\langle f, \mu\rangle = \langle S^{n_j} f, \mu\rangle = \int_{\mathrm{B}} \left|\left\langle x, T'^{n_j} x'\right\rangle\right| \mu(\mathrm{d}x') = \int_{\mathrm{B}} \left|\left\langle T^{n_j} x, x'\right\rangle\right| \mu(\mathrm{d}x')$$

for every $j \in \mathbb{N}$. Define $g_j(x') := |\langle T^{n_j} x, x'\rangle|$. Then $g_j \to 0$ pointwise on B, by hypothesis. By the Dominated Convergence Theorem, $\int_{\mathrm{B}} g_j \,\mathrm{d}\mu \to 0$. This shows that

$\langle f, \mu \rangle = 0$ and the claim is proved. Now, since S is power bounded, the Cesàro-sums $\mathrm{A}_n[S]f$ must converge to 0 in the norm of $\mathrm{C}(\mathrm{B})$ (cf. the proof of Theorem 8.1). But this is exactly what was to be proved.

The contraction case being settled, suppose now that T is just power-bounded. Define $|||y||| := \sup_{n\geq 0} \|T^n y\|$ for $y \in X$. Then $|||\cdot|||$ is a norm on X satisfying $\|y\| \leq |||y||| \leq M\|y\|$, with $M := \sup_n \|T^n\|$. With respect to this new norm T is a contraction. Moreover

$$|||x'||| = \sup\{|\langle y, x'\rangle| \, : \, |||y||| \leq 1\} \leq \sup\{|\langle y, x'\rangle| \, : \, \|y\| \leq 1\} = \|x'\|$$

for any $x' \in X'$. Hence the dual unit ball with respect to the new norm contains the dual unit ball with respect to the old norm, and so the contraction case can be applied to conclude the proof. $\square$

Finally, the missing implication (v)⇒(iv) in Theorem 12.14 is a special case of the following general fact, first observed in [Koopman and von Neumann (1932)] in a continuous setting.

Lemma 12.16 (Koopman–von Neumann). *For $x = (x_n)_n \in \ell^\infty$ the following conditions are equivalent.*

(i) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |x_i| = 0.$

(ii) *There exists a subset $N \subset \mathbb{N}$ with $d(N) = 1$ and $\lim_{n\in N} x_n = 0$.*

Proof. We may suppose that $x_n \in \mathbb{R}_+$ for all $n \in \mathbb{N}_0$. (i)⇒(ii). Let $J_k := \{n \in \mathbb{N}_0 : x_n \geq 1/k\}$. Then $J_1 \subset J_2 \subset \ldots$ and $d(J_k) = 0$, since

$$\frac{1}{k} \frac{\operatorname{card}(J_k \cap [0,n))}{n} \leq \frac{1}{n} \sum_{j=0}^{n-1} x_j \to 0 \qquad (n \to \infty).$$

Therefore, we can choose integers $1 \leq n_1 < n_2 < \ldots$ such that

$$\frac{\operatorname{card}(J_k \cap [0,n))}{n} < \frac{1}{k} \qquad (n \geq n_k).$$

Now we define

$$J := \bigcup_{k\geq 1} J_k \cap [n_k, \infty)$$

and claim that $d(J) = 0$. Indeed, let $m \in \mathbb{N}$ such that $n_k \leq m < n_{k+1}$. Then $J \cap [0,m) \subset J_k \cap [0,m)$ (because the J_k increase with k) and so

$$\frac{\operatorname{card}(J \cap [0,m))}{m} \leq \frac{\operatorname{card}(J_k \cap [0,m))}{m} \leq \frac{1}{k}$$

since $m \geq n_k$. This shows $d(J) = 0$ as claimed. Defining $N := \mathbb{N}_0 \setminus J$, it is clear by construction that $\lim_{n\in N} x_n = 0$.

(ii)⇒(i). Let $\varepsilon > 0$ and $c := \sup\{x_n : n \in \mathbb{N}_0\}$. Choose $n_\varepsilon \in \mathbb{N}$ such that $n > n_\varepsilon$ and $n \in N$ imply that $x_n < \varepsilon$. If $n > n_\varepsilon$ we conclude that

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1} x_j \le \frac{cn_\varepsilon}{n} + \frac{1}{n}\sum_{j=n_\varepsilon}^{n-1} x_j &\le \frac{cn_\varepsilon}{n} + \varepsilon\frac{n-n_\varepsilon}{n} + \frac{1}{n}\sum_{j\in[n_\varepsilon,n)\setminus N} x_j \\ &\le \frac{cn_\varepsilon}{n} + \varepsilon + c\frac{\operatorname{card}([0,n)\setminus N)}{n}.\end{aligned}$$

Since $d(\mathbb{N}\setminus N) = 0$, this implies

$$\limsup_{n\to\infty} \frac{1}{n+1}\sum_{j=0}^{n} x_j \le \varepsilon,$$

and since ε was arbitrary, the proof is complete. □

Remark 12.17. The construction in the first part of the proof of the lemma may seem miraculous, but it is not. Indeed, the set N should satisfy $N \setminus [0,n_k) \subset J_k^{\mathrm{c}}$ for some increasing sequence n_k tending to ∞. This is equivalent to $N^{\mathrm{c}} \cup [0,n_k) \supset J_k$ which in turn is equivalent to $N^{\mathrm{c}} \supset J_k \cap [n_k,\infty)$. This being true for every k is again equivalent to

$$N^{\mathrm{c}} \supset \bigcup_{k\ge 1} J_k \cap [n_k,\infty)$$

Since N^{c} should be as small as possible, the definition of $J = N^{\mathrm{c}}$ is natural.

We conclude this section with a spectral characterisation of the case $X = X_{\mathrm{s}}$ which is a direct corollary of Theorem 12.10.

Proposition 12.18. *Let $T \in \mathscr{L}(X)$ generate a relatively weakly compact semigroup, and let $X = X_{\mathrm{r}} \oplus X_{\mathrm{s}}$ be the corresponding JdLG-decomposition. Then the following assertions are equivalent.*

(i) $X = X_{\mathrm{s}}$;

(ii) $\sigma_p(T) \cap \mathbb{T} = \emptyset$.

An operator T which satisfies the equivalent conditions of the proposition is called **almost weakly stable**.

Concluding Remarks

According to the JdLG-decomposition, the operator theory of an MDS or a TDS can be reduced to the two extreme cases of operators with *discrete spectrum* and *almost weakly stable* operators. When we discuss mixing properties of dynamical systems in Lecture 14, we shall see that the latter case appears precisely when the MDS is “weakly mixing”. In the first case, the dynamical system turns out to be isomorphic to a group rotation. This is the content of the Halmos–von Neumann theorem, the topic of the coming lecture.

The Jacobs–de Leeuw–Glicksberg decomposition is one of the most powerful tools in the asymptotic analysis of dynamical systems, both discrete and continuous. Although it has a long (and glorious) history already, recently a new aspect has emerged: it can be seen as an instance of a quite general (heuristic) phenomenon called "dichotomy between structure and randomness", see [Tao (2007)]. In this terminology, the stable part X_{s} is considered to be "random" and the reversible part X_{r} is "structured". The name "structured" comes from the fact that the operator on the reversible part is completely determined by its unimodular eigenvalues and the corresponding eigenvectors. On the opposite, on the stable part the behaviour of T^n is unknown for numbers n from a set of density zero, with no further information about the distribution of these numbers.

Supplement: Grothendieck's Theorem

In this supplement we present an important example of weakly compact operator semigroups, generalising the case of rotations on compact groups.

Proposition 12.19. *Let S be a compact semitopological semigroup, and define $(L_x f)(y) := f(xy)$ for $f \in \mathrm{C}(S)$ and $x, y \in S$. Then*

$$\{L_x f \, : \, x \in S\} \subset \mathrm{C}(S)$$

is weakly compact for each $f \in \mathrm{C}(S)$. Consequently, $\mathscr{T} := \{L_x : x \in S\}$ is a relatively weakly compact semigroup on $\mathrm{C}(S)$, and each operator L_x, $x \in S$, is mean ergodic on $\mathrm{C}(S)$.

In classical terms, the proposition says that on a compact semitopological semigroup every continuous function is weakly almost periodic. The proof of Proposition 12.19 rests on the following important theorem from [Grothendieck (1952)].

Theorem 12.20 (Grothendieck). *Let K be a compact topological space, and let $A \subset \mathrm{C}(K)$ be a uniformly bounded subset. Then A is compact in the topology of pointwise convergence if and only if it is compact in the weak topology of $\mathrm{C}(K)$.*

To simplify terminology we use the terms "p-compact, p-continuous, etc." when we mean compact, continuous etc. with respect to the topology of pointwise convergence.

Proof of Proposition 12.19: Since the left multiplication with $x \in S$ is continuous, $L_x f \in \mathrm{C}(S)$ whenever $f \in \mathrm{C}(S)$. Hence L_x is a bounded linear operator on $\mathrm{C}(S)$. (Indeed, $L_x = T_{\varphi_x}$ in the terminology of previous lectures.) Since $L_{xy} = L_y L_x$, the set $\mathscr{S}$ is a semigroup. Denote by p the topology of pointwise convergence on $\mathrm{C}(S)$. If $f \in \mathrm{C}(S)$, then $S \longrightarrow \mathrm{C}(S)$, $x \longmapsto L_x f$ is p-continuous, since also the right multiplications on S are all continuous. Since the p-topology is Hausdorff, the set $\mathscr{T} f = \{L_x f \, : \, x \in S\}$ is p-compact. Since this set is obviously bounded, Grothendieck's theorem implies that $\mathscr{T} f$ is weakly compact. By Lemma 12.1, $\mathscr{T}$

is relatively weakly compact as claimed. Finally, every operator L_x is mean ergodic, by Remark 12.2.2. □

We shall now prove Grothendieck's Theorem 12.20 following the lines of [Berglund et al. (1989), Appendix A].

The idea of the proof is readily sketched. If A is weakly compact, then it is p-compact since the identity mapping $\mathrm{Id} : (\mathrm{C}(K), \sigma) \longrightarrow (\mathrm{C}(K), p)$ is continuous. For the converse implication, one employs the Eberlein–Šmulian theorem C.9 reducing the question to weak *sequential* compactness.

Hence one starts with a sequence $(f_n)_n \subset A$ and tries to find a weakly convergent subsequence. Since the sequence is uniformly bounded and since continuous functionals on K can be identified with measures (Riesz' theorem), the Dominated Convergence Theorem reduces the question further: it suffices to show that there is $f \in A$ and a subsequence $(f_{n_k})_k$ such that $f_{n_k}(x) \longrightarrow f(x)$ for all $x \in K$.

Now, by p-compactness of A, there is a p-cluster point f of $(f_n)_n$:

$$f \in \bigcap_{m\in\mathbb{N}} \mathrm{cl}_p\big\{f_n : n \geq m\big\} \cap A.$$

For every point $x \in K$ one can therefore choose a subsequence of $(f_n)_n$ converging at x to $f(x)$. The major difficulty is to show that there is a subsequence that does it for all points $x \in K$ simultaneously.

To obtain this one uses a "metrisation trick". Introduce the semi-metric d on K by

$$d(x,y) := |f(x) - f(y)| + \sum_{n\in\mathbb{N}} \frac{1}{2^n} |f_n(x) - f_n(y)| \qquad (x, y \in K).$$

This semi-metric induces the coarsest topology that makes all functions f, f_n continuous. This topology is not Hausdorff in general, but one can pass to the quotient space $L := K/{\sim}$ identifying points with d-distance zero. Since all functions f, f_n are continuous on the original topology of K, L is a continuous image of K and hence a compact metric space. In particular, L contains a countable dense set, and pulling this back to K yields a countable set $M \subset K$ which is d-dense in K.

As said above, for each $x \in M$ there is a subsequence of $(f_n)_n$ that converges to f at x. Since M is countable, a standard diagonal procedure yields a subsequence that does it for all $x \in M$ simultaneously. By renaming the sequence, we can therefore suppose without loss of generality that

$$f_n(x) \to f(x) \qquad (x \in M).$$

Now let $x \in K$ be an arbitrary point. Since M is d-dense in K, there is a sequence $(x_m)_m \subset M$ such that $d(x, x_m) \to 0$ as $m \to \infty$. It is now sufficient to prove the following lemma.

Lemma 12.21. *In the situation described above, every subsequence of $(f_n)_n$ has a subsequence that converges to $f(x)$ at x.*

Proof. Renaming again, we have to find a subsequence $(f_{n_k})_k$ such that $f_{n_k}(x) \to f(x)$ as $k \to \infty$. Since K is compact, there is a cluster point $y \in K$ of the sequence $(x_m)_m$. Since the mapping $d(x,\cdot)$ is continuous on K, it follows that $d(x,y) = 0$, in particular $f_n(x) = f_n(y)$ for all $n \in \mathbb{N}$.

Since A is p-compact there is a p-cluster point $g \in A$ of the sequence $(f_n)_n$. Passing to a subsequence we may then suppose that $\lim_n f_n(y) = g(y)$. Since point evaluations are p-continuous (by definition) $g(x_m)$ is a cluster point of the sequence $(f_n(x_m))_n$, which converges to $f(x_m)$. So we conclude that $g(x_m) = f(x_m)$ for each $m \in \mathbb{N}$. On the other hand, since g is continuous, $g(y)$ is a cluster point of $(g(x_m))_m$. But $g(x_m) = f(x_m) \to f(x)$ and hence $g(y) = f(x)$. Putting the pieces together we obtain

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} f_n(y) = g(y) = f(x)$$

as we wanted. $\square$

Exercises

1. Let A be an algebra over $\mathbb{R}$ and let $S \subset A$ be multiplicative, i.e., $S \cdot S \subset S$. Show that $\mathrm{conv}(S)$ is multiplicative, too. Show that if $st = ts$ holds for all $s,t \in S$ then it also holds for all $s,t \in \mathrm{conv}(S)$.

2. Prove the analogue of Lemma 12.1 for the strong operator topology.

3. Let H be a Hilbert space and let $P : H \longrightarrow H$ be a projection. Prove that P is an orthogonal projection if and only if $\|P\| \le 1$.

4. Let X be a Banach space and let $G \subset \mathscr{L}(X)$ be a (weakly) compact Abelian group. For $\chi \in G^*$ let

$$P_\chi x := \int_G \overline{\chi(g)}\,(gx)\,\mathrm{d}g \qquad (x \in X)$$

as in Section 12.1. Show that $P_\chi P_\eta = \delta_{\chi\eta} P_\chi$ for any $\chi, \eta \in G^*$, and further that if $X = H$ is a Hilbert space, then P_χ is self-adjoint, and hence

$$H = \bigoplus_{\chi \in G^*} \operatorname{ran} P_\chi$$

is an orthogonal decomposition of H.

5. Let G be a compact group and $a \in G$. Suppose that $\{a^n : n \in \mathbb{Z}\}$ is dense in G. Show that $\{a^n : n \in \mathbb{N}\}$ is dense in G. (Hint: Consider the compact Abelian semigroups $H_k := \mathrm{cl}\{a^n : n \ge k\}$ and use Theorem 11.2; or cf. Lecture 3).

6. If X' is separable and T generates a weakly relatively compact semigroup $\mathscr{T}_T$ on X, then there exists a sequence $(n_j)_{j\in\mathbb{N}} \subset \mathbb{N}$ with density 1 such that

$$T^{n_j} x \to 0 \text{ weakly for all } x \in X_{\mathrm{s}},$$

i.e., $(n_j)_{j\in\mathbb{N}}$ can be chosen independently of x.
(Hint: Observe first that the closed unit ball in $\mathscr{L}_w(X)$ is metrisable for the metric

$$d(T,S) := \sum_{j,k=1}^{\infty} \frac{\langle (T-S)x_k, x_j' \rangle}{2^{j+k}\, \|x_k\| \cdot \|x_j'\|}$$

for some fixed dense subsets $\{x_k : k \in \mathbb{N}\} \subset X \setminus \{0\}$ and $\{x_j' : j \in \mathbb{N}\} \subset X' \setminus \{0\}$. Conclude from $0 \in \mathrm{cl}_w\{T^n : n \in \mathbb{N}_0\}$ that there exists a subsequence $(n_j)_j$ such that $T^{n_j} \to 0$ in the weak operator topology. To obtain a sequence with density one imitate the proof of the Jones–Lin theorem for the function $f(x') := \sum_{k=1}^{\infty} 2^{-k} \frac{1}{\|x_k\|} \langle x_k, x' \rangle$.)

7. For a subset $A \subset \mathbb{N}_0$ the **upper** and **lower** density are defined as

$$\overline{d}(A) := \limsup_{n\to\infty} \frac{1}{n} \operatorname{card}(A \cap [0,n)) \quad \text{and} \quad \underline{d}(A) := \liminf_{n\to\infty} \frac{1}{n} \operatorname{card}(A \cap [0,n)).$$

Show that $\overline{d}(A)$ and $\underline{d}(A)$ do not change if we change A by finitely many elements.

8. Prove the following assertions.

a) For $\alpha, \beta \in [0,1]$, $\alpha \le \beta$ there is a set $A \subset \mathbb{N}$ with $\underline{d}(A) = \alpha$ and $\overline{d}(A) = \beta$.

b) Suppose that $A \subset \mathbb{N}$ **relatively dense** (i.e., has bounded gaps; see Definition 3.10), then $\underline{d}(A) > 0$. The converse implication, however, is not true.

c) Let $\mathbb{N} = A_1 \cup A_2 \cup \cdots \cup A_k$. Then there is $1 \le j \le k$ with $\overline{d}(A_j) > 0$. The assertion is not true if $\overline{d}$ is replaced by $\underline{d}$.

d) If $d(A) = d(B) = 1$, then $d(A \cap B) = 1$.

Lecture 13
Dynamical Systems with Discrete Spectrum

> I want to know how God created this world. I am not interested in this or that phenomenon, in the spectrum of this or that element. I want to know His thoughts, the rest are details.[1]
>
> Albert Einstein

In the previous lecture we have obtained important information on the structure of weakly compact semigroups of linear operators on Banach spaces. The central example is the closure $\mathscr{S}$ in the weak operator topology of a semigroup $\mathscr{T}_T := \{T^n : n \in \mathbb{N}_0\}$ generated by a single operator $T \in \mathscr{L}(X)$, X a Banach space. In such a compact semitopological semigroup we find a smallest ideal $K(\mathscr{S})$, which is actually a group. The unit element Q of this group is a projection inducing the JdLG-decomposition $X = X_{\mathrm{r}} \oplus X_{\mathrm{s}}$.

In this lecture we confine our attention to the case when $X = X_{\mathrm{r}}$, i.e., $Q = \mathrm{I}$ is the identity operator. Equivalently, $K(\mathscr{S}) = \mathscr{S}$ consists of invertible operators only. By Corollary 12.11 from Lecture 12 this implies that T has **discrete spectrum**, i.e., X is spanned by the eigenvectors of T corresponding to unimodular eigenvalues, i.e.,

$$X = \overline{\mathrm{lin}}\{x : \exists\, \lambda \in \mathbb{T} : Tx = \lambda x\}. \tag{13.1}$$

(cf. Remark 12.12.) On the other hand, it is easy to see using Lemma 12.1 that if (13.1) holds and T is power-bounded, then $\mathrm{cl}_w \mathscr{T}_T$ is a weakly compact semigroup with $X = X_{\mathrm{r}}$ (see Exercise 13.5). In particular, a power-bounded operator with discrete spectrum is *mean ergodic*, by Theorem 8.8 (iii).

The aim of this lecture is now to examine dynamical systems whose induced operator has discrete spectrum. Our chief example is again the group rotation.

Example 13.1. Let G be a compact Abelian group and let $T := T_{\varphi_a}$ be the operator induced by the rotation with $a \in G$. Since every character $\chi \in G^*$ is an eigenfunction

[1] Source: Eric Weisstein's World of Physics;
http://scienceworld.wolfram.com/biography/Einstein.html

of T corresponding to the eigenvalue $\chi(a) \in \mathbb{T}$ and $\operatorname{lin} G^*$ is dense in $\mathrm{C}(G)$ (Proposition 11.13), T has discrete spectrum on $\mathrm{C}(G)$. A fortiori, T has discrete spectrum also on $\mathrm{L}^p(G, \mathfrak{B}_0(G), \mu)$ for every $1 \le p < \infty$, where μ is the Haar measure on G.

A TDS $(K;\varphi)$ (MDS $(\Omega, \Sigma, \mu; \varphi)$) is said to have **discrete spectrum** if its induced operator T on $\mathrm{C}(K)$ (on $\mathrm{L}^2(\Omega, \Sigma, \mu)$) has discrete spectrum[1]. In this case, T is a bijective isometry and hence satisfies $\sigma(T) \subset \mathbb{T}$ (Exercise 13.3). If in addition the TDS is minimal (the MDS is ergodic), then the eigenspaces of T are all one-dimensional, by Theorem 4.15 (or Proposition 7.11). We shall show that the dynamical system is completely determined (up to isomorphism) by this spectral information. Indeed, it will turn out that the dynamical system is essentially a minimal/ergodic rotation on a compact group. To begin with, let us have a closer look at this model situation.

13.1 Monothetic Groups

Recall from Lecture 12 that a topological group G is called **monothetic** if there is an element $a \in G$ such that $\{a^n : n \in \mathbb{N}_0\}$ is dense in G; in this case a is called a **generating element** of G. By Exercise 11.3 monothetic groups are necessarily Abelian. For a compact group G with Haar measure μ and an element $a \in G$ the following statements are equivalent:

(i) The group G is monothetic with generating element a.

(ii) The cyclic subgroup $\langle a \rangle := \{a^n : n \in \mathbb{Z}\}$ is dense in G.

(iii) The TDS (G, φ_a) is minimal.

(iv) The MDS $(G, \mathfrak{B}_0(G), \mu; \varphi_a)$ is ergodic.

(This follows from Theorem 9.11 and Exercise 12.5.) We now characterise the duals of compact monothetic groups.

Proposition 13.2. *Let G be a compact group. If G is monothetic with generating element $a \in G$, then G^* is (algebraically) isomorphic to the group*

$$\mathrm{X}(a) := \{\chi(a) : \chi \in G^*\} \subset \mathbb{T}.$$

Conversely, if G^* is algebraically isomorphic to a subgroup of $\mathbb{T}$, then G is monothetic.

Proof. Suppose that a generates G. Then the set $\mathrm{X}(a)$ is a subgroup of $\mathbb{T}$ and the evaluation mapping $\chi \longmapsto \chi(a)$ is a surjective group homomorphism. It is even injective since $\{a^n : n \in \mathbb{N}_0\}$ is dense in G and characters are continuous.

[1] It turns out that one could use any L^p-space with $1 \le p < \infty$ in this definition. See Remark 13.10 below.

For the converse, suppose that $\gamma : G^* \longrightarrow \mathbb{T}$ is an injective homomorphism, i.e., a character of G^*. By Pontryagin's Theorem 11.19 there is $a \in G$ with $\gamma(\chi) = \chi(a)$ for all $\chi \in G^*$. We claim that $\langle a \rangle$ is dense in G. Indeed, if $H := \mathrm{cl}\langle a \rangle \neq G$, then by Corollary11.11 there exists a character $\chi \neq \mathbf{1}$ of G being constant 1 on $\langle a \rangle$. But then γ is not injective, a contradiction. □

The set $\mathrm{X}(a)$ can also be interpreted in a different way.

Lemma 13.3. *Let G be a compact Abelian group with Haar measure μ. Let $a \in G$, φ_a the rotation by a, and let $T := T_{\varphi_a}$ be the induced operator. Then*

$$\mathrm{X}(a) = \sigma_p(T),$$

when T is considered either as an operator on the space $\mathrm{C}(G)$ or on the space $\mathrm{L}^p(G, \mathfrak{B}_0(G), \mu)$, $1 \leq p < \infty$.

Proof. For every $\chi \in G^*$ one has $(T\chi)(x) = \chi(ax) = \chi(a)\chi(x)$ for any $x \in G$. Hence $\chi(a)$ is an eigenvalue of T. This shows that $\mathrm{X}(a) \subset \sigma_p(T)$. The converse inclusion follows from Exercise 13.6. □

The next result characterises isomorphic minimal/ergodic group rotations.

Proposition 13.4. *For two monothetic compact groups G and H with generating elements $a \in G$ and $b \in H$ the following statements are equivalent.*

(i) *The TDSs $(G; \varphi_a)$ and $(H; \varphi_b)$ are isomorphic.*

(ii) *The MDSs $(G, \mathfrak{B}_0(G), \mu_G; \varphi_a)$ and $(H, \mathfrak{B}_0(H), \mu_H; \varphi_b)$ are isomorphic.*

(iii) *There is a topological group isomorphism $\Phi : G \longrightarrow H$ with $\Phi(a) = b$.*

(iv) $\mathrm{X}(a) = \mathrm{X}(b)$.

Proof. (iii)⇒(i),(ii) If Φ is a topological group isomorphism with $\Phi(a) = b$, then Φ is also an isomorphism of the dynamical systems.

(i) or (ii)⇒(iv) If the dynamical systems are isomorphic, the induced operators on the corresponding spaces are similar. Hence they have the same spectral properties, in particular they have the same point spectrum. So (iv) follows from Lemma 13.3.

(iv)⇒(iii) If $\mathrm{X}(a) = \mathrm{X}(b)$, then by Proposition 13.2 the discrete groups G^* and H^* are isomorphic. Recall that this isomorphism is given by

$$G^* \ni \chi \longmapsto \chi(a) = \eta(b) \longmapsto \eta \in H^*.$$

By Pontryagin's Theorem 11.19 this implies that G and H are topologically isomorphic under $\Phi : G \longrightarrow H$, where $\Phi(g) \in H$ is the unique element with $\eta(\Phi(g)) = \chi(g)$ if $\chi(a) = \eta(b)$, $\chi \in G^*$, $\eta \in H^*$. Now $\Phi(a) = b$, and the assertion follows. □

Example 13.5. Let H be an arbitrary subgroup of $\mathbb{T}_{\mathrm{d}}$. By Proposition 13.2, the compact group $G := H^*$ is monothetic with some generating element $a \in G$. As shown before (see Proposition 3.5), the group rotation $(G;\varphi_a)$ is minimal and has discrete spectrum. Its induced operator has point spectrum

$$\sigma_p(T) = \{\chi(a) : \chi \in G^* = H^{**} = H\} = H,$$

by Propositions 11.23 and 13.2.

13.2 Minimal TDSs with Discrete Spectrum

We now confine our attention to the topological case. Our aim is to give a complete description (up to isomorphism) of minimal TDS with discrete spectrum.

If two TDSs are isomorphic, the induced operators on the corresponding $\mathrm{C}(K)$-spaces are similar and hence have the same spectral properties. In particular, they have the same point spectrum. This shows that the point spectrum of the induced operator is an **isomorphism invariant** of TDSs. On the other hand, Proposition 13.4 and Lemma 13.3 show that within the class of minimal group rotations, the point spectrum completely determines the system up to isomorphism. Therefore we say that it is a **complete isomorphism invariant** for this class. Our aim in this section is to show that it is a complete isomorphism invariant even for the (larger) class of minimal TDS with discrete spectrum. This is done by showing that every minimal TDS with discrete spectrum is essentially a minimal group rotation.

In the proof we shall need the following straightforward consequence of Theorem 4.10.

Lemma 13.6. *Two TDSs $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ are isomorphic if and only if there exists a unital Banach algebra isomorphism*

$$\Phi : \mathrm{C}(K_1) \longrightarrow \mathrm{C}(K_2)$$

with $\Phi \circ T_{\varphi_1} = T_{\varphi_2} \circ \Phi$. In this case Φ is induced by the isomorphism of the TDSs.

We now can state the major result.

Theorem 13.7. *Let $(K;\varphi)$ be a minimal TDS with discrete spectrum. Then it is isomorphic to a minimal rotation on a compact monothetic group G.*

Proof. By hypothesis the induced operator $T = T_\varphi$ on $\mathrm{C}(K)$ generates a compact group

$$G := \mathrm{cl}_w\{T^n : n \in \mathbb{N}_0\} = \mathrm{cl}_s\{T^n : n \in \mathbb{N}_0\} \subset \mathscr{L}_s(\mathrm{C}(K))$$

(see Proposition 12.9). We now show that $(K;\varphi)$ is isomorphic to $(G;\varphi_T)$, where φ_T is the left rotation by $T \in G$. The operator T and hence every $S \in G$ is a Banach algebra homomorphism on $\mathrm{C}(K)$. Therefore, by Theorem 4.10, there exist *unique* continuous mappings

$$\psi_S : K \longrightarrow K$$

such that $$Sf = f \circ \psi_S \qquad \text{for every } S \in G,\ f \in \mathrm{C}(K).$$

We also obtain

$$\varphi = \psi_T \quad \text{and} \quad \psi_{S_1S_2} = \psi_{S_1} \circ \psi_{S_2} \qquad \text{for all } S_1, S_2 \in G.$$

Pick $x_0 \in K$ and define

$$\Psi : G \longrightarrow K \quad \text{by} \quad \Psi(S) := \psi_S(x_0) \text{ for } S \in G.$$

We show that this map yields an *isomorphism* between $(G;\varphi_T)$ and $(K;\varphi)$. Indeed, let $S, R \in G$, $f \in \mathrm{C}(K)$ and $\varepsilon > 0$ be given. Then

$$|f(\Psi(S)) - f(\Psi(R))| = |f(\psi_S(x_0)) - f(\psi_R(x_0))| = |Sf(x_0) - Rf(x_0)| \le \varepsilon,$$

whenever $\|Sf - Rf\|_\infty \le \varepsilon$. This shows the continuity of $f \circ \Psi$ for every $f \in \mathrm{C}(K)$ and — by Lemma 4.9 — also the continuity of Ψ.

Observe that $\Psi(G)$ is a closed φ-invariant subset of K. Since $(K;\varphi)$ is minimal, it follows that $\Psi(G) = K$, i.e., Ψ is surjective.

To prove injectivity, suppose that $\Psi(S_1) = \Psi(S_2)$ for some $S_1, S_2 \in G$. Then

$$\psi_{S_1}(\varphi^n(x_0)) = \psi_{S_1}(\psi_T^n(x_0)) = \psi_T^n(\psi_{S_1}(x_0)) = \psi_T^n(\psi_{S_2}(x_0)) = \psi_{S_2}(\varphi^n(x_0))$$

for all $n \in \mathbb{N}_0$. By minimality, $\overline{\mathrm{orb}}_+(x_0) = K$. Hence from continuity of ψ_{S_1} and ψ_{S_2} we conclude that $\psi_{S_1} = \psi_{S_2}$, which yields $S_1 = S_2$.

Since $\varphi = \psi_T$, the diagram

$$\begin{tikzcd} G \arrow[r, "\varphi_T"] \arrow[d, "\Psi"'] & G \arrow[d, "\Psi"] \\ K \arrow[r, "\varphi"'] & K \end{tikzcd}$$

commutes, and hence the proof is complete. □

The above representation theorem and the characterisation of isomorphic minimal group rotations in Proposition 13.4 yield the next theorem.

Theorem 13.8. *Two minimal TDSs $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ with discrete spectrum are isomorphic if and only if the corresponding induced operators have the* same *point spectrum.*

Proof. Of course, operators induced by isomorphic TDSs always have the same point spectrum. For the converse, suppose that $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ are minimal TDSs with the induced operators $T_1 := T_{\varphi_1}$, $T_2 := T_{\varphi_2}$ with discrete spectrum and the same point spectrum. The two TDSs are isomorphic to some group rotations $(G_1;\varphi_{a_1})$ and $(G_2;\varphi_{a_2})$ by Theorem 13.7. But the induced operators of these TDSs have the same point spectrum, so by Proposition 13.4 they are isomorphic. □

Combining this theorem with Example 13.5 we see that for each subgroup $H \subset \mathbb{T}$ there is a (up to isomorphism) unique minimal TDS $(K;\varphi)$ with discrete spectrum whose induced operator has exactly H as its point spectrum.

13.3 Ergodic MDSs with Discrete Spectrum

If $(\Omega,\Sigma,\mu;\varphi)$ is an MDS with discrete spectrum, then the induced operator $T := T_\varphi$ is a bijective isometry on $\mathrm{L}^2(\Omega,\Sigma,\mu)$, hence unitary. Analogously to what we have done above, we show that the isomorphism of two ergodic MDSs with discrete spectrum depends only on the point spectrum of the induced operators, and each ergodic MDS with discrete spectrum is isomorphic to an ergodic rotation on a compact group.

The approach is the following: Given an MDS $(\Omega,\Sigma,\mu;\varphi)$ with discrete spectrum, we construct a closed T-invariant subalgebra $\mathscr{A}$ of $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$, on which T acts as an operator with discrete spectrum. By the Gelfand–Naimark Theorem, we may represent $\mathscr{A}$ as $\mathrm{C}(K)$ for some compact space K. Thus the MDS turns into a TDS on this space, and we can use the already established TDS-result. The main point is to choose the subalgebra $\mathscr{A}$ (i.e., the space $\mathrm{C}(K)$) carefully such that the TDS becomes minimal and has discrete spectrum.

The following result from [Halmos and von Neumann (1942)] is the fundamental step in the "classification" of ergodic MDSs with discrete spectrum.

Theorem 13.9 (Halmos–von Neumann). *Let $(\Omega,\Sigma,\mu;\varphi)$ be an ergodic MDS such that the induced operator $T := T_\varphi$ has discrete spectrum in $\mathrm{L}^p(\Omega,\Sigma,\mu)$ for some $1 \le p < \infty$. Then it is isomorphic to a rotation on a compact monothetic group endowed with the normalised Haar measure.*

Proof. Every eigenfunction $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$ of T belongs to $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$ since all eigenvalues λ are unimodular and

$$T|f| = |Tf| = |\lambda|\cdot|f| = |f| = c\cdot\mathbf{1}$$

holds for ergodic MDSs (see Proposition 7.8). Since the product of two eigenfunctions is again an eigenfunction, the linear hull of

$$\bigl\{f \in \mathrm{L}^p(\Omega,\Sigma,\mu) : Tf = \lambda f \text{ for some } |\lambda| = 1\bigr\}$$

is a conjugation-invariant subalgebra of $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$, and its closure $\mathscr{A}$ in $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$ is a commutative unital C^*-algebra. By the Gelfand–Naimark Theorem 4.6 there exists a compact space K and an isometric $*$-isomorphism $\Phi : \mathscr{A} \longrightarrow \mathrm{C}(K)$. The restriction of T to $\mathscr{A}$ is an algebra homomorphism on $\mathscr{A}$, therefore so is $S := \Phi \circ T \circ \Phi^{-1}$ on $\mathrm{C}(K)$. By Theorem 4.10, S is induced by some continuous mapping $\psi : K \longrightarrow K$.

We show that $(K;\psi)$ is a minimal TDS with discrete spectrum. Since $T|_{\mathscr{A}}$ has discrete spectrum on the space $\mathscr{A}$, also S has discrete spectrum on the space $\mathrm{C}(K)$ and is mean ergodic thereon. We have $1 = \dim \mathrm{Fix}(T) = \dim \mathrm{Fix}(S)$, i.e., ψ is uniquely ergodic. The measure ν on K defined by

$$\langle f, \nu \rangle := \left\langle \Phi^{-1} f, \mu \right\rangle \quad (f \in \mathrm{C}(K))$$

is ψ-invariant and strictly positive, because μ is φ-invariant and strictly positive. This means that ψ is strictly ergodic, and by Proposition 9.10 the TDS $(K;\psi)$ is minimal.

We can now apply Theorem 13.8 to $(K;\psi)$ and obtain the existence of a compact monothetic group G with generating element a such that $(K;\psi)$ is isomorphic under some homeomorphism $\Theta : G \longrightarrow K$ to the rotation TDS $(G;\varphi_a)$. This isomorphism induces an algebra isomorphism $T_\Theta : \mathrm{C}(K) \longrightarrow \mathrm{C}(G)$. We can sum up the above in the following commutative diagram

$$\begin{tikzcd}
\mathscr{A} \arrow[r, "\Phi"] \arrow[d, "T"'] & \mathrm{C}(K) \arrow[r, "T_\Theta"] \arrow[d, "S"'] & \mathrm{C}(G) \arrow[d, "L_a"] \\
\mathscr{A} \arrow[r, "\Phi"'] & \mathrm{C}(K) \arrow[r, "T_\Theta"'] & \mathrm{C}(G)
\end{tikzcd}$$

where $(L_a)f(g) := f(ag)$ for $f \in \mathrm{C}(G)$ (the induced operator of the rotation φ_a). Now $\mathscr{A}$, $\mathrm{C}(K)$ and $\mathrm{C}(G)$ are dense subspaces in $\mathrm{L}^p(\Omega,\Sigma,\mu)$, $\mathrm{L}^p(K,\mathfrak{B}_0,\nu)$ and $\mathrm{L}^p(G)$, respectively (where we take the Haar measure on G). Since $(T'_\Theta)^{-1}\nu$ is a φ_a-invariant probability measure on G, it coincides with the Haar measure, the unique invariant probability measure ($(G;\varphi_a)$ is minimal). Therefore T_Θ is isometric for the corresponding L^p-norms. The induced operators T, L_a and T_Θ are all isometric lattice isomorphisms if considered on the corresponding L^p-spaces (see Lecture 7). Similarly, Φ is isometric by the construction of ν. The positivity of Φ can be seen as follows: if $0 \le f \in \mathscr{A}$, then there is $g \in \mathscr{A}$ with $g\overline{g} = f$ (see Exercise 13.4), so $\Phi(f) = \Phi(g)\overline{\Phi(g)} \ge 0$. By denseness we can extend Φ to a positive isometry (hence a lattice isomorphism) $\widetilde{\Phi} : \mathrm{L}^p(\Omega,\Sigma,\mu) \longrightarrow \mathrm{L}^p(K,\mathfrak{B}_0,\mu)$. We obtain that the diagram

$$\begin{tikzcd}
\mathrm{L}^p(\Omega,\Sigma,\mu) \arrow[r, "T_\Theta \circ \widetilde{\Phi}"] \arrow[d, "T_\varphi"'] & \mathrm{L}^p(G) \arrow[d, "T_\psi"] \\
\mathrm{L}^p(\Omega,\Sigma,\mu) \arrow[r, "T_\Theta \circ \widetilde{\Phi}"] & \mathrm{L}^p(G)
\end{tikzcd}$$

is commutative, hence by Proposition 7.13 and Remark 7.14 the claim is proved. □

Remark 13.10. Theorem 13.9 shows in particular that if the induced operator of an ergodic MDS has discrete spectrum on L^p for some $p \in [1,\infty)$, then this holds for all p in this range. (This is actually true also in non-ergodic case, but we shall not prove this here.)

As in the topological case we deduce from the above theorem that ergodic MDSs with discrete spectrum are completely determined by their point spectrum.

Corollary 13.11. *Two ergodic MDSs with induced operators having discrete spectrum in* L^p, $1 \le p < \infty$, *are isomorphic if and only if the induced operators have the same point spectrum.*

13.4 Examples

1. Rotations on Tori

The n-tori $G := \mathbb{T}^n$ are monothetic groups for $n \in \mathbb{N}$. Indeed, Kronecker's Theorem 11.24 describes precisely their generating elements. However, we can even form "infinite dimensional" tori as well: let I be a non-empty set, then the product $G := \mathbb{T}^I$ with the product topology and the pointwise operations is a compact group. The next theorem shows when this group is monothetic. The proof relies on the existence of sufficiently many rationally independent elements in $\mathbb{T}$.

Theorem 13.12. *Consider the compact group* $G := \mathbb{T}^I$, *I a non-empty set. Then G is monothetic if and only if* $\mathrm{card}(I) \le \mathrm{card}(\mathbb{T})$.

Proof. If G is monothetic, then by Proposition 13.2 we have that $\mathrm{card}(G^*) \le \mathrm{card}(\mathbb{T})$. Since the projections $\pi_\iota : \mathbb{T}^I \longrightarrow \mathbb{T}$, $\iota \in I$, are all different characters of G, we obtain $\mathrm{card}(I) \le \mathrm{card}(G^*) \le \mathrm{card}(\mathbb{T})$.

Now suppose that $\mathrm{card}(I) \le \mathrm{card}(\mathbb{T})$. Take a Hamel basis $\mathscr{B}$ of $\mathbb{R}$ over $\mathbb{Q}$ such that $1 \in \mathscr{B}$. Then $\mathrm{card}(\mathscr{B}) = \mathrm{card}(\mathbb{T})$, so there is an injective function $f : I \longrightarrow \mathscr{B} \setminus \{1\}$. Define $a := (a_\iota)_{\iota \in I} := (\mathrm{e}^{2\pi \mathrm{i} f(\iota)})_{\iota \in I} \in \mathbb{T}^I$. If $U \subset \mathbb{T}^I$ is a non-empty open rectangle, then by Kronecker's Theorem 11.24 we obtain $a^n \in U$ for some $n \in \mathbb{Z}$. □

The above product construction is very special and makes heavy use of the structure of $\mathbb{T}$. In fact, products of monothetic groups in general may not be monothetic (see Exercise 13.1). Notice also that these monothetic tori $\mathbb{T}^I$ are all connected (as the product of connected spaces).

2. Dyadic Integers

Consider the measure space $([0,1), \mathfrak{B}, \lambda)$, λ the Lebesgue measure on $[0,1)$, and define

$$\varphi(x) := x - \frac{2^k - 3}{2^k} \quad \text{if } x \in \Big[\frac{2^k - 2}{2^k}, \frac{2^k - 1}{2^k}\Big) \quad (k \in \mathbb{N}).$$

Then $([0,1), \mathfrak{B}, \lambda; \varphi)$ is an MDS (see Exercise 13.9). Let $T := T_\varphi$ be its induced operator which is unitary on $\mathrm{L}^2([0,1), \mathfrak{B}, \lambda)$. Define

$$g_m := \mathbf{1}_{[0,2^{-m})}$$

and
$$f_m := \sum_{k=0}^{2^m-1} \mathrm{e}^{-\frac{2\pi \mathrm{i} k}{2^m}} T^k g_m \qquad (m \in \mathbb{N}_0).$$

Notice that $T^{2^m} g_m = g_m$ and that $f_m \neq 0$ is an eigenvector of T:

$$T f_m = \sum_{k=0}^{2^m-1} \mathrm{e}^{-\frac{2\pi \mathrm{i} k}{2^m}} T^{k+1} g_m = \mathrm{e}^{\frac{2\pi \mathrm{i}}{2^m}} \sum_{k=1}^{2^m} \mathrm{e}^{-\frac{2\pi \mathrm{i} k}{2^m}} T^k g_m = \mathrm{e}^{\frac{2\pi \mathrm{i}}{2^m}} f_m,$$

that is $\mathrm{e}^{\frac{2\pi \mathrm{i}}{2^m}} \in \sigma_p(T)$. Since T is multiplicative, it also follows that $T f_m^n = \mathrm{e}^{\frac{2\pi \mathrm{i} n}{2^m}} f_m^n$. Moreover, as T is unitary, f_m^n and $f_{m'}^{n'}$ are orthogonal in L^2 whenever $m \neq m'$ or $0 \leq n \neq n' \leq 2^m - 1$. Some computation gives

$$g_m = \frac{1}{2^m} \sum_{k=0}^{2^m-1} f_m^k.$$

So, as $\operatorname{lin}\{g_m : m \in \mathbb{N}_0\}$ is dense in L^2, $\operatorname{lin}\{f_m^k : m \in \mathbb{N}_0,\, 0 \leq k < 2^m - 1\}$ is also dense, hence $\{f_m^k : m \in \mathbb{N}_0,\, 0 \leq k < 2^m - 1\}$ forms an orthonormal basis in L^2. This shows that T has discrete spectrum and that

$$\sigma_p(T) = \Big\{ \mathrm{e}^{\frac{2\pi \mathrm{i} k}{2^m}} : m \in \mathbb{N}_0,\, 0 \leq k \leq 2^m - 1 \Big\}.$$

By the usual argument of expanding a vector $f \in \operatorname{Fix}(T)$ in the above orthonormal basis, we can prove that f is constant, hence by Theorem 7.8 the MDS $([0,1), \mathfrak{B}, \lambda; \varphi)$ is ergodic (cf. proof of Proposition 11.23). Now the Halmos–von Neumann Theorem 13.9 tells that there is compact group such that the MDS is isomorphic to a group rotation. In our case, however, this group can be described concretely. Those who solved Exercise 11.8 might have noticed that $\sigma_p(T)$ is actually algebraically isomorphic to the dual group of the *dyadic integers* $\mathbb{A}$ (see Exercise 3.4). In fact, for $\mathbf{1} \in \mathbb{A}$, $\mathbf{1} = (1,0,0,\dots)$ one has

$$\sigma_p(T) = \Big\{ \mathrm{e}^{\frac{2\pi \mathrm{i} k}{2^m}} : m \in \mathbb{N}_0,\, 0 \leq k \leq 2^m - 1 \Big\} = \Big\{ \chi(\mathbf{1}) : \chi \in \mathbb{A}^* \Big\} = \sigma_p(T_{\varphi_{\mathbf{1}}}).$$

So by Corollary 13.11 $([0,1), \mathfrak{B}, \lambda; \varphi)$ is isomorphic to $(\mathbb{A}, \Sigma, \mu; \varphi_{\mathbf{1}})$, Σ the (product) Borel algebra, μ the Haar measure on $\mathbb{A}$ and $\varphi_{\mathbf{1}}$ the rotation by $\mathbf{1}$ (notice that the rotation on $\mathbb{A}$ is ergodic because it is minimal as $\operatorname{orb}_+(\mathbf{1})$ is dense in $\mathbb{A}$).

In contrast to the tori in Example 13.4.1, this group $\mathbb{A}$ is **totally disconnected**, i.e., every non-empty connected subset of $\mathbb{A}$ is a singleton (because it is a product of discrete spaces).

3. Bernoulli Shifts

The Bernoulli shift $B(p_0, p_1, \ldots, p_{k-1}) = (\mathscr{W}_k^+, \Sigma, \mu; \tau)$, discussed in Example 5.1.5, is ergodic, but — except the trivial case — does not have discrete spectrum. This can be seen as follows. Denote by $T := T_\tau$ the induced operator on L^2 and recall from Proposition 6.12 that

$$\lim_{n\to\infty} \int_{\mathscr{W}_k^+} T^n \mathbf{1}_A \cdot \mathbf{1}_B \,\mathrm{d}\mu = \lim_{n\to\infty} \mu(\tau^{*n} A \cap B) = \mu(A)\mu(B) = \int_{\mathscr{W}_k^+} \mathbf{1}_A \cdot \mathbf{1}_B \,\mathrm{d}\mu$$

holds for all $A, B \in \Sigma$. It is now a routine argument to show that this holds for any $f, g \in \mathrm{L}^2(\mathscr{W}_k^+, \Sigma, \mu)$ replacing $\mathbf{1}_A$ and $\mathbf{1}_B$. If $f \in \ker(\lambda \mathrm{I} - T)$ for some $\lambda \in \mathbb{T}$, then

$$\lim_{n\to\infty} \lambda^n \langle f, g\rangle = \langle f, g\rangle$$

follows for all $g \in \mathrm{L}^2$. But this can only happen if $f = 0$ or $\lambda = 1$. This shows that $\sigma_p(T) \cap \mathbb{T} = \{1\}$, and the ergodicity implies $\dim \mathrm{Fix}(T) = 1$, so T cannot have discrete spectrum (except $\dim \mathrm{L}^2 = 1$). The phenomenon observed here will be the topic of the next lecture.

Final remark

At the end of this lecture, let us return to *general* TDSs $(K; \varphi)$ and MDSs $(\Omega, \Sigma, \mu; \varphi)$. We take $X = \mathrm{C}(K)$ in the TDS case and $X = \mathrm{L}^p(\Omega, \Sigma, \mu)$ in the MDS case and consider the induced operator $T = T_\varphi$ thereon. The subspace

$$X_\mathrm{r} = \overline{\mathrm{lin}}\big\{ f \in X : Tf = \lambda f \text{ for some } \lambda \in \mathbb{T} \big\}$$

becomes a T-invariant, closed subalgebra/sublattice of X and the restriction T_r has discrete spectrum. If the original system is minimal/ergodic, so is the restricted one. By Theorem 13.7 and Theorem 13.9 the restricted system is isomorphic to a rotation on a monothetic compact group and is called the **Kronecker factor** of $(K; \varphi)$ and $(\Omega, \Sigma, \mu; \varphi)$, respectively.

Exercises

1. Show that a monothetic group is commutative. Describe all finite monothetic groups.

2. Prove that $(\mathbb{T}; \varphi_a)$ and $(\mathbb{T}; \varphi_{1/a})$, $a \in \mathbb{T}$, are isomorphic.

3. Show that the spectrum of a bijective isometry on a Banach space is contained in $\mathbb{T}$.

4. Let $\mathscr{A}$ be a conjugation invariant, closed subalgebra of $\mathrm{C}(K)$. Show that a positive element $f \in \mathscr{A}$ has (real) square root $g \in \mathscr{A}$, i.e., $g = \overline{g}$ and $g^2 = f$. (Hint: prove this first for $\|f\|_\infty \leq 1$ using the binomial series

$$(1+x)^{1/2} = \sum_{n=0}^{\infty} \binom{1/2}{n} x^n$$

which is absolutely convergent for $-1 \leq x \leq 1$.)

5. Let $T \in \mathscr{L}(X)$, X a Banach space, be with discrete spectrum, i.e., satisfying (13.1). Show that if T is power bounded, then it generates a relatively weakly compact semigroup. (Hint: Use Lemma 12.1. Show that if T is Cesàro-bounded, then T is mean ergodic with

$$\overline{\mathrm{ran}}(\mathrm{I}-T) = \overline{\mathrm{lin}}\{x \in X : \exists\, 1 \neq \lambda \in \mathbb{T} : Tx = \lambda x\})$$

6. Let T be a power-bounded operator on a Banach space X and let $A \subset \mathbb{T}$ such that the linear hull of

$$\bigcup_{\lambda \in A} \ker(\lambda \mathrm{I} - T)$$

is dense in X. Show that $\sigma_p(T) \subset A$.

7. Prove that if H is a closed subgroup of $\mathbb{T}$, then $H = \mathbb{T}$ or H is finite cyclic group.

8. Let $(\Omega, \Sigma, \mu; \varphi)$ be an ergodic MDS whose induced operator $T := T_\varphi$ has discrete spectrum on $H := \mathrm{L}^2(\Omega, \Sigma, \mu)$. Prove that if $\dim H = \infty$, then $\sigma(T) = \mathbb{T}$. (We note that the same is true even without the assumption of “discrete specrum”.)

$*$ **9.** Work out the details of Example 13.4.1.

Lecture 14
Mixing Dynamical Systems

> Verschiedene Weine zu mischen mag falsch sein, aber alte und neue Weisheit mischen sich ausgezeichnet.[1]
>
> Bertolt Brecht[2]

Whereas in the previous lecture we studied the reversible part in the Jacobs–de Leeuw–Glicksberg (JdLG) decomposition, we now turn to the stable part. In Theorem 12.14 we have already seen an abstract description but in this lecture we confine our considerations on MDSs and their induced operators on the L^p-spaces with $1 \leq p < \infty$. Note that in this case the semigroup generated by the induced operator is indeed relatively weakly compact, a necessary condition for the JdLG-decomposition.

It turns out that an MDS with trivial (i.e., one-dimensional) reversible part shows an interesting mixing behaviour. To clarify this we recall that by Proposition 9.1 an MDS $(\Omega,\Sigma,\mu;\varphi)$ is ergodic if and only if

$$\frac{1}{n}\sum_{j=0}^{n-1}\mu(\varphi^{*j}B\cap A)\to\mu(A)\mu(B) \qquad \text{for all } A,B\in\Sigma.$$

Halmos [Halmos (1956), p. 36] gives the following interpretation: We may view the transformation φ as a way of mixing the content of a glas Ω (of total volume $\mu(\Omega)=1$) filled with 50% wine and 50% water. If A is the region originally occupied by the wine and B is any other part of the glass, then, after n repetitions of the mixing procedure, the relative amount of wine in B is $\mu(\varphi^{*n}B\cap A)/\mu(B)$. The ergodicity of φ implies that *on the average* this relative amount comes eventually close to 50%. However, for a real mixing situation, one should be able to make the much stronger statement that, after sufficiently many repetitions of the mixing, every part B of the glass should contain approximately 50% of the wine.

---

[1] It may be mistaken to mix different wines, but old and new wisdom mix very well.

[2] Der Kaukasische Kreidekreis, Szene 1; Edition Suhrkamp; (The Caucasian Chalk Circle, Heinemann 1996; Translation: Stefan S. Brecht, James Stern)

This leads to the following mathematical concept.

14.1 Strong mixing

Consider an MDS $(\Omega,\Sigma,\mu;\varphi)$ and its induced operator $T := T_\varphi$ on $\mathrm{L}^p(\Omega,\Sigma,\mu)$. Here $p \in [1,\infty)$ is fixed, but we shall see that all results in this lecture are actually independent of our particular choice of p.

Definition 14.1. An MDS $(\Omega,\Sigma,\mu;\varphi)$ is called **strongly mixing** (or just **mixing**) if $T^n \to \mu \otimes \mathbf{1}$ in the weak operator topology, i.e.,

$$\langle T^n f, g\rangle \to \langle f,\mathbf{1}\rangle \cdot \langle \mathbf{1}, g\rangle = \int_\Omega f\,\mathrm{d}\mu \cdot \int_\Omega g\,\mathrm{d}\mu \tag{14.1}$$

for all $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$, $g \in \mathrm{L}^q(\Omega,\Sigma,\mu)$.

(Here and in the following we use of the notation $\mu \otimes \mathbf{1}$ for the operator $f \longmapsto \int_\Omega f\,\mathrm{d}\mu \cdot \mathbf{1}$. Moreover, q is the conjugate exponent, i.e., $1/p + 1/q = 1$.)

We note that a strongly mixing MDS is ergodic since the reversible part in the corresponding JdLG-decomposition is be trivial. Indeed, the condition (14.1) implies that

$$\langle T^n(f - \langle f,\mathbf{1}\rangle \cdot \mathbf{1}), g\rangle \to 0$$

for all $g \in \mathrm{L}^q$, i.e., $T^n(f - \langle f,\mathbf{1}\rangle \cdot \mathbf{1}) \to 0$ weakly. Hence $f - \langle f,\mathbf{1}\rangle \cdot \mathbf{1} \in X_{\mathrm{s}}$, which means that $X_{\mathrm{r}} = \mathrm{Fix}(T) = \mathrm{lin}\{\mathbf{1}\}$. (This also shows that a nontrivial ergodic group rotation is never strongly mixing.)

Here is a characterisation of strongly mixing MDSs that parallels the one for ergodic systems in Section 9.1.

Proposition 14.2. *For an MDS $(\Omega,\Sigma,\mu;\varphi)$ the following assertions are equivalent.*

(i) $(\Omega,\Sigma,\mu;\varphi)$ *is strongly mixing.*

(ii) $\mu(\varphi^{*n}A \cap B) \longrightarrow \mu(A)\mu(B)$ *for every* $A, B \in \Sigma$.

(iii) $\mu(\varphi^{*n}A \cap A) \longrightarrow \mu(A)^2$ *for every* $A \in \Sigma$.

Proof. (i) $\Leftrightarrow$ (ii) follows by specialising $f = \mathbf{1}_A$, $g = \mathbf{1}_B$ in the definition of strong mixing and by a standard density argument. The implication (ii) $\Rightarrow$ (iii) is trivial.

Assume now (iii). Then for every $A \in \Sigma$, $k, m \in \mathbb{N}_0$, $f := \mathbf{1}_{\varphi^{*k}A}$, $g := \mathbf{1}_{\varphi^{*m}A}$ and $n \geq k - m$ we have

$$\int_\Omega \Big(T^n \mathbf{1}_{\varphi^{*k}A}\Big) \cdot \mathbf{1}_{\varphi^{*m}A}\,\mathrm{d}\mu = \int_\Omega \Big(T^{n+k-m}\mathbf{1}_A\Big) \cdot \mathbf{1}_A\,\mathrm{d}\mu \to \mu(A)^2 = \langle f,\mathbf{1}\rangle \cdot \langle \mathbf{1}, g\rangle$$

as $n \to \infty$. Since T is power bounded, the convergence

$$\langle T^n f, g\rangle \to \langle f, \mathbf{1}\rangle \cdot \langle \mathbf{1}, g\rangle . \tag{14.2}$$

holds even for $f, g \in X_A := \overline{\mathrm{lin}}\{\mathbf{1}_{\varphi^{*n}A} : n \in \mathbb{N}_0\} \subset \mathrm{L}^2(\Omega, \Sigma, \mu)$. Now if $f \in X_A$ and $g \in X_A^\perp \subset \mathrm{L}^2(\Omega, \Sigma, \mu)$, then one has $\langle T^n f, g\rangle = 0 = \langle \mathbf{1}, g\rangle$. So we actually have (14.2) for all $g \in \mathrm{L}^2(\Omega, \Sigma, \mu)$. Specialising $g := \mathbf{1}_B$ the assertion in (ii) follows. □

Remark 14.3. Using straightforward density arguments it suffices in (ii) to take A and B from a dense subalgebra of Σ.

Example 14.4 (Bernoulli and Markov shifts). It has been shown already in Proposition 6.12 that Bernoulli shifts are strongly mixing. In Section 9.1 we have seen that irreducible Markov shifts are ergodic. However, not every such Markov shift needs to be strongly mixing. An easy example is provided by the Markov shift associated with the transition matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

on two points. The Markov measure μ on $\mathscr{W}_2^+$ puts equal weight of $1/2$ on the two points

$$x_1 := (0,1,0,1,\dots) \quad \text{and} \quad x_2 := (1,0,1,0,\dots)$$

and the system oscillates between those two under the shift τ, so this system is ergodic. On the other hand, if $A = \{x_1\}$, then $\mu(\varphi^{*n}A \cap A)$ is either 0 or $1/2$, depending whether n is odd or even. Hence this Markov shift is not strongly mixing.

Now let $(\mathscr{W}_k^+, \Sigma, \mu(S,p); \tau)$ be the Markov shift on $L := \{0, \dots, k-1\}$ associated with a row-stochastic matrix $S = (s_{ij})_{0 \le i,j < k}$ and a fixed probability vector $p = (p_j)_{0 \le j < k}$. Since ergodicity is necessary for strong mixing, we suppose in addition that S is irreducible (cf. Theorem 9.4). As in Section 9.1 we introduce $Q := \lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} P^j$. Then Q is strictly positive and each row of Q is equal to p.

Proposition. *In the situation described above, the following assertions are equivalent.*

(i) *The Markov shift $(\mathscr{W}_k^+, \Sigma, \mu(S,p); \tau)$ is strongly mixing.*

(ii) *$S^n \to Q$ as $n \to \infty$.*

(iii) *There is $n \ge 0$ such that S^n is strictly positive.*

(iv) $\sigma(S) \cap \mathbb{T} = \{1\}$.

Proof. Since Q is strictly positive, (ii) implies (iii). That (iv) implies (ii) can be seen by using the ergodic decomposition $\mathbb{C}^k = \mathrm{Fix}(S) \oplus \mathrm{ran}(\mathrm{I} - S)$ and noting that restricted to the second direct summand S has spectral radius strictly less than 1. For the implication (iii)⇒(iv) we may suppose without loss of generality that S itself is strictly positive. Indeed, suppose we know that $\sigma(A) \cap \mathbb{T} = \{1\}$ for every strictly positive row-stochastic matrix A, then taking $A = S^n$ we obtain

$$\sigma(S)^n \cap \mathbb{T} = \sigma(S^n) \cap \mathbb{T} = \{1\}$$

for all large $n \in \mathbb{N}$. This implies that eventually $\lambda^n = \lambda^{n+1} = 1$ for every peripheral eigenvalue λ of S, hence (iv) holds.

So suppose that S is strictly positive. Then there is $\varepsilon \in (0,1)$ and a row-stochastic matrix S' such that $S = (1-\varepsilon)S' + \varepsilon E$, where $E = (1/k)\mathbf{1}\cdot\mathbf{1}^\top$ is the matrix having each entry equal to $1/k$. Since row-stochastic matrices act on row vectors as contractions for the 1-norm, we have

$$\|xS - yS\|_1 = (1-\varepsilon)\left\|(x-y)S'\right\|_1 \le (1-\varepsilon)\|x-y\|_1$$

for all probability vectors $x,y \in \mathbb{C}^k$. Banach's fixed-point theorem yields that $\lim_n xS^n$ must exist for every probability vector x, hence for *every* $x \in \mathbb{C}^k$. Therefore 1 is the only peripheral eigenvalue of S.

Suppose now that (ii) holds and let

$$E = \{i_0\} \times \cdots \times \{i_l\} \times \prod L, \quad F = \{j_0\} \times \cdots \times \{j_r\} \times \prod L$$

for certain $i_0, \ldots, i_l, j_0, \ldots, j_r \in L$. As in Section 9.1,

$$\mu([\tau^n \in F] \cap E) = p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} [S^{n-l}]_{i_l j_0} s_{j_0 j_1} \ldots s_{j_{r-1} j_r}$$

for $n > l$. By (ii) this converges to

$$p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} p_{j_0} s_{j_0 j_1} \ldots s_{j_{r-1} j_r} = \mu(F)\mu(E).$$

Since the above sets E,F as above form a dense subalgebra of Σ, (i) follows. Conversely, if (i) holds, then taking $r = l = 0$ in the above we obtain

$$p_{i_0}[S^n]_{i_0 j_0} = \mu([\tau^n \in F] \cap E) \to \mu(E)\mu(F) = p_{i_0} p_{j_0}.$$

Since p is strictly positive, (ii) follows. □

If the matrix S satisfies the equivalent statements from above, it is called **primitive**. Another term often used is **aperiodic**; this is due to the fact that a row-stochastic matrix is primitive if and only if the greatest common divisor of the lengths of cycles in the associated transition graph over L is equal to 1 (see [Billingsley (1979), p.106] for more information).

Let us return to the more theoretical aspects. The following striking result was proved in [Blum and Hanson (1960)]. We state it without proof.

Theorem 14.5 (Blum–Hanson, 1960). *Let $(\Omega,\Sigma,\mu;\varphi)$ be an MDS and $T := T_\varphi$ its induced operator on $\mathrm{L}^p(\Omega,\Sigma,\mu)$, $1 \le p < \infty$. The MDS is strongly mixing if and only if for every subsequence $(n_k)_{k\in\mathbb{N}}$ of $\mathbb{N}$ and every $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$ the (strong) convergence*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} T^{n_j} f = \int_\Omega f \,\mathrm{d}\mu \cdot \mathbf{1}$$

holds for every $f \in \mathrm{L}^p(\Omega,\Sigma,\mu;\varphi)$.

Notice that the equivalence stated in this theorem is rather elementary if strong convergence is replaced by weak convergence of subsequences. This follows from the fact that a sequence $(a_n)_{n\in\mathbb{N}}$ in $\mathbb{C}$ is convergent if and only if each of its subsequences is Cesàro convergent (see Exercise 14.1).

Remarks 14.6. 1) If we specialise the above theorem to subsequences of the form $(kn)_{n\in\mathbb{N}}$ we see that T^k and T have the same mean ergodic projection $P = \mu \otimes \mathbf{1}$.

2) Theorem 14.5 is actually of pure operator-theoretic content. In [Jones and Kuftinec (1971)] and [Akcoglu and Sucheston (1972)] the following was proved.
Let T be a contraction on a Hilbert space H and denote by P the corresponding mean ergodic projection. Then the following assertions are equivalent.

(i) $T^n \to P$ *in the weak operator topology.*

(ii) $\displaystyle\frac{1}{n}\sum_{j=0}^{n-1} T^{n_j} \to P$ *in the strong operator topology for every subsequence* $(n_j)_{j\in\mathbb{N}}$ *of* $\mathbb{N}$.

Recently, Müller and Tomilov have shown that the statement fails in general for power bounded operators on Hilbert spaces [Müller and Tomilov (2007)].

The above characterisations of mixing involve knowledge about all powers T^n or, equivalently, φ^n, and usually are not easy to verify. That is why Katok and Hasselblatt [Katok and Hasselblatt (1995), p.748] write:

> "... It [strong mixing] is, however, one of those notions, that is easy and natural to define but very difficult to study...".

Recalling that ergodicity can be characterised by the purely spectral condition "1 is a simple eigenvalue of T" it is natural to ask for a *spectral characterisation* of strong mixing. Up to now, this is an open problem. However, a natural spectral condition leads to (and in fact is equivalent to) a weaker notion of mixing. This finally brings us back to the stable part of the JdLG-decomposition.

14.2 Weak mixing

We begin with a definition based on spectral properties only. Again we fix $p \in [1,\infty)$ but shall see that all results are actually independent of our choice.

Definition 14.7. An MDS $(\Omega,\Sigma,\mu;\varphi)$ is called **weakly mixing** if 1 is a simple and the unique peripheral eigenvalue of the induced operator $T = T_\varphi$, i.e., $\operatorname{Fix} T = \operatorname{lin}\{\mathbf{1}\}$ and $\sigma_p(T) \cap \mathbb{T} = \{1\}$ for the induced operator T on $\mathrm{L}^p(\Omega,\Sigma,\mu)$.

Note that this property is indeed independent of $p \in [1,\infty)$. This follows from the fact that under the condition $\mathrm{Fix}(T) = \mathrm{lin}\{\mathbf{1}\}$ every L^p-eigenfunction of T associated with some peripheral eigenvalue $\lambda \in \mathbb{T}$ must be contained in L^∞, see beginning of the proof of Theorem 13.9. Moreover, an MDS $(\Omega,\Sigma,\mu;\varphi)$ with induced operator $T := T_\varphi$ is weakly mixing if and only if the JdLG-decomposition associated with T yields $X_\mathrm{r} = \mathrm{lin}\{\mathbf{1}\}$, or, equivalently the projection Q onto X_r is $\mu \otimes \mathbf{1}$.

A strongly mixing MDS is weakly mixing and a weakly mixing MDS is ergodic. We shall see below that none of the converse implications holds in general.

For some characterisations of weak mixing we need the notion of the **product** of two MDSs, which is defined analogously to the TDS case (cf. Lecture 2). If $(\Omega,\Sigma,\mu;\varphi)$ and $(\Omega',\Sigma',\mu';\psi)$ are MDSs, their **product** is the MDS

$$(\Omega\times\Omega', \Sigma\otimes\Sigma', \mu\otimes\mu'; \varphi\times\psi)$$

with $\varphi\times\psi : (x,y) \longmapsto (\varphi(x),\psi(y))$. Here is now the main result characterising weak mixing in many different ways thereby explaining its connection to the concept of mixing.

Theorem 14.8. *For an MDS $(\Omega,\Sigma,\mu;\varphi)$ with induced operator $T = T_\varphi$ on $\mathrm{L}^p(\Omega,\Sigma,\mu)$, $1 \le p < \infty$, the following assertions are equivalent.*

(i) *The MDS $(\Omega,\Sigma,\mu;\varphi)$ is weakly mixing.*

(ii) *For every $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$, $T^{n_i} f \to \int_\Omega f\,\mathrm{d}\mu \cdot \mathbf{1}$ weakly for some subsequence $(n_i) \subset \mathbb{N}$.*

(iii) *For every $f \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$ there is a subsequence $(n_i) \subset \mathbb{N}$ such that $\langle T^{n_i} f, g\rangle \to \left(\int f\,\mathrm{d}\mu\right)\cdot\left(\int g\,\mathrm{d}\mu\right)$ holds for all $g \in L^\infty(\Omega,\Sigma,\mu)$.*

(iv) *For every $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$, $T^{n_i} f \to \int_\Omega f\,\mathrm{d}\mu \cdot \mathbf{1}$ weakly for some subsequence $(n_i) \subset \mathbb{N}$ having density 1.*

(v) *For every $A \in \Sigma$ there is a subsequence $(n_i) \subset \mathbb{N}$ with density 1 such that $\mu(\varphi^{*n_i} A \cap B) \to \mu(A)\cdot\mu(B)$ for all $B \in \Sigma$.*

(vi) *For each $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$*

$$\frac{1}{n}\sum_{k=0}^{n-1}\left|\left\langle T^k f, g\right\rangle - \langle f,\mathbf{1}\rangle\cdot\langle\mathbf{1},g\rangle\right| \underset{n\to\infty}{\longrightarrow} 0$$

uniformly for $g \in \mathrm{L}^q(\Omega,\Sigma,\mu)$ with $\|g\|_q \le 1$.

(vii) *$\varphi\times\varphi$ is weakly mixing.*

(viii) *$\varphi\times\varphi$ is ergodic.*

(ix) *$\varphi\times\psi$ is ergodic for every ergodic MDS $(\Omega',\Sigma',\mu';\psi)$.*

Proof. First, recall that (i) is equivalent to the fact that the projection onto X_r in the JdLG-decomposition has the form $Q = \mu\otimes\mathbf{1}$.

Suppose (i), and decompose $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$ as $f = Qf + (\mathrm{I}-Qf)$. Now we have that

$$T^n f = Qf + T^n(\mathrm{I}-Q)f = \int_\Omega f\,\mathrm{d}\mu \cdot \mathbf{1} + T^n(\mathrm{I}-Q)f$$

holds for all $n \in \mathbb{N}$. This implies

$$\langle T^n f, g\rangle - \langle f, \mathbf{1}\rangle \cdot \langle \mathbf{1}, g\rangle = \langle T^n(\mathrm{I}-Q)f, g\rangle \quad (g \in \mathrm{L}^q(\Omega,\Sigma,\mu)).$$

Since $(\mathrm{I}-Q)f \in X_\mathrm{s}$, the implication (i) $\Rightarrow$ (ii) as well as the equivalences (ii) $\Leftrightarrow$ (iv) $\Leftrightarrow$ (vi) follow from Theorem 12.14.

Suppose now that (iv) holds and let Q be the projection onto X_r. Since $(\mathrm{I}-Q)f \in X_\mathrm{s}$, by Theorem 12.14 and using Exercise 12.8.d, there is a subsequence $(n_j) \subset \mathbb{N}$ of density 1 such that $T^{n_j}(\mathrm{I}-Q)f \to 0$ and $T^{n_j} f \to (\mu \otimes \mathbf{1})f$ holds simultaneously in the weak topology as $j \to \infty$. This implies $Q = \mu \otimes \mathbf{1}$, hence (i) follows.

The implication (ii) $\Rightarrow$ (iii) follows by specialising to $f, g \in \mathrm{L}^\infty$. The converse is a straightforward density argument.

(iv) $\Rightarrow$ (v) Specialise to characteristic functions.

(v) $\Rightarrow$ (i) We have to show that $X = \mathrm{lin}\{\mathbf{1}\} \oplus X_\mathrm{s}$. It suffices to show that $f \in X_\mathrm{s}$ for every $f \in X = \mathrm{L}^p(\Omega,\Sigma,\mu)$ with $\langle f, \mathbf{1}\rangle = 0$. Suppose first that f is a step function, i.e., a linear combination of characteristic functions $\mathbf{1}_{A_i}$, $i = 1,\ldots,k$. By assumption (v) and by Exercise 12.8.d there is a subsequence $(n_j) \subset \mathbb{N}$ with density 1 such that $\mu(\varphi^{*n_j} A_i \cap B) \to \mu(A_i)\cdot\mu(B)$ for all $B \in \Sigma$ and $i = 1,\ldots,k$ as $j \to \infty$. This yields $\langle T^{n_j} f, g\rangle \to 0$ for every characteristic function g. Since step funtions are dense in $\mathrm{L}^q(\Omega,\Sigma,\mu)$, $T^{n_j} f \to 0$ weakly, and hence $f \in X_\mathrm{s}$. An arbitrary $f \in \mathrm{L}^p(\Omega,\Sigma,\mu)$ satisfying $\langle f, \mathbf{1}\rangle = 0$ can be approximated by step functions f_n which also satisfy $\langle f_n, \mathbf{1}\rangle = 0$. Since X_s is closed, we conclude that $f \in X_\mathrm{s}$.

(v) $\Rightarrow$ (vii) We will show the validity of (v) for the MDS $(\Omega\times\Omega, \Sigma\otimes\Sigma, \mu\otimes\mu; \varphi\times\varphi)$ to show that it is weakly mixing. For this purpose take $A_1, A_2, B_1, B_2 \in \Sigma$. Take $(n_j)_{j\in\mathbb{N}}$ such that $\mu(\varphi^{*n_j} A_1 \cap B_1) \to \mu(A_1)\mu(B_1)$ and $\mu(\varphi^{*n_j} A_2 \cap B_2) \to \mu(A_2)\mu(B_2)$ as $j \to \infty$. Then we have

$$(\mu\otimes\mu)[((\varphi\times\varphi)^{*n_j} A_1 \times A_2) \cap (B_1 \times B_2)] = \mu(\varphi^{*n_j} A_1 \cap B_1)\cdot\mu(\varphi^{*n_j} A_2 \cap B_2)$$

which converge as $j \to \infty$ to

$$\mu(A_1)\mu(B_1)\mu(A_2)\mu(B_2) = (\mu\otimes\mu)(A_1\times A_2)\cdot(\mu\otimes\mu)(B_1\times B_2).$$

This implies (vii) for the MDS $(\Omega\times\Omega, \Sigma\otimes\Sigma, \mu\otimes\mu; \varphi\times\varphi)$, which is, therefore, weakly mixing by the already proved equivalence.

(vii) $\Rightarrow$ (viii) is clear, since every weakly mixing MDS is ergodic.

To prove the implication (v) $\Rightarrow$ (ix) we set $a_n := \mu(\varphi^{*n} A_1 \cap B_1)$ and $b_n := \mu'(\psi^{*n} A_2 \cap B_2)$ for given $A_1, B_1 \in \Sigma$ and $A_2, B_2 \in \Sigma'$. Then by the assumption (vii) there is (n_j) of density 1 with $a_{n_j} \to \mu(A_1)\mu(B_1)$. On the other, as ψ is ergodic, the sequence (b_n) has Cesàro limit $\mu'(A_2)\cdot\mu'(B_2)$ (see Proposition 9.1). By using Exercise 14.2 we obtain that

$$\begin{aligned}
&\frac{1}{n}\sum_{j=0}^{n-1}(\mu\otimes\mu')\big((\varphi\times\psi)^{*j}(A_1\times A_2)\cap(B_1\times B_2)\big)\\
&\quad=\frac{1}{n}\sum_{j=0}^{n-1}\mu(\varphi^{*j}A_1\cap B_1)\mu'(\psi^{*j}A_2\cap B_2)\\
&\quad=\frac{1}{n}\sum_{j=0}^{n-1}a_jb_j\to(\mu\otimes\mu')\big((A_1\times A_2)\cap(B_1\times B_2)\big)\quad(n\to\infty).
\end{aligned}$$

Proposition 9.1 yields that the product MDS is ergodic.

(ix) $\Rightarrow$ (viii) Observe that if the product MDS is ergodic, so is $(\Omega,\Sigma,\mu;\varphi)$. This follows from the fact that if $A\in\Sigma$ is φ-invariant, then the product $A\times A\in\Sigma\otimes\Sigma$ is $\varphi\times\varphi$-invariant. Now (viii) follows by taking $\psi=\varphi$.

(viii) $\Rightarrow$ (i) Assume that $Tf=\lambda f$, $\lambda\in\mathbb{T}$ and $0\neq f\in\mathrm{L}^1(\Omega,\Sigma,\mu)$. Then we have $T\bar{f}=\bar{\lambda}\bar{f}$ and hence for the function $f\otimes\bar{f}:(x,y)\longmapsto f(x)\cdot\bar{f}(y)$, $(x,y)\in\Omega\times\Omega$, we obtain $T_{\varphi\times\varphi}(f\otimes\bar{f})=\lambda f\otimes\bar{\lambda}\bar{f}=|\lambda|^2(f\otimes\bar{f})=f\otimes\bar{f}$. But 1 is a simple eigenvalue of $T_{\varphi\times\varphi}$ with eigenvector $\mathbf{1}\otimes\mathbf{1}$. Therefore we conclude $f=c\mathbf{1}$ and $\lambda=1$, i.e. φ is weakly mixing. □

Remark 14.9. If $\mathrm{L}^p(\Omega,\Sigma,\mu)$ is separable, one can replace the above conditions (ii) and (iv) by the following equivalent ones. For the proof one can use Exercise 12.6.

(ii') $T^{n_i}\to\mu\otimes\mathbf{1}$ *in the weak operator topology for some subsequence* $(n_i)\subset\mathbb{N}$.

(iv') $T^{n_i}\to\mu\otimes\mathbf{1}$ *in the weak operator topology for some subsequence* $(n_i)\subset\mathbb{N}$ *having density* 1.

Remark 14.10. Let $(\Omega,\Sigma,\mu;\varphi)$ be a weakly mixing MDS, let $k\in\mathbb{N}$ and consider its k^{th}-iterate $(\Omega,\Sigma,\mu;\varphi^k)$. The operator induced by this latter MDS on $\mathrm{L}^p(\Omega,\Sigma,\mu)$ is $S:=T^k$, when $T:=T_\varphi$ is induced by φ. By the spectral mapping theorem we have $\sigma_p(T)^k=\sigma_p(S)=\{1\}$. Moreover, since $\operatorname{Fix}S=\operatorname{lin}\{x:Sx=\lambda x,\ \lambda^k=1\}=1$ is one-dimensional, we see that φ^k is weakly mixing, and in particular ergodic.

An MDS $(\Omega,\Sigma,\mu;\varphi)$ with the property that all its k^{th} iterates are ergodic is called **strongly ergodic**. By Theorem 14.8, every weakly mixing system is strongly ergodic (cf. also Exercise 14.3).

Again, a nontrivial ergodic group rotation yields an example of a not weakly mixing systems. It had been, however, an open problem for quite some time to find MDSs which are weakly but not strongly mixing. Eventually, Chacon in [Chacon (1967)] constructed examples via his so-called *stacking method* (see also [Chacon (1969)] and Petersen [Petersen (1989), Section 4.5]).

Curiously enough, the *existence* of such systems had been established much earlier, based on Baire category arguments. First, Halmos [Halmos (1944)] showed in 1944 that the set of weakly mixing transformations is residual (i.e., its complement is of first category) in the complete metric space G of all measure preserving transformations with respect to the metric coming from the strong operator topology. Four years later, Rohlin [Rohlin (1948)] proved that the set of all strongly mixing transformations is of first category in G. Thus, in this sense, the generic MDS is weakly but not strongly mixing.

14.3 Weak mixing of all orders

For our mixing concepts we considered pairs of sets $A, B \in \Sigma$ or functions f, g. In this section we show that, intuitively speaking, weakly mixing systems even mix every finite number of sets. The results again do not depend on the parameter $p \in [1, \infty)$ and we take $p = 2$. In this context the following lemma plays an essential role.

Lemma 14.11 (van der Corput). *Let H be a Hilbert space and $\{u_n\}_{n=1}^{\infty} \subset H$ with $\|u_n\| \leq 1$. For $j \in \mathbb{N}$ define*

$$\gamma_j := \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=0}^{N-1} \langle u_n, u_{n+j} \rangle \right|.$$

Then $\lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} \gamma_j = 0$ *implies* $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} u_n = 0$.

For its proof we refer to the supplement below.

Lemma 14.12. *Let $(\Omega, \Sigma, \mu; \varphi)$ be an MDS. Then for every $f \in X_{\mathrm{s}}$ one has*

$$\int_{\Omega} f \,\mathrm{d}\mu = 0,$$

where X_{s} is the stable part in the JdLG-decomposition of the induced operator $T = T_{\varphi}$.

Proof. Since $f \in X_{\mathrm{s}}$, there exists $(n_j)_{j=1}^{\infty}$ such that $\lim_{j\to\infty} T^{n_j} f = 0$ weakly. By the φ-invariance of μ we obtain

$$\int_{\Omega} f \,\mathrm{d}\mu = \int_{\Omega} T f \,\mathrm{d}\mu = \int_{\Omega} T^{n_j} f \,\mathrm{d}\mu \to 0. \qquad \square$$

We now state the main result of this section. If forms part of the recurrence theory developed by Furstenberg [Furstenberg (1981)]. Slightly more on this topic will appear in our final lecture.

Theorem 14.13. *Let $(\Omega, \Sigma, \mu; \varphi)$ be a weakly mixing MDS and let $T := T_{\varphi}$ be the induced operator on $\mathrm{L}^2(\Omega, \Sigma, \mu)$. Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (T^n f_1)(T^{2n} f_2) \cdots (T^{(k-1)n} f_{k-1}) = \int_{\Omega} f_1 \,\mathrm{d}\mu \cdots \int_{\Omega} f_{k-1} \,\mathrm{d}\mu \cdot \mathbf{1} \tag{14.3}$$

holds for every $k \geq 2$ and every $f_1, \dots, f_{k-1} \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)$.

Note that since we take the functions $f_1, \dots, f_{k-1}$ from $\mathrm{L}^\infty(\Omega, \Sigma, \mu)$, the above products remain in $\mathrm{L}^2(\Omega, \Sigma, \mu)$.

Proof. We prove the theorem by induction on $k \geq 2$. For $k = 2$, the assertion is just the mean ergodic theorem for ergodic systems, see Proposition 9.1. By induction, suppose that the assertion holds for some $k \geq 2$ and take $f_1, \dots f_k \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)$. By the JdLG–decomposition we may suppose that $f_k \in X_\mathrm{s}$ or $f_k \in X_\mathrm{r} = \mathrm{lin}\{\mathbf{1}\}$. In the second case, f_k is a constant function hence fixed under T, so the assertion reduces to the induction hypothesis. Hence we may suppose that $f_k \in X_\mathrm{s}$.

By Lemma 14.12 we have to show that the limit in (14.3) (with $k-1$ replaced by k) is zero. To achieve this we use the van der Corput Lemma 14.11. Define $u_n := T^n f_1 \cdot T^{2n} f_2 \cdots T^{kn} f_k$. Then the invariance of μ and the multiplicativity of T imply

$$\begin{aligned}
\langle u_n, u_{n+j} \rangle &= \int_\Omega (T^n f_1 \cdot T^{2n} f_2 \cdots T^{kn} f_k) \cdot (T^{n+j} \overline{f_1} \cdot T^{2n+2j} \overline{f_2} \cdots T^{kn+kj} \overline{f_k}) \,\mathrm{d}\mu \\
&= \int_\Omega (f_1 \cdot T^n f_2 \cdots T^{(k-1)n} f_k) \cdot (T^j \overline{f_1} \cdot T^{n+2j} \overline{f_2} \cdots T^{(k-1)n+kj} \overline{f_k}) \,\mathrm{d}\mu \\
&= \int_\Omega (f_1 T^j \overline{f_1}) \cdot T^n (f_2 T^{2j} \overline{f_2}) \cdots T^{(k-1)n} (f_k T^{kj} \overline{f_k}) \,\mathrm{d}\mu.
\end{aligned}$$

Thus, by the induction hypothesis, the Cesàro means of $\langle u_n, u_{n+j} \rangle$ converge to

$$\int_\Omega f_1 T^j \overline{f_1} \,\mathrm{d}\mu \cdot \int_\Omega f_2 T^{2j} \overline{f_2} \,\mathrm{d}\mu \cdots \int_\Omega f_k T^{kj} \overline{f_k} \,\mathrm{d}\mu.$$

So we obtain

$$\gamma_j := \limsup_{N \to \infty} \Big| \frac{1}{N} \sum_{n=0}^{N-1} \langle u_n, u_{n+j} \rangle \Big| = \Big| \langle f_1, T^j \overline{f_1} \rangle \Big| \cdots \Big| \langle f_k, T^{kj} \overline{f_k} \rangle \Big|.$$

Since $f_k \in X_\mathrm{s}$, this implies

$$\begin{aligned}
\limsup_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \gamma_j &\leq \|f_1\|^2 \dots \|f_{k-1}\|^2 \limsup_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \Big| \langle f_k, T^{kj} \overline{f_k} \rangle \Big| \\
&\leq \|f_1\|^2 \dots \|f_{k-1}\|^2 \, k \lim_{n \to \infty} \frac{1}{kn} \sum_{j=0}^{k(n-1)} \Big| \langle f_k, T^j \overline{f_k} \rangle \Big| = 0.
\end{aligned}$$

An application of the van der Corput lemma concludes the argument. □

Remark 14.14. The same proof works for commuting weakly mixing transformations $\varphi_1, \dots \varphi_{k-1}$ on (Ω, Σ, μ) with induced operators $T_1, \dots, T_{k-1}$. More precisely, for every $f_1, \dots, f_{k-1} \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)$ one has

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} T_1^n f_1 \cdot T_2^n f_2 \cdot \ldots \cdot T_{k-1}^n f_{k-1} = \int_\Omega f_1\,\mathrm{d}\mu \cdot \ldots \cdot \int_\Omega f_{k-1}\,\mathrm{d}\mu \cdot \mathbf{1}.$$

(The proof is left as Exercise 14.5.) In particular, by Remark 14.10 one can take arbitrary powers of T (different from $T^0 = \mathrm{I}$) instead of $T^n, T^{2n}, \ldots, T^{(k-1)n}$ in the above theorem.

Taking characteristic functions in the above theorem we obtain the following result explaining the title of this section.

Corollary 14.15. *Every weakly mixing MDS* $(\Omega,\Sigma,\mu;\varphi)$ *is* **weakly mixing of all orders***, i.e.,*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} \mu(A_0 \cap \varphi^{*n} A_1 \cap \ldots \cap \varphi^{*(k-1)n} A_{k-1}) = \mu(A_0)\mu(A_1)\cdots\mu(A_{k-1})$$

holds for every $k \in \mathbb{N}$ *and every* $A_0, \ldots, A_{k-1} \in \Sigma$.

It is a long-standing open problem whether the analogous statement is true for strongly mixing systems, even for $k = 3$.

Concluding remarks

Let us summarise the important relations between these many notions in the following diagram:

$$\begin{tikzcd}
\text{strongly mixing} \arrow[r] & \text{weakly mixing} \arrow[r] \arrow[dl, leftrightarrow] \arrow[d, leftrightarrow] & \text{strongly ergodic} \arrow[r] & \text{ergodic} \\
\begin{array}{c}\text{weak mixing}\\ \text{of all orders}\end{array} & \varphi\times\varphi \text{ weakly mixing} \arrow[r, leftrightarrow] & \varphi\times\varphi \text{ ergodic} \arrow[ur] &
\end{tikzcd}$$

Supplement: Proof of the van der Corput lemma

We now present the proof of the van der Corput lemma. It is elementary, but technical.

Proof of Lemma 14.11. We shall use the notation $o(1)$ to denote terms converging to 0 as $N \to \infty$. First observe that for a fixed $J \in \mathbb{N}$ we have

$$\frac{1}{N}\sum_{n=0}^{N-1} u_n = \frac{1}{N}\sum_{n=0}^{N-1} u_{n+j} + o(1)$$

for every $0 \le j \le J-1$ and hence

$$\frac{1}{N}\sum_{n=0}^{N-1} u_n = \frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{J}\sum_{j=0}^{J-1} u_{n+j} + o(1).$$

By the Cauchy-Schwarz inequality

$$\Big\|\frac{1}{N}\sum_{n=0}^{N-1} u_n\Big\| \le \frac{1}{N}\sum_{n=0}^{N-1}\Big\|\frac{1}{J}\sum_{j=0}^{J-1} u_{n+j}\Big\| + o(1) \le \Big(\frac{1}{N}\sum_{n=0}^{N-1}\Big\|\frac{1}{J}\sum_{j=0}^{J-1} u_{n+j}\Big\|^2\Big)^{1/2} + o(1)$$
$$\le \Big(\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{J^2}\sum_{j_1,j_2=0}^{J-1}\big|\langle u_{n+j_1}, u_{n+j_2}\rangle\big|\Big)^{1/2} + o(1).$$

As $N \to \infty$ we thus obtain

$$\limsup_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1} u_n\Big\|^2 \le \frac{1}{J^2}\sum_{j_1,j_2=0}^{J-1}\gamma_{|j_1-j_2|}. \tag{14.4}$$

We now claim that

$$\frac{1}{J^2}\sum_{j_1,j_2=0}^{J-1}\gamma_{|j_1-j_2|} \le 2\frac{1}{J}\sum_{j=0}^{J-1}\gamma_j. \tag{14.5}$$

To show this we observe that

$$\sum_{j_1,j_2=0}^{J-1}\big(\gamma_{|j_1-j_2|} + \gamma_{J-1-|j_1-j_2|}\big) = J(\gamma_0+\gamma_{J-1}) + 2(J-1)(\gamma_1+\gamma_{J-2}) +$$
$$\ldots + 2\cdot 2(\gamma_{J-2}+\gamma_1) + 2(\gamma_{J-1}+\gamma_0) \le 2(J+1)\sum_{j=0}^{J-1}\gamma_j.$$

Therefore we have

$$\frac{1}{J^2}\sum_{j_1,j_2=0}^{J-1}\gamma_{|j_1-j_2|} \le \frac{(J+1)}{J^2}\sum_{j=0}^{J-1}\gamma_j \le \frac{2}{J}\sum_{j=0}^{J-1}\gamma_j,$$

so (14.5) is proved and together with (14.4) implies the assertion. □

Remark 14.16. Using analogous arguments one can prove the stronger assertion

$$\limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=0}^{N-1} u_n\right\|^2 \le \limsup_{J\to\infty}\frac{1}{J}\sum_{n=0}^{J-1}\gamma_j.$$

See [Tao (2008b)] for more on the van der Corput lemma.

Exercises

1. Prove that a sequence $(a_n)_{n\in\mathbb{N}}$ in $\mathbb{C}$ converges if and only each of its subsequences is Cesàro convergent.

2. Let $(a_n)_{n=1}^{\infty} \subset \mathbb{C}$ converge to a along a subsequence $(n_j)_{j=1}^{\infty} \subset \mathbb{N}$ with density 1. Assume further that $(b_n)_{n=1}^{\infty} \subset \mathbb{C}$ satisfies $\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} b_j = b$. Then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} a_j b_j = ab.$$

Note that the Cesàro limit is not multiplicative in general!
(Hint: Use the Koopman–von Neumann lemma and a 3ε-argument.)

3. Give an example of an ergodic MDS $(\Omega,\Sigma,\mu;\varphi)$ such that the MDS $(\Omega\times\Omega,\Sigma\otimes\Sigma,\mu\otimes\mu;\varphi\times\varphi)$ is not ergodic.

4. Interpret weak mixing and mixing of all orders (i.e., Theorem 14.8 and Corollary 14.15) in the case of mixing wine and water.

5. Prove Remark 14.14.

6.Let φ be weakly mixing. Then φ^k is weakly mixing for every $k \in \mathbb{N}$.
(Hint: To show $X_{\mathrm{s}}(T) = X_{\mathrm{s}}(T^k)$ for the induced operator $T = T_{\varphi}$, use arguments as at the end of the proof of Theorem 14.13.)

7. Let S be an irreducible row-stochastic $k \times k$-matrix with fixed probability vector p. Show that if the associated Markov shift $(\mathscr{W}_k^+,\Sigma,\mu(S,p);\tau)$ is weakly mixing, then it is actually strongly mixing.

Lecture 15
A Glimpse on Arithmetic Progressions

> ...then shalt thou count to three, no more, no less. Three shall be the number thou shalt count, and the number of the counting shall be three. Four shalt thou not count, neither count thou two, excepting that thou then proceed to three. Five is right out. Once the number three, being the third number, be reached...
>
> Book of Armaments 2:9–21
> Monty Python[1]

In this final lecture we shall discuss some of the applications of ergodic theory to (combinatorial) number theory. An **arithmetic progression** of length $k \in \mathbb{N}$ is a set A of natural numbers of the form

$$\{a, a+d, a+2d, \dots, a+(k-1)d\} = \{a+jk : 0 \le j < k\}$$

for some $a, d \in \mathbb{N}$. The number a is called the **starting point** and d is called the **distance** of this arithmetic progression. An arithmetic progression of length k is also called a k-term arithmetic progression.

One aspect of combinatorial number theory deals with the question how "big" a subset $A \subset \mathbb{N}$ must be in order to ensure that it contains infinitely many arithmetic progressions of a certain length, or even of arbitrary length. The following result of van der Waerden [van der Waerden (1927)] is the classical example for a statement of this kind.

Theorem 15.1 (van der Waerden). *If $\mathbb{N} = A_1 \cup \ldots \cup A_m$, then there is $j \in \{1, \dots, m\}$ such that*

$$\forall k \; \exists a \; \exists n : \quad a, a+n, a+2n, \dots, a+(k-1)n \in A_j.$$

In other words, if you colour the natural numbers with finitely many colours, then there are arbitrarily long *monochromatic* arithmetic progressions, i.e., progressions

[1] Monty Python and the Holy Grail, Movie, 1975.

all of whose members carry the same colour. (However, in general one cannot find a monochomatic progression of *infinite* length.)

Many years later, a major extension was proved by Szemerédi in [Szemerédi (1975)] using the concept of the *upper density*

$$\overline{d}(A) = \limsup_{n\to\infty} \frac{\operatorname{card}(A \cap \{1,\ldots,n\})}{n}$$

of a subset $A \subset \mathbb{N}$ (cf. Section 12.2).

Theorem 15.2 (Szemerédi). *Any subset $A \subset \mathbb{N}$ with $\overline{d}(A) > 0$ contains arbitrarily long arithmetic progressions.*

In the 1970's it was discovered that the results of van der Waerden and Szemerédi (and many others) could be obtained also by "dynamical" methods, i.e., by transferring them to statements about dynamical systems. Major contributions came from Furstenberg who was able to recover van der Waerden's theorem by methods from topological dynamics (see [Furstenberg (1981)], [Pollicott and Yuri (1998)]). Later on he used ergodic theory to obtain and extend the result of Szemerédi.

His ideas building heavily on the structure theory of ergodic MDSs were further developed by many authors and finally led to one of the most striking results in this area so far.

Theorem 15.3 (Green–Tao). *The set of prime numbers contains arbitrarily long arithmetic progressions.*

(See [Green and Tao (2008)] and in particular Tao's ICM lecture [Tao (2007)] for the connection with ergodic theory.)

The discovery of such a beautiful structure within the chaos of prime numbers was preceded and accompanied by various numerical experiments. Starting from the arithmetic progression $7, 37, 67, 97, 127, 157$ (of length 6 and distance 30) the actual world record (since May 17, 2008 by R. Chermoni and J. Wroblewskiis a progression of length 25 starting at

$$6\ 171\ 054\ 912\ 832\ 631$$

with distance $366\ 384 \times 223\ 092\ 870$.

However striking, the Green–Tao theorem still falls short of the following audacious conjecture formulated by Erdős and Turán in 1936.

Conjecture 15.4 (Erdős–Turán). Let $A \subset \mathbb{N}$ be such that $\sum_{a\in A}(1/a) = +\infty$. Then A contains arbitrarily long arithmetic progressions.

Unfortunately, the Green–Tao theorem is far beyond the scope of these lectures. Instead, we shall describe the major link between "density results" (like Szemerédi's) and ergodic theory and apply it to obtain weaker, but nevertheless stunning results (the theorems of Roth and Furstenberg–Sárközy, see below). Again, our major operator-theoretic tool is the JdLG-decomposition.

15.1 The Furstenberg Correspondence Principle

Let us fix a subset $A \subset \mathbb{N}$ with $\overline{d}(A) > 0$. For given k consider the following statement.

$$\text{There exist } a, d \in \mathbb{N} \text{ such that } a, a+d, a+2d, \ldots a+(k-1)d \in A. \qquad (\mathrm{AP}_k)$$

Our goal is to construct an associated dynamical system that allows us to reformulate (AP_k) in ergodic theoretic terms.

Consider the compact metric space $\mathscr{W}_2^+ = \{0,1\}^{\mathbb{N}_0}$. A subset $B \subset \mathbb{N}$ can be identified with a point in $\mathscr{W}_2^+$ via its characteristic function:

$$B \subset \mathbb{N} \quad \longleftrightarrow \quad \mathbf{1}_B \in \mathscr{W}_2^+.$$

We define

$$K := \overline{\{\varphi^n \mathbf{1}_A \,:\, n \geq 1\}} \subset \mathscr{W}_2^+$$

where $\varphi := \tau$ is the left shift on $\mathscr{W}_2^+$. Then K is a closed τ-invariant subset of $\mathscr{W}_2^+$, i.e., $(K;\varphi)$ is a (forward transitive) TDS. The set

$$M := \left\{ (x_n)_{n=0}^{\infty} \in K : \ x_0 = 1 \right\}$$

is open and closed in K, and hence is every set $\varphi^{-j}(M)$, $j \in \mathbb{N}_0$. Note that we have

$$n \in A \quad \text{ if and only if } \quad \varphi^n(\mathbf{1}_A) \in M. \tag{15.1}$$

We now translate (AP_k) into a property of this dynamical system. By (15.1), we obtain for $a \in \mathbb{N}$ and $n \in \mathbb{N}_0$ the following equivalences.

$$\begin{aligned} &a, a+n, \ldots, a+(k-1)n \in A \\ &\qquad \iff \varphi^a \mathbf{1}_A \in M,\ \varphi^n \varphi^a \mathbf{1}_A \in M, \ldots,\ \varphi^{(k-1)n} \varphi^a \mathbf{1}_A \in M \\ &\qquad \iff \varphi^a \mathbf{1}_A \in M \cap \varphi^{-n}(M) \cap \ldots \cap \varphi^{-(k-1)n}(M). \end{aligned}$$

Since each $\varphi^{-j}(M)$ is open and $\{\varphi^a \mathbf{1}_A \,:\, a \in \mathbb{N}\}$ is dense in K, we have that (AP_k) is equivalent to

$$(15.2) \qquad \exists n \in \mathbb{N} : \ M \cap \varphi^{-n}(M) \cap \ldots \cap \varphi^{-(k-1)n}(M) \neq \emptyset.$$

The strategy to prove (15.2) is now to turn the TDS into an MDS by choosing an invariant measure μ in such a way that for some $n \in \mathbb{N}$ the set

$$M \cap \varphi^{-n}(M) \cap \ldots \cap \varphi^{-(k-1)n}(M)$$

has positive measure. Now, since $\overline{d}(A) > 0$, there is a subsequence $(n_j)_{j \in \mathbb{N}} \subset \mathbb{N}$ such that

$$\lim_{j\to\infty}\frac{1}{n_j}\sum_{k=1}^{n_j}\delta_{\{k\}}(A)=\overline{d}(A)>0.$$

With this subsequence we define a sequence of probability measures $(\mu_j)_{j\in\mathbb{N}}$ on K by

$$\mu_j(B):=\frac{1}{n_j}\sum_{k=1}^{n_j}\delta_{\{\varphi^k\mathbf{1}_A\}}(B),\qquad (B\in\mathfrak{B}(K),j\in\mathbb{N}).$$

By using (15.1) we have

$$\mu_j(M)=\frac{1}{n_j}\sum_{k=1}^{n_j}\delta_{\{k\}}(A)\to\overline{d}(A)\quad \text{as } j\to\infty. \tag{15.3}$$

The metrisability of the compact set $\mathrm{M}_1(K)$ for the weak*-topology imply that there exists a subsequence (which we again denote by $(\mu_j)_{j\in\mathbb{N}}$) weakly* converging to some probability measure μ on K. Since M is open and closed, the characteristic function $\mathbf{1}_M : K\longrightarrow\mathbb{R}$ is continuous, hence (15.3) implies that $\mu(M)=\overline{d}(A)$.

To show that μ is φ-invariant, take $f\in\mathrm{C}(K)$ and note that

$$\begin{aligned}\int_K(f\circ\varphi)\,\mathrm{d}\mu_j-\int_K f\,\mathrm{d}\mu_j&=\frac{1}{n_j}\sum_{k=1}^{n_j}f(\varphi^{k+1}(\mathbf{1}_A))-f(\varphi^k(\mathbf{1}_A))\\&=\frac{1}{n_j}\big(f(\varphi^{n_j+1}(\mathbf{1}_A))-f(\varphi(\mathbf{1}_A))\big)\end{aligned}$$

for every $j\in\mathbb{N}$. The first term in this chain of equations converges to

$$\int_K(f\circ\varphi)\,\mathrm{d}\mu-\int_K f\,\mathrm{d}\mu$$

as $j\to\infty$ by construction of μ; the last one converges to 0 since $n_j\to\infty$. Therefore μ is φ-invariant (cf. Lemma 8.7).

In this way we obtain an MDS $(K,\mathfrak{B},\mu;\varphi)$ such that $\mu(M)=\overline{d}(A)>0$. Furthermore, we may suppose without loss of generality that this MDS is *ergodic*. Indeed, we know from Lecture 9 that the set $\mathrm{M}^1_\varphi(K)$ of all φ-invariant probability measures on K is the weak*-closure of

$$\operatorname{conv}\{\mu : \mu\ \varphi\text{-invariant ergodic probability measure on } K\}.$$

Since $\mathbf{1}_M\in\mathrm{C}(K)$ and since by the above there is at least one $\mu\in\mathrm{M}^1_\varphi(K)$ with $\int_K\mathbf{1}_M\,\mathrm{d}\mu=\mu(M)>0$, there must also be an ergodic measure with this property.

Consider now the induced linear operator $T:=T_\varphi$ on $\mathrm{L}^2(K,\mathfrak{B},\mu)$. The existence of a k-term arithmetic progression in A, i.e., (AP$_k$), is certainly ensured by

$$\exists n\in\mathbb{N}:\ \ \mathbf{1}_M\cdot T^n\mathbf{1}_M\cdots T^{(k-1)n}\mathbf{1}_M\neq 0$$

or the even stronger statement

$$\limsup_{n\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_K \mathbf{1}_M \cdot (T^n \mathbf{1}_M) \cdots (T^{(k-1)n} \mathbf{1}_M) \,\mathrm{d}\mu \;>\; 0. \tag{15.4}$$

Hence we have established the following result from [Furstenberg (1977)].

Theorem 15.5 (Furstenberg Correspondence Principle). *If for every ergodic MDS $(\Omega, \Sigma, \mu; \varphi)$ the induced operator $T := T_\varphi$ on $\mathrm{L}^2(\Omega, \Sigma, \mu)$ satisfies the condition*

$$\limsup_{n\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_\Omega f \cdot (T^n f) \cdots (T^{(k-1)n} f) \,\mathrm{d}\mu > 0 \tag{15.5}$$

for all $0 < f \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)$, then (AP_k) holds for every $A \subset \mathbb{N}$ with $\overline{d}(A) > 0$.

(Here and in the following we write $f > 0$ when we mean $f \geq 0$ and $f \neq 0$.) In [Furstenberg (1977)] it was shown that (15.5) is indeed true for every ergodic MDS and every $k \in \mathbb{N}$. This completed the ergodic-theoretic proof of Szemerédi's theorem. As a matter of fact, Furstenberg proved the truth of the apparently stronger version of (15.5) when $\liminf_{n\to\infty}$ replaces $\limsup_{n\to\infty}$. As we shall see in a moment, this makes no difference because the *limit* actually *exists*.

15.2 A Multiple Ergodic Theorem

Furstenberg, in his correspondence principle, worked with expressions of the form

$$\frac{1}{N} \sum_{n=0}^{N-1} \int_\Omega f \cdot (T^n f) \cdots (T^{(k-1)n} f) \,\mathrm{d}\mu = \int_\Omega f \cdot \frac{1}{N} \sum_{n=0}^{N-1} (T^n f) \cdots (T^{(k-1)n} f) \,\mathrm{d}\mu.$$

Quite recently, B. Host and B. Kra in [Host and Kra (2005b)] discovered a surprising fact regarding the multiple Cesàro sums on the right-hand side.

Theorem 15.6 (Host–Kra). *Let $(\Omega, \Sigma, \mu; \varphi)$ be an ergodic MDS, and consider the induced operator $T := T_\varphi$ on $\mathrm{L}^2(\Omega, \Sigma, \mu)$. Then the limit of the multiple ergodic averages*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (T^n f_1) \cdot (T^{2n} f_2) \cdots (T^{(k-1)n} f_{k-1}) \tag{15.6}$$

exists in $\mathrm{L}^2(\Omega, \Sigma, \mu)$ for every $f_1, \ldots, f_{k-1} \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)$ and $k \geq 2$.

Note that for $k = 2$ the Host–Kra theorem is just von Neumann's Mean Ergodic Theorem 8.1 (and holds even for arbitrary $f_1 \in \mathrm{L}^2(\Omega, \Sigma, \mu)$). We shall give a proof of the Host–Kra theorem for the case $k = 3$. (The case $k \geq 4$ would need further efforts beyond the scope of these lectures.)

Our major tool is the JdLG-decomposition of $\mathrm{L}^2(\Omega, \Sigma, \mu) = X_\mathrm{r} \oplus X_\mathrm{s}$ associated with the induced operator T. Let

$$Q : \mathrm{L}^2(\Omega,\Sigma,\mu) \longrightarrow X_{\mathrm{r}}$$

be the associated projection. Since every T^n is positive and Q is contained in the weak operator closure of $\{T^n : n \geq 0\}$, the projection Q remains positive. In addition, Q satisfies $Q\mathbf{1} = \mathbf{1}$, and therefore $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$ is Q-invariant. Finally, Q is strictly positive, i.e., if $f \geq 0$ and $Qf = 0$, then $f = 0$. This follows from the identity

$$\int_\Omega Qf \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu \qquad (f \in \mathrm{L}^2(\Omega,\Sigma,\mu)).$$

Furthermore, Q has the following important property.

Lemma 15.7. *The space $X_{\mathrm{r}} \cap \mathrm{L}^\infty(\Omega,\Sigma,\mu)$ is a closed subalgebra of $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$. Moreover, the projection Q is a "conditional expectation" operator, meaning that*

$$Q(f \cdot g) = (Qf) \cdot g \tag{15.7}$$

for every $f \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$ and $g \in X_{\mathrm{r}} \cap \mathrm{L}^\infty(\Omega,\Sigma,\mu)$.

Proof. Since Q is positive and fixes $\mathbf{1}$, its restriction to $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$ is bounded. The space $X_{\mathrm{r}} \cap \mathrm{L}^\infty = Q(\mathrm{L}^\infty)$ is therefore a closed subspace of $\mathrm{L}^\infty(\Omega,\Sigma,\mu)$. We show that Q satisfies (15.7). It then follows readily that $X_{\mathrm{r}} \cap \mathrm{L}^\infty$ is an algebra.

Take $f \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$. If $g \in X_{\mathrm{r}}$ is such that $Tg = \lambda g$ for some $\lambda \in \mathbb{T}$, then $T^n(fg) = \lambda^n(T^n f)g$ for each n. As Q is a weak accumulation point of $\{T^n : n \geq 0\}$ there is a net $(n_\alpha)_\alpha$ such that $T^{n_\alpha} \to Q$ weakly. Since $\mathbb{T}$ is compact, by passing to a subnet we may suppose that $\lambda^{n_\alpha} \to \eta$ for some $\eta \in \mathbb{T}$. Hence $Q(fg) = \eta(Qf)g$. Since η does not depend on f, we can iterate and obtain $Q(fg) = \eta^n(Qf)g$ for every $n \in \mathbb{N}$; this readily implies that $Q(fg) = (Qf)g$. Hence (15.7) is true for all g in the linear hull of all unimodular eigenvectors of T. Since the operators $g \longmapsto Q(fg)$ and $g \longmapsto (Qf)g$ are L^2-bounded, they must agree on the closure of this linear hull, i.e., on X_{r}. A fortiori, (15.7) is true for all $g \in X_{\mathrm{r}} \cap \mathrm{L}^\infty$ as claimed. □

In order to prove that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (T^n f) \cdot (T^{2n} g) \tag{15.8}$$

exists for every $f, g \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$, we may split f according to the JdLG-decomposition and consider the cases

$$(1) \quad f \in X_{\mathrm{s}} \qquad \text{and} \qquad (2) \quad f \in X_{\mathrm{r}}$$

separately. Case (1) is covered by the following lemma, which actually yields slightly more information. Notice that the proof is very similar to that of "weakly mixing of all orders", i.e., Theorem 14.13.

Lemma 15.8. *Let $f, g \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$ such that $f \in X_{\mathrm{s}}$ or $g \in X_{\mathrm{s}}$. Then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}(T^n f)\cdot(T^{2n}g) = 0.$$

Proof. The proof is an application of the van der Corput Lemma. We let $u_n := (T^n f)\cdot(T^{2n}g)$ and write

$$\begin{aligned}\big(u_n \,|\, u_{n+m}\big) &= \int_\Omega (T^n f)\cdot(T^{2n}g)\cdot(T^{n+m}\overline{f})\cdot(T^{2n+2m}\overline{g})\,\mathrm{d}\mu\\ &= \int_\Omega (fT^m\overline{f})\cdot T^n(gT^{2m}\overline{g})\,\mathrm{d}\mu\end{aligned}$$

by the invariance of μ. Compute now

$$\frac{1}{N}\sum_{n=0}^{N-1}\big(u_n \,|\, u_{n+m}\big) = \int\limits_\Omega (fT^m\overline{f})\frac{1}{N}\sum_{n=0}^{N-1}T^n(gT^{2m}\overline{g})\,\mathrm{d}\mu \to \int\limits_\Omega fT^m\overline{f}\,\mathrm{d}\mu\cdot\int\limits_\Omega gT^{2m}\overline{g}\,\mathrm{d}\mu$$

as $N\to\infty$, since μ is ergodic. Therefore we have

$$\begin{aligned}\gamma_m := \lim_{N\to\infty}\Big|\frac{1}{N}\sum_{n=0}^{N-1}\big(u_n \,|\, u_{n+m}\big)\Big| &= \Big|\int_\Omega (fT^m\overline{f})\,\mathrm{d}\mu\cdot\int_\Omega (gT^{2m}\overline{g})\,\mathrm{d}\mu\Big|\\ &= \Big|\big(f \,|\, T^m f\big)\big(g \,|\, T^{2m}g\big)\Big|.\end{aligned}$$

If $f\in X_{\mathrm{s}}$, by Theorem 12.14 we obtain that

$$0 \le \frac{1}{N}\sum_{m=0}^{N-1}\gamma_m \le \frac{1}{N}\sum_{m=0}^{N-1}\big|\big(f \,|\, T^m f\big)\big|\cdot\|g\|_\infty^2 \to 0 \quad \text{as} \quad N\to\infty.$$

In the case that $g\in X_{\mathrm{s}}$, the reasoning is similar. Now, the van der Corput Lemma 14.11 implies that

$$\frac{1}{N}\sum_{n=0}^{N-1}(T^n f)\cdot(T^{2n}g) = \frac{1}{N}\sum_{n=0}^{N-1}u_n \to 0. \qquad \square$$

Proof of Theorem 15.6, case $k=3$: By Lemma 15.8 we may suppose that $f\in X_{\mathrm{r}}$. Let $g\in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$ be fixed, and define $(S_N)_{N\in\mathbb{N}}\subset\mathscr{L}(X)$ by

$$S_N f := \frac{1}{N}\sum_{n=0}^{N-1}(T^n f)\cdot(T^{2n}g).$$

If f is an eigenfunction of T with eigenvalue $\lambda\in\mathbb{T}$, then

$$S_N f = f\frac{1}{N}\sum_{n=0}^{N-1}(\lambda T^2)^n g$$

holds. Since the operator λT^2 is mean ergodic as well (see Example 8.10), we have that $(S_N f)_{N\in\mathbb{N}}$ converges as $N \to \infty$. This yields that $(S_N f)_{N\in\mathbb{N}}$ converges for every

$$f \in \operatorname{lin}\{h : Th = \lambda h \text{ for some } \lambda \in \mathbb{T}\},$$

which is a dense subset of X_{r} by Theorem 12.10. Since the estimate $\|S_N\| \le \|g\|_\infty$ is valid for all $N \in \mathbb{N}$, we obtain the convergence of $(S_n)_{n\in\mathbb{N}}$ on the whole of X_{r}. □

In order to obtain Theorem 15.2 it remains to show that the limit in the Host–Kra theorem is positive whenever $0 < f = f_1 = \ldots = f_{k-1}$.

Proposition 15.9. *In the setting of Theorem 15.6*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f \cdot (T^n f) \cdot (T^{2n} f) \cdots (T^{(k-1)n} f) > 0$$

whenever $0 < f \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)$.

Again, we prove this result for $k = 3$ only. Our argument follows Furstenberg [Furstenberg (1981), Theorem 4.27].

Proof. It suffices to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_\Omega f \cdot (T^n f) \cdot (T^{2n} f) \, \mathrm{d}\mu > 0. \tag{15.9}$$

We have $f = f_{\mathrm{s}} + f_{\mathrm{r}}$ with $f_{\mathrm{r}} := Qf \in X_{\mathrm{r}}$ and $f_{\mathrm{s}} = f - Qf \in X_{\mathrm{s}}$. Since Q is strictly positive and $f > 0$, $f_{\mathrm{r}} > 0$ as well. Let $g, h \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)$ be arbitrary. Since $X_{\mathrm{r}} \perp X_{\mathrm{s}}$ in L^2 (see Example 12.6), it follows from Lemma 15.7 that

$$\int_\Omega f \cdot (T^n g) \cdot (T^{2n} h) \, \mathrm{d}\mu = 0$$

whenever two of the functions f, g, h are in X_{r} and the remaining one is from X_{s}. By Lemma 15.8 we have

$$\frac{1}{N} \sum_{n=0}^{N-1} \int_\Omega f \cdot (T^n g) \cdot (T^{2n} h) \, \mathrm{d}\mu \to 0$$

if at least one of the functions g, h is from X_{s}. So we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_\Omega f \cdot (T^n f) \cdot (T^{2n} f) \, \mathrm{d}\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_\Omega f_{\mathrm{r}} \cdot (T^n f_{\mathrm{r}}) \cdot (T^{2n} f_{\mathrm{r}}) \, \mathrm{d}\mu.$$

We may therefore suppose without loss of generality that $f = f_{\mathrm{r}}$.

Take now $0 < \varepsilon < \int_\Omega f^3 \, \mathrm{d}\mu$ and suppose that $\|f\|_\infty \le 1$, again without loss of generality. We show that there is some subset $B \subset \mathbb{N}$ with bounded gaps (also called *relatively dense* or *syndetic*; cf. Definition 3.10.b) such that

$$\int_\Omega f \cdot (T^n f) \cdot (T^{2n} f)\,\mathrm{d}\mu > \int_\Omega f^3\,\mathrm{d}\mu - \varepsilon \qquad \text{for all } n \in B. \tag{15.10}$$

(Recall that this property of B means that there is $L \in \mathbb{N}$ such that every interval of length L intersects B.)

Since $f \in X_{\mathrm{r}}$, there exists ψ of the form $\psi = c_1 g_1 + \ldots + c_j g_j$ for some eigenfunctions $g_1, \ldots, g_j$ of T corresponding to eigenvalues $\lambda_1, \ldots, \lambda_j \in \mathbb{T}$ such that $\|f - \psi\|_2 < \frac{\varepsilon}{6}$. Observe first that for every $\delta > 0$ there is a relatively dense set $B \subset \mathbb{N}$ such that

$$|\lambda_1^n - 1| < \delta,\ \ldots,\ |\lambda_j^n - 1| < \delta \quad \text{for all } n \in B.$$

Indeed, if we consider the rotation by $(\lambda_1, \ldots, \lambda_j)$ on $\mathbb{T}^j$, this is precisely the almost periodicity of the point $(1, \ldots, 1) \in \mathbb{T}^j$ (see also Proposition 3.15). Now, by taking a suitable $\delta > 0$ depending on the coefficients $c_1, \ldots c_j$ we obtain

$$\|T^n \psi - \psi\|_2 < \frac{\varepsilon}{12}$$

and
$$\|T^n f - f\|_2 \le \|T^n f - T^n \psi\|_2 + \|T^n \psi - \psi\|_2 + \|\psi - f\|_2 < \frac{\varepsilon}{2}$$

for all $n \in B$. Moreover, for $n \in B$ one has

$$\begin{aligned}\|T^{2n} f - f\|_2 &\le \|T^{2n} f - T^{2n}\psi\|_2 + \|T^{2n}\psi - T^n \psi\|_2 + \|T^n \psi - \psi\|_2 + \|\psi - f\|_2 \\ &\le 2\|f - \psi\|_2 + 2\|T^n \psi - \psi\|_2 < \frac{\varepsilon}{2}.\end{aligned}$$

Altogether we obtain

$$\begin{aligned}\Big|\int_\Omega f \cdot T^n f \cdot T^{2n} f\,\mathrm{d}\mu - \int_\Omega f^3\,\mathrm{d}\mu\Big| &\le \int_\Omega f \cdot T^n f \cdot |T^{2n} f - f|\,\mathrm{d}\mu + \int_\Omega f^2 \cdot |T^n f - f|\,\mathrm{d}\mu \\ &\le \|f\|_\infty^2 \big(\|T^{2n} f - f\|_2 + \|T^n f - f\|_2\big) < \varepsilon,\end{aligned}$$

and (15.10) is proved. Since $B \cap [jL, (j+1)L] \neq \emptyset$ for all $j \in \mathbb{N}_0$, it follows that

$$\begin{aligned}\frac{1}{LN}\sum_{n=0}^{NL} \int_\Omega f \cdot T^n f \cdot T^{2n} f\,\mathrm{d}\mu &\ge \frac{1}{LN}\sum_{n \in B, n=0}^{NL} \int_\Omega f \cdot T^n f \cdot T^{2n} f\,\mathrm{d}\mu \\ &\ge \frac{1}{L}\Big(\int_\Omega f^3\,\mathrm{d}\mu - \varepsilon\Big) > 0.\end{aligned}$$ □

Combining the above results leads to the classical theorem of Roth, a precursor of Szemerédi's Theorem 15.2.

Theorem 15.10 (Roth, 1952). *Let $A \subset \mathbb{N}$ have positive upper density. Then A contains infinitely many arithmetic progressions of length* 3.

The full proof of Szemerédi's theorem is based on an inductive procedure involving "higher-oder" decompositions of a JdLG-flavour. See for instance [Furstenberg (1977), Furstenberg (1981)], [Petersen (1989)], [Tao (2008a)], [Green (2008)] or [Einsiedler and Ward (2009)]. However, the required techniques are beyond the scope of these lectures.

15.3 The Furstenberg–Sárközy Theorem

We continue in the same spirit, but instead of arithmetic progressions we now consider pairs of the form $a, a+n^2$ for $a, n \in \mathbb{N}$.

Theorem 15.11 (Furstenberg–Sárközy). *Let $A \subset \mathbb{N}$ have positive upper density. Then there exist $a, n \in \mathbb{N}$ such that $a, a+n^2 \in A$.*

For a more general statement we refer to is from [Furstenberg (1981), Proposition 3.19]. In order to prove the above theorem we first need an appropriate correspondence principle.

Theorem 15.12 (Furstenberg Correspondence Principle for Squares). *If for every ergodic MDS $(\Omega, \Sigma, \mu; \varphi)$ and its induced operator $T := T_\varphi$ on $\mathrm{L}^2(\Omega, \Sigma, \mu)$ the condition*

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_\Omega f \cdot (T^{n^2} f)\,\mathrm{d}\mu \;>\; 0 \tag{15.11}$$

is satisfied for all $0 < f \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)$, then Theorem 15.11 holds.

The proof is analogous to the one of Theorem 15.5. To show (15.11) we begin with the following "nonconventional" ergodic theorem.

Theorem 15.13 (Mean Ergodic Theorem for Squares). *Let T be a contraction on a Hilbert space H. Then the Cesàro square averages*

$$S_N x := \frac{1}{N} \sum_{n=0}^{N-1} T^{n^2} x$$

converge for every $x \in H$ as $N \to \infty$.

Proof. We leave the complete proof as a project for Phase 2 and suppose here in addition that T is an isometry. By the JdLG-decomposition it suffices to prove the assertion for $x \in H_\mathrm{r}$ and $x \in H_\mathrm{s}$ separately.
Take first $x \in H_\mathrm{r}$ satisfying $Tx = \mathrm{e}^{2\pi\mathrm{i}\alpha} x$ for some $\alpha \in [0,1)$. Then we have

$$S_N x = \frac{1}{N} \sum_{n=0}^{N-1} \mathrm{e}^{2\pi\mathrm{i}\alpha n^2} x.$$

If α is irrational, this sequence converges to 0 by the equidistribution of $(\alpha n^2)_n$ (see Propositions 9.14 and 10.20). If $\alpha \in \mathbb{Q}$, say $d\alpha \in \mathbb{N}_0$ for some $d \in \mathbb{Z} \setminus \{0\}$, then one can see that

$$\frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi i\alpha n^2} \to \frac{1}{d}\sum_{j=0}^{d-1} e^{2\pi i\alpha j^2} \quad (N\to\infty).$$

Since all S_N are contractive and since the linear span of all eigenvectors corresponding to unimodular eigenvalues is dense in H_{r} (see Theorem 12.10), we obtain convergence for every $x \in H_{\mathrm{r}}$.

Take now $x \in H_{\mathrm{s}}$ and define $u_n := T^{n^2}x$. In order to use the van der Corput lemma we calculate

$$\bigl(u_n \,|\, u_{n+m}\bigr) = \bigl(T^{n^2}x \,|\, T^{(n+m)^2}x\bigr) = \bigl(x \,|\, T^{2nm+m^2}x\bigr) = \bigl(T^{*m^2}x \,|\, (T^{2m})^n x\bigr),$$

where we have used that T is an isometry. This implies for any fixed $m \in \mathbb{N}$ that

$$\begin{aligned}\gamma_m = \limsup_{N\to\infty}\Bigl|\frac{1}{N}\sum_{n=0}^{N-1}\bigl(u_n \,|\, u_{n+m}\bigr)\Bigr| &\le \limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\Bigl|\bigl(T^{*m^2}x \,|\, (T^{2m})^n x\bigr)\Bigr| \\ &\le 2m\lim_{N\to\infty}\frac{1}{2mN}\sum_{j=0}^{2mN-1}\Bigl|\bigl(T^{*m^2}x \,|\, T^j x\bigr)\Bigr| = 0,\end{aligned}$$

since $x \in H_{\mathrm{s}}$ so by Theorem 12.14 the limit on the right-hand side is 0. The van der Corput Lemma 14.11 concludes the proof. □

So we see that, in particular, for an MDS $(\Omega,\Sigma,\mu;\varphi)$ and its induced operator $T := T_\varphi$ the limit

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\cdot T^{n^2} f \tag{15.12}$$

exists in $\mathrm{L}^2(\Omega,\Sigma,\mu)$ for every $f \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$. To finish the proof of Theorem 15.11 it remains to show that it is positive whenever f is.

Proposition 15.14. *Let* $(\Omega,\Sigma,\mu;\varphi)$ *be an ergodic MDS. Then the limit* (15.12) *is positive for every* $0 < f \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$.

The proof is analogous to that of Proposition 15.9 and again uses the decomposition $\mathrm{L}^2(\Omega,\Sigma,\mu) = X_{\mathrm{s}} \oplus X_{\mathrm{r}}$, the vanishing of the above limit on X_{s} and a relative density argument on X_{r}.

Remarks 15.15. 1) A sequence $(n_k)_{k\in\mathbb{N}}$ is called a **Poincaré sequence** (see, e.g., [Furstenberg (1981), Def. 3.3.6]) if for every MDS $(\Omega,\Sigma,\mu;\varphi)$ and $M \in \Sigma$ with $\mu(M) > 0$ one has

$$\mu(M \cap \varphi^{-n_k}(M)) > 0 \quad \text{for some } k \in \mathbb{N}.$$

Poincaré's Theorem 6.6 tells that $(n)_{n\in\mathbb{N}}$ is a Poincaré sequence, hence the terminology. One can prove that Proposition 15.14 remains valid even for not necessarily ergodic MDSs, so we obtain that $(n^2)_{n\in\mathbb{N}}$ is a Poincaré sequence.

2) The above results remain true if one replaces the polynomial $p(n) = n^2$ by an arbitrary polynomial p with integer coefficients and $p(0) = 0$, see Furstenberg [Furstenberg (1981), pp. 66–75].

Final remarks

There are many generalisations and modifications of the above results. We only quote few examples. The first is from [Furstenberg and Weiss (1996)].

Theorem 15.16 (Furstenberg, Weiss). *Let $(\Omega,\Sigma,\mu;\varphi)$ be an MDS, and consider the induced linear operator $T := T_\varphi$ on $X = \mathrm{L}^2(\Omega,\Sigma,\mu)$. Then the limit of the multiple ergodic averages*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T^{an} f \cdot T^{bn} g \tag{15.13}$$

exists in $\mathrm{L}^2(\Omega,\Sigma,\mu)$ for every $f,g \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$ and every (distinct) $a,b \in \mathbb{N}$.

Remark 15.17. In [Furstenberg and Weiss (1996)] it is also shown that the averages

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T^{n} f \cdot T^{n^2} g$$

converge in $\mathrm{L}^2(\Omega,\Sigma,\mu)$ for every $f,g \in \mathrm{L}^\infty(\Omega,\Sigma,\mu)$. Together with the positivity of this limit for $f = g > 0$ and the Furstenberg correspondence principle this yields the existence of (many) triples $a, a+n, a+n^2$ in any $A \subset \mathbb{N}$ with $\overline{d}(A) > 0$.

Host and Kra [Host and Kra (2005a)] proved convergence of multiple ergodic averages for strongly ergodic systems with powers $p_1(n),\dots,p_{k-1}(n)$ replacing $n,\dots,(k-1)n$ in (15.6), where $p_1,\dots,p_{k-1}$ are arbitrary polynomials with integer coefficients. This leads to the existence of a subset of the form $\{a, a+p_1(n),\dots,a+p_{k-1}(n)\}$ in every set $A \subset \mathbb{N}$ with positive upper density.

Remark 15.18. The results of Szemerédi and Furstenberg–Sárközy presented in this lecture are actually true in a slightly strengthened form. One can replace the upper density $\overline{d}$ by the so-called **upper Banach-density** defined as

$$\overline{Bd}(A) := \limsup_{m,n\to\infty} \frac{\operatorname{card}(A \cap \{m,\dots,m+n\})}{n+1}.$$

The actual change to be carried out in order to obtain these generalisations is in Furstenberg's Correspondence Principle(s), see [Furstenberg (1981)].

Appendix A
Topology

A.1 Metric spaces

A **metric space** is a pair (Ω, d) consisting of a non-empty set Ω and a function $d : \Omega \times \Omega \longrightarrow \mathbb{R}$ which describes the distance between any two points of Ω, and for which we require the following properties:

(i) $d(x,y) \geq 0$, and $d(x,y) = 0$ if and only if $x = y$.

(ii) $d(x,y) = d(y,x)$.

(iii) $d(x,y) \leq d(x,z) + d(z,y)$ **(triangle inequality)**.

The function d is called a **metric** on Ω. If instead of (i) we require only $d(x,x) = 0$ for all $x \in \Omega$, we obtain the notion of a **semi-metric** (cf. B.9). For $A \subset \Omega$ and $x \in \Omega$ we define

$$d(x,A) := \inf\{d(x,y) \,:\, y \in A\}$$

called the **distance** of x from A. The triangle inequality implies that

$$|d(x,A) - d(y,A)| \leq d(x,y) \qquad (x, y \in \Omega).$$

By a **ball** with center x and radius $r > 0$ we mean either of the sets

$$\mathrm{B}(x,r) := \{y \in \Omega \,:\, d(x,y) < r\},$$
$$\overline{\mathrm{B}}(x,r) := \{y \in \Omega \,:\, d(x,y) \leq r\}.$$

A set $O \subseteq X$ is called **open** if for all $x \in O$ there is a ball $B \subseteq O$ with centre x and radius $r > 0$. A set $A \subseteq \Omega$ is called **closed** if $\Omega \setminus A$ is open. The ball $\mathrm{B}(x,r)$ is open, and the ball $\overline{\mathrm{B}}(x,r)$ is closed for any $x \in \Omega$ and $r > 0$. Two trivially open, and at the same time closed sets are the empty set $\emptyset$ and the set Ω itself.

A sequence (x_n) in Ω is **convergent to the limit** $x \in \Omega$ (we write: $x_n \to x$), if for all $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ with $d(x_n, x) < \varepsilon$ for $n \geq n_0$. For a subset $A \subset \Omega$ the following assertions are equivalent:

(i) A is closed;

(ii) if $x \in \Omega$ and $d(x,A) = 0$ then $x \in A$;

(iii) if $(x_n)_n \subset A$ and $x_n \to x \in \Omega$, then $x \in A$.

A **Cauchy-sequence** (x_n) in Ω is a sequence with the property that for all $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ with $d(x_n, x_m) < \varepsilon$ for $n, m \geq n_0$. A convergent sequence is always a Cauchy-sequence. A metric space is called **complete** if the converse implication also holds.

A.2 Topological spaces

Starting with a non-empty set Ω we would like to define open sets. For this purpose observe that open sets in Euclidean spaces or as defined in A.1 satisfy the following:

a) $\emptyset$ and Ω are open.

b) If O_1 and O_2 are open, so is their intersection $O_1 \cap O_2$.

c) An arbitrary union of open sets is open.

We now take these three properties as the defining characteristics for the family of open sets in Ω. More precisely, assume that $\mathscr{O} \subseteq \mathscr{P}(\Omega)$ satisfies

(i) $\emptyset, \Omega \in \mathscr{O}$;

(ii) If $O_1, \ldots, O_n \in \mathscr{O}$, then $O_1 \cap \cdots \cap O_n \in \mathscr{O}$; (i.e., $\mathscr{O}$ is $\cap$-stable, cf. page 191)

(iii) If $O_\iota \in \mathscr{O}$, $\iota \in I$, then $\bigcup_{\iota \in I} O_\iota \in \mathscr{O}$.

We call $\mathscr{O}$ the **family of open sets** in Ω, and we say that $(\Omega, \mathscr{O})$ (or simply Ω) is a **topological space**. We also call $\mathscr{O}$ itself the topology on Ω. **Closed sets** are then the complements of open sets. Finite unions and arbitrary intersections of closed sets are closed.

A topological space is called **metrisable** if there exists a metric that induces the topology. Not every topological space is metrisable, cf. Section A.9 below.

If $\mathscr{O}, \mathscr{O}'$ are both topologies on Ω and $\mathscr{O}' \subseteq \mathscr{O}$, then we say that the topology $\mathscr{O}$ is **finer** than $\mathscr{O}'$ (or $\mathscr{O}'$ is **coarser** than $\mathscr{O}$).

Example A.1. a) Let Ω be non-empty, and $\mathscr{O} := \{\emptyset, \Omega\}$. Then $\mathscr{O}$ satisfies the properties (i)–(iii), thus $(\Omega, \mathscr{O})$ is a topological space, whose topology is called the **trivial topology**. Note that it is the coarsest among all topologies on Ω.

b) The other extreme case is when we set $\mathscr{O} := \mathscr{P}(\Omega)$, the largest possible choice. The so defined topology is the **discrete topology**. The explanation for this terminology is that for all points in $x \in \Omega$ the singleton $\{x\}$ is open (and also closed). The discrete topology is the finest among all possible topologies on Ω. In this case all sets are closed and open at the same time.

c) If (Ω, d) is a metric space, the open sets from A.1 define a topology $\mathscr{O}_d$ on Ω, and we say that the metric induces the topology on Ω. In this case there are many different metrics that induce the same topology, and these metrics we call **equivalent metrics**.

d) For Ω is a non-empty set, the function $d : \Omega \times \Omega \longrightarrow \mathbb{R}$ defined by $d(x,y) = 0$ for $x = y$, $d(x,y) = 1$ for $x \neq y$ is a metric and it induces the discrete topology.

A **neighbourhood** of a point $x \in \Omega$ is a set U such that there is an open set $O \subseteq \Omega$ with $x \in O \subseteq U$. An open set is a neighbourhood of all of its points. If A is a neighbourhood of x, then x is called an interior point of A. The set of all interior points of A is denoted by $\mathring{A}$ and is called the **interior** of A. The **closure** $\overline{A}$ of a subset $A \subseteq \Omega$ is

$$\bigcap_{\substack{A \subseteq F \subseteq \Omega \\ F \text{ closed}}} F,$$

which is obviously the smallest closed set that contains A. If (Ω, d) is a metric space and $A \subset \Omega$, then $\overline{A} = \{x : d(x,A) = 0\}$, and $x \in \overline{A}$ iff x is the limit of a sequence in A.

To define a topology it is not necessary to specify all the open sets. We may as well proceed similarly to the metric case, by replacing the family of open balls

$$\{\mathrm{B}(x,r) : x \in \Omega, \, r > 0\}$$

by a suitable system of neighbourhoods. A **base** $\mathscr{B} \subseteq \mathscr{O}$ for the topology $\mathscr{O}$ on Ω, is a system which has the property that all open sets can be written as the union of base-elements. For example, the family of open balls is a base for the topology induced by the metric on the metric space (Ω, d). A topological space is called **second countable** if it has a countable base for its topology.

A topological space Ω is called **Hausdorff** if any two points $x, y \in \Omega$ can be separated by disjoint open neighbourhoods, i.e., there are $U, V \in \mathscr{O}$ with $U \cap V = \emptyset$ and $x \in U$, $y \in V$. In a Hausdorff space a singleton $\{x\}$ $x \in \Omega$ is closed. Discrete spaces are Hausdorff, while trivial topological spaces are not unless Ω is a singleton. More generally metric spaces are Hausdorff. If $\mathscr{O}$, $\mathscr{O}'$ are two topologies on Ω, $\mathscr{O}$ finer than $\mathscr{O}'$ and $\mathscr{O}'$ Hausdorff, then also $\mathscr{O}$ is Hausdorff. In these lectures we will always consider Hausdorff spaces, even if this is not stated explicitly.

If Ω' is a non-empty subset of Ω, then the **subspace topology** on Ω' is given by $\mathscr{O}_{\Omega'} := \{\Omega' \cap O : O \in \mathscr{O}\}$. A subspace of a Hausdorff space is Hausdorff. An **isolated point** of Ω' is a point $y \in \Omega'$ for which $\{y\}$ is open in the subspace topology of Ω'. The non-isolated points of Ω' are called **accumulation points**. A point $x \in \Omega$ is the **cluster point** of the sequence $(x_n)_{n \in \mathbb{N}}$ in Ω if any neighbourhood of x contains infinitely many members of the sequence.

A subset A of a topological space Ω is called **dense** in Ω if $\overline{A} = \Omega$. A topological space Ω is called **separable** if there is a countable set $A \subset \Omega$ which is dense in Ω. A subspace of a separable metric space is separable, but a similar statement for general topological spaces is false.

A.3 Continuity

Given $(\Omega,\mathscr{O})$, $(\Omega',\mathscr{O}')$ two topological spaces, a function $f:\Omega\longrightarrow\Omega'$ is called **continuous** if the inverse image $f^{-1}(O)$ of each open set $O\in\mathscr{O}'$ is open in Ω; we sometimes say that $f:(\Omega,\mathscr{O})\longrightarrow(\Omega',\mathscr{O}')$ is continuous. Replacing open sets by closed sets yields the same notion. The function f is **continuous at** $x\in\Omega$ if for all (open) neighbourhood V of $f(x)$ in Ω', there is U an (open) neighbourhood of x with $f(U)\subseteq V$.

For metric spaces Ω, Ω' continuity is the same as **sequential continuity**, i.e., the property that for $x_n\in\Omega$, x_n convergent to x, one has the convergence $f(x_n)\to f(x)$.

If Ω is endowed with the discrete topology then all functions $f:\Omega\longrightarrow\Omega'$ are continuous. The same is true if Ω' has the trivial topology.

A bijective continuous transformation whose inverse is also continuous is called a **homeomorphism**. The Hausdorff property is homeomorphism-invariant, whereas completeness of metric spaces is not (cf. also B.9).

A mapping $f:(\Omega,d)\longrightarrow(\Omega',d')$ between two metric spaces is called **uniformly continuous** if for each $\varepsilon>0$ there is $\delta>0$ such that $d(x,y)<\delta$ implies $d(f(x),f(y))<\varepsilon$ for all $x,y\in\Omega$. If $A\subset\Omega$ then the distance function $x\longmapsto d(x,A)$ from Ω to $\mathbb{R}$ is uniformly continuous.

A.4 Inductive and projective topologies

Let Ω_ι, $\iota\in I$ be topological spaces and $f_\iota:\Omega_\iota\longrightarrow\Omega$ for some non-empty set Ω. Define

$$\mathscr{O}_{\mathrm{ind}}:=\{A\subseteq\Omega\,:\,f_\iota^{-1}(A)\text{ for all }\iota\in I\}.$$

Then $(\Omega,\mathscr{O}_{\mathrm{ind}})$ is a topological space. The topology $\mathscr{O}_{\mathrm{ind}}$ is called the **inductive topology** on Ω with respect to $(f_\iota)_{\iota\in I}$, and it is the finest topology such that all the mappings f_ι become continuous. A function $g:\Omega\longrightarrow Z$, $(\Omega',\mathscr{O}')$ a topological space is continuous, if and only if all the functions $g\circ f_\iota:\Omega_\iota\longrightarrow\Omega'$, $\iota\in I$ are continuous.

The projective topology is defined on the other way round: Suppose that Ω_ι are topological spaces with $f_\iota:\Omega\longrightarrow\Omega_\iota$ mappings for some set Ω, $\iota\in I$. The **projective topology** $\mathscr{O}_{\mathrm{proj}}$ with resect to $(f_\iota)_{\iota\in I}$ is the coarsest for which the functions f_ι become continuous. For the existence of this coarsest topology consider the family

$$\mathscr{B}_{\mathrm{proj}}:=\{f_\iota^{-1}(O_\iota)\,:\,O_\iota\in\mathscr{O}\text{ for all }\iota\in I\}\subseteq\mathscr{P}(\Omega).$$

One can show that $\mathscr{B}_{\mathrm{proj}}$ is a base for a topology which has the required properties.

A function $g:\Omega'\longrightarrow\Omega$, $(\Omega',\mathscr{O}')$ a topological space is continuous, if and only if all the functions $f_\iota\circ g:\Omega'\longrightarrow\Omega_\iota$, $\iota\in I$ are continuous.

An example is the *subspace topology*: If $\Omega' \subseteq \Omega$, $(\Omega, \mathscr{O})$ a topological space, then the subspace topology on Ω' is exactly the projective topology with respect to the natural imbedding $J : \Omega' \longrightarrow \Omega$.

$$\begin{tikzcd}
\Omega \arrow[r, "g"] & \Omega' \\
\Omega_\iota \arrow[u, "f_\iota"] \arrow[ur, dotted, "g \circ f_\iota"'] &
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
\Omega' \arrow[r, "g"] & \Omega \arrow[d, "f_\iota"] \\
& \Omega_\iota \arrow[ul, dotted, "f_\iota \circ g"]
\end{tikzcd}$$

Fig. A.1 Continuity for the inductive respectively for the projective topology

A.5 Product spaces

Let $(\Omega_\iota)_{\iota \in I}$ a non-empty family of non-empty topological spaces. The **product topology** on

$$\Omega := \prod_{\iota \in I} \Omega_\iota = \left\{ x : I \longrightarrow \bigcup \Omega_\iota \ : \ x(\iota) \in \Omega_\iota \right\}$$

is the projective topology with respect to the canonical projections $\pi_\iota : \Omega \longrightarrow \Omega_\iota$. Instead of $x(\iota)$ we usually write x_ι. A base for this topology is formed by the open rectangles

$$A_{\iota_1,\ldots,\iota_n} := \{x = (x_\iota)_{\iota \in I} \ : \ x_{\iota_i} \in O_{\iota_i} \text{ for } i = 1, \ldots, n\}$$

for $\iota_1, \ldots, \iota_n \in I$, $n \in \mathbb{N}$ and O_{ι_i} open in Ω_{ι_i}. For the product of two (or finitely many) spaces we also use the notation $\Omega \times \Omega'$ and the like. A space is Hausdorff if and only if the diagonal $\{(x,x) \ : \ x \in \Omega\}$ is closed in the product space $\Omega \times \Omega$.

If I is countable and Ω_n, $n \in I$ are all metrisable spaces then so is their product $\prod_{n \in I} \Omega_n$. The convergence in this product space is just the coordinatewise convergence. If $\Omega_\iota = \Omega$ for all $\iota \in I$, then we use the notation Ω^I for the product space.

A.6 Quotient spaces

Let $(\Omega, \mathscr{O})$ be a topological space and $\sim$ an equivalence relation on Ω, with quotient mapping

$$q : \Omega \longrightarrow \Omega/\!\sim,$$

sending each $x \in \Omega$ to its equivalence class. The inductive topology on $\Omega/\!\sim$ with respect to q is called the **quotient topology**. A set $A \subseteq \Omega/\!\sim$ is open in $\Omega/\!\sim$ if and only if $\bigcup A$, the union of the elements in A is open in Ω.

For example consider $\Omega = [0,1]$ and the equivalence relation $\sim$ whose equivalence classes are $\{0,1\}$, $\{x\}$, $x \in (0,1)$. Then the quotient $[0,1]/\sim$ is homeomorphic to the unit circle $\mathbb{T}$, under the mapping $(x \longmapsto e^{2\pi i \cdot x}) : [0,1]/\sim \longrightarrow \mathbb{T}$. This example is just the same as factorising $\mathbb{R}$ by $\mathbb{Z}$: take $\Omega = \mathbb{R}$ and define $x \sim y$ if $x - y \in \mathbb{Z}$. Then we have $\mathbb{R}/\mathbb{Z} = \mathbb{R}/\sim$ homeomorphic to $\mathbb{T}$.

A.7 Spaces of Continuous Functions

For Ω a topological space the set $\mathrm{C}(\Omega)$ or $\mathrm{C}^{\mathrm{b}}(\Omega)$ of all continuous functions respectively bounded, continuous functions $f : \Omega \longrightarrow \mathbb{K}$ with pointwise multiplication and addition is an algebra over $\mathbb{K}$ ($\mathbb{K}$ stands for $\mathbb{R}$ or $\mathbb{C}$).

A sequence of bounded functions on Ω is **uniformly convergent** to $f : \Omega \longrightarrow \mathbb{K}$ if

$$\sup_{x \in \Omega} |f_n(x) - f(x)| \to 0, \quad \text{for } n \to \infty.$$

If each f_n is continuous and the sequence $(f_n)_n$ converges to f uniformly, the function f is continuous. The function $(f \longmapsto \|f\|_\infty) : \mathrm{C}^{\mathrm{b}}(\Omega) \longrightarrow \mathbb{R}$ is a norm and turns $\mathrm{C}^{\mathrm{b}}(\Omega)$ into a Banach space, even a Banach algebra. If (Ω, d) is a metric space, then the space $\mathrm{BUC}(\Omega, d)$ of bounded uniformly continuous functions on Ω is a closed subspace of $\mathrm{C}^{\mathrm{b}}(\Omega)$.

For general topological spaces the space $\mathrm{C}(\Omega)$ may be quite "small". For example, if Ω carries the trivial topology, the only continuous functions thereon are the constant ones. In "good" topological spaces the continuous functions **separate the points**, i.e., for every $x, y \in \Omega$ such that $x \neq y$ there is $f \in \mathrm{C}(\Omega)$ such that $f(x) \neq f(y)$. (Such spaces are necessarily Hausdorff.) Even better it is, when the continuous functions **separate closed sets**. This means that for every pair of disjoint closed subsets $A, B \subset \Omega$ there is a function $f \in \mathrm{C}^{\mathrm{b}}(\Omega)$ such that

$$0 \le f \le 1, \quad f(A) \subset \{0\}, \quad f(B) \subset \{1\}.$$

Metric spaces do have this property: Let (Ω, d) be a metric space. For each $A \subset \Omega$ the function $x \longmapsto d(x, A)$ is continuous; moreover, it is equal to zero precisely on $\overline{A}$. Hence if A, B are disjoint closed subsets of Ω then the function

$$f(x) := \frac{d(x,B)}{d(x,A) + d(x,B)} \qquad (x \in \Omega)$$

separates A from B. (Note that f is even uniformly continuous.)

A.8 Compactness

A topological space $(\Omega, \mathscr{O})$, $\mathscr{O}$ the family of open sets in Ω, is called **compact** if it is Hausdorff and every open cover of Ω has a finite subcover. This latter condition is equivalent to the **finite intersection property**: every family of closed subsets of Ω, every finite subfamily of which has non-empty intersection, has itself non-empty intersection. A subset $\Omega' \subseteq \Omega$ is compact if Ω' with the subspace topology is compact. A compact set in a Hausdorff space is closed, and a closed subset in a compact space is compact. A **relatively compact** set is set whose closure is compact.

By the Heine-Borel theorem, a subset of $\mathbb{R}^d$ is compact iff it is closed and bounded.

The continuous image of a compact space is compact, if it is Hausdorff. Moreover, if Ω is compact and Ω' is Hausdorff, a mapping $\varphi : \Omega \longrightarrow \Omega'$ is already a homeomorphism if it is continuous and bijective. In particular, if Ω is compact for some topology $\mathscr{O}$ and if $\mathscr{O}'$ is another topology on Ω, coarser than $\mathscr{O}$ but still Hausdorff, then $\mathscr{O} = \mathscr{O}'$.

Theorem A.2 (Tychonov). *Suppose $(\Omega_\iota)_{\iota\in I}$is a family of non-empty topological spaces. Then the product space $\Omega = \prod_{\iota\in I} \Omega_\iota$ is compact if and only if each Ω_ι, $\iota \in I$ is compact.*

A metric space Ω is compact if and only if it is **sequentially compact**, that is, each sequence $(x_n) \subseteq \Omega$ has a convergent subsequence. A compact metric space is complete, separable and has a countable base, and every continuous function on it is bounded and uniformly continuous.

A Hausdorff topological space $(\Omega, \mathscr{O})$ is called **locally compact** if each of its points has a compact neighbourhood. It follows then that the topology has a base consisting of relatively compact, open sets. A compact space is (trivially) locally compact. The **support** of a function $f : \Omega \longrightarrow \mathbb{C}$ is the set

$$\operatorname{supp} f := \overline{\{x \in \Omega \,:\, f(x) \neq 0\}}.$$

The set of all continuous functions with compact support is a vector space, denoted by $\mathrm{C_c}(\Omega)$. On locally compact spaces, continuous functions separate closed from compact subsets:

Lemma A.3 (Urysohn's lemma). *Let $(\Omega, \mathscr{O})$ be a locally compact space, and let A, B disjoint closed subsets of Ω, with B compact. Then there exists a continuous function $f : \Omega \longrightarrow [0,1]$ such that f has compact support, $f(A) \subseteq \{0\}$, and $f(B) \subseteq \{1\}$.*

Let $\Omega = K$ be compact. Then $\mathrm{C}(K) = \mathrm{C_c}(K) = \mathrm{C^b}(K)$ is a Banach algebra with respect to the uniform norm. Urysohn's Lemma shows that $\mathrm{C}(K)$ separates closed sets.

A.9 Metrisability

There are various sufficient conditions for the existence of a metric on a topological space that induces the given topology. The most convenient for us is the following: Suppose that $(K, \mathscr{O})$ is a compact space and there is a countable family of functions f_n, $n \in \mathbb{N}$ that separates the points of K. Then the function

$$d(x,y) := \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|f_n(x) - f_n(y)|}{1 + |f_n(x) - f_n(y)|}$$

is continuous on $K \times K$ and it is a metric. Thus the topology $\mathscr{O}_d$ induced by d is coarser than the original one. This means that $\mathrm{Id} : (K, \mathscr{O}) \longrightarrow (K, \mathscr{O}_d)$ is a continuous mapping, but then because of the compactness of K it is also a homeomorphism, hence $\mathscr{O}_d = \mathscr{O}$.

More generally, a compact space is metrisable if and only if it has a countable base if and only if $\mathrm{C}(K)$ is separable.

A.10 Category

A subset A of a topological space Ω is called **nowhere dense** if its closure $\overline{A}$ has empty interior: $(\overline{A})^\circ = \emptyset$. A is called of **first category** in Ω if it is the union of countably many nowhere dense subsets of Ω. A is called of **second category** in Ω if it is not of first category. Countable unions of sets of first category are of first category.

One should have the picture in mind that sets of first category are small, whereas sets of second category are large. Typically one expects that "fat" sets, for example non-empty open sets are somehow large. This requirement is the defining characteristic of Baire spaces: $(\Omega, \mathscr{O})$ is called a **Baire space** if every non-empty open subset of Ω is of second category in Ω.

Theorem A.4. *Each locally compact space and each complete metric space is a Baire space.*

A countable intersection of open sets in a topological space is called a G_δ **set**; analogously, F_σ **sets** are those that can be written as countable union of closed sets. The following is an easy consequence of the definitions:

Theorem A.5. *Let Ω be a Baire-space.*

a) *An F_σ set is of first category if and only if it has empty interior.*
b) *A G_δ set is of first category if and only if it is nowhere dense.*
c) *A countable intersection of dense G_δ sets is dense.*

Note that in a metric space every closed subset A is G_δ since $A = \bigcap_{n \in \mathbb{N}} \{x : d(x,A) < 1/n\}$.

A.11 Polish spaces

A topological space Ω is called a **Polish space** if it is separable and its topology comes from some complete metric. A locally compact space is Polish if and only if it has a countable base for its topology. A compact space is Polish if and only if it is metrisable.

Appendix B
Measure and Integration Theory

We begin with some general set-theoretic notions. Let Ω be a set. Then its **power set** is denoted by

$$\mathscr{P}(\Omega) := \{A : A \subset \Omega\}.$$

Given $A \subset \Omega$ its **complement** is denoted by $A^{\mathrm{c}} := \Omega \setminus A$, and its **characteristic function** $\mathbf{1}_A$ is defined by

$$\mathbf{1}_A(x) := \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$$

for $x \in \Omega$. One often writes $\mathbf{1}$ in place of $\mathbf{1}_\Omega$ if the reference set Ω is understood. For a sequence $(A_n)_n \subset \mathscr{P}(\Omega)$ we write $A_n \searrow A$ if

$$A_n \supset A_{n+1} \quad (n \in \mathbb{N}) \quad \text{and} \quad \bigcap_{n \in \mathbb{N}} A_n = A.$$

Similarly, $A_n \nearrow A$ is short for

$$A_n \subset A_{n+1} \quad (n \in \mathbb{N}) \quad \text{and} \quad \bigcup_{n \in \mathbb{N}} A_n = A.$$

A family $(A_\iota)_\iota \subset \mathscr{P}(\Omega)$ is called **pairwise disjoint** if $\iota \neq \eta$ implies that $A_\iota \cap A_\eta = \emptyset$. A subset $\mathcal{E} \subset \mathscr{P}(\Omega)$ is often called a **set system** on Ω. A set system is called $\cap$-stable ($\cup$-stable, $\setminus$-stable) if $A, B \in \mathcal{E}$ implies that $A \cap B$ ($A \cup B$, $A \setminus B$) belongs to $\mathcal{E}$ as well. If $\mathcal{E}$ is a set system, then any mapping $\mu : \mathcal{E} \longrightarrow [0, \infty]$ is called a (positive) **set function**. Such a set function is called **σ-additive** if

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

whenever $(A_n)_{n \in \mathbb{N}} \subset \mathcal{E}$ is pairwise disjoint and $\bigcup_n A_n \in \mathcal{E}$. Here we adopt the convention that

$$a + \infty = \infty + a = \infty \qquad (-\infty < a \leq \infty).$$

A similar rule holds for sums $a + (-\infty)$ where $a \in [-\infty, \infty)$. The sum $\infty + (-\infty)$ is not defined. Other conventions for computations with the values $\pm\infty$ are:

$$0 \cdot (\pm\infty) = (\pm\infty) \cdot 0 = 0, \quad \alpha \cdot (\pm\infty) = (\pm\infty) \cdot \alpha = \pm\infty \quad \beta \cdot (\pm\infty) = (\pm\infty) \cdot \beta = \mp\infty$$

for $\infty < \beta < 0 < \alpha$. If $f : \Omega \longrightarrow \Omega'$ is a mapping and $B \subset \Omega'$ then we denote

$$[f \in B] := f^{-1}(B) := \{x \in \Omega : f(x) \in B\}.$$

Likewise, if $P(x_1, \dots, x_n)$ is a property of n-tuples $(x_1, \dots, x_n) \in (\Omega')^n$ and $f_1, \dots, f_n : \Omega \longrightarrow \Omega'$ are mappings, then we write

$$[P(f_1, \dots f_n)] := \{x \in \Omega : P(f_1(x), \dots, f_n(x)) \text{ holds}\}.$$

E.g., for $f, g : \Omega \longrightarrow \Omega'$ we abbreviate $[f = g] := \{x \in \Omega : f(x) = g(x)\}$.

B.1 σ-Algebras

Let Ω be any set. A **σ-algebra** is a collection $\Sigma \subset \mathscr{P}(\Omega)$ of subsets of Ω, such that the following hold:

1) $\emptyset, \Omega \in \Sigma$.
2) If $A, B \in \Sigma$ then $A \cup B, A \cap B, A \setminus B \in \Sigma$.
3) If $(A_n)_{n\in\mathbb{N}} \subset \Sigma$, then $\bigcup_{n\in\mathbb{N}} A_n, \bigcap_{n\in\mathbb{N}} A_n \in \Sigma$.

If a set system Σ satisfies merely 1) and 2), it is called an **algebra**, and if Σ satisfies just 2) and $\emptyset \in \Sigma$, then it is called a **ring**. A pair (Ω, Σ) with Σ being a σ-algebra on Ω is called a **measurable space**.

An arbitrary intersection of σ-algebras over the same set Ω is again a σ-algebra. Hence for $\mathcal{E} \subset \mathscr{P}(\Omega)$ one can form

$$\sigma(\mathcal{E}) := \bigcap \{\Sigma : \mathcal{E} \subset \Sigma \subset \mathscr{P}(\Omega), \Sigma \text{ a } \sigma\text{-algebra}\},$$

the σ-algebra **generated** by $\mathcal{E}$. It is the smallest σ-algebra that contains all sets from $\mathcal{E}$. If $\Sigma = \sigma(\mathcal{E})$, we call $\mathcal{E}$ a **generator** of Σ.

If Ω is a topological space, the σ-algebra generated by all open sets is called the **Borel σ-algebra** $\mathfrak{B}(\Omega)$. By 1) and 2), $\mathfrak{B}(\Omega)$ contains all closed sets as well. A set belonging to $\mathfrak{B}(\Omega)$ is called a **Borel set** or **Borel measurable**.

Lemma B.1. *Let Ω be a topological space, and let $A \subset \Omega$ with the subspace topology. Then $\mathfrak{B}(A) = \{A \cap B : B \in \mathfrak{B}(\Omega)\}$.*

Consider the example that $\Omega = [-\infty, \infty]$ is the **extended real line**. This becomes a compact metric space via the (order-preserving) homeomorphism arctan :

$[-\infty, \infty] \longrightarrow [-\pi/2, \pi/2]$. The subspace topology of $\mathbb{R}$ coincides with its natural topology. The Borel algebra $\mathfrak{B}([-\infty, \infty])$ is generated by $\{(\alpha, \infty] : \alpha \in \mathbb{R}\}$.

A **Dynkin system** (also called λ**-system**) on a set Ω is a subset $\mathcal{D} \subset \mathscr{P}(\Omega)$ with the following properties:

1) $\Omega \in \mathcal{D}$.
2) If $A, B \in \mathcal{D}$ and $A \subset B$ then $B \setminus A \in \mathcal{D}$.
3) If $(A_n)_n \subset \mathcal{D}$ and $A_n \nearrow A$, then $A \in \mathcal{D}$.

Theorem B.2 (Dynkin). *If $\mathcal{D}$ is a Dynkin system and $\mathcal{E} \subset \mathcal{D}$ is $\cap$-stable, then $\sigma(\mathcal{E}) \subset \mathcal{D}$.*

The proof is in [Bauer (1990), p.8] and [Billingsley (1979), Thm. 3.2].

B.2 Measures

Let Ω be a set and $\Sigma \subset \mathscr{P}(\Omega)$ a σ-algebra of subsets of Ω. A (positive) **measure** is a σ-additive set function

$$\mu : \Sigma \longrightarrow [0, \infty].$$

In this case the triple (Ω, Σ, μ) is called a **measure space** and the sets in Σ are called **measurable sets**. If $\mu(\Omega) < \infty$, the measure is called **finite**. If $\mu(\Omega) = 1$, it is called a **probability measure** and (Ω, Σ, μ) is called a **probability space**. Suppose $\mathcal{E} \subset \Sigma$ is given and there is a sequence $(A_n)_n \subset \mathcal{E}$ such that

$$\mu(A_n) < \infty \quad (n \in \mathbb{N}) \quad \text{and} \quad \Omega = \bigcup_n A_n;$$

then the measure μ is called σ**-finite with respect to** $\mathcal{E}$. If $\mathcal{E} = \Sigma$, we simply call it σ-finite.

From the σ-additivity of the measure one derives the following properties:

a) **(Finite Additivity)** $\mu(\emptyset) = 0$ and

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B) \qquad (A, B \in \Sigma).$$

b) **(Monotonicity)** $A, B \in \Sigma, \quad A \subset B \quad \Longrightarrow \quad \mu(A) \le \mu(B)$.

c) **(σ-Subadditivity)** $(A_n)_n \subset \Sigma \quad \Longrightarrow \quad \mu(\bigcup_{n \in \mathbb{N}} A_n) \le \sum_{n=1}^{\infty} \mu(A_n)$.

See [Billingsley (1979), p.134] for the elementary proofs.

An application of Dynkin's theorem yields the uniqueness theorem.

Theorem B.3 (Uniqueness Theorem). [Billingsley (1979), Thm. 10.3]
Let $\Sigma = \sigma(\mathcal{E})$ with $\mathcal{E}$ being $\cap$-stable. Let μ, ν be two measures on Σ, both σ-finite with respect to $\mathcal{E}$. If μ and ν coincide on $\mathcal{E}$, they are equal.

B.3 Construction of Measures

An **outer measure** on a set Ω is a mapping

$$\mu^* : \mathscr{P}(\Omega) \longrightarrow [0,\infty]$$

such that $\mu^*(\emptyset) = 0$ and μ^* is monotone and σ-subadditive.

Theorem B.4 (Carathéodory). [Billingsley (1979), Thm. 11.1] *Let μ^* be an outer measure on the set Ω. Define*

$$\mathscr{M}(\mu^*) := \{E \subset \Omega \,:\, \mu^*(A) = \mu^*(A \cap E) + \mu^*(A \setminus E) \quad \forall A \subset \Omega\}.$$

Then $\mathscr{M}(\mu^*)$ is a σ-algebra and $\mu^*\big|_{\mathscr{M}(\mu^*)}$ is a measure on it.

The set system $\mathcal{E} \subset \mathscr{P}(\Omega)$ is called a **semi-ring** if it satisfies the following two conditions:

1) $\mathcal{E}$ is $\cap$-stable and $\emptyset \in \mathcal{E}$.
2) If $A, B \in \mathcal{E}$ then $A \setminus B$ is a disjoint union of members of $\mathcal{E}$.

An example of such a system is $\mathcal{E} = \{(a,b] \,:\, a \le b\} \subset \mathscr{P}(\mathbb{R})$. If $\mathcal{E}$ is a semi-ring then the system of all disjoint unions of members of $\mathcal{E}$ is a ring.

Theorem B.5 (Hahn). [Billingsley (1979), p.140] *Let $\mathcal{E}$ be a semi-ring on a set Ω and let $\mu : \mathcal{E} \longrightarrow [0,\infty]$ be σ-additive on $\mathcal{E}$. Then $\mu^* : \mathscr{P}(\Omega) \longrightarrow [0,\infty]$ defined by*

$$\mu^*(A) := \inf\Big\{ \sum\nolimits_{n\in\mathbb{N}} \mu(E_n) \,:\, (E_n)_n \subset \mathcal{E},\, A \subset \bigcup\nolimits_n E_n \Big\} \qquad (A \in \mathscr{P}(\Omega))$$

is an outer measure. Moreover, $\sigma(\mathcal{E}) \subset \mathscr{M}(\mu^*)$ and $\mu^*|_{\mathcal{E}} = \mu$.

One may summarise these results in the following way: if a set function on a semi-ring $\mathcal{E}$ is σ-additive on $\mathcal{E}$ then it has a extension to a measure on $\sigma(\mathcal{E})$. If in addition Ω is σ-finite with respect to $\mathcal{E}$, then this extension is unique.

Sometimes, for instance in the construction of infinite products, it is convenient to work with the following criterion.

Lemma B.6. [Billingsley (1979), Thm. 10.2] *Let $\mathcal{E}$ be an algebra on a set Ω, and let $\mu : \mathcal{E} \longrightarrow [0,\infty)$ be a finitely additive set function with $\mu(\Omega) < \infty$. Then μ is σ-additive on $\mathcal{E}$ if and only if for each decreasing sequence $(A_n)_n \subset \mathcal{E}$, $A_n \searrow \emptyset$, one has $\mu(A_n) \to 0$.*

B.4 Measurable Functions and Mappings

Let (Ω, Σ) and (Ω', Σ') be measurable spaces. A mapping $\varphi : \Omega \longrightarrow \Omega'$ is called **measurable** if

$$[\varphi \in A] \in \Sigma \qquad (A \in \Sigma').$$

(It suffices to check this condition for each A from a generator of Σ'.) We denote by

$$\mathfrak{M}(\Omega;\Omega') = \mathfrak{M}(\Omega,\Sigma;\Omega,\Sigma')$$

the set of all measurable mappings between Ω and Ω'. For the special case $\Omega' = [0,\infty]$ we write

$$\mathfrak{M}_+(\Omega) := \{f : \Omega \longrightarrow [0,\infty] \,:\, f \text{ is measurable}\}.$$

Example: For $A \in \Sigma$ its characteristic function $\mathbf{1}_A$ is measurable, since one has $[\mathbf{1}_A \in B] = \emptyset, A, A^c, \Omega$, depending on whether or not 0 respectively 1 is contained in B.

Example: If Ω, Ω' are topological spaces and $\varphi : \Omega \longrightarrow \Omega'$ is continuous, then it is $\mathfrak{B}(\Omega) - \mathfrak{B}(\Omega')$ measurable.

Lemma B.7. [Lang (1993), p.117] *Let Ω' be a metric space and $\Sigma = \mathfrak{B}(\Omega')$ its Borel algebra. If $\varphi_n : \Omega \longrightarrow \Omega'$ is measurable for each $n \in \mathbb{N}$ and $\varphi_n \to \varphi$ pointwise, then φ is measurable as well.*

The following lemma summarises the basic properties of positive measurable functions.

Lemma B.8. [Billingsley (1979), Section 13] *Let (Ω,Σ,μ) be a measure space. Then the following assertions hold.*

a) *If $f,g \in \mathfrak{M}_+(\Omega), \alpha \geq 0$, then $fg, f+g, \alpha f \in \mathfrak{M}_+(\Omega)$.*
b) *If $f,g \in \mathfrak{M}(\Omega;\mathbb{R})$ and $\alpha,\beta \in \mathbb{R}$, then $fg, \alpha f + \beta g \in \mathfrak{M}(\Omega;\mathbb{R})$.*
c) *$f,g : \Omega \longrightarrow [-\infty,\infty]$ are measurable then $-f, \min\{f,g\}, \max\{f,g\}$ are measurable.*
d) *If $f_n : \Omega \longrightarrow [-\infty,\infty]$ is measurable for each $n \in \mathbb{N}$ then $\sup_n f_n, \inf_n f_n$ are measurable.*

A **simple function** on a measure space (Ω,Σ,μ) is a linear combination of characteristic functions of measurable sets. Positive measurable functions can be approximated by simple functions:

Lemma B.9. [Billingsley (1979), Thm. 13.5] *Let $f : \Omega \longrightarrow [0,\infty]$ be measurable. Then there exists a sequence of simple functions $(f_n)_n$ such that*

$$0 \leq f_n \leq f_{n+1} \nearrow f \qquad (\textit{pointwise as } n \to \infty).$$

If f is bounded, the convergence is uniform.

B.5 The Integral of Positive Measurable Functions

Given a measure space (Ω, Σ, μ) and a positive simple function

$$f = \sum_{j=1}^{n} \alpha_j \mathbf{1}_{A_j}$$

on Ω, one defines its **integral** by

$$\int_\Omega f \, \mathrm{d}\mu := \sum_{j=1}^{n} \alpha_j \mu(A_j).$$

By using common refinements one can show that this definition is independent of the actual representation of f as a linear combination of characteristic functions. For a general $f \in \mathfrak{M}_+(\Omega)$ one defines

$$\int_\Omega f \, \mathrm{d}\mu := \lim_n \int_\Omega f_n \, \mathrm{d}\mu$$

where $(f_n)_n$ is an arbitrary sequence of simple functions with $0 \le f_n \nearrow f$ pointwise. (This is the way of [Bauer (1990), Chapter 11] and [Rana (2002), Section 5.2]; [Billingsley (1979), Section 15] takes a similar, but slightly different route.)

Theorem B.10. *The integral for positive measurable functions has the following properties.*

a) **(Action on Characteristic Functions)** *($A \in \Sigma$)*

$$\int_\Omega \mathbf{1}_A \, \mathrm{d}\mu = \mu(A).$$

b) **(Additivity and homogeneity)** *($f, g \in \mathfrak{M}_+(\Omega), \alpha \ge 0$)*

$$\int_\Omega (f + \alpha g) \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu + \alpha \int_\Omega g \, \mathrm{d}\mu.$$

c) **(Monotonicity)** *($f, g \in \mathfrak{M}_+(\Omega)$)*

$$f \le g \quad \Rightarrow \quad \int_\Omega f \, \mathrm{d}\mu \le \int_\Omega g \, \mathrm{d}\mu.$$

d) **(Beppo Levi, Monotone Convergence Theorem)** *Let $(f_n)_{n\in\mathbb{N}} \subset \mathfrak{M}_+(\Omega)$ such that $0 \le f_1 \le f_2 \le \dots$ and $f_n \to f$ pointwise, then*

$$\int_\Omega f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu.$$

e) **(Fatou's lemma)** *Let $(f_n)_{n\in\mathbb{N}} \subset \mathfrak{M}_+(\Omega)$. Then*

$$\int_\Omega \liminf_{n\to\infty} f_n \,\mathrm{d}\mu \le \liminf_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu.$$

Let $1 \le p \le \infty$. Then its **dual exponent** is the unique number $q = p' \in [1,\infty]$ such that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Theorem B.11 (Hölder's Inequality). *Let (Ω,Σ,μ) be a measure space, let $f,g \in \mathfrak{M}_+(\Omega)$, and let $1 < p < \infty$ with dual exponent q. Then $fg, f^p, g^q \in \mathfrak{M}_+(\Omega)$ as well and*

$$\int_\Omega fg \,\mathrm{d}\mu \le \left(\int_\Omega f^p \,\mathrm{d}\mu\right)^{1/p} \left(\int_\Omega g^q \,\mathrm{d}\mu\right)^{1/q}.$$

See [Haase (2007)] for a nice proof.

B.6 Push-forward Measures and Measures with Density

If (Ω,Σ,μ) is a measure space, (Ω',Σ') is a measurable space and $\varphi : \Omega \longrightarrow \Omega'$ is measurable, then a measure is defined on Σ' by

$$[\varphi_*\mu](B) := \mu[\varphi \in B] \qquad (B \in \Sigma).$$

The measure $\varphi_*\mu$ is called the **image** of μ under φ, or the **push-forward** of μ along φ. If μ is finite or a probability measure, so is $\varphi_*\mu$. If $f \in \mathfrak{M}_+(\Omega')$ then

$$\int_{\Omega'} f \,\mathrm{d}(\varphi_*\mu) = \int_\Omega (f\circ\varphi)\,\mathrm{d}\mu.$$

Let (Ω,Σ,μ) be a measure space and $f \in \mathfrak{M}_+(\Omega)$. Then by

$$(f\mu)(A) := \int_A f \,\mathrm{d}\mu := \int_\Omega \mathbf{1}_A f \,\mathrm{d}\mu \qquad (A \in \Sigma)$$

a new measure $f\mu$ on Σ is defined. We call f the **density function** of $f\mu$. One has

$$\int_\Omega g \,\mathrm{d}(f\mu) = \int_\Omega gf \,\mathrm{d}\mu.$$

for all $g \in \mathfrak{M}_+(\Omega')$. [Billingsley (1979), Thm. 16.10 and 16.12].

Let μ,ν be two measures on Σ. We say that ν is **absolutely continuous** with respect to μ, written $\nu \ll \mu$, if $A \in \Sigma$, $\mu(A) = 0$ implies $\nu(A) = 0$. Clearly, if $\nu = f\mu$ with a density f, then ν is absolutely continuous with respect to μ. The converse is true under σ-finiteness conditions.

Theorem B.12 (Radon–Nikodym I). *Let (Ω,Σ,μ) be a σ-finite measure space, and let ν be a σ-finite measure on Σ, absolutely continuous with respect to μ. Then there is $f \in \mathfrak{M}_+(\Omega)$ such that $\nu = f\mu$.*

In [Billingsley (1979), Thm. 32.2] and [Bauer (1990), Satz 17.10] the proof is based on the so-called "Hahn decomposition" of signed measures; the Hilbert space approach of von Neumann is reproduced in [Rudin (1987), 6.10].

B.7 Product Spaces

If (Ω_1, Σ_1) and (Ω_2, Σ_2) are measurable spaces, then on the product space $\Omega_1 \times \Omega_2$ we define the **product σ-algebra**

$$\Sigma_1 \otimes \Sigma_2 := \sigma\{A \times B : A \in \Sigma_1, B \in \Sigma_2\}.$$

If $\mathcal{E}_j$ is a generator of Σ_j with $\Omega_j \in \mathcal{E}_j$ for $j = 1, 2$, then

$$\mathcal{E}_1 \times \mathcal{E}_2 := \{A \times B : A \in \mathcal{E}_1, B \in \mathcal{E}_2\}$$

is a generator of $\Sigma_1 \otimes \Sigma_2$. As a consequence we obtain:

Lemma B.13. *Let Ω_1, Ω_2 be second countable topological (e.g., separable metric) spaces. Then*

$$\mathfrak{B}(\Omega_1 \otimes \Omega_2) = \mathfrak{B}(\Omega_1) \otimes \mathfrak{B}(\Omega_2).$$

If (Ω, Σ) is another measurable space, then a mapping $f = (f_1, f_2) : \Omega \longrightarrow \Omega_1 \times \Omega_2$ is measurable if and only if the projections $f_1 = \pi_1 \circ f, f_2 = \pi_2 \circ f$ are both measurable.

If $f : (\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2) \longrightarrow (\Omega', \Sigma')$ is measurable, then $f(x, \cdot) : \Omega_2 \longrightarrow \Omega'$ is measurable, for every $x \in \Omega_1$, see [Billingsley (1979), Theorem 18.1].

Theorem B.14 (Tonelli). [Billingsley (1979), Theorem 18.3] *Let $(\Omega_j, \Sigma_j, \mu_j)$, $j = 1, 2$, be σ-finite measure spaces and $f \in \mathfrak{M}_+(\Omega_1 \times \Omega_2)$. Then the functions*

$$\begin{aligned} F_1 : &\Omega_1 \longrightarrow [0, \infty], \quad x \longmapsto \int_{\Omega_2} f(x, \cdot) \, \mathrm{d}\mu_2 \\ F_2 : &\Omega_2 \longrightarrow [0, \infty], \quad y \longmapsto \int_{\Omega_1} f(\cdot, y) \, \mathrm{d}\mu_1 \end{aligned}$$

are measurable and there is a unique measure $\mu_1 \otimes \mu_2$ such that

$$\int_{\Omega_1} F_1 \, \mathrm{d}\mu_1 = \int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}(\mu_1 \otimes \mu_2) = \int_{\Omega_2} F_2 \, \mathrm{d}\mu_2.$$

The measure $\mu_1 \otimes \mu_2$ is called the **product measure** of μ_1, μ_2. Note that for the particular case $F = f_1 \otimes f_2$, with

$$(f_1 \otimes f_2)(x_1, x_2) := f_1(x_1) \cdot f_2(x_2) \qquad (f_j \in \mathfrak{M}_+(\Omega_j), x_j \in \Omega_j \quad (j = 1, 2)),$$

we obtain

$$\int_{\Omega_1\times\Omega_2} (f_1\otimes f_2)\,\mathrm{d}(\mu_1\otimes\mu_2) = \left(\int_{\Omega_1} f_1\,\mathrm{d}\mu_1\right)\left(\int_{\Omega_2} f_2\,\mathrm{d}\mu_2\right).$$

Infinite Products and Ionescu Tulcea's Theorem

For a measurable space (Ω,Σ) we denote by $\mathrm{M}^+(\Omega,\Sigma)$ the set of all positive and by $\mathrm{M}^1(\Omega,\Sigma)$ the set of all probability measures on (Ω,Σ). There is a natural σ-algebra $\tilde{\Sigma}$ on $\mathrm{M}^+(\Omega,\Sigma)$, the smallest such that each mapping

$$\mathrm{M}^+(\Omega,\Sigma)\longrightarrow[0,\infty],\quad \nu\longmapsto\nu(A)\qquad (A\in\Sigma)$$

is measurable.

Let (Ω_j,Σ_j), $j=1,2$ be measurable spaces. A **measure kernel** from Ω_1 to Ω_2 is a measurable mapping $\mu:\Omega_2\longrightarrow\mathrm{M}^+(\Omega_1,\Sigma_1)$. Such a kernel μ can also be interpreted as a mapping of two variables

$$\mu:\Omega_2\times\Sigma\longrightarrow[0,\infty],$$

and we shall do so when it seems convenient. If $\mu(y,\cdot)\in\mathrm{M}^1(\Omega_1,\Sigma_1)$ for each $y\in\Omega_2$ then μ is called a **probability kernel**.

Let (Ω,Σ) be another measurable space, and let $\mu:\Omega_2\longrightarrow\mathrm{M}^+(\Omega_1,\Sigma_1)$ be a kernel. Then there is an induced operator

$$T_\mu:\mathfrak{M}_+(\Omega\times\Omega_1)\longrightarrow\mathfrak{M}_+(\Omega\times\Omega_2)$$
$$(T_\mu f)(x,x_2):=\int_{\Omega_1} f(x,x_1)\,\mu(x_2,\mathrm{d}x_1)\quad (x\in\Omega,x_2\in\Omega_2).$$

The operator T_μ is additive and positively homogeneous, and if $f_n\nearrow f$ pointwise on Ω_1 then $T_\mu f_n\nearrow T_\mu f$ pointwise on Ω_2. Moreover,

$$T_\mu(f\otimes g)=(f\otimes\mathbf{1})\cdot T_\mu(\mathbf{1}\otimes g)\qquad (f\in\mathfrak{M}_+(\Omega),g\in\mathfrak{M}_+(\Omega_2)).$$

Conversely, each operator $T:\mathfrak{M}_+(\Omega\times\Omega_1)\longrightarrow\mathfrak{M}_+(\Omega\times\Omega_2)$ with these properties is of the form T_μ, for some kernel μ.

If $\mu:\Omega_2\longrightarrow\mathrm{M}^+(\Omega_1)$ and $\nu:\Omega_3\longrightarrow\mathrm{M}^+(\Omega_2)$ are kernels, then $T_\nu\circ T_\mu=T_\eta$ for

$$\eta(x_3,A):=\int_{\Omega_2}\mu(x_2,A)\nu(x_3,\mathrm{d}x_2)\qquad (x_3\in\Omega_3,A\in\Sigma_1).$$

Kernels can be used to construct measures on infinite products. Let (Ω_n,Σ_n), $n\in\mathbb{N}$, be measurable spaces, and let $\Omega:=\prod_{n\in\mathbb{N}}\Omega_n$ be the Cartesian product, with the projections $\pi_n:\Omega\longrightarrow\Omega_n$. The natural σ-algebra on Ω is

$$\bigotimes_n\Sigma_n:=\sigma\{\pi_n^{-1}(A_n)\,:\,n\in\mathbb{N},A_n\in\Sigma_n\}.$$

A generating algebra is

$$\mathscr{A} := \Big\{ A_n \times \prod_{k>n} \Omega_k \, : \, n \in \mathbb{N}, A_n \in \Sigma_1 \otimes \ldots \otimes \Sigma_n \Big\},$$

the algebra of **cylinder sets**.

Theorem B.15 (Ionescu Tulcea). [Ethier and Kurtz (1986), p.504] *Let (Ω_n, Σ_n), $n \in \mathbb{N}$, be measurable spaces, let*

$$\mu_n : \Omega_1 \times \cdots \times \Omega_{n-1} \longrightarrow \mathrm{M}^1(\Omega_n) \qquad (n \in \mathbb{N}, n \geq 2)$$

be probability kernels, and let μ_1 be a probability measure on Ω_1. Let

$$X^{(n)} := \Omega_1 \times \cdots \times \Omega_n \qquad \textit{with} \qquad \Sigma^{(n)} = \Sigma_1 \otimes \ldots \otimes \Sigma_n.$$

Let, for $n \geq 1$, $\quad T_n : \mathfrak{M}_+(X^{(n)}, \Sigma^{(n)}) \longrightarrow \mathfrak{M}_+(X^{(n-1)}, \Sigma^{(n-1)})$ $\quad$ *be given by*

$$(T_n f)(x^{(n-1)}) = \int_{\Omega_n} f(x^{(n-1)}, x_n) \, \mu_n(x^{(n-1)}, \mathrm{d}x_n) \qquad (x^{(n-1)} \in X^{(n-1)}).$$

Then there is a unique probability measure ν on $X^{(\infty)} := \prod_{n \in \mathbb{N}} \Omega_n$ such that

$$\int_{X^{(\infty)}} f(x_1, \ldots, x_n) \, \mathrm{d}\nu(x_1, \ldots) = T_1 T_2 \ldots T_n f \qquad (f \in \mathfrak{M}_+(\Omega_1 \times \cdots \times \Omega_n))$$

for every $n \in \mathbb{N}$.

An important special case of the Ionescu Tulcea theorem is the construction of the infinite product measure. Here one has a probability measure ν_n on (Ω_n, Σ_n), for each $n \in \mathbb{N}$. If one applies the Ionescu Tulcea theorem with $\mu_n \equiv \nu_n$, then the ν of the theorem satisfies

$$(\pi_1, \ldots, \pi_n)_* \nu = \nu_1 \otimes \ldots \otimes \nu_n \qquad (n \in \mathbb{N}).$$

We write $\nu := \bigotimes_n \nu_n$ and call it the **product** of the ν_n. For products of uncountably many probability spaces see [Hewitt and Stromberg (1969), Chapter 22].

B.8 Null Sets

Let (Ω, Σ, μ) be a measure space. A set $A \subset \Omega$ is called a **null set** if there is a set $N \in \Sigma$ such that $A \subset N$ and $\mu(N) = 0$. (In general a null set need not be measurable). Null sets have the following properties:

a) If A is a null set and $B \subset A$ then B is also a null set.
b) If A_n is a null set $(n \in \mathbb{N})$, then $\bigcup_n A_n$ is a null set.

Lemma B.16. [Billingsley (1979), Theorem 15.2] *Let (Ω, Σ, μ) be a measure space and let $f : \Omega \longrightarrow [-\infty, \infty]$ be measurable. Then the following assertions hold.*

a) $\int_\Omega |f| \, \mathrm{d}\mu = 0$ *if and only if the set* $[\, f \neq 0\,] = [\,|f| > 0\,]$ *is a null set.*
b) *If* $\int_\Omega |f| \, \mathrm{d}\mu < \infty$*, then the set* $[\,|f| = \infty\,]$ *is a null set.*

One says that two functions f, g are equal **μ-almost everywhere** (abbreviated by "$f = g$ a.e." or "$f \sim_\mu g$") if the set $[\, f \neq g\,]$ is a null set. More generally, let P be a property of points of Ω. Then P is said to hold **almost everywhere** or for **μ-almost all** $x \in \Omega$ if the set

$$\{x \in \Omega \,:\, P \text{ does not hold for } x\}$$

is a μ-null set. If μ is understood, we leave out the reference to it.

For each set Ω', the relation $\sim_\mu$ ("is equal μ-almost everywhere to") is an equivalence relation on the space of mappings from Ω to Ω'. For such a mapping f we sometimes denote by $[f]$ its equivalence class, in situations when notational clarity is needed. If μ is understood, we write simply $\sim$ instead of $\sim_\mu$. By choosing $\Omega = \{0,1\}$ an equivalence relation on Σ is induced via

$$A \sim B \quad \stackrel{\text{Def.}}{\iff} \quad \mathbf{1}_A = \mathbf{1}_B \quad \mu\text{-a.e.} \quad \iff \quad \mu(A \triangle B) = 0.$$

The space of equivalence classes $\Sigma/\!\sim$ is called the **measure algebra**. For a set $A \in \Sigma$ we sometimes write $[\,A\,]$ for its equivalence class with respect to $\sim$, but usually we omit the brackets and simply write A again. Clearly, if $f = g$ μ-a.e. then $[\, f \in B\,] \sim [\, g \in B\,]$ for every $B \subset \Omega'$. The usual set-theoretic operations can be induced on the elements of $\Sigma/\!\sim$ by setting

$$[A] \cap [B] := [A \cap B], \quad \bigcup\nolimits_n [A_n] := \big[\bigcup\nolimits_n A_n\big] \; \ldots.$$

Also, one defines

$$\mu[A] := \mu(A) = \int_\Omega \mathbf{1}_A \, \mathrm{d}\mu \qquad (A \in \Sigma)$$

and writes $\emptyset := [\emptyset]$ again. Hence on the measure algebra, $\mu(A) = 0$ if and only if $A = \emptyset$.

B.9 Convergence in Measure

Let (Ω, Σ, μ) be a σ-finite measure space and (X, d) a complete metric space, with its Borel σ-algebra. Let

$$\mathfrak{M}_s(\Omega, \Sigma; X) := \{ f \in \mathfrak{M}(\Omega; X) \,:\, f(\Omega) \text{ is separable}\}.$$

Note that $\mathfrak{M}_s(\Omega; X) = \mathfrak{M}(\Omega; X)$ if X is separable. Choose a complete metric d on X such that d induces the topology and such that d is uniformly bounded. (For example, if d is *any* complete metric inducing the topology, one can replace d by $d/(d+1)$ to obtain an equiavlent metric which is also bounded.) Using Lemma B.13 one sees that the mapping

$$d(f,g) : \Omega \times \Omega \longrightarrow [0,\infty), \quad (x,y) \longmapsto d(f(x),g(y))$$

is product measurable. For a fixed $A \in \Sigma$ with $\mu(A) < \infty$ we define a semi-metric on $\mathfrak{M}_s(\Omega;X)$ by

$$d_A(f,g) := \int_A d(f,g)\,\mathrm{d}\mu \qquad (f,g \in \mathfrak{M}_s(\Omega;X)).$$

Clearly $d_A(f,g) = 0$ if and only if $f = g$ almost everywhere on A. One has $f_n \to f$ with respect to d_A if and only if

$$\mu\big([d(f_n,f) > \varepsilon] \cap A\big) \to 0 \quad \text{for each } \varepsilon > 0.$$

Convergence in d_Ω is called **convergence globally in measure**.

Let (Ω,Σ,μ) be a σ-finite measure space, and choose $\Omega_n \in \Sigma$ of finite measure and such that $\Omega = \bigcup_n \Omega_n$. Let

$$D(f,g) := \sum\nolimits_{n=1}^{\infty} 2^{-n} d_{\Omega_n}(f,g) \qquad (f,g \in \mathfrak{M}_s(\Omega;X)).$$

Then D is a semi-metric on $\mathfrak{M}_s(\Omega;X)$. The convergence with respect to D is called **convergence (locally) in measure**. Note that $D = d_\Omega$ if μ is finite.

Theorem B.17. *Let (Ω,Σ,μ) be a σ-finite measure space and X a completely metrizable space.*

a) *The semi-metric D on $\mathfrak{M}_s(\Omega;X)$ is complete.*
b) *$D(f,g) = 0$ if and only if $f = g$ μ-almost everywhere.*
c) *$f_n \to f$ in measure if and only if every subsequence of $(f_n)_n$ has a subsequence which converges to f pointwise almost everywhere.*
d) *$D(f_n,f) \to 0$ if and only if $d_A(f_n,f) \to 0$ for all $A \in \Sigma$, $\mu(A) < \infty$.*

Note that c) shows that a choice of an equivalent (complete bounded) metric on E leads to an equivalent semi-metric on $\mathfrak{M}_s(\Omega;E)$. We do not know of a good reference for Theorem B.17. In [Bauer (1990), Chap. 20] one finds all decisive details, although formulated for the case $E = \mathbb{R}$. The case of a probability space is treated in [Kallenberg (2002), Lemmas 4.2 and 4.6].

Theorem B.18 (Egoroff). [Rana (2002), 8.2.4] *Let (Ω,Σ,μ) be a finite measure space and X a complete metric space. Let $(f_n)_n \subset \mathfrak{M}(\Omega;X)$ and $f : \Omega \longrightarrow X$. Then $f_n \to f$ pointwise almost everywhere if and only if for each $\varepsilon > 0$ there is $A \in \Sigma$ with $\mu(A^{\mathrm{c}}) < \varepsilon$ and $f_n \to f$ uniformly on A.*

We denote by

$$\mathrm{L}^0(\Omega;X) := \mathrm{L}^0(\Omega,\Sigma,\mu;X) := \mathfrak{M}_s(\Omega;X)/\!\sim$$

the space of equivalence classes of measurable, separably-valued mappings modulo equality almost everywhere. By a) and b) of the theorem above, D induces a complete metric on $\mathrm{L}^0(\Omega;X)$.

By restricting to characteristic functions, i.e., to the case $X = \{0,1\}$, this induces a (complete!) metric on the measure algebra $\Sigma/\sim$. If $\mu(\Omega) = 1$, this metric is given by

$$d([A],[B]) = d_\Omega(\mathbf{1}_A, \mathbf{1}_B) = \mu(A \triangle B) \qquad (A, B \in \Sigma).$$

B.10 The Lebesgue-Bochner Spaces

Let (Ω, Σ, μ) be a σ-finite measure space and X be a Banach space with norm $\|\cdot\|_X$. Then $\mathrm{L}^0(\Omega;X)$ is an F-space, i.e., a topological vector space, completely metrisable by a translation invariant metric. A function $f : \Omega \longrightarrow X$ is called a **step function** if it is of the form

$$f = \sum_{j=1}^{n} \mathbf{1}_{A_j} \otimes x_j = \sum_{j=1}^{n} \mathbf{1}_{A_j}(\cdot) x_j$$

for some finitely many $x_j \in X$, $A_j \in \Sigma$, $\mu(A_j) < \infty$ $(j = 1, \dots, n)$. We denote by

$$\mathrm{St}(\Omega;X) := \mathrm{lin}\{\mathbf{1}_A \otimes x : x \in X,\ A \in \Sigma,\ \mu(A) < \infty\}$$

the space of all X-valued step functions. An X-valued function is called **μ-measurable** if there is a sequence of step functions converging to f pointwise μ-almost everywhere.

Lemma B.19. [Lang (1993), pp. 124 and 142] *Let (Ω, Σ, μ) be a σ-finite measure space, let X be a Banach space, and let $f : \Omega \longrightarrow X$ be a mapping. Then $[f] \in \mathrm{L}^0(\Omega;X)$ if and only if f is μ-measurable, if and only if there is a sequence $(f_n)_n \subset \mathrm{St}(\Omega;X)$ of step functions such that $f_n \to f$ a.e. and $\|f_n(\cdot)\|_X \le 2\|f(\cdot)\|_X$ a.e., for all $n \in \mathbb{N}$.*

A consequence of this lemma together with Theorem B.17 is that $\mathrm{St}(\Omega;X)$ is dense in the complete metric space $\mathrm{L}^0(\Omega;X)$.

For $f \in \mathrm{L}^0(\Omega;X)$ we define

$$\|f\|_\infty := \inf\{t > 0 : \mu[\,\|f(\cdot)\|_X > t\,] = 0\}$$

and we set

$$\mathrm{L}^\infty(\Omega;X) := \mathrm{L}^\infty(\Omega, \Sigma, \mu; X) := \{f \in \mathrm{L}^0(\Omega;X) : \|f\|_\infty < \infty\}.$$

Then $\|\cdot\|_\infty$ defines a complete norm on $\mathrm{L}^\infty(\Omega;X)$. We simply write $\mathrm{L}^\infty(\Omega)$ when we deal with scalar-valued functions.

Let $1 \le p < \infty$. For $f \in \mathrm{L}^0(\Omega;X)$ we define

$$\|f\|_p := \left(\int_\Omega \|f(\cdot)\|_X^p \, \mathrm{d}\mu \right)^{\frac{1}{p}}$$

and $\mathrm{L}^p(\Omega;X) := \mathrm{L}^p(\Omega,\Sigma,\mu;X) := \{f \in \mathrm{L}^0(\Omega;X) : \|f\|_p < \infty\}$. We simply write $\mathrm{L}^p(\Omega)$ when dealing with scalar-valued functions.

Theorem B.20. *Let (Ω,Σ,μ) be a σ-finite measure space, X a Banach space and $1 \le p < \infty$. Then the following assertions hold.*

a) *$\|\cdot\|_p$ is a complete norm on $\mathrm{L}^p(\Omega;X)$.*

b) *The embedding $\mathrm{L}^p(\Omega;X) \subset \mathrm{L}^0(\Omega;X)$ is continuous.*

c) *If $f_n \to f$ in $\mathrm{L}^p(\Omega;X)$ then there is $g \in \mathrm{L}^p(\Omega;\mathbb{R})$ and a subsequence $(f_{n_k})_k$ such that $\|f_{n_k}(\cdot)\|_X \le g$ a.e., for all $k \in \mathbb{N}$, and $f_{n_k} \to f$ pointwise a.e..*

d) *$\mathrm{St}(\Omega;X)$ is dense in $\mathrm{L}^p(\Omega;X)$.*

e) **(LDC)** *If $(f_n)_n \subset \mathrm{L}^p(\Omega;X)$ $f_n \to f$ in measure and there is $g \in \mathrm{L}^p(\Omega;\mathbb{R})$ such that $\|f_n(\cdot)\|_X \le g$ a.e., for all $n \in \mathbb{N}$, then $f \in \mathrm{L}^p(\Omega;X)$, and $\|f_n - f\|_p \to 0$.*

The abbreviation “LDC” stands for *L*ebesgue’s *D*ominated *C*onvergence theorem.

The (Bochner-)Integral

We want to integrate functions from $\mathrm{L}^1(\Omega,\Sigma,\mu;X)$. In the case $X = \mathbb{C}$ one can use the already defined integral for positive measurable functions, and this is how it is done in most of the textbooks. However, this does not work for Banach space-valued functions. Therefore we take a different route and shall see eventually that in the case $X = \mathbb{C}$ we recover the common definition.

For a step function $f = \sum_{j=1}^n \mathbf{1}_{A_j} \otimes x_j$ we define its **integral** by

$$\int_\Omega f \,\mathrm{d}\mu := \sum_{j=1}^n \mu(A_j) x_j.$$

This is independent of the representation of f and hence defines a linear mapping

$$\left[f \longmapsto \int_\Omega f \,\mathrm{d}\mu\right] : \mathrm{St}(\Omega;X) \longrightarrow X.$$

Since obviously

$$\Big\| \int_\Omega f \mathrm{d}\mu \Big\|_X \le \int_\Omega \|f(\cdot)\|_X \,\mathrm{d}\mu = \|f\|_1 \qquad (f \in \mathrm{St}(\Omega;X)),$$

this mapping can be extended by continuity to all of $\mathrm{L}^1(\Omega;X)$ to a linear contraction

$$\left[f \longmapsto \int_\Omega f \,\mathrm{d}\mu\right] : \mathrm{L}^1(\Omega;X) \longrightarrow X.$$

It is easy to see that for $f : \Omega \longrightarrow [0,\infty)$ this definition of the integral and the one for positive measurable functions coincide. This shows that for complex-valued functions our definition of the integral leads to the same as the one usually given in more elementary treatments.

If Y is another Banach space and $T : X \longrightarrow Y$ is a bounded linear mapping, then

$$\int_\Omega (T \circ f)\, \mathrm{d}\mu = T \int_\Omega f \,\mathrm{d}\mu \qquad (f \in \mathrm{L}^1(\Omega; X)).$$

Applying linear functionals yields

$$\Big\| \int_\Omega f \,\mathrm{d}\mu \Big\|_X \le \int_\Omega \|f(\cdot)\|_X \,\mathrm{d}\mu \qquad (f \in \mathrm{L}^1(\Omega; X)).$$

Theorem B.21 (Averaging Theorem). [Lang (1993), Thm. 5.15] *Let $S \subset X$ be a closed subset, and let $f \in \mathrm{L}^1(\Omega; X)$. If*

$$\frac{1}{\mu(A)} \int_A f \,\mathrm{d}\mu \in S$$

for all $A \in \Sigma$ such that $0 < \mu(A) < \infty$, then $f \in S$ almost everywhere.

As a corollary one obtains that if $\int_A f = 0$ for all A with finite measure, then $f = 0$ almost everywhere.

B.11 Approximations

Let (Ω, Σ, μ) be a measure space. Directly from Lemma B.9 we see that the set of simple functions is dense in $\mathrm{L}^\infty(\Omega, \Sigma, \mu; \mathbb{R})$, and we know already that $\mathrm{St}(\Omega, \Sigma; X)$ is dense in $\mathrm{L}^p(\Omega; X)$ if X is a Banach space and $p < \infty$. Here we are interested in more refined statements, involving step functions

$$\mathrm{St}(\Omega, \mathcal{E}; X) := \mathrm{lin}\{\mathbf{1}_B \otimes x : B \in \mathcal{E}, x \in X\}$$

with respect to a generator $\mathcal{E}$ of Σ.

Lemma B.22. [Billingsley (1979), Thm. 11.4] *Let $\mathcal{E} \subset \Sigma$ be a ring with $\sigma(\mathcal{E}) = \Sigma$. Fix $C \in \mathcal{E}$ with $\mu(C) < \infty$ and define*

$$\mathcal{E}_C := \{B \in \mathcal{E} : B \subset C\} = \{B \cap C : B \in \mathcal{E}\}.$$

Then for each $A \in \Sigma$ and each $\varepsilon > 0$ there is $B \in \mathcal{E}_C$ such that $\mu((A \cap C) \triangle B) < \varepsilon$.

Based on the lemma, one can prove the following.

Theorem B.23. *Let (Ω, Σ, μ) be a measure space and let $\mathcal{E} \subset \Sigma$ be a ring that generates Σ and consists exclusively of sets of finite measure. Furthermore, suppose that Ω is σ-finite with respect to $\mathcal{E}$. Then the following assertions hold.*

a) *$\{[B] : B \in \mathcal{E}\}$ is dense in $\Sigma/\!\sim$.*

b) *If X is a Banach space then* $\mathrm{St}(\Omega,\mathcal{E};X)$ *is dense in* $\mathrm{L}^0(\Omega;X)$.

c) *If X is a Banach space and* $1 \le p < \infty$ *then* $\mathrm{St}(\Omega,\mathcal{E};X)$ *is dense in* $\mathrm{L}^p(\Omega;X)$.

Fubini's Theorem

As an application we consider two σ-finite measure spaces $(\Omega_j,\Sigma_j,\mu_j)$, $j=1,2$, and their product

$$(\Omega,\Sigma,\mu)=(\Omega_1\times\Omega_2,\Sigma_1\otimes\Sigma_2,\mu_1\otimes\mu_2).$$

Let $\mathcal{R}:=\{A_1\times A_2 : A_j\in\Sigma_j,\mu(A_j)<\infty\ (j=1,2)\}$ be the set of **measurable rectangles**. Then $\mathcal{R}$ is a semi-ring, and its generated ring $\mathcal{E}$ satisfies the conditions of Theorem B.23. Since $\mathcal{E}$ consists of disjoint unions of members of $\mathcal{R}$, we obtain:

Corollary B.24. *Let X be a Banach space and* $1\le p<\infty$. *The space*

$$\mathrm{lin}\{\mathbf{1}_{A_1}\otimes\mathbf{1}_{A_2}\otimes x\ :\ x\in X, A_j\in\Sigma_j,\mu(A_j)<\infty\ (j=1,2)\}$$

is dense in $\mathrm{L}^p(\Omega;X)$.

Using this and Tonelli's theorem, one proves Fubini's theorem.

Theorem B.25 (Fubini). [Lang (1993), Thm. 8.4] *Let X be a Banach space and* $f\in\mathrm{L}^1(\Omega_1\times\Omega_2;X)$. *Then for μ_1-almost every $x\in\Omega_1$,* $f(x,\cdot)\in\mathrm{L}^1(\Omega_2;X)$ *and with*

$$F:=\left(x\longmapsto\int_{\Omega_2}f(x,\cdot)\,\mathrm{d}\mu_2\right)$$

(defined almost everywhere on Ω_1) one has $F\in\mathrm{L}^1(\Omega_1;X)$*; moreover,*

$$\int_{\Omega_1}F\,\mathrm{d}\mu_1=\int_{\Omega_1}\int_{\Omega_2}f(x,y)\,\mathrm{d}\mu_2(y)\,\mathrm{d}\mu_1(x)=\int_{\Omega_1\otimes\Omega_2}f\,\mathrm{d}(\mu_1\otimes\mu_2).$$

B.12 Complex Measures

A **complex measure** on a measurable space (Ω,Σ) is a mapping $\mu:\Sigma\longrightarrow\mathbb{C}$ which is σ-additive and satisfies $\mu(\emptyset)=0$. If the range of μ is contained in $\mathbb{R}$, μ is called a **signed measure**. For a complex measure μ one defines its **total variation** $|\mu|$ by

$$|\mu|(A):=\inf\left\{\sum_{n=1}^{\infty}|\mu(A_n)|\ :\ (A_n)_n\subset\Sigma \text{ pairwise disjoint}, A=\bigcup_n A_n\right\}$$

for $A\in\Sigma$. Then $|\mu|$ is a positive finite measure, see [Rudin (1987), Thm. 6.2]. With respect to the norm $\|\mu\|_1:=|\mu|(\Omega)$, the space of complex measures on (Ω,Σ) is a Banach space.

Let μ be a complex measure on a measurable space (Ω,Σ), and let X be a Banach space. For a step function $f = \sum_{j=1}^{n} \mathbf{1}_{A_j} \otimes x_j \in \mathrm{St}(\Omega,\Sigma,|\mu|;X)$ one defines

$$\int_{\Omega} f \,\mathrm{d}\mu = \sum_{j=1}^{n} \mu(A_j)x_j$$

as usual, and shows (using finite additivity) that this does not depend on the representation of f. Moreover, one obtains

$$\Big\| \int_{\Omega} f \,\mathrm{d}\mu \Big\|_X \le \int_{\Omega} \|f(\cdot)\|_X \,\mathrm{d}|\mu| = \|f\|_{\mathrm{L}^1(|\mu|)},$$

whence the integral has a continuous linear extension to all of $\mathrm{L}^1(\Omega,\Sigma,|\mu|;X)$.

Let (Ω,Σ,μ) be a measure space. Then for $f \in \mathrm{L}^1(\Omega;\mathbb{C})$ by

$$(f\mu)(A) := \int_{\Omega} \mathbf{1}_A f \,\mathrm{d}\mu \qquad (A \in \Sigma)$$

a complex measure is defined, with $|f\mu| = |f|\,\mu$.

Theorem B.26 (Radon–Nikodym II). [Rudin (1987), Thm. 6.10] *Let (Ω,Σ,μ) be a σ-finite measure space. The mapping $(f \longmapsto f\mu)$ is an isometric isomorphism between $\mathrm{L}^1(\Omega;\mathbb{C})$ and the space of complex measures ν on Σ with the property that $|\nu|$ is absolutely continuous with respect to μ.*

Appendix C
Functional Analysis

In this appendix we review some notions and facts from functional analysis but refer to [Conway (1990)], [Dunford and Schwartz (1958)], [Rudin (1987)], [Rudin (1991)], [Megginson (1998)] or [Schaefer (1980)] for more information.

C.1 Banach Spaces

Let X be a vector space over $\mathbb{K}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. A **semi-norm** on X is a mapping $\|\cdot\| : X \longrightarrow \mathbb{R}$ satisfying

$$\|x\| \geq 0, \quad \|\lambda x\| = |\lambda| \, \|x\|, \quad \text{and} \quad \|x+y\| \leq \|x\| + \|y\|$$

for all $x, y \in X$, $\lambda \in \mathbb{K}$. (The rightmost condition is called the **triangle inequality**.)

A semi-norm is a **norm** if $\|x\| = 0$ happens if and only if $x = 0$. A norm on a vector space defines a metric via

$$d(x,y) := \|x - y\| \qquad (x, y \in X)$$

and hence a topology, called the **norm topology**. It follows directly from the axioms of the norm that the mappings

$$X \times X \longrightarrow X, \quad (x,y) \longmapsto x+y \quad \text{and} \quad \mathbb{K} \times X \longrightarrow X, \quad (\lambda, x) \longmapsto \lambda x$$

are continuous. From the triangle inequality it follows that

$$\big| \|x\| - \|y\| \big| \leq \|x - y\| \qquad (x, y \in X),$$

which implies that also the norm mapping itself is continuous.

A **normed space** is a vector space together with a norm on it, and a **Banach space** is a normed space such that the metric induced by the norm is complete. If X is a normed space, its **closed unit ball** is defined as

$$\mathrm{B}_X := \{x \in X : \|x\| \le 1\}.$$

The set B_X is compact if and only if X is finite-dimensional.

Recall that a metric space is **separable** if it contains a countable dense set. A normed space X is separable if and only if there is a countable set A such that its linear span $\mathrm{lin} A$ is dense in X.

A subset $A \subset X$ of a normed space is called **(norm) bounded** if there is $c > 0$ such that $\|a\| \le c$ for all $a \in A$. Every compact subset is norm bounded, since the norm mapping is continuous.

A Banach space X with an associative bilinear mapping $X \times X \longrightarrow X$ (usually called a *multiplication*) is called a **Banach algebra** if $\|xy\| \le \|x\| \|y\|$ for all $x, y \in X$. The Banach algebra is called **commutative** if $xy = yx$ for all $x, y \in X$. If there is $\mathrm{e} \in X$ such that $\mathrm{e}x = x\mathrm{e} = x$ for all $x \in X$, then e is called a **unit element**, and X is called a **unital Banach algebra** if it is a Banach algebra with unit e and such that $\|\mathrm{e}\| = 1$.

An **involution** on a complex Banach algebra $\mathscr{A}$ is a map $x \longmapsto x^*$ satisfying

$$(x^*)^* = x, \quad (x+y)^* = x^* + y^*, \quad (\lambda x)^* = \overline{\lambda} x^*, \quad (xy)^* = y^* x^*$$

for all $x, y \in \mathscr{A}$, $\lambda \in \mathbb{C}$. It is called a **C^*-algebra** if in addition $\|x\|^2 = \|x^* x\|$ for all $x \in \mathscr{A}$.

C.2 Linear Operators

A linear map $T : X \longrightarrow Y$ between two normed spaces is called a **bounded linear mapping** if $T(\mathrm{B}_X)$ is a norm bounded set in Y. The linear mapping T is bounded if and only if it is continuous if and only if there is a norm estimate of the form

$$\|Tx\| \le c\,\|x\| \qquad (x \in X) \tag{C.1}$$

for some $c \ge 0$ independent of $x \in X$. The smallest c such that (C.1) holds is called the **operator norm** of T, denoted by $\|T\|$. One has

$$\|T\| = \sup\{\|Tx\| : x \in \mathrm{B}_X\}.$$

This defines a norm on the space

$$\mathscr{L}(X;Y) := \{T : X \longrightarrow Y : T \text{ is a bounded linear mapping}\},$$

which is a Banach space if Y is a Banach space. If X, Y, Z are normed spaces and $T \in \mathscr{L}(X;Y)$, $S \in \mathscr{L}(Y;Z)$, then $ST := S \circ T \in \mathscr{L}(X;Z)$ with $\|ST\| \le \|S\| \|T\|$. The identity operator I is neutral with respect to multiplication of operators, and clearly $\|\mathrm{I}\| = 1$. In particular, the space

$$\mathscr{L}(X) := \mathscr{L}(X;X)$$

is a unital Banach algebra.

Bounded linear mappings are also called **operators**. Associated with a linear mapping $T : X \longrightarrow Y$ is its **kernel** $\ker T := \{x \in X : Tx = 0\}$ and its **range** $\operatorname{ran} T := \{Tx : x \in X\}$. If T is bounded, its kernel is a closed subspace of X. One has $\ker T = \{0\}$ if and only if T is injective.

A subset $M \subset \mathscr{L}(X;Y)$ of operators that is a bounded set in the normed space $\mathscr{L}(X;Y)$ is often called **uniformly (norm) bounded**. The following result — sometimes called the Banach–Steinhaus theorem — gives an important characterisation. (See [Rudin (1987), Thm. 5.8] for a proof.)

Theorem C.1 (Principle of Uniform Boundedness). *Let X,Y be Banach spaces and $M \subset \mathscr{L}(X;Y)$. Then M is uniformly bounded if and only if for each $x \in X$ the set $\{Tx : T \in M\}$ is a bounded subset of Y.*

A bounded linear mapping $T \in \mathscr{L}(X;Y)$ is called a **contraction** if $\|T\| \le 1$. It is called called **isometric** or an **isometry** if $\|Tx\| = \|x\|$ for all $x \in X$. Isometries are contractive and injective. An operator $P \in \mathscr{L}(X)$ is called a **projection** if $P^2 = P$. In this case, $Q := \mathrm{I} - P$ is also a projection, and a direct sum decomposition of X

$$X = \operatorname{ran} P \oplus \operatorname{ran}(\mathrm{I} - P) = \operatorname{ran} P \oplus \ker P$$

into closed subspaces is induced. Conversely, whenever X is a Banach space and one has a direct sum decomposition $X = F \oplus E$ into *closed* linear subspaces, then the associated projections $X \longrightarrow F$, $X \longrightarrow E$ are bounded. This is a consequence of the **Closed Graph Theorem** [Rudin (1987), p.114, Ex.16]

An operator $T \in \mathscr{L}(X;Y)$ is called **invertible** or an **isomorphism** (of normed spaces) if T is bijective and T^{-1} is also bounded. If X,Y are Banach spaces, the boundedness of T^{-1} is automatic:

Theorem C.2 (Inverse Mapping Theorem). *If X,Y are Banach space and $T \in \mathscr{L}(X;Y)$ is bijective, then T is an isomorphism.*

See [Rudin (1987), Thm.5.10] for a proof.

C.3 Duals, Bi-Duals and Adjoints

For a normed space X its **dual space** is defined as $X' := \mathscr{L}(X;\mathbb{K})$. Since $\mathbb{K}$ is complete, this is always a Banach space. Elements of X' are called (bounded linear) **functionals**. One frequently writes

$$\langle x, x' \rangle \quad \text{in place of} \quad x'(x)$$

for $x \in X, x' \in X'$ and calls the mapping

$$\langle\cdot,\cdot\rangle : X \times X' \longrightarrow \mathbb{K}$$

the **canonical duality**.

Theorem C.3 (Hahn–Banach). *Let X be a Banach space, $F \subset X$ a linear subspace and $f' \in F'$. Then there exists $x' \in X'$ such that $x' = f'$ on F and $\|x'\| = \|f'\|$.*

For a proof see [Rudin (1987), Theorem 5.16]. A consequence of the Hahn–Banach theorem is that X' *separates the points* of X. Another consequence is that the norm of X can be computed as

$$\|x\| = \sup\left\{ \left|\left\langle x, x' \right\rangle\right| \,:\, \left\|x'\right\| \le 1 \right\}. \tag{C.2}$$

A third consequence is that X is separable whenever X' is.

Given normed spaces X, Y and an operator $T \in \mathscr{L}(X;Y)$ we define its **adjoint** operator $T' \in \mathscr{L}(Y';X')$ by $T'y' := y' \circ T$, $y' \in Y'$. Using the canonical duality this just means

$$\left\langle Tx, y' \right\rangle = \left\langle x, T'y' \right\rangle \qquad (x \in X, y' \in Y').$$

The map $(T \longmapsto T') : \mathscr{L}(X;Y) \longrightarrow \mathscr{L}(Y';X')$ is linear and isometric, and one has $(ST)' = T'S'$ for $T \in \mathscr{L}(X;Y)$ and $S \in \mathscr{L}(Y;Z)$.

The space X'' is called the **bi-dual** of X. The mapping

$$X \longrightarrow X'', \quad x \longmapsto \left(x' \longmapsto \left\langle x, x' \right\rangle \right)$$

is called the **canonical embedding**. The canonical embedding is a linear isometry by (C.2), and hence one often considers X to be a subspace of X''. Given $T \in \mathscr{L}(X;Y)$ one has $T''|_X = T$. If the canonical embedding is surjective, in sloppy notation: $X = X''$, the space X is called **reflexive**.

C.4 The weak* topology

Let X be a Banach space and X' its dual. The coarsest topology on X' that makes all mappings

$$x' \longmapsto \left\langle x, x' \right\rangle, \quad x \in X$$

continuous, is called the **weak* topology** (or: $\sigma(X',X)$-topology) on X' (cf. also Appendix A.4). A fundamental system of open and convex neighbourhoods of $y' \in X'$ is given by

$$\left\{ x' \in X' \,:\, \max_{1\le j\le d} \left|\left\langle x_j, x' - y' \right\rangle\right| < \varepsilon \right\} \qquad (d \in \mathbb{N},\, x_1, \ldots, x_d \in X,\, \varepsilon > 0).$$

Thus a sequence $(x'_n)_{n\in\mathbb{N}} \subset X'$ converges weakly* (that is, in the weak* topology) to $x' \in X'$ if and only if $\langle x, x'_n \rangle \to \langle x, x' \rangle$ as $n \to \infty$, for every $x' \in X'$.

Theorem C.4 (Banach–Alaoglu). *Let X be a Banach space. Then the dual unit ball* $\mathrm{B}_{X'} = \{x' \in X' : \|x'\| \leq 1\}$ *is weakly* compact.*

The proof is a more or less straightforward application of Tychonov's Theorem A.2, see [Rudin (1991), Thm. 3.15]. Since the weak* topology is usually not metrisable, one cannot in general test continuity of mappings or compactness of sets via criteria using sequences. Therefore, the following theorem is often useful (see [Dunford and Schwartz (1958), Thm.V.5.1]).

Theorem C.5. *Let X be a Banach space. Then the weak* topology on the dual unit ball $\mathrm{B}_{X'}$ is metrisable if and only if X is separable.*

Recall that X can be considered (via the canonical embedding) a norm closed subspace of X''. The next theorem shows in particular that X is $\sigma(X'', X')$-dense in X''.

Theorem C.6 (Goldstine). *Let X be a Banach space. Then the closed unit ball B_X of X is $\sigma(X'', X')$-dense in the closed unit ball of X''.*

A proof is in [Dunford and Schwartz (1958), Thm.V.4.5].

C.5 The weak topology

We now interchange the roles of X and X' with respect to the previous section. Let X be a Banach space and X' its dual. The coarsest topology on X that makes all mappings

$$x \longmapsto \langle x, x' \rangle, \quad x' \in X'$$

continuous, is called the **weak topology** (or: $\sigma(X, X')$-topology) on X (cf. also Appendix A.4). A fundamental system of open (and convex) neighbourhoods of $y \in X$ is given by

$$\Big\{x \in X : \max_{1 \leq j \leq d} \big|\langle x - y, x'_j \rangle\big| < \varepsilon\Big\} \qquad (d \in \mathbb{N},\ x'_1, \ldots, x'_d \in X',\ \varepsilon > 0).$$

Thus a sequence $(x_n)_{n \in \mathbb{N}} \subset X$ converges weakly (that is, in the weak topology) to $x \in X$ if and only if $\langle x_n, x' \rangle \to \langle x, x' \rangle$ as $n \to \infty$, for every $x' \in X'$. Note that the weak topology on X is the subspace topology when X is considered as a subspace $X \subset X''$ via the canonical embedding, and X'' is endowed with the $\sigma(X'', X')$-topology.

Theorem C.7 (Mazur). *Let X be a Banach space and let $A \subset X$ be convex. Then the norm closure of A coincides with its weak closure.*

See [Rudin (1991), Theorem 3.12] or [Dunford and Schwartz (1958), Thm. V.3.18].

As the weak*-topology, the weak topology is not metrisable in general. Fortunately, there is a series of deep results on metrisability and compactness for the weak topology, facilitating its use. The first is the analogue of Theorem C.5.

Theorem C.8. *The weak topology on the closed unit ball of a Banach space X is metrisable if and only if the dual space X' is norm separable.*

A proof is in [Dunford and Schwartz (1958), Theorem V.5.2]. The next theorem is extremely useful since it allows sequences in testing of weak compactness.

Theorem C.9 (Eberlein–Šmulian). *Let X be a Banach space. Then $A \subset X$ is weakly compact if and only if A is weakly sequentially compact.*

For a proof see [Dunford and Schwartz (1958), Thm.V.6.1]. Although, as mentioned above, the weak topology is usually far from being metrisable, the following holds.

Theorem C.10. *The weak topology on a weakly compact subset of a separable Banach space is metrisable.*

A proof is in [Dunford and Schwartz (1958), V.6.3]. The next result is often useful when considering Banach-space valued (weak) integrals. For the definition of the closed convex hull see Section C.6 below.

Theorem C.11 (Kreĭn–Šmulian). *The closed convex hull of a weakly compact subset of a Banach space is weakly compact.*

For a proof see [Dunford and Schwartz (1958), Thm. V.6.4]. Finally we state a characterisation of reflexivity, which is a mere combination of the Banach–Alaoglu theorem and Goldstine's theorem.

Theorem C.12. *A Banach space is reflexive if and only if its closed unit ball is weakly compact.*

For a proof see [Dunford and Schwartz (1958), Thm. V.4.7]. In particular, every bounded set is relatively weakly compact if and only if the Banach space is reflexive.

C.6 Convex sets and their extreme points

A subset $A \subset X$ of a real vector space X is called **convex** if, whenever $a, b \in A$ then also $ta + (1-t)b \in A$ for all $t \in [0,1]$. Geometrically this means that the straight line between a and b is contained in A whenever a and b are. Inductively one shows that a convex set contains arbitrary **convex combinations**

$$t_1 a_1 + \cdots + t_n a_n, \qquad 0 \le t_1, \ldots, t_n \le 1,\ t_1 + \cdots + t_n = 1$$

whenever $a_1, \ldots, a_n \in A$. Intersecting convex sets yields a convex set, and so for any set $B \subset X$ there is a smallest convex set containing B, its **convex hull**

$$\operatorname{conv}(B) = \bigcap \big\{ A : A \supset B,\ \text{convex} \big\}.$$

Alternatively, $\operatorname{conv}(B)$ is the set of all convex combinations of elements of B.

A vector space X endowed with a topology (see Appendix A.2) such that the operations of addition $X \times X \longrightarrow X$ and scalar multiplication $\mathbb{K} \times X \longrightarrow X$ are continuous, is called a **topological vector space**. In a topological vector space, the closure of a convex set is convex. We denote by

$$\overline{\mathrm{conv}}(B) := \overline{\mathrm{conv}(B)}$$

the **closed convex hull** of $B \subset X$, which is the smallest closed convex set containing B.

A topological vector space is called **locally convex** if it is Hausdorff and the topology has a base consisting of convex sets. Examples are the norm topology of a Banach space X, the weak topology on X and the weak* topology on X'.

Let $A \subset X$ be a convex set. A point $p \in A$ is called an **extreme point** of A if it cannot be written as a convex combinations of two points of A *distinct from* p. In other words, whenever $a, b \in A$ and $t \in [0,1]$ such that $p = ta + (1-t)b$, then $a = b = p$. Let us denote the set of extreme points of a convex set A with $\mathrm{ex}\, A$.

Theorem C.13 (Kreĭn–Milman). *Let X be a locally convex space and let $\emptyset \neq K \subset X$ be compact and convex. Then K is the closed convex hull of its extreme points: $K = \overline{\mathrm{conv}}(\mathrm{ex}\, K)$. In particular,* $\mathrm{ex}\, K$ *is not empty.*

For a proof see [Rudin (1991), Thm. 3.23]. Since the weak* topology on X', X a Banach space, is locally convex, the Kreĭn–Milman theorem applies in particular to weakly* compact convex subsets of X'.

Theorem C.14 (Milman). *Let X be a locally convex space and let $\emptyset \neq K \subset X$ be compact. If $\overline{\mathrm{conv}} K$ is also compact, then K contains the extreme points of $\overline{\mathrm{conv}} K$.*

A proof is in [Rudin (1991), Thm. 3.25]. Combining Milman's result with the Kreĭn–Šmulian theorem C.11 yields the following.

Corollary C.15. *Let X be a Banach space and let $K \subset X$ be weakly compact. Then K contains the extreme points of $\overline{\mathrm{conv}} K$.*

C.7 The Strong and the Weak Operator Topology on $\mathscr{L}(X)$

Let X be a Banach space. Besides the norm topology there are two other canonical topologies on $\mathscr{L}(X)$. The **strong operator topology** is the coarsest topology making all evaluation mappings

$$\mathscr{L}(X) \longrightarrow X, \quad T \longmapsto Tx, \qquad x \in X$$

continuous. A fundamental system of open neighbourhoods of $T \in \mathscr{L}(X)$ is given by

$$\left\{S \in \mathscr{L}(X) : \max_{1\le j\le d} \left\|(T-S)x_j\right\| < \varepsilon\right\} \qquad (d \in \mathbb{N},\ x_1,\dots,x_d \in X,\ \varepsilon > 0).$$

Thus a sequence of operators $(T_n)_{n\in\mathbb{N}} \subset \mathscr{L}(X)$ converges strongly to T if and only if

$$\lim_{n\to\infty} T_n x = Tx \quad \text{for every } x \in X.$$

We denote by $\mathscr{L}_s(X)$ the space $\mathscr{L}(X)$ endowed with the strong operator topology. It is a corollary of the Principle of Uniform Boundedness C.1 that if $(T_n)_n \subset \mathscr{L}(X)$ is such that

$$Tx := \lim_{n\to\infty} T_n x$$

exists for every $x \in X$, then $T \in \mathscr{L}(X)$ and $\sup_n \|T_n\| < \infty$. The following simple property of strong operator convergence is very useful.

Proposition C.16. *For a sequence $(T_n)_{n\in\mathbb{N}} \subset \mathscr{L}(X)$ and $T \in \mathscr{L}(X)$ the following assertions are equivalent.*

(i) *$(T_n)_n$ converges to T strongly.*

(ii) *$(T_n)_n$ is uniformly bounded and* $\lim_{n\to\infty} T_n x = Tx$ *for all x from a dense set $D \subset X$.*

(iii) $\lim_{n\to\infty} T_n x = Tx$ *uniformly on compact sets of X.*

The strong operator topology is metrisable on (norm-)bounded sets if X is separable.

The **weak operator topology** is the coarsest topology making all evaluation mappings

$$\mathscr{L}(X) \longrightarrow \mathbb{K}, \quad T \longmapsto \langle Tx, x'\rangle, \qquad x \in X,\, x' \in X'$$

continuous. A fundamental system of open neighbourhoods of $T \in \mathscr{L}(X)$ is given by

$$\left\{S \in \mathscr{L}(X) : \max_{1\le j\le d} \left|\left\langle (T-S)x_j, x'_j\right\rangle\right| < \varepsilon\right\}$$
$$(d \in \mathbb{N},\ x_1,\dots,x_d \in X,\ \ x'_1,\dots,x'_d \in X',\ \varepsilon > 0).$$

Thus a sequence of operators $(T_n)_{n\in\mathbb{N}}$ converges to T in the weak operator topology if and only if

$$\lim_{n\to\infty} \langle T_n x, x'\rangle = \langle Tx, x'\rangle \quad \text{for every } x \in X,\ x' \in X'.$$

The space $\mathscr{L}(X)$ considered with the weak operator topology is denoted by $\mathscr{L}_\sigma(X)$. The weak operator topology is metrisable on (norm-)bounded sets if X' is separable.

The following proposition shows that the notions of boundedness coincide for all operator topologies.

Proposition C.17. *For a family of operators $\mathscr{T} \subset \mathscr{L}(X)$, X a Banach space, the following assertions are equivalent.*

(i) *$\mathscr{T}$ is bounded for the weak operator topology.*

(ii) *$\mathscr{T}$ is bounded for the strong operator topology.*

(iii) *$\mathscr{T}$ is uniformly bounded, i.e.,* $\sup\{\|T\|, T \in \mathscr{T}\} < \infty$.

We now characterise compactness in the weak operator topology of a subset in $\mathscr{L}(X)$.

Proposition C.18. *For a family of operators $\mathscr{T} \subset \mathscr{L}(X)$, X a Banach space, the following assertions are equivalent.*

(i) *$\mathscr{T}$ is relatively compact in $\mathscr{L}_\sigma(X)$.*

(ii) *$\mathscr{T}x := \{Tx : T \in \mathscr{T}\}$ is relatively weakly compact in X, for all $x \in X$.*

(iii) *$\mathscr{T}$ is bounded, and $\mathscr{T}x$ is relatively weakly compact in X, for all x in some norm-dense subset of X.*

Proof. The implication (i)⇒(ii) follows directly from the continuity of the mapping

$$\mathscr{L}_\sigma(X) \longrightarrow (X, \sigma(X, X')), \quad T \mapsto Tx$$

for every $x \in X$. For the converse implication (ii)⇒(i) note that the mapping

$$(\mathscr{T}, \mathscr{L}_\sigma(X)) \longrightarrow \prod_{x \in X} \overline{\mathscr{T}x}^{\sigma(X,X')}, \qquad T \longmapsto (Tx)_{x \in X}$$

is a homeomorphic embedding. Hence (i) follows from Tychonov's theorem A.2. The implication (ii)⇒(iii) follows from the uniform boundedness principle.

Let us prove the implication (iii)⇒(ii). Let $D \subset X$ be a norm-dense subset of X such that $\mathscr{T}x$ is relatively weakly compact for every $x \in D$. Fix $x \in X$ and a sequence $(x_n)_{n\in\mathbb{N}} \subset D$ converging to x. By the Eberlein–Šmulian theorem C.9 it is enough to show that every sequence within $\mathscr{T}x$ has a weakly convergent subsequence. This is done by a standard diagonal argument.

Take a sequence $(T_n)_n \subset \mathscr{T}$. Since $\mathscr{T}x_1$ is relatively weakly compact, there exists a subsequence $\pi_1 : \mathbb{N} \longrightarrow \mathbb{N}$ and a vector $z_1 \in X$ such that $T_{\pi_1(n)}x_1 \to z_1$ weakly. Since $\mathscr{T}x_2$ is relatively weakly compact, there exists a subsequence $\pi_2 : \mathbb{N} \longrightarrow \mathbb{N}$ and a vector $z_2 \in X$ such that $T_{\pi_1\pi_2(n)}x_2 \to z_2$. Continuing in this manner and finally taking the diagonal sequence $n_k := \pi_1 \ldots \pi_k(k)$, $k \in \mathbb{N}$, yields that $T_{n_k}x_m \to z_m$ weakly as $k \to \infty$, for *every* $m \in \mathbb{N}$.

Now let $C := \sup\{\|T\| : T \in \mathscr{T}\} < \infty$ (by assumption). For each $x' \in X'$ and $n, m \in \mathbb{N}$ we thus have

$$\left|\left\langle T_{n_k}x_m - T_{n_k}x_n, x'\right\rangle\right| \le C\,\|x_n - x_m\|\,\|x'\|.$$

Taking the limit $k \to \infty$ and then the supremum over all $x' \in \mathrm{B}_{X'}$ we obtain $\|z_n - z_m\| \le C\,\|x_n - x_m\|$ for all $n, m \in \mathbb{N}$. Since X is a Banach space there exists $z \in X$ such that $\|z - z_n\| \to 0$. A 3ε-argument now shows that $T_{n_k}x \to z$ weakly, as desired.

C.8 Multiplication on $\mathscr{L}(X)$

We now consider continuity of the multiplication

$$(S,T) \longmapsto ST, \qquad S,T \in \mathscr{L}(X)$$

for the operator topologies considered above.

Proposition C.19. *The multiplication on $\mathscr{L}(X)$ is*

1) *jointly continuous for the norm topology in $\mathscr{L}(X)$;*
2) *jointly continuous on bounded sets for the strong operator topology;*
3) *separately but (in general) not jointly continuous for the weak and strong operator topology.*

The following example shows that the multiplication is not jointly continuous for the weak operator topology.

Example C.20. Let T be the right shift on $\ell^2(\mathbb{Z})$ given by $T((x_k)_{k\in\mathbb{Z}}) := (x_{k-1})_{k\in\mathbb{Z}}$. Then $(T^n)_n$ converges to 0 weakly for the weak operator topology (use the density of the space of finite sequences in $\ell^2(\mathbb{Z})$). The same holds for the left shift T^{-1}, but $(TT^{-1})^n = \mathrm{I}$ does not converge to 0.

C.9 Spectral Theory

Let X be a non-trivial complex Banach space and $T \in \mathscr{L}(X)$. The **resolvent set** $\rho(T)$ of T is the set of all $\lambda \in \mathbb{C}$ such that the operator $\lambda\mathrm{I} - T$ is invertible. The function

$$\rho(T) \longrightarrow \mathscr{L}(X), \quad \lambda \longmapsto R(\lambda,T) := (\lambda\mathrm{I} - T)^{-1}$$

is called the **resolvent** of T. Its complement $\sigma(T) := \mathbb{C} \setminus \rho(T)$ is called the **spectrum** of T. The resolvent set is an open subset of $\mathbb{C}$, and given $\lambda_0 \in \rho(T)$ one has

$$R(\lambda,A) = \textstyle\sum_{n=0}^{\infty}(\lambda_0 - \lambda)^n R(\lambda_0,T)^{n+1} \quad \text{for} \quad |\lambda - \lambda_0| < \|R(\lambda_0,T)\|^{-1}.$$

In particular, $\mathrm{dist}(\lambda_0, \sigma(T)) \geq \|R(\lambda_0,T)\|^{-1}$, showing that the norm of the resolvent blows up when λ_0 approaches a spectral point.

Every $\lambda \in \mathbb{C}$ with $|\lambda| > \|T\|$ is contained in $\rho(T)$, and in this case $R(\lambda,T)$ is given by the Neumann series

$$R(\lambda,T) := \textstyle\sum_{n=0}^{\infty} \lambda^{-(n+1)} T^n.$$

In particular, one has $r(T) \leq \|T\|$, where

$$r(T) := \sup\big\{|\lambda| \ : \ \lambda \in \sigma(T)\big\}$$

is called the **spectral radius** of T. An important result states that the spectrum is always a *non-empty* compact subset of $\mathbb{C}$ and the spectral radius can be computed by the formula

$$r(T) = \lim_{n\to\infty} \|T^n\|^{1/n} = \inf_{n\in\mathbb{N}} \|T^n\|^{1/n}.$$

(See [Rudin (1991), Thm 10.13] for a proof.) An operator is called **power-bounded** if $\sup_n \|T^n\| < \infty$. From the spectral radius formula it follows that the spectrum of a power-bounded operator is contained in the closed unit disc. For isometries one has more precise information.

Proposition C.21. *If X is a complex Banach space and $T \in \mathscr{L}(X)$ is an isometry, then exactly one of the following two cases holds:*

1) *T is not surjective and $\sigma(T) = \{z : |z| \le 1\}$ is the full closed unit disc;*
2) *T is an isomorphism and $\sigma(T) \subset \mathbb{T}$.*

The spectrum can be divided into the following parts. The **point spectrum** is

$$\sigma_p(T) := \big\{\lambda \in \mathbb{C} : (\lambda \mathrm{I} - T) \text{ is not injective } \big\},$$

and each $\lambda \in \sigma_p(T)$ is called an **eigenvalue** of T. For $\lambda \in \sigma_p(T)$ the closed subspace

$$\ker(\lambda \mathrm{I} - T)$$

is called the corresponding **eigenspace**. Every nonzero element $0 \neq x \in \ker(\lambda \mathrm{I} - T)$ is called a corresponding **eigenvector**. An eigenvalue λ is called a **simple eigenvalue** if its eigenspace is one-dimensional.

The point spectrum is a subset of the **approximate point spectrum**

$$\sigma_a(T) := \{\lambda \in \mathbb{C} : \lambda \mathrm{I} - T \text{ is not injective or } \mathrm{ran}(\lambda - T) \text{ is not closed}\},$$

and each $\lambda \in \sigma_a(T)$ is called an **approximate eigenvalue** of T. A number $\lambda \in \mathbb{C}$ is contained in $\sigma_a(T)$ if and only if there is a sequence $(x_n)_n \subset X$ such that $\|x_n\| = 1$ for all $n \in \mathbb{N}$ and $\|(\lambda \mathrm{I} - T)x_n\| \to 0$. (Such a sequence is called a corresponding **approximate eigenvector**. One has $\sigma_p(T) \cup \partial\sigma(T) \subset \sigma_a(T)$, where $\partial\sigma(T)$ denotes the topological boundary of $\sigma(T)$ in $\mathbb{C}$. In particular, $\sigma_a(T)$ is not empty.

Finally, the **residual spectrum** of T is

$$\sigma_r(T) := \big\{\lambda \in \mathbb{C} : \mathrm{ran}(\lambda \mathrm{I} - T) \text{ is not dense in } X\big\} = \sigma_p(T').$$

Then $\sigma(T) = \sigma_a(T) \cup \sigma_r(T)$, but this union is not necessarily disjoint.

C.10 Important Examples

The following Banach spaces are most important for our lectures (see also the Appendices A and B).

Example C.22 (Continuous Functions).
Let K be a compact topological space and let $\mathrm{C}(K)$ be the space of continuous functions on K. The space $\mathrm{C}(K)$ is a unital Banach algebra with respect to the **uniform norm**

$$\|f\|_\infty := \sup\big\{|f(x)| : x \in K\big\} \qquad (f \in \mathrm{C}(K)).$$

By the Riesz Representation Theorem [Rudin (1987), Thm. 6.19] its dual $\mathrm{C}(K)'$ can be identified with the set $\mathrm{M}(K)$ of complex Baire measures on K, endowed with the total variation norm (see also Appendix B.12 and Lecture 5).The set

$$\mathrm{M}^1(K) = \big\{\mu \in \mathrm{M}(K) : \mu \geq 0, \mu(K) = 1\big\}$$

of probability measures is a convex subset of $\mathrm{M}(K)$, compact for the weak* topology (use the Banach–Alaoglu theorem), and its extreme points are precisely the **Dirac measures**

$$\mathrm{ex}\,\mathrm{M}^1(K) = \big\{\delta_x : x \in K\big\}$$

(see [Conway (1990), Thm. 8.4] for a proof).

Example C.23 (Lebesgue Spaces, $p = \infty$).
Let (Ω, Σ, μ) be a measure space. The space

$$\mathrm{L}^\infty(\Omega, \Sigma, \mu)$$

of equivalence classes of bounded measurable functions is a unital Banach algebra with respect to the **essential-sup norm**

$$\|f\|_\infty := \inf\big\{c > 0 : |f| \leq c \quad \mu\text{-almost everywhere}\big\} \qquad (f \in \mathrm{L}^\infty(\Omega, \Sigma, \mu)),$$

see also Appendix B.10. If μ is the counting measure then $\mathrm{L}^\infty(\Omega, \Sigma, \mu) = \ell^\infty(\Omega)$ is simply the space of all bounded functions on Ω, and the essential-sup norm coincides with the uniform norm.

Example C.24 (Lebesgue Spaces, $1 \leq p < \infty$).
Let (Ω, Σ, μ) be a measure space and $1 \leq p < \infty$. The space

$$\mathrm{L}^p(\Omega, \Sigma, \mu)$$

of equivalence classes of p-integrable measurable functions is a Banach space with respect to the p-norm

$$\|f\|_p := \left(\int_\Omega |f|^p \,\mathrm{d}\mu\right)^{1/p} \qquad (f \in \mathrm{L}^p(\Omega, \Sigma, \mu)),$$

see also Appendix B.10. The **dual exponent** of p is $q \in (1, \infty]$ with $1/p + 1/q = 1$. Hölder's inequality establishes a duality

$$\langle f, g\rangle_{p,q} := \int_\Omega fg \,\mathrm{d}\mu \qquad (f \in \mathrm{L}^p(\Omega, \Sigma, \mu),\, g \in \mathrm{L}^q(\Omega, \Sigma, \mu))$$

such that $\left|\langle f,g\rangle_{p,q}\right| \le \|f\|_p \, \|g\|_q$ (see also Theorem B.11). Via this duality one has

- $\mathrm{L}^p(\Omega,\Sigma,\mu)' = \mathrm{L}^q(\Omega,\Sigma,\mu)$, $1 < p < \infty$;
- $\mathrm{L}^1(\Omega,\Sigma,\mu)' = \mathrm{L}^\infty(\Omega,\Sigma,\mu)$, if (Ω,Σ,μ) is σ-finite.

For a proof see [Rudin (1987), Thm. 6.16]. The space $\mathrm{L}^1(\Omega,\Sigma,\mu)$ is reflexive if and only if it is finite-dimensional.

References

Akcoglu and Sucheston (1972). Akcoglu, M., Sucheston, L. : *On operator convergence in Hilbert space and in Lebesgue space*. Period. Math. Hungar. **2**, 1972, 235–244, collection of articles dedicated to the memory of Alfréd Rényi, I.

Akcoglu (1975). Akcoglu, M. A. : *A pointwise ergodic theorem in L_p-spaces*. Canad. J. Math. **27** (5), 1975, 1075–1082.

Baker (1984). Baker, A. : *A concise introduction to the theory of numbers*. Cambridge University Press, Cambridge, 1984.

Bauer (1981). Bauer, H. : *Probability theory and elements of measure theory*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1981.

Bauer (1990). Bauer, H. : *Maß- und Integrationstheorie*. de Gruyter Lehrbuch. Walter de Gruyter & Co., Berlin, 1990.

Berglund et al. (1989). Berglund, J. F., Junghenn, H. D., Milnes, P. : *Analysis on semigroups*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc., New York, 1989.

Billingsley (1979). Billingsley, P. : *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-Chichester-Brisbane, 1979.

Birkhoff (1912). Birkhoff, G. D. : *Quelques théorèmes sur le mouvement des systèmes dynamiques.* S. M. F. Bull. **40**, 1912, 305–323.

Birkhoff (1931). Birkhoff, G. D. : *Proof of the ergodic theorem.* Proc. Nat. Acad. Sci. U. S. A. **17**, 1931, 656–660.

Blum and Hanson (1960). Blum, J. R., Hanson, D. L. : *On the mean ergodic theorem for subsequences*. Bull. Amer. Math. Soc. **66**, 1960, 308–311.

Bogachev (2007). Bogachev, V. I. : *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007.

Boltzmann (1885). Boltzmann, L. : *Ueber die Eigenschaften monocyklischer und anderer damit verwandter Systeme.* J. reine angew. Math. **98**, 1885, 68–94.

Borel (1909). Borel, E. : *Les probabilités denombrables et leurs applications arithmétiques.* Palermo Rend. **27**, 1909, 247–271.

Brush (1971). Brush, S. G. : *Milestones in mathematical physics. Proof of the impossibility of ergodic systems: the 1913 papers of Rosenthal and Plancherel*. Transport Theory Statist. Phys. **1** (4), 1971, 287–298.

Burkholder (1962). Burkholder, D. L. : *Semi-Gaussian subspaces*. Trans. Amer. Math. Soc. **104**, 1962, 123–131.

Chacon (1964). Chacon, R. V. : *A class of linear transformations*. Proc. Amer. Math. Soc. **15**, 1964, 560–564.

Chacon (1967). Chacon, R. V., 1967. : *A geometric construction of measure preserving transformations*. In: Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 2. Univ. California Press, Berkeley, Calif., pp. 335–360.

Chacon (1969). Chacon, R. V. : *Weakly mixing transformations which are not strongly mixing*. Proc. Amer. Math. Soc. **22**, 1969, 559–562.
Conway (1990). Conway, J. B. : *A Course in Functional Analysis. 2nd ed.* Graduate Texts in Mathematics, 96. New York etc.: Springer-Verlag, 1990.
de Leeuw and Glicksberg (1959). de Leeuw, K., Glicksberg, I. : *Almost periodic compactifications*. Bull. Amer. Math. Soc. **65**, 1959, 134–139.
de Leeuw and Glicksberg (1961). de Leeuw, K., Glicksberg, I. : *Applications of almost periodic compactifications*. Acta Math. **105**, 1961, 63–97.
Denker et al. (1976). Denker, M., Grillenberger, C., Sigmund, K. : *Ergodic theory on compact spaces*. Lecture Notes in Mathematics, Vol. 527. Springer-Verlag, Berlin, 1976.
Dunford and Schwartz (1958). Dunford, N., Schwartz, J. T. : *Linear Operators. I. General Theory*. With the assistance of W. G. Bade and R. G. Bartle. Pure and Applied Mathematics, Vol. 7. Interscience Publishers, Inc., New York, 1958.
Ehrenfest (1912). Ehrenfest, P. u. T. : *Begriffliche Grundlagen der statistischen Auffassung in der Mechanik.* Encykl. d. Math. Wissensch. IV 2 II, Heft 6, 1912.
Einsiedler and Ward (2009). Einsiedler, M., Ward, T. : *Ergodic Theory: with a view towards Number Theory*. book manuscript, 2009.
URL `http://www.mth.uea.ac.uk/ergodic/`
Ellis (1957). Ellis, R. : *Locally compact transformation groups*. Duke Math. J. **24**, 1957, 119–125.
Engel and Nagel (2000). Engel, K.-J., Nagel, R. : *One-Parameter Semigroups for Linear Evolution Equations.* Graduate Texts in Mathematics. 194. Berlin: Springer-Verlag, 2000.
Ethier and Kurtz (1986). Ethier, S. N., Kurtz, T. G. : *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.
Furstenberg (1961). Furstenberg, H. : *Strict ergodicity and transformation of the torus*. Amer. J. Math. **83**, 1961, 573–601.
Furstenberg (1977). Furstenberg, H. : *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*. J. Analyse Math. **31**, 1977, 204–256.
Furstenberg (1981). Furstenberg, H. : *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, Princeton, N.J., 1981.
Furstenberg and Weiss (1996). Furstenberg, H., Weiss, B., 1996. : *A mean ergodic theorem for* $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$. In: Convergence in ergodic theory and probability (Columbus, OH, 1993). Vol. 5 of Ohio State Univ. Math. Res. Inst. Publ. de Gruyter, Berlin, pp. 193–227.
Gallavotti (1975). Gallavotti, G. : *Ergodicity, ensembles, irreversibility in Boltzmann and beyond*. J. Statist. Phys. **78**, 1975, 1571–1589.
Garsia (1965). Garsia, A. M. : *A simple proof of E. Hopf's maximal ergodic theorem*. J. Math. Mech. **14**, 1965, 381–382.
Garsia (1970). Garsia, A. M. : *Topics in almost everywhere convergence*. Vol. 4 of Lectures in Advanced Mathematics. Markham Publishing Co., Chicago, Ill., 1970.
Gottschalk and Hedlund (1955). Gottschalk, W. H., Hedlund, G. A. : *Topological dynamics*. American Mathematical Society Colloquium Publications, Vol. 36. American Mathematical Society, Providence, R. I., 1955.
Green (2008). Green, B. : *Lectures on Ergodic Theory, Part III*, 2008, http://www.dpmms.cam.ac.uk/ bjg23/ergodic-theory.html.
Green and Tao (2008). Green, B., Tao, T. : *The primes contain arbitrarily long arithmetic progressions*. Ann. of Math. (2) **167** (2), 2008, 481–547.
Grothendieck (1952). Grothendieck, A. : *Critères de compacité dans les espaces fonctionnels généraux.* Am. J. Math. **74**, 1952, 168–186.
Haase (2007). Haase, M. : *Convexity inequalities for positive operators*. Positivity **11** (1), 2007, 57–68.
Halmos (1944). Halmos, P. R. : *Approximation theories for measure preserving transformations*. Trans. Amer. Math. Soc. **55**, 1944, 1–18.
Halmos (1956). Halmos, P. R. : *Lectures on ergodic theory*. Publications of the Mathematical Society of Japan, no. 3. The Mathematical Society of Japan, 1956.

Halmos and von Neumann (1942). Halmos, P. R., von Neumann, J. : *Operator methods in classical mechanics. II*. Ann. of Math. (2) **43**, 1942, 332–350.
Hewitt and Ross (1979). Hewitt, E., Ross, K. A. : *Abstract harmonic analysis. Vol. I*, 2nd Edition. Vol. 115 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1979.
Hewitt and Stromberg (1969). Hewitt, E., Stromberg, K. : *Real and abstract analysis. A modern treatment of the theory of functions of a real variable*. Second printing corrected. Springer-Verlag, New York, 1969.
Hindman and Strauss (1998). Hindman, N., Strauss, D. : *Algebra in the Stone-Čech compactification*. Vol. 27 of de Gruyter Expositions in Mathematics. Walter de Gruyter & Co., Berlin, 1998.
Hlawka (1979). Hlawka, E. : *Theorie der Gleichverteilung*. Bibliographisches Institut, Mannheim, 1979.
Hofmann and Morris (2006). Hofmann, K. H., Morris, S. A. : *The structure of compact groups*, augmented Edition. Vol. 25 of de Gruyter Studies in Mathematics. Walter de Gruyter & Co., Berlin, 2006.
Hofmann and Mostert (1966). Hofmann, K. H., Mostert, P. S. : *Elements of compact semigroups*. Charles E. Merr ll Books, Inc., Columbus, Ohio, 1966.
Hopf (1954). Hopf, E. : *The general temporally discrete Markoff process*. J. Rational Mech. Anal. **3**, 1954, 13–45.
Host and Kra (2005a). Host, B., Kra, B. : *Convergence of polynomial ergodic averages*. Israel J. Math. **149**, 2005a, 1–19, probability in mathematics.
Host and Kra (2005b). Host, B., Kra, B. : *Nonconventional ergodic averages and nilmanifolds*. Ann. of Math. (2) **161** (1), 2005b, 397–488.
Ionescu Tulcea (1964). Ionescu Tulcea, A. : *Ergodic properties of isometries in L^p spaces,* $1 < p < \infty$. Bull. Amer. Math. Soc. **70**, 1964, 366–371.
Ionescu Tulcea (1965). Ionescu Tulcea, A. : *On the category of certain classes of transformations in ergodic theory*. Trans. Amer. Math. Soc. **114**, 1965, 261–279.
Jacobs (1956). Jacobs, K. : *Ergodentheorie und fastperiodische Funktionen auf Halbgruppen*. Math. Z. **64**, 1956, 298–338.
Jones and Kuftinec (1971). Jones, L., Kuftinec, V. : *A note on the Blum-Hanson theorem*. Proc. Amer. Math. Soc. **30**, 1971, 202–203.
Jones and Lin (1976). Jones, L. K., Lin, M. : *Ergodic theorems of weak mixing type*. Proc. Amer. Math. Soc. **57** (1), 1976, 50–52.
Kallenberg (2002). Kallenberg, O. : *Foundations of modern probability*, 2nd Edition. Probability and its Applications (New York). Springer-Verlag, New York, 2002.
Katok and Hasselblatt (1995). Katok, A., Hasselblatt, B. : *Introduction to the modern theory of dynamical systems*. Vol. 54 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1995.
Kelley (1975). Kelley, J. L. : *General topology*. Springer-Verlag, New York, 1975.
Kern et al. (1977). Kern, M., Nagel, R., Palm, G. : *Dilations of positive operators: construction and ergodic theory*. Math. Z. **156** (3), 1977, 265–277.
Kolmogoroff (1933). Kolmogoroff, A. : *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Ergebnisse der Math. 2, Nr. 3, IV + 62 S , 1933.
Kolmogoroff (1925). Kolmogoroff, A. N. : *Sur les fonctions harmoniques conjugées et les séries de Fourier*. Fund. Math. **7**, 1925, 23–28.
Koopman and von Neumann (1932). Koopman, B., von Neumann, J. : *Dynamical systems of continuous spectra.* Proc. Natl. Acad. Sci. USA **18**, 1932, 255–263.
Krengel (1985). Krengel, U. : *Ergodic theorems*. Vol. 6 of de Gruyter Studies in Mathematics. Walter de Gruyter & Co., Berlin, 1985.
Kuipers and Niederreiter (1974). Kuipers, L., Niederreiter, H. : *Uniform distribution of sequences*. Wiley-Interscience [John Wiley & Sons], New York, 1974.
Kuratowski (1966). Kuratowski, K. : *Topology. Vol. I*. New edition, revised and augmented. Translated from the French by J. Jaworowski. Academic Press, New York, 1966.

Lang (1993). Lang, S. : *Real and Functional Analysis. 3rd ed.* Graduate Texts in Mathematics 142. Springer-Verlag, New York, 1993.

Lang (2005). Lang, S. : *Algebra. 3rd ed.* Vol. 211 of Graduate Text in Mathematics. Springer-Verlag, 2005.

Mathieu (1988). Mathieu, M. : *On the origin of the notion 'Ergodic Theory'*. Expo. Math. **6**, 1988, 373–377.

Megginson (1998). Megginson, R. E. : *An Introduction to Banach Space Theory.* Graduate Texts in Mathematics. 183. New York, NY: Springer-Verlag, 1998.

Müller and Tomilov (2007). Müller, V., Tomilov, Y. : *Quasisimilarity of power bounded operators and Blum-Hanson property*. J. Funct. Anal. **246** (2), 2007, 385–399.

Nagel and Palm (1982). Nagel, R., Palm, G. : *Lattice dilations of positive contractions on L^p-spaces*. Canad. Math. Bull. **25** (3), 1982, 371–374.

Namioka (1974). Namioka, I. : *Separate continuity and joint continuity*. Pacific J. Math. **51**, 1974, 515–531.

Petersen (1989). Petersen, K. : *Ergodic theory*. Vol. 2 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 1989.

Phelps (1966). Phelps, R. R. : *Lectures on Choquet's theorem*. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1966.

Pollicott and Yuri (1998). Pollicott, M., Yuri, M. : *Dynamical systems and ergodic theory*. Vol. 40 of London Mathematical Society Student Texts. Cambridge University Press, Cambridge, 1998.

Raimi (1964). Raimi, R. A. : *Minimal sets and ergodic measures in $\beta N - N$*. Bull. Amer. Math. Soc. **70**, 1964, 711–712.

Rana (2002). Rana, I. K. : *An introduction to measure and integration*, 2nd Edition. Vol. 45 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2002.

Reed and Simon (1972). Reed, M., Simon, B. : *Methods of Modern Mathematical Physics. 1: Functional Analysis.* New York-London: Academic Press, Inc., 1972.

Rohlin (1948). Rohlin, V. : *A "general" measure-preserving transformation is not mixing*. Doklady Akad. Nauk SSSR (N.S.) **60**, 1948, 349–351.

Royden (1963). Royden, H. L. : *Real analysis*. The Macmillan Co., New York, 1963.

Rudin (1976). Rudin, W. : *Principles of mathematical analysis*, 3rd Edition. McGraw-Hill Book Co., New York, 1976.

Rudin (1987). Rudin, W. : *Real and Complex Analysis. 3rd ed.* New York, NY: McGraw-Hill, 1987.

Rudin (1990). Rudin, W. : *Fourier Analysis on Groups. Paperback edition.* Wiley Classics Library; A Wiley-Interscience Publication. New York etc.: John Wiley & Sons., 1990.

Rudin (1991). Rudin, W. : *Functional Analysis. 2nd ed.* International Series in Pure and Applied Mathematics. New York, NY: McGraw-Hill, 1991.

Schaefer (1974). Schaefer, H. H. : *Banach lattices and positive operators*. Springer-Verlag, New York, 1974.

Schaefer (1980). Schaefer, H. H. : *Topological Vector Spaces. 4th corr. printing.* Graduate Texts in Mathematics, 3. New York-Heidelberg-Berlin: Springer-Verlag, 1980.

Silva (2008). Silva, C. E. : *Invitation to ergodic theory*. Vol. 42 of Student Mathematical Library. American Mathematical Society, Providence, RI, 2008.

Stein (1961a). Stein, E. M. : *On limits of seqences of operators*. Ann. of Math. (2) **74**, 1961a, 140–170.

Stein (1961b). Stein, E. M. : *On the maximal ergodic theorem*. Proc. Nat. Acad. Sci. U.S.A. **47**, 1961b, 1894–1897.

Stein (1993). Stein, E. M. : *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals. With the assistance of Timothy S. Murphy.* Princeton Mathematical Series. 43. Princeton, NJ: Princeton University Press, 1993.

Szemerédi (1975). Szemerédi, E. : *On sets of integers containing no k elements in arithmetic progression*. Acta Arith. **27**, 1975, 199–245, collection of articles in memory of Juriĭ Vladimirovič Linnik.

Tao (2007). Tao, T., 2007. : *The dichotomy between structure and randomness, arithmetic progressions, and the primes*. In: International Congress of Mathematicians. Vol. I. Eur. Math. Soc., Zürich, pp. 581–608.

Tao (2008a). Tao, T. : *Topics in Ergodic Theory*, 2008a, http://terrytao.wordpress.com/category/254a-ergodic-theory/.

Tao (2008b). Tao, T., 2008b. : *The van der Corput trick, and equidistribution on nilmanifolds*. In: Topics in Ergodic Theory. http://terrytao.wordpress.com/2008/06/14/the-van-der-corputs-trick-and-equidistribution-on-nilmanifolds.

van der Waerden (1927). van der Waerden, B. L. : *Beweis einer Baudetschen Vermutung.* Nieuw Archief **15**, 1927, 212–216.

von Neumann (1932a). von Neumann, J. : *Einige Sätze über messbare Abbildungen*. Ann. of Math. (2) **33** (3), 1932a, 574–586.

von Neumann (1932b). von Neumann, J. : *Proof of the quasi-ergodic hypothesis.* Proc. Nat. Acad. Sci. USA **18**, 1932b, 70–82.

Walters (1982). Walters, P. : *An introduction to ergodic theory*. Vol. 79 of Graduate Texts in Mathematics. Springer-Verlag, New York, 1982.

Weyl (1916). Weyl, H. : *Über die Gleichverteilung von Zahlen mod. Eins*. Math. Ann. **77** (3), 1916, 313–352.

Wiener (1939). Wiener, N. : *The ergodic theorem*. Duke Math. J. **5** (1), 1939, 1–18.

Wigner (1960). Wigner, E. : *The unreasonable effectiveness of mathematics in the natural sciences*. Comm. Pure Appl. Math. **13**, 1960, 1–14.

Willard (2004). Willard, S. : *General topology*. Dover Publications Inc., Mineola, NY, 2004.

Yosida (1978). Yosida, K. : *Functional analysis*, 5th Edition. Springer-Verlag, Berlin, 1978.

Yosida and Kakutani (1939). Yosida, K., Kakutani, S. : *Birkhoff's ergodic theorem and the maximal ergodic theorem*. Proc. Imp. Acad., Tokyo **15**, 1939, 165–168.

Letter 1

Dear Participants,

with Lecture 1, uploaded on October 20 at

http://isem.mathematik.tu-darmstadt.de/isem/Lecture1

the ISEM 2008/09 on ERGODIC THEORY is on its way.

We welcome all participants and their local coordinators and hope that you all will enjoy the mathematics and the non-local communication via our website.

The ISEM 08/09 differs from the eleven previous seminars in two respects.

1. "Time ist discrete" this year. While we still consider dynamical systems and time evolution, there will be no differential equations and one-parameter semigroups.

2. "Four lectures at three different places" will write the lectures, but try to produce a homogeneous course.

More information on how the ISEM works will follow in due course.

The task of writing up and publishing the first week’s exercises is assigned to the T¨bingen group.

Please, don’t forget about last week’s homework and edit your personal data on the ISEM website if necessary.

We are looking forward to a productive and enjoyable cooperation.

The virtual lecturers

--
Tanja Eisner
Balint Farkas
Markus Haase
Rainer Nagel
--

Letter 2

Dear Participants,

with Lecture 2, posted on 29 October, we start the systematic investigation of dynamical systems.

Last week we explained that it is not realistic to require the exact knowledge of the states and why it is reasonable to consider observables instead of the states.

Anyhow, in the present and the following lecture we are interested in the behaviour of single states as time passes. The natural questions to ask are:

Starting from a particular state can we reach another given one? Can we reach all states (at least approximately)? Or, if we start from two different states that are near to each other, will they remain close? Will a state return to itself, and when yes how often?

Already the formulation of these questions suggests that the correct setting is that of topological spaces, in which "approximately" and "near" have precise meaning. Lecture 2 and 3 therefore are devoted to an introduction into "topological dynamics".

In the first lecture we provide an abundance of examples which will frequently reappear in this course.

As announced we ask the teams from Budapest and Dresden to post their solutions to this week’s exercises on our web page.

Best regards

ISem-Team

Letter 3

Dear Participants,

in Lecture 3, posted today, we continue the study of topological dynamical systems and investigate how a particular state evolves in time. Does it go away "to infinity" or will it return to itself form time to time which is the most interesting behaviour (to us).

Consider the for instance the one point compactification $\mathbf{N} \cup \{\infty\}$ of the set of natural numbers and the map that increases the integers by 1, $n \mapsto n+1$, leaving infinity fixed, $\infty \mapsto \infty$. Clearly any finite point n in this system approaches the point ∞ as time passes, whereas ∞ is the only fixed point in this TDS. All these properties are very much connected through the general fact that in every TDS we always find a non-empty closed invariant subset A which is "minimal" in the TDS.

G.D.Birkhoff called the elements of such minimal TDSs recurrent (returning) points. In our example, ∞ is the only point that returns to itself from time to time.

Birkhoff then showed that any recurrent point is even regularly recurrent, in our terminology almost periodic, meaning that the waiting times between two subsequent approximate returns remain bounded. Clearly, in the above TDS ∞ is even a periodic (actually a fixed) point.

The present lecture is centered around the recurrence theorem of Birkhoff and some simple applications to number theory.

This week, some exercises are marked with a small $\star$, meaning that either the exercise is only indirectly connected to the subject of the lecture (like Exercise 3.12 on topology) or the exercise is mainly for self-amusement (like Exercise 3.1 and Exercise 3.7).

The new exercises are assigned to the teams from Lille, Lyon, Paris and Darmstadt.

Best regards,
ISem-Team

Letter 4

Dear Participants,

With this new lecture we change our perspective: the dynamics described by a (nonlinear) map phi on a compact (hence "small") space leads to a linear operator T on the (in general: infinite dimensional) Banach space $C(K)$ of all continuous functions on K.

This corresponds to what we mean by "operator-theoretic approach" and we will demonstrate its value in later lectures.

The main aim of the present lecture is to see that by performing this transition no information is lost. More precisely, we shall establish a kind of "dictionary" that translates the basic properties of the TDS on K into properties of the linear system on $C(K)$.

As a matter of fact, functional analytic notions and results become increasingly important. In particular we need a couple of "standard" results about $C(K)$-spaces that we state and use without proof. Readers not familiar with them are encouraged to take them for granted and to see how they are used and where they are important.

We also provide a new Appendix B with some background from functional analysis and operator theory. Very little of the material there is needed at the moment, so do not worry too much.

We, the virtual lecturers, would like to express at this point our gratitude (and our solidarity) to all who still follow us into the fascinating field of ergodic theory. We know very well the difficulties caused by so many different mathematical disciplines (topology, measure theory, functional analysis, even algebra and number theory) coming into play. But isn't it beautiful to see all these interacting?

The exercises are assigned to the teams from Bucharest, Iasi and Ulm.

Best regards,
Your ISem-Team

Letter 5

Dear Participants,

after the one week break and a series of celebrations in Tübingen (for some of us), the virtual lecturers are back at work and in particular to the ISem. With this lecture we leave topological dynamics towards dynamical systems on (finite) measure spaces. The core notion is that of a probability measure, invariant under a measurable self-map of the underlying set.

The historical reasons why we consider such a situation lie in the attempts of 19th century scholars to derive the second law of thermodynamics from mechanical principles. Central to this attempts is the notion of a (thermodynamical) equilibrium, which is a statistical phenomenon. Trying to put this into rigorous mathematics leads to the notion of an invariant measure.

In today's lecture we look mainly at examples for this situation, postponing a more systematic study to subsequent lectures. To clarify the link to the topological dynamics considered until now, we have included a short section on measures on compact spaces. Here – once more – we had to quote some results from the literature, and we ask you to accept this with patience and benevolence.

The exercises go to the groups of Delft, Helsinki and Prague.

With best regards,
The ISEM-team

Letter 6

Dear Participants,

we start our systematic study of measure-preserving systems with two central notions: recurrence and ergodicity.

In topological dynamics, recurrence is a property of an individual point. However, due to the presence of null sets, in measurable dynamics pointwise notions are meaningless. We therefore have to consider subsets (of positive measure) of the state space, or better: equivalence classes of subsets (differing only by null sets). This is the same procedure as in the functional analysis approach to L^p-spaces, and we hope that you get accustomed quickly to this kind of identifications.

We shall prove and discuss Poincar’s famous recurrence theorem, one of the cornerstones of modern ergodic theory. Then we pass to the concept of ergodicity, which is the analogue of minimality in the measure-theoretic context. We shall see later that it is the mathematical concept that comes close to Boltzmann’s orginal "Ergodenhypothese" (which is wrong as it stands).

To show ergodicity of a concretely given MDS is one of the major issues in ergodic theory. We shall see several examples of ergodic MDSs during the course, but for most of them we need more tools than we have at the moment. So, again, we have to appeal to your patience.

The exercises are assigned to Karlsruhe and Delhi.

Have fun,
The ISem-team

Letter 7

Dear Participants,

ORDER is the key to this lecture and to our operator theoretic approach to MDSs. For this we need a function space over the state space and the corresponding induced linear operator reflecting all relevant properties of the underlying dynamical system.

In the case of a TDS the Banach algebra $C(K)$ and the induced algebra homomorphism did the job (see Lecture 4). For MDSs, the appropriate space is the space L^1 of all integrable functions (or L^p if you prefer some other p). On this space, a natural order is given making it a Banach lattice and the induced operator becomes a lattice homomorphism. We will introduce (and occasionally use) notions and arguments from the theory of abstract ordered vector spaces, a theory which is typical for the Tübingen school of functional analysis under H.H. Schaefer, and therefore is at the heart of some of his students still around.

If you are unfamiliar with (or dislike) this abstract order, you may "throw it away" and think of and work with the concrete order of functions.

The main message of this lecture is that two MDSs are (algebra) isomorphic if and only if their corresponding linear dynamical systems on L^1 are (lattice) isomorphic. Hence an operator theoretic approach to ergodic theory should work what we shall show in the following lectures.

The exercises go to the teams from Torun and Columbia.

Best wishes,

ISem-Team

Letter 8

Dear Participants,

Just before the Christmas break we arrive at the first major theorem of ergodic theory: John von Neumann proved in 1931 what is now called his mean ergodic theorem thereby (simultaneously with Birkhoff and his individual ergodic theorem) putting Boltzmann's "Ergodenhypothese" onto solid mathematical ground.

However, von Neumann's theorem (and its proof) immediately indicates that it belongs to the theory of bounded linear operators on Hilbert or Banach spaces. Therefore we choose this framework to introduce "mean ergodic operators" by the convergence of the arithmetic means (called Cesaro means) of their powers. It is a (small) miracle that a simple algebraic identity combined with some weak compactness implies strong convergence, i.e. mean ergodicity.

We discuss this property systematically and illustrate it with a series of examples. The ties to ergodic theory will become evident only in the following lecture, i.e., next year.

The supplement should help you to bridge the long gap before the next lecture will be posted (on 7 January, 2009).

The exercises are assigned to the teams from Ferrara, Lecce and Ljubljana.

We wish you a merry Christmas and a happy New Year! We hope to see you all again in 2009 (virtually).

Best wishes,

ISem-Team

Letter 9

Dear Participants,

Welcome to the ISEM again. With our best wishes for 2009 we send you Lecture 9 which, indeed, is just a second part of Lecture 8. There we studied mean ergodic operators on general Banach spaces, now we look at this property for the operators induced by MDSs or TDSs. While mean ergodicity is more or less automatic in the MDS case, it is a rather strong property for TDS. In this case, the operators induced on $\mathrm{C}(G)$ by rotations on a compact group G are our main (and up to now only interesting) examples of mean ergodic operators.

However, at the end of this lecture the pendulum swings back from the operator theory to elementary number theory and we obtain Weyl's Equidistribution Theorem as a corollary of our abstract results.

This week there are again optional exercises marked by a small star.

The exercises are assigned to the teams from Pisa, Szeged and Zaragoza.

Best regards,

Your ISem-Team

Letter 10

Dear Participants,

The question "whether time mean equals space mean" was the motivating force at the very beginning, and it took almost 50 years until G.D. Birkhoff in 1931 made a theorem out of Boltzmann’s "Ergodenhypothese". In this lecture we prove an operator theoretic generalisation of Birkhoff’s theorem and show its connection to probability and number theory.

The main difficulty (and challenge) lies in the fact that the methods of linear functional analysis do not suffice to prove the main result. Instead, one has to involve a non-linear mapping, the so-called maximal operator, and furthermore lattice-theoretic notions prevail.

These elegant methods are also the principle tools in the theory of singular integrals and modern Fourier analysis, and dear to the Dutch part of the ISem team.

The exercises are assigned to the teams from Shanghai, Sydney and Valencia.

Have fun,

Your ISem team

Letter 11

Dear Participants,

as we have already seen, ERGODIC THEORY has many connections to other mathematical disciplines. It is motivated by problems in mathematical physics. It builds on topology, measure theory and functional analysis. It interprets probability theory and has applications to number theory. Now it is the turn of ALGEBRA (group theory, to be precise) to enter the stage.

Our main objective is to define and study "trigonometric polynomials" on a general compact Abelian group. This problem belongs to the so-called 'duality theory of locally compact groups', a part of abstract harmonic analysis. Since we do not assume that you are all familiar with it, we devote to it the major part of the present lecture, including an ample supplement. (We tried to be self-contained, yet we ask you to accept without proof the existence of a Haar measure.)

Although we describe some simple applications to dynamical systems (mean ergodicity of group rotations) and number theory (Kronecker's theorem for the n-torus), the most impressive ones will be presented only in the following two lectures. Hence, for the time being, you may consider the material as a contribution to your general mathematical education.

At this point we take the opportunity to thank PETRA CSOMÓS for having been the "office" of this year's ISem. Since last October she patiently has answered all your e-mails, cared for the web site and helped to run the ISem smoothly. She is now looking forward to new duties (and pleasures), but we hope to see her (with her baby) at the workshop in Blaubeuren again.

Köszönjük Petra, and good luck.

This week the Tübingen team volunteered to solve the exercises, we ask Darmstadt to help them.

Best wishes

ISem-Team

Letter 12

Dear Participants,

this week's ISEM lecture features one of the core results of the operator-theoretic approach to ergodic theory: the Jacobs-de Leeuw-Glicksberg decomposition. Developed around 1960, this theorem has quickly become central to asymptotic theory of dynamical systems, discrete and continuous.

Our approach to it (which is actually the one of de Leeuw and Glicksberg themselves) is based on the abstract semigroup theory provided in the previous lecture, in particular on the existence of a smallest ideal, which is a compact group, in each compact semitopological semigroup.

The supplement of the lecture contains a classical example (based on a theorem of Grothendieck) satisfying the hypotheses of the decomposition theorem. Originally we thought that we would need this example for the main results, and so we worked out a proof. In the end it turned out that it was not necessary, but since the result and its proof are nice (as we think), we included it nevertheless. You may skip it without harm, though.

This week we still do quite abstract operator theory, but this will change soon when we classify dynamical systems "with discrete spectrum" and discuss mixing in MDSs. All the hard work will pay off eventually; so that was the crux of the matter!

The exercises of this lecture are assigned to the team from Dresden.

Have fun,

ISem-Team

Letter 13

Dear Participants,

so far in these lectures, a frequent advice was: "look at the group rotation". Indeed, rotations on compact groups served us well to illustrate many interesting phenomena occuring in dynamical systems. However it was only in the last lecture that (compact Abelian) groups showed their full power. Here the Haar measure proved to be of fundamental importance. In this context, an anonymous poet noted:

Said a mathematician named Haar,
"Von Neumann can't see very far.
He missed a great treasure
- They call it Haar measure -
Poor Johnny's just not up to par."

Von Neumann was a universal mathematician: his work ranges from set theory via algebra to analysis (and aerodynamics, nuclear physics (including the atomic bomb) and operator algebras and game theory and computer science and ...). Together with his student and later colleague Paul Halmos (both, like Alfred Haar, of Hungarian orgin) he found one of the central results in the structural theory of MDSs.

This theorem states that group rotations are universal examples for measure preserving systems. More precisely, every ergodic MDS with discrete spectrum is isomorphic to a group rotation on a compact group. In view of the JdLG-decomposition from the last lecture, we see that group rotations are essential building blocks of MDSs.

You will therefore understand that we cannot agree at all to the last line of the limerick above.

Exercises go to the team from Budapest.

Have fun,

ISem-Team

Letter 14

Dear Participants,

Ergodic theory proposes several mathematical models for the daily life phenomenon of "mixing". We look at three different concepts, and you should notice the conflict between "natural" and "mathematically simple and interesting". The guideline is again the JdLG-decomposition and now, instead of the reversible part as in Lecture 13, the stable part will play the main role.

With this lecture we hope to have laid down a solid foundation of (the operator theoretic aspects of) ergodic theory and the final Lecture 15 next week can be considered as a dessert. Or: as an appetizer for more interesting topics to study in phase 2 and 3 of our seminar.

Best wishes

ISem-Team

Letter 15

Dear Participants,

With this lecture we arrive at the end of our course. In fact, we even return to its very beginning (cf. recurrence) where we promised applications of ergodic theory to certain number theoretic problems. Such applications blossom since H. Furstenberg in 1977 established his correspondence principle building a bridge from (combinatorial) number theory to ergodic theory (and vice versa). To allow you a glimpse from this bridge onto (parts of) the mathematical landscape was one of our main motivations, and Lecture 15 is entirely devoted to this task.

With this lecture we also arrive at the end of the beginning. Soon we will continue with Phase 2 of the seminar in which we propose a list of projects in the spirit of the course. You will be invited to apply for such a project and, if accepted, you are supposed to work on the project with 2 or 3 team members (from different places). These projects will then be presented at the final workshop from June 7 to 13, 2009 at the Heinrich Fabri Institut Blaubeuren (max. capacity 50 participants).

At this point we want to express our sincere thanks to all participants, to the silent ones as well to those having solved exercises and made comments (special cheers to Dresden!). It was a pleasure to work with you.

Your virtual lecturers

Balint, Markus, Rainer, Tanja