Ergodic Theory in the Perspective of Functional Analysis

13 Lectures

by Roland Derndinger, Rainer Nagel, Günther Palm

(uncompleted version)

In July 1984 this manuscript has been accepted for publication in the series "Lecture Notes in Mathematics" by Springer-Verlag. Due to a combination of unfortunate circumstances none of the authors was able to perform the necessary final revision of the manuscript.

Tübingen, July 1987

I. What is Ergodic Theory?

The notion “ergodic” is an artificial creation, and the newcomer to “ergodic theory” will have no intuitive understanding of its content: “elementary ergodic theory” neither is part of high school- or college- mathematics (as does “algebra”) nor does its name explain its subject (as does “number theory”). Therefore it might be useful first to explain the name and the subject of “ergodic theory”. Let us begin with the quotation of the first sentence of P. Walters’ introductory lectures (1975, p. 1):

“Generally speaking, ergodic theory is the study of transformations and flows from the point of view of recurrence properties, mixing properties, and other global, dynamical, properties connected with asymptotic behavior.”

Certainly, this definition is very systematic and complete (compare the beginning of our Lectures III. and IV.). Still we will try to add a few more answers to the question: “What is Ergodic Theory ?”

Naive answer: A container is divided into two parts with one part empty and the other filled with gas. Ergodic theory predicts what happens in the long run after we remove the dividing wall.

First etymological answer: εργοδης=difficult.

Historical answer:

1880	-	Boltzmann, Maxwell	-	ergodic hypothesis
1900	-	Poincaré	-	recurrence theorem
1931	-	von Neumann	-	mean ergodic theorem
1931	-	Birkhoff	-	individual ergodic theorem
1958	-	Kolmogorov	-	entropy as an invariant
1963	-	Sinai	-	billiard flow is ergodic
1970	-	Ornstein	-	entropy classifies Bernoulli shifts
1975	-	Akcoglu	-	individual L^p-ergodic theorem

Naive answer of a physicist: Ergodic theory proves that time mean equals space mean.

I.E. Farquhar’s [1964] answer: “Ergodic theory originated as an offshot of the work of Boltzmann and of Maxwell in the kinetic theory of gases. The impetus provided by the physical problem led later to the development by pure mathematicians of ergodic theory as a branch of measure theory, and, as is to be expected, the scope of this mathematical theory extends now far beyond the initial field of interest. However, the chief physical problems to which ergodic theory has relevance, namely, the justification of the methods of statistical mechanics and the relation between reversibility and irreversibility have been by no means satisfactorily solved, and the question arises of how far the mathematical theory contributes to the elucidation of these physical problems.”

Physicist's answer:

Reality	Physical model	Mathematical consequences
A gas with n particles at time $t = 0$ is given.	The "state" of the gas is a point x in the "state space" $X = \mathbb{R}^{6n}$.	
Time changes	Time change is described by the Hamiltonian differential equations. Their solutions yield a mapping $\varphi : X \to X$, such that the state x_0 at time $t = 0$ becomes the state $x_1 = \varphi(x_0)$ at time $t = 1$.	**Theorem of Liouville:** φ preserves the (normalized) Lebesgue measure μ on X.
the long run behavior is observed.	**Definition:** An observable is a function $f : X \to \mathbb{R}$, where $f(x)$ can be regarded as the outcome of a measurement, when the gas is in the state $\in X$. **Problem:** Find $\lim f(\varphi^n(x))$!	
1st objection: Time change is much faster than our observations. **2nd objection:** In practice, it is impossible to determine the state x.	**Modified problem:** Find the time mean $M_t f(x) := \lim \frac{1}{n} \sum_{i=0}^{n-1} f(\varphi^i(x))$! **Additional hypothesis** (ergodic hypothesis): Each particular motion will pass through every state consistent with its energy (see P.u.T. Ehrenfest 1911).	**"Theorem" 1:** If the ergodic hypothesis is satisfied, we have $M_t f(x) = \int f \, \mathrm{d}\mu =$ space mean, which is independent of the state x. **"Theorem" 2:** The ergodic hypothesis is "never" satisfied.

Ergodic theory looks for better ergodic hypothesis and better "ergodic theorems".

Commonly accepted etymological answer:

ἔργον = energy

–ὁδός = –path (P. and T. Ehrenfest 1911, p. 30)

"Correct" etymological answer:

ἔργον	=	energy	
–ῶδης	=	–like	(Boltzmann 1884/85, see also III.)

K. Jacobs' [1965] answer:
"... als Einführung für solche Leser gedacht, die gern einmal erfahren möchten, womit sich diese Theorie mit dem seltsamen, aus den griechischen Wörtern εργον (Arbeit) und οδος (Weg) zusammengesetzten Namen eigentlich beschäftigt. Die Probleme der Ergodentheorie kreisen um einen Begriff, der einerseits so viele reizvolle Spezialfälle umfaßt, daß sowohl der Polyhistor als auch der stille Genießer auf ihre Kosten kommen, andererseits so einfach ist, daß sich die zentralen Ergebnisse und Probleme der Ergodentheorie leicht darstellen lassen; diese einfach zu formulierenden Fragestellungen erfordern jedoch bei naherer Untersuchung oft derartige Anstrengungen, daß harte Arbeiter hier ihr rechtes Vergnügen finden werden."

J. Dieudonne's [1977] answer:
"Le point de départ de la théorie ergodique provient du développement de la mécanique statistique et de la theorie cinétique des gaz, où l'expérience suggère und tendence à l'"uniformite": si l'on considère à un instant donné un mélange hétérogène de plusieurs gaz, l'évolution du mélange au cours du temps tend à le rendre homogéne."

W. Parry's [1981] answer:
"Ergodic Theory is difficult to characterize, as it stands at the junction of so many areas, drawing on the techniques and examples of probability theory, vector fields on manifolds, group actions on homogeneous spaces, number theory, statistical mechanics, etc..."" (e.g. functional analysis; added by the authors).

Elementary mathematical answer:
Let X be a set, $\varphi : X \to X$ a mapping. The induced operator T_φ maps functions $f : X \to \mathbb{R}$ into $T_\varphi f := f \circ \varphi$. Ergodic theory investigates the asymptotic behavior of φ^n and T_φ^n for $n \in \mathbb{N}$.

Our answer:
More structure is needed on the set X, usually at least a topological or a measure theoretical structure. In both cases we can study the asymptotic behavior of the powers T^n of the linear operator $T = T_\varphi$, defined either on the Banach space $C(X)$ of all continuous functions on X or on the Banach space $L^1(X, \Sigma, \mu)$ of all μ-integrable functions on X.

II. Dynamical Systems

Many of the answers presented in Lecture I indicate that ergodic theory deals with pairs (X, φ) where X is a set whose points represent the "states" of a physical system while φ is a mapping from X into X describing the change of states after one time unit. The first step towards a mathematical theory consists in finding out which abstract properties of the physical state spaces will be essential. It is evident that an "ergodic theory" based only on set-theoretical assumptions is of little interest. Therefore we present three different mathematical structures which can be imposed on the state space X and the mapping φ in order to yield "dynamical systems" that are interesting from the mathematical point of view. The parallel development of the corresponding three "ergodic theories" and the investigation of their mutual interaction will be one of the characteristics of the following lectures.

II.1 Definition:

(i) $(X, \Sigma, \mu; \varphi)$ is a *measure-theoretical dynamical system* (briefly: MDS) if (X, Σ, μ) is a probability space and $\varphi : X \to X$ is a bi-measure-preserving transformation.

(ii) $(X; \varphi)$ is a *topological dynamical system* (TDS) if X is a compact space and $\varphi : X \to X$ is homeomorphism.

(iii) $(E; T)$ is a *functional-analytic dynamical system* (FDS) if E is a Banach space and $T : E \to E$ is a bounded linear operator.

Remarks:

1. The term "bi-measure-preserving" for the transformation $\varphi : X \to X$ in (i) is to be understood in the following sense: There exists a subset X_0 of X with $\mu(X_0) = 1$ such that the restriction $\varphi_0 : X_0 \to X_0$ of φ is bijective, and both φ_0 and its inverse are measurable and measure-preserving for the induced σ-algebra $\Sigma_0 := \{A \cap X_0 : A \in \Sigma\}$.
2. If φ is bi-measure-preserving with respect to μ, we call μ a φ-invariant measure.
3. As we shall see in (II.4) every MDS and TDS leads to an FDS in a canonical way. Thus a theory of FDSs can be regarded as a joint generalization of the topological theory of TDSs and the probabilistic theory of MDSs. In most of the following chapters we will either start from or aim for a formulation of the main theorem(s) in the language of FDSs.
4. DDSs *("differentiable dynamical systems")* will not be investigated in these lectures (see Bowen [1975], Smale [1967], [1980]).

Before proving any results we present in this lecture the fundamental (types of) examples of dynamical systems which will frequently reappear in the ensuing text. The reader is invited to apply systematically every definition and result to at least some of these examples.

II.2. Rotations:

(i) Let $\Gamma = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle, Σ its Borel algebra, and m the normalized Lebesgue measure on Γ. Choose an $a \in \Gamma$ and define

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

Clearly, $(\Gamma; \varphi_a)$ is a TDS, and $(\Gamma, \mathcal{B}, m; \varphi_a)$ an MDS.

(ii) A more abstract version of the above example is the following: Take a compact group G with Borel algebra $\mathcal{B}$ and normalized Haar measure m. Choose $h \in G$ 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, and $(G, \mathcal{B}, m; \varphi_h)$ an MDS.

II.3. Shifts:

(i) "Dough-kneading" leads to the following bi-measure-preserving transformation

or in a more precise form: if $X := [0,1]^2$, $\mathcal{B}$ the Borel algebra on X, m the Lebesgue measure, and

$$\varphi(x,y) := \begin{cases} (2x, \frac{y}{2}) & \text{for } 0 \leqslant x \leqslant \frac{1}{2}, \\ (2x-1, \frac{(y+1)}{2} & \text{for } \frac{1}{2} < x \leqslant 1, \end{cases}$$

we obtain an MDS, but no TDS for the natural topology on X.

(ii) "Coin-throwing" may also be described in the language of dynamical systems: Assume that somebody throws a dime once a day from eternity to eternity. An adequate mathematical description of such an "experiment" is a point

$$x = (x_n)_{n \in \mathbb{Z}}$$

in the space $\hat{X} := \{0,1\}^{\mathbb{Z}}$, which is compact for the product topology.

$$\begin{array}{rl} \text{Tomorrow, the point} & (x_n) = (\ldots\ldots, x_{-1}, x_0, x_1, x_2, \ldots) \\ & \qquad\qquad\qquad\qquad\qquad \updownarrow \\ \text{will be} & (x_{n+1}) = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots\ldots) \end{array}$$

where the arrow points to the current outcome of the dime-throwing experiment. Therefore, time evolution corresponds to the mapping

$$\tau : \hat{X} \to \hat{X}, \quad (x_n)_{n \in \mathbb{Z}} \mapsto (x_{n+1})_{n \in \mathbb{Z}}.$$

$(\hat{X}; \tau)$ is a TDS, and τ is called the (left)*shift* on $\hat{X}$. Let us now introduce a probability measure $\hat{\mu}$ on $\hat{X}$ telling which events are probable and which not. If we assume firstly, that this measure should be determined by its values on the (measurable) rectangles in $\hat{X}$ (see A.17), and secondly, that the probability of the outcome should not change with time, we obtain that $\hat{\mu}$ is a shift invariant probability measure on the product σ-algebra $\hat{\Sigma}$ on $\hat{X}$, and that $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ is an MDS.

On $\hat{X}$ there are many τ-invariant probability measures, but in our concrete case, it is reasonable to assume further that today's outcome is independent of all the previous results, and that the two possible results of "coin throwing" have equal probabilities $p(0) = p(1) = \frac{1}{2}$. Then $(\hat{X}, \hat{\Sigma}, \hat{\mu})$ is the product space $(\{0,1\}, \mathcal{P}\{0,1\}, p)^{\mathbb{Z}}$ (see A.17).

Exercise: Show that (i) and (ii) are the "same" ! (Hint: see (VI.D.2))

(iii) Again we present an abstract version of the previous examples. Let (X, Σ, p) be a probability space, where $X := \{0, \ldots, k-1\}$, $k > 1$, is finite, Σ the power set of X and $p = (p_0, \ldots, p_{k-1})$ a probability measure on X.

Take $\widehat{X} = X^{\mathbb{Z}}$, the product σ-algebra $\widehat{\Sigma}$ on X, the product measure $\widehat{\mu}$ and the shift τ on $\widehat{X}$. Then we obtain an MDS $(\widehat{X}, \widehat{\Sigma}, \widehat{\mu}; \tau)$, called the *Bernoulli shift* with distribution p and denoted by $B(p_0, \ldots, p_{k-1})$.

II.4. Induced operators:
Very important examples of FDSs arise from TDSs and MDSs as follows:
(1) Let $(X; \varphi)$ be a TDS and let $C(X)$ be the Banach space of all (real- or complex-valued) continuous functions on X (see B.18). Define the "induced operator"

$$T_\varphi : f \mapsto f \circ \varphi \quad \text{for } f \in C(X).$$

It is easy to see that T_φ is an isometric linear operator on $C(X)$, and hence $(C(X); T_\varphi)$ is an FDS. Moreover, we observe that T_φ is a lattice isomorphism (see C.5) and thus a positive operator on the Banach lattice $C(X)$ (see C.1 and C.2). On the other hand, if we consider the complex space $C(X)$ as a C^*-algebra (see C.6 and C.7) it is clear that T_φ is a $*$-algebra isomorphism (see C.8).

(2) Let $(X, \Sigma, \mu; \varphi)$ be an MDS and consider the function spaces $L^p(X, \Sigma, \mu)$, $1 \leqslant p \leqslant \infty$ (see B.20). Define

$$T_\varphi : f \mapsto f \circ \varphi \quad \text{for } f \in L^p(X, \Sigma, \mu),$$

or more precisely: $T_\varphi \check{f} := \widetilde{f \circ \varphi}$ where $\check{f}$ denotes the equivalence class in $L^p(X, \Sigma, \mu)$ corresponding to the function f. Again, the "induced operator" T_φ is an isometric (resp. unitary) linear operator on $L^p(X, \Sigma, \mu)$ (resp. on $L^2(\mu)$) since φ is measure-preserving, and hence $(L^p(X, \Sigma, \mu); T_\varphi)$ is an FDS. As above, T_φ is a lattice isomorphism if we consider $L^p(X, \Sigma, \mu)$ as a Banach lattice (see C.1 and C.2). Finally, the space $L^\infty(X, \Sigma, \mu)$ is a commutative C^*-algebra and the induced operator T_φ on $L^\infty(X, \Sigma, \mu)$ is a $*$-algebra isomorphism.

Remark: Via the representation theorem of Gelfand-Neumark the case $(L^\infty(\mu); T_\varphi)$ in (2) may be reduced to the situation of (1) above (see **??**). Therefore we are able to switch from measure-theoretical to functional-analytic or to topological dynamical systems. This flexibility is important in order to tackle a given problem with the most adequate methods.

II.5. Stochastic matrices:
An FDS that is not induced by a TDS or an MDS can be found easily: Take $(E; T)$, where E is $\mathbb{R}^k = C(\{0, \ldots, k-1\})$ and T is a $k \times k$-matrix. We single out a particular case or special interest in probability theory: Let T be *stochastic*, i.e. $T = (a_{ij})$ such that $0 \leqslant a_{ij}$ and $\sum_{j=0}^{k-1} a_{ij} = 1$ for $i = 0, 1, \ldots, k-1$. Then $(E; T)$ is an FDS and $T\mathbf{1} = \mathbf{1}$ where $\mathbf{1} = (1, ..., 1)$. The matrix T has the following interpretation in probability theory. We consider $X = \{0, 1, \ldots, k-1\}$ as the "state space" of a certain system, and T as a description of time evolution of the states in the following senses a_{ij} denotes the probability that the system moves from state i to state j in one time step and is called the *"transition probability"* from i to j. Thus T (resp. $(E; T)$) can be regarded as a "stochastic" version of a dynamical system. Indeed, if every row and every column of T contains a 1 (and therefore only zeros in the other places), then the system is "deterministic" in the sense that T is induced by a mapping (permutation) $\varphi : X \to X$ (resp. $(E; T)$ is induced by a TDS (X, φ)).

II.6. Markov shifts:
Let $T : \mathbb{R}^k \to \mathbb{R}^k$ be a stochastic matrix (a_{ij}) as in (II.5). Let $\mu = (p_0, \ldots, p_{k-1})^\top$ be an invariant probability vector, i.e.

$$p_i \geqslant 0, \quad \sum_{i=0}^{k-1} p_i = 1$$

and μ is invariant under the adjoint of T, i.e. $\sum_{i=0}^{k-1} a_{ij} p_i = p_j$ for all j (it is well known and also follows from (IV.5) and (IV.4).e that there are such non-trivial invariant vectors). We call μ the *probability distribution* at time 0, and the probabilistic interpretation of the entries a_{ij} (see II.5) gives us a natural way of defining probabilities on

$$\hat{X} := \{0, 1, \ldots, k-1\}^{\mathbb{Z}} = \{(x_i)_{i \in \mathbb{Z}} : x_i \in \{0, 1, \ldots, k-1\}\}$$

with the product σ-algebra $\hat{\Sigma}$. For $0 \leqslant l \leqslant k-1$, $\mathrm{pr}[x_0 = l]$ denotes the probability that $x \in \hat{X}$ is in the state l at time 0. We define

$$\begin{aligned} \mathrm{pr}[x_0 = l] &:= p_l \\ \mathrm{pr}[x_0 = l,\ x_1 = m] &:= p_l a_{lm} \\ \mathrm{pr}[x_0 = l_0,\ x_1 = l_1, \ldots, x_t = l_t] &:= p_{l_0} a_{l_0 l_1} a_{l_1 l_2} \cdots a_{l_{t-1} l_t}. \end{aligned}$$

Moreover, since μ is invariant,

$$\begin{aligned} &\mathrm{pr}[x_1 = l] = \sum_{i=0}^{k-1} \mathrm{pr}[x_0 = i, x_1 = l] = \sum_{i=0}^{k-1} p_i a_{il} = p_l = \mathrm{pr}[x_0 = l], \\ &\mathrm{pr}[x_t = l] = p_l = \mathrm{pr}[x_0 = l], \quad \text{and finally} \end{aligned}$$

$$(*) \qquad \begin{aligned} &\mathrm{pr}[x_s = l_0, x_{s+1} = l_1, \ldots, x_{s+t} = l_t] = p_{l_0} a_{l_0 l_1} a_{l_1 l_2} \cdots a_{l_{t-1} l_t} = \\ &\mathrm{pr}[x_0 = l_0, x_1 = l_1, \ldots, x_t = l_t] \quad \text{for any choice of } s \in \mathbb{Z},\ t \in \mathbb{N}_0 \\ &\qquad \text{and } l_0, \ldots, l_t \in \{0, \ldots, k-1\} \end{aligned}$$

The equation $(*)$ gives a probability measure on each algebra $\mathcal{F}_m := \{A \in \hat{\Sigma} : A = \bigcap_{i=-m}^{m} [x_i \in A_i], A_i \subseteq X\}$. By (A.17) this determines exactly one probability measure μ on the product σ-algebra $\hat{\Sigma}$ on $\hat{X}$. This measure μ is obviously invariant under the shift

$$\tau : (x_n) \mapsto (x_{n+1})$$

on $\hat{X}$. Therefore $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ is an MDS, called the *Markov shift* with *invariant distribution* μ and *transition matrix* T.

Note that the examples (II.5) and (II.6), although they describe the same stochastic process, are quite different, because the operator T of (II.5) is not induced by a transformation of the state space $\{0, 1, \ldots, k-1\}$, whereas in (II.6) the shift τ is defined on the state space $\{0, 1, \ldots, k-1\}^{\mathbb{Z}}$. We have refined (i.e. enlarged) the state space of (II.5) to make the model "deterministic".

An analogous construction can be carried out in the infinite-dimensional case for so-called Markov-operators (see App. U and X), or for transition probabilities (see Bauer).

This construction is well-known in the theory of Markov processes; its functional-analytic counterpart, the so-called *dilation*, will be presented in App. U.

Exercise: The Bernoulli shift $B(p_0, \ldots, p_{k-1})$ is a Markov shift. What is its invariant distribution and its transition matrix?

II.D Discussion

II.D.1. Non-bijective dynamical systems:
It is clear, that the Definitions (II.1.i,ii) make sense not only for bijective but also for arbitrary measure-preserving, resp. continuous transformations, but we prefer to sacrifice this greater generality for the sake of simplicity. Such non-bijective transformations also induce FDSs by a procedure similar to that in (II.4). Examples are the mappings

$$\varphi : [0,1] \to [0,1] \quad \text{defined by}$$

$$\varphi(t) := \begin{cases} 2t & \text{for } 0 \leqslant t \leqslant \frac{1}{2} \\ 2-2t & \text{for } \frac{1}{2} < t \leqslant 1 \end{cases}$$

or

$$\varphi(t) := 4t(1-t).$$

II.D.2. Banach algebras vs. Banach lattices:
The function spaces used in ergodic theory, i.e. $C(X)$ and $L^p(X, \Sigma, \mu)$, are Banach lattices and the induced operators T_φ are lattice isomorphisms (see II.4 and App.C). Therefore, the vector lattice structure seems to be adequate for a simultaneous treatment of topological and measure-theoretical dynamical systems. If you prefer Banach algebras and algebra isomorphisms, you have to consider the operators T_φ on the spaces $C(X)$ and $L^\infty(X, \Sigma, \mu)$.

II.D.3. Real vs. complex Banach spaces:
Since order structure and positivity makes sense only for real Banach spaces, one could be inclined to study only spaces of *real* valued functions. But methods from spectral theory play a central role in ergodic theory and require *complex* Banach spaces. However, no real trouble is caused, since the complex Banach spaces $C(X)$ and $L^p(X, \Sigma, \mu)$ decompose canonically into real and imaginary parts, and we restrict our attention to the real part whenever we use the order relation. Moreover, the induced operator T_φ (like any positive linear operator) is uniquely determined by its restriction to this real part.

II.D.4. Null sets in (X, Σ, μ):
In the measure-theoretical case some technical problems may be caused by the sets $A \in \Sigma$ with $\mu(A) = 0$. But in ergodic theory, it is customary (and reasonable, as can be understood from the physicist's answer in Lecture I: A is a set of "states" having probability 0) to identify measurable sets which differ only by such a null set. From now on, this will be done without explicit statement. For example, we will say that a measurable function f is constant if

$$f(x) = c$$

for all $x \in X \backslash A$, $\mu(A) = 0$.
The reader familiar with the "function" spaces $L^p(X, \Sigma, \mu)$ realizes that we identify the function with its equivalence class in $L^p(\mu)$, but still keep the terminology

of functions. These subtleties should not disturb the beginner since no serious mistakes can be made (see A.7 and B.20).

II.D.5. Which FDSs are TDSs?

We have seen in II.4 that to every TDS $(X;\varphi)$ canonically corresponds the FDS $(C(X), T_\varphi)$. Since this correspondence occurs frequently in our operator-theoretical approach to ergodic theory, it is important to know which FDSs arise in this way. More precisely: Which operators

$$T : C(X) \to C(X)$$

are induced by a homeomorphism

$$\varphi : X \to X$$

in the sense that $T = T_\varphi$? A complete answer is given as follows.

Theorem: Consider the real Banach space $C(X)$ and $T \in \mathscr{L}(C(X))$. Then the following assertions are equivalent:

(i) T is a lattice isomorphism satisfying $T\mathbf{1} = \mathbf{1}$.
(ii) T is an algebra isomorphism.
(iii) $T = T_\varphi$ for a (unique) homeomorphism φ on X.

Proof. Clearly, (iii) implies (i) and (ii).

(ii) $\Rightarrow$ (iii): Let $D := \{\delta_x : x \in X\}$ be the weak* compact set of all Dirac measures on X. This coincides with the set of all normalized multiplicative linear forms on $C(X)$, and from (C.9) it follows that X is homeomorphic to D. Since T is an algebra isomorphism its adjoint T' maps D on D. The restriction of T' to D defines a homeomorphism φ on X having the desired properties.

(i) $\Rightarrow$ (iii): The proof requires some familarity with Banach lattices. We refer to Schaefer 1974, III.9.1 for the details as well as for the "complex" case of the theorem. ∎

II.D.6. Which FDSs are MDSs?

Due to the existence of null sets (and null functions) the analogous problem in the measure-theoretical context is more difficult: Which operators

$$T : L^p(X, \Sigma, \mu) \to L^p(X, \Sigma, \mu)$$

are induced by a bi-measure-preserving transformation

$$\varphi : X \to X$$

in the sense that $T = T_\varphi$? Essentially, it turns out that the appropriate operators are again the Banach lattice isomorphisms, but we will return to this problem in Lecture VI.

II.D.7. Discrete vs. continuous time:

Applying φ (or T) in a dynamical system may be interpreted as movement from the state x at time t to the state $\varphi(x)$ at time $t + \Delta t$. Therefore, repeated application of φ means advancing in time with a discrete time scale in steps of Δt. Intuitively it is more realistic to consider a continuous time scale, and in our mathematical model the transformation φ and the group homomorphism

$$n \mapsto \varphi^n$$

defined on $\mathbb{Z}$ should be replaced by a continuous group of transformations, i.e. a group homomorphism

$$t \mapsto \varphi_t$$

from $\mathbb{R}$ into an appropriate set of transformations on X. Observe that the "composition rule"

$$\varphi^{n+m} = \varphi^n \circ \varphi^m, \quad n, m \in \mathbb{Z},$$

in the discrete model is replaced by

$$\varphi_{t+s} = \varphi_t \circ \varphi_s, \quad t, s \in \mathbb{R}.$$

Adding some continuity or measurability assumptions one obtains "continuous dynamical systems" (e.g. Rohlin [1966], Chapt. II.). We prefer the simpler discrete model, since we are mainly interested in the asymptotic behavior of the system as t tends to infinity.

II.D.8. From a differential equation to a dynamical system:
In (II.D.7) we briefly discussed the problem "discrete vs.continuous time". Clearly, a "continuous dynamical system" $(X; (\varphi_t)_{t\in\mathbb{R}})$ gives rise to many "discrete dynamical systems" $(X; \varphi)$ by setting $\varphi := \varphi_t$ for any $t \in \mathbb{R}$. We present here a short introduction into the so-called "classical dynamical systems" which arise from differential equations and yield continuous dynamical systems, also called "flows".

Let $X \subseteq \mathbb{R}^n$ be a compact smooth manifold and $f(x)$ a C^1-vector field on X. We consider the autonomous ordinary differential equation

$$(*) \qquad \dot{x} = \frac{\mathrm{d}x}{\mathrm{d}t} = f(x)$$

(or in coordinates: $\dot{x}_i = f_i(x_1, \dots, x_n)$, $i = 1, \dots, n$). It is known that for every $x \in X$ the equation $(*)$ has a unique solution $\varphi_t(x)$ that satisfies $\varphi_0(x) = x$. The uniqueness of the solution implies the group property $\varphi_{t+s} = \varphi_t \circ \varphi_s$ for all $t, s \in \mathbb{R}$, and, in addition, the mapping

$$\begin{aligned} \Phi : \times\mathbb{R} &\to X \\ (x, t) &\mapsto \varphi_t(x) \end{aligned}$$

is continuous (see Nemyckii-Stepanov [1960]). Therefore, $(X; (\varphi_t)_{t\in\mathbb{R}})$ is a continuous topological dynamical system.

II.D.9 Examples:
(i) Let $\Gamma^2 = \mathbb{R}^2/\mathbb{Z}^2$ be the 2-dimensional torus and let

$$\begin{aligned} \dot{x} &= 1 \\ \dot{y} &= \alpha \end{aligned}$$

with $\alpha \neq 0$. The flow (φ_t) on Γ^2 is given by

$$\varphi_t\big(\tbinom{x}{y}\big) = \begin{pmatrix} (x+t) \mod 1 \\ (y+\alpha t) \mod 1 \end{pmatrix}.$$

(ii) Take the space $X = \Gamma^2$ as in (i) and define

$$\begin{aligned} \dot{x} &= F\big(\tbinom{x}{y}\big) \\ \dot{y} &= \alpha \cdot F\big(\tbinom{x}{y}\big) \end{aligned}$$

where F is C^1-function which is 1-periodic in each variable. Assume that F is strictly positive on X. The solution curves of this motion agree with those of (i), but the "speed" is changed.

For applications the above definition of a "continuous topological dynamical system" has three disadvantages: first, the manifold X (the "state" space) is not always compact, second, if X is not compact, in general not every-solution of $(*)$ can be continued for all times t (e.g. the scalar equation $\dot{x} = x^2$), and finally, it is often necessary to consider non-autonomous differential equations, i.e. the C^1-vector field f is defined on $X \times \mathbb{R}$ where X is a manifold. All of these difficulties can be overcome by generalizing the above definition (see Sell [1971].

Next, we want to consider "classical measure-theoretical dynamical systems". The problem of finding a φ_t-invariant measure, defined by a continuous density, is solved by the Liouville theorem (see Nemyckii-Stepanov [1960]). We only present a special case.

Many equations of classical mechanics can be written as a Hamiltonian system of differential equations. Let $q = (q_1, \ldots, q_n)$ (coordinates) and $p = (p_1, \ldots, p_n)$ (moments) be a coordinate system in $\mathbb{R}^{2n}$ and $H(p, q)$ a C^2-function which does not depend on time explicitly. The equations

$$(**) \qquad \begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q} \end{aligned}$$

define a flow on $\mathbb{R}^{2n}$ called the "Hamiltonian flow". The divergence of the vector field $(**)$ vanishes:

$$\frac{\partial}{\partial q}\Big(\frac{\partial H}{\partial p}\Big) + \frac{\partial}{\partial p}\Big(\frac{\partial H}{\partial q}\Big) = 0.$$

Therefore, the measure $\mathrm{d}q_1 \ldots \mathrm{d}q_n \, \mathrm{d}p_1 \ldots \mathrm{d}p_n$is invariant under the induced flow. But the considered state space is not compact and the invariant measure is not finite.

To avoid this difficulty we observe that

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\partial H}{\partial q}\dot{q} + \frac{\partial H}{\partial p}\dot{p} = \frac{\partial H}{\partial q}\frac{\partial H}{\partial p} + \frac{\partial H}{\partial q}\Big(-\frac{\partial H}{\partial q}\Big) = 0$$

i.e. H is a first integral of $(**)$ (conservation of energy!). This means that $X_E := \{(p, q) \in \mathbb{R}^{2n} : H(p, q) = E\}$ for every $E \in \mathbb{R}$ is invariant under the flow. X_E turns out to be a compact smooth manifold for typical values of the constant E, and we obtain on it an "induced" measure by a method similar to the construction of the 1-dimensional Lebesgue measure from the 2-dimensional Lebesgue measure. This induced measure is (φ_t)-invariant and finite, and we obtain "continuous measure-theoretical dynamical systems".

Example linear harmonic oscillator: Let $X = \mathbb{R}^2$ and let $\binom{p}{q}$ be the canonical coordinates on X. For simplicity, we suppose that the constants of the oscillator are all 1. The Hamiltonian function is the sum of the kinetic and the potential energy and therefore

$$H(p,q) = H_{\text{kin}}(p) + H_{\text{pot}}(q) = \frac{1}{2}p^2 + \frac{1}{2}q^2$$

The system $(**)$ becomes

$$\begin{aligned} \dot{q} &= \ \ p \\ \dot{p} &= -q \end{aligned}$$

and the solution with initial value $\binom{p}{q}$ is

$$\varphi_t\big(\tbinom{p}{q}\big) = \begin{pmatrix} \sqrt{p^2+q^2}\,\sin(t+\beta) \\ \sqrt{p^2+q^2}\,\cos(t+\beta) \end{pmatrix},$$

where $\beta \in [0, 2\pi)$ is defined by $\sqrt{p^2+q^2} \cdot \sin\beta = q$ and $\sqrt{p^2+q^2} \cdot \cos\beta = p$. Now, let us consider the surface $H(p,q) = \frac{1}{2}p^2 + \frac{1}{2}q^2 =: E =$constant.

Obviously, E must be positive. For $E = 0$ we have the (invariant) trivial manifold $\{\binom{0}{0}\}$. For $E > 0$ the $(\varphi_t)_{t\in\mathbb{R}}$-invariant manifold

$$X_E := \Big\{\tbinom{p}{q} \in \mathbb{R}^2 : H(p,q) = E\Big\}$$

is the circle about 0 with radius $\sqrt{2E}$, and therefore compact. The "induced" invariant measure on X_E is the 1-dimensional Lebesgue measure, and the induced flow agrees with a flow of rotations on this circle.

II.D.10. Dilating an FDS to an MDS:
We have indicated in (II.D.6) that rather few FDSs on Banach spaces $L^1(\mu)$ are induced by MDSs. But in (II.6) we presented an ingenious way of reducing the study of certain FDSs to the study of MDSs. These constructions are solutions of the following problem:

Let T be a bounded linear operator on $E = L^1(X, \Sigma, \mu)$, $\mu(X) = 1$. Can we find an MDS $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \varphi)$ and operators J and Q, such that the diagram

$$\begin{tikzcd} L^1(X,\Sigma,\mu) \arrow[r, "T^n"] \arrow[d, "J"'] & L^1(X,\Sigma,\mu) \\ L^1(\hat{X},\hat{\Sigma},\hat{\mu}) \arrow[r, "\hat{T}_\varphi^n"'] & L^1(\hat{X},\hat{\Sigma},\hat{\mu}) \arrow[u, "Q"'] \end{tikzcd}$$

commutes for all $n = 0, 1, 2, \ldots$?

If we want the MDS $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \varphi)$ to reflect somehow the "ergodic" behaviour of the FDS $(L^1(X, \Sigma, \mu); T)$, it is clear that the operators J and Q must preserve the order structure of the L^1-spaces (see II.4). Therefore, we call $(L^1(\hat{X}, \hat{\Sigma}, \hat{\mu}); \hat{T}_\varphi)$, resp. $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \varphi)$, a *lattice dilation* of $(L^1(X, \Sigma, \mu); T)$ if - in the diagram above - J is an isometric lattice homomorphism (with $J\mathbf{1} = \hat{\mathbf{1}}$), and Q is a positive contraction. From these requirements it follows that T has to be positive with $T\mathbf{1} = \mathbf{1}$ and $T'\mathbf{1} = \mathbf{1}$. In App. U we show that these conditions are even sufficient.

III. Recurrent, Ergodic and Minimal Dynamical Systems

"Ergodic theory is the study of transformations from the point of view of recurrence properties" (Walters [1975], p. 1). Sometimes, you meet such properties in daily life: If you walk in a park just after it has snowed, you will have to step into your own footprints after a finite number of steps. The more difficult problem of the reappearance of certain celestial phenomena led Poincaré to the first important result of ergodic theory at the end of the last century.

III.1 Definition:
Let $(X, \Sigma, \mu; \varphi)$ be an MDS and take $A \in \Sigma$. A point $x \in A$ is called recurrent to A if there exists $n \in \mathbb{N}$ such that $\varphi^n(x) \in A$.

III.2 Theorem (Poincaré, 1890):
Let $(X, \Sigma, \mu; \varphi)$ be an MDS and take $A \in \Sigma$. Almost every point of A is (infinitely often) recurrent to A.

Proof. For $A \in \Sigma$, $\varphi^{-n}A$ is the set of all points that will be in A at time n (i.e. $\varphi^n(x) \in A$). Therefore, $A_{\text{rec}} := A \cap (\varphi^{-1}A \cup \varphi^{-2}A \cup \dots)$ is the set of all points of A which are recurrent to A.

If $B := A \cup \varphi^{-1}A \cup \varphi^{-2}A \cup \dots$ we obtain $\varphi^{-1}B \subseteq B$ and $A \backslash A_{\text{rec}} = B \backslash \varphi^{-1}B$. Since φ is measure-preserving and μ finite, we conclude

$$\mu(A \backslash A_{\text{rec}}) = \mu(B) - \mu(\varphi^{-1}B) = 0,$$

and thus the non-recurrent points of A form a null set. For the statement in brackets, we notice that $(X, \Sigma, \mu; \varphi^k)$ is an MDS for every $k \in \mathbb{N}$. The above results implies

$$\mu(A_k) = 0 \quad \text{for} \quad A_k := \big\{x \in A : (\varphi^k)^n(x) \notin A \text{ for } n \in \mathbb{N}\big\}.$$

Hence, $A_\infty := \bigcup_{k=1}^{\infty} A_k$ is a null set, and the points of $A \backslash A_\infty$ are infinitely often recurrent to A. ∎

We explained in the physicist's answer in Lecture I that the dynamics can be described by the MDS $(X, \Sigma, \mu; \varphi)$ on the state space

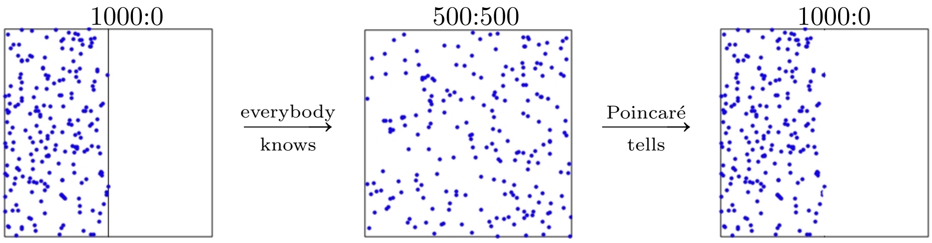
$$X := \{\text{coordinates of the possible locations and impulses of the 1000 molecules in the box}\}$$

As the set A to which recurrence is expected we choose

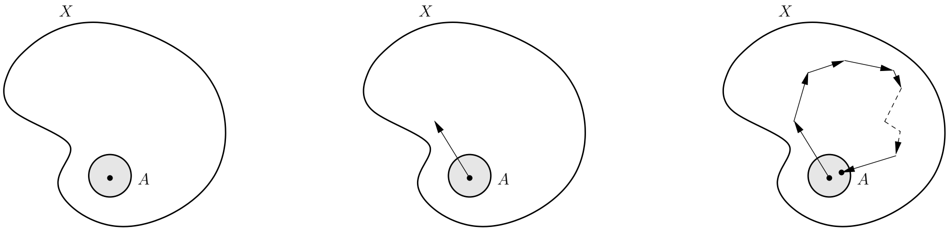
$$A := \{\text{all 1000 molecules are located on the left hand side}\}.$$

Since $\mu(A) > 0$, we obtain from Poincaré's recurrence theorem a surprising conclusion contradicting somehow our daily life experience.

gas container



state space



"Ergodic theory is the study of transformations from the point of view of *mixing properties*" (Walters [1975] p. 1), where "mixing" can even be understood literally (see Lecture IX). In a sense, ergodicity and minimality are the weakest possible "mixing properties" of dynamical systems. Another, purely mathematical motivation for the concepts to be introduced below is the aim of defining (and then classifying) the "indecomposable" objects, e.g. simple groups, factor von Neumann algebras, irreducible polynomials, prime numbers, etc.. From these points of view the following basic properties (III.3) and (III.6) appear quite naturally.

III.3 Definition:
An MDS $(X, \Sigma, \mu; \varphi)$ is called *ergodic* if there are no non-trivial φ-invariant sets $A \in \Sigma$, i.e. $\varphi(A) = A$ implies $\mu(A) = 0$ or $\mu(A) = 1$.

It is obvious that an MDS which is not ergodic is "reducible" in the sense that it can be decomposed into the "sum" of two MDSs. Therefore the name "irreducible" instead of "ergodic" would be more intuitive and more systematic. Still, the use of the word "ergodic" may be justified by the fact that ergodicity in the above sense implies the validity of the classical "ergodic hypothesis": time mean equal space mean (see III.D.6), and therefore gave rise to "ergodic theory" as a mathematical theory. Our first proposition contains a very useful criterion for ergodicity and shows for the first time the announced duality between properties of the transformation $\varphi : X \to X$ and the induced operator $T_\varphi : L^p(\mu) \to L^p(\mu)$.

III.4 Proposition:
For an MDS $(X, \Sigma, \mu; \varphi)$ the following statements are equivalent:

(a) $(X, \Sigma, \mu; \varphi)$ is ergodic.
(b) The fixed space $F := \{f \in L^p(X, \Sigma, \mu) : T_\varphi f = f\}$ of T_φ is one-dimensional, or: 1 is a simple eigenvalue of $T_\varphi \in L^p(\mu)$ for $1 \leqslant p \leqslant \infty$.

Proof. We observe, first, that the constant functions are always contained in F, hence 1 is an eigenvalue of T_φ. Moreover, we shall see that the proof does not depend on the choice of p.

(b) $\Rightarrow$ (a): If $A \in \Sigma$ is φ-invariant, then $\mathbf{1}_A \in F$ and $\dim F \geqslant 2$.

(a) $\Rightarrow$ (b): For any $f \in F$ and any $c \in \mathbb{R}$ the set

$$[f > c] := \{x \in X : f(x) > c\}$$

is φ invariant, and hence trivial. Let $c_0 := \sup\{c \in \mathbb{R} : \mu[f > c] = 1\}$. Then for $c < c_0$ we have $\mu[f \leqslant c] = 0$, and therefore $\mu[f < c_0] = 0$. For $c > c_0$ we have $\mu[f > c] \neq 1$, hence $\mu[f > c] = 0$, and therefore $\mu[f > c_0] = 0$, too. This implies $f = c_0$ a.e.. ∎

III.5 Examples:

(i) The *rotation* $(\Gamma, \mathcal{B}, m; \varphi_a)$ is ergodic, iff $a \in \Gamma$ is not a root of unity: If $a^n = 1$ for some $n \in \mathbb{N}$, then $\mathbf{1}$ and $f : z \to z^n$ are in Γ, and so φ_a is not ergodic. On the other hand, if $a^n \neq 1$ for all $n \in \mathbb{N}$, assume $T_{\varphi_a} f = f$ for some $f \in L^2(m)$. Since the functions f_n, $n \in \mathbb{Z}$, with $f(z) = z^n$ form an orthonormal basis in $L^2(\mu)$ we obtain

$$f = \sum_{n=-\infty}^{\infty} b_n f_n \quad \text{and} \quad T_{\varphi_a} f = \sum_{n=-\infty}^{\infty} b_n T_{\varphi_a} f_n = \sum_{n=-\infty}^{\infty} b_n a^n f_n.$$

The comparison of the coefficients yields $b_n(a^n - 1) = 0$ for all $n \in \mathbb{Z}$, hence $b_n = 0$ for all $n \in \mathbb{N}$, i.e. f is constant.

(ii) The Bernoulli shift $B(p_0, \ldots, p_{k-1})$ is ergodic: Let $A \in \widehat{\Sigma}$ be τ-invariant with $0 < \widehat{\mu}(A)$ and let $\varepsilon > 0$. By definition of the product σ-algebra, there exists $B \in \widehat{\Sigma}$ depending only on a finite number of coordinates such that $\widehat{\mu}(A \triangle B) < \varepsilon$, and therefore $|\widehat{\mu}(A) - \widehat{\mu}(B)| < \varepsilon$. Choose $n \in \mathbb{N}$ large enough such that $C := \tau^n B$ depends on different coordinates than B. Since μ is the product measure, we obtain $\widehat{\mu}(B \cap C) = \widehat{\mu}(B) \cdot \widehat{\mu}(C) = \widehat{\mu}(B)^2$, and $\tau(A) = A$ gives $\widehat{\mu}(A \triangle B) = \widehat{\mu}(\tau^n(A \triangle B)) = \widehat{\mu}(A \triangle C)$. We have $A \triangle (B \cap C) \subseteq (A \triangle B) \cup (A \triangle C)$ and therefore $\widehat{\mu}(A \triangle (B \cap C)) < 2\varepsilon$. This implies

$$\begin{aligned} |\widehat{\mu}(A) - \widehat{\mu}(A)^2| &\leqslant |\widehat{\mu}(A) - \widehat{\mu}(B \cap C)| + |\widehat{\mu}(B \cap C) - \widehat{\mu}(A)^2| \\ &\leqslant \widehat{\mu}(A \triangle (B \cap C)) + |\widehat{\mu}(B)^2 - \widehat{\mu}(A)^2| \\ &= \widehat{\mu}(A \triangle (B \cap C)) + |\widehat{\mu}(B) - \widehat{\mu}(A)| \cdot |\widehat{\mu}(B) + \widehat{\mu}(A)| \\ &\leqslant 4\varepsilon, \quad \text{which proves } \widehat{\mu}(A) = \widehat{\mu}(A)^2 = 1. \end{aligned}$$

In the last third of this lecture we introduce the concept of "irreducible" TDSs. Formally, this will be done in complete analogy to III.3, but due to the fact that in general the complement of a closed φ-invariant set is not closed, the result will be quite different.

III.6 Definition:
A TDS $(X; \varphi)$ is called *minimal*, if there are no non-trivial φ-invariant closed sets $A \subseteq X$, i.e. $\varphi(A) = A$, A closed, implies $A = \varnothing$ or $A = X$.

Again, "irreducible" seems to be the more adequate term (see III.D.11) but "minimal" is the term used by the topological dynamics specialists. It is motivated by property (ii) in the following proposition.

III.7 Proposition:

(i) If $(X;\varphi)$ is minimal, then the fixed space $F := \{f \in C(X) : T_\varphi f = f\}$ is one-dimensional.

(ii) If $(X;\varphi)$ is a TDS, then there exists a non-empty φ-invariant, closed subset Y of X such that $(Y;\varphi)$ is minimal.

Proof. We observe that the orbit $\{\varphi^n(x) : n \in \mathbb{Z}\}$ of any point $x \in X$ and also its closure are φ-invariant sets. Therefore, $(X;\varphi)$ is minimal iff the orbit of every point $x \in X$ is dense in X.

(i) For $f \in F$ we obtain $f(x) = f(\varphi^n(x))$ for all $x \in X$ and $n \in \mathbb{Z}$. If $(X;\varphi)$ is minimal, the continuity of f implies $f =$ constant.

(ii) The proof of this assertion is a nice, but standard application of Zorn's lemma and the finite intersection property of compact spaces. ■

III.8 Examples:

(i) Take $X = [0,1]$ and $\varphi(x) = x^2$. Then $(X;\varphi)$ is not minimal (since $\varphi(0) = 0$) but $\dim F = 1$

(ii) A property analogous to (III.7.ii) is not valid for MDSs: in $([0,1], \mathcal{B}, m; \mathrm{id})$ there exists no "minimal" invariant subset with positive measure.

(iii) The rotation $(\Gamma;\varphi_a)$ is minimal iff $\in \Gamma$ is not a root of unity: If $a^{n_0} = 1$ for some $n_0 \in \mathbb{N}$, then $\{z \in \Gamma : z^{n_0} = 1\}$ is closed and φ-invariant. For the other implication, we show that the orbit of every point in Γ is dense. To do this we need only prove that $\{1, a, a^2, \dots\}$ is dense in Γ. Choose $\varepsilon > 0$. 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 $0 < |a^l - a^k| < \varepsilon$. $0 < |a^l - a^k| = |1 - a^{k-l}| = |a^{(k-l)n} - a^{(k-l)(n+1)}| < \varepsilon$ 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 Γ, we proved that there is at least one power of a in every ε-segment of Γ.

(iv) The shift τ on $\{0, 1, \dots, k-1\}$ is not minimal, since $\tau(x) = x$ for $x = (\dots, 0, 0, 0, \dots)$.

We state once more that ergodicity and minimality are the most fundamental properties of our measure-theoretical or topological dynamical systems. On the other hand they gave us the first opportunity to demonstrate how dynamical properties of a map $\varphi : X \to X$ are reflected by (spectral) properties of the induced linear operator T_φ (see III.4 and III.7.i). In particular, it can be expected that the set $P\sigma(T_\varphi)$ of all eigenvalues of T_φ has great significance in ergodic theory (see Lectures VIII and IX). Here we show only the effect of ergodicity or minimality on the structure of the point spectrum $P\sigma(T_\varphi)$.

III.9 Proposition:
Let $(X;\varphi)$ be a minimal TDS (resp. $(X,\Sigma,\mu;\varphi)$ an ergodic MDS). Then the point spectrum $P\sigma(T_\varphi)$ of the induced operator T_φ on $C(X)$ (resp. $L^p(X,\Sigma,\mu)$) is a subgroup of Γ, and each eigenvalue is simple.

Proof. Since T_φ is a bijective isometry the spectrum of T_φ is contained in Γ. Let $T_\varphi f = \lambda f$, $\|f\| = 1 = |\lambda|$. Since T_φ is a lattice homomorphism we conclude

$$T_\varphi|f| = |T_\varphi f| = |\lambda f| = |\lambda| \cdot |f| = |f|,$$

and hence $|f| = \mathbf{1}$ by (III.7.i), resp. (III.4), i.e. every normalized eigenfunction is unimodular and the product of two such eigenfunctions is non-zero. Since T_φ is also an algebra homomorphism (on $L^\infty(X)$, resp. $C(X)$) we conclude from $T_\varphi f = \lambda_1 f \neq 0$ and $T_\varphi g = \lambda_2 g \neq 0$ that

$$T_\varphi(f \cdot g^{-1}) = T_\varphi f \cdot T_\varphi g^{-1} = \lambda_1 \cdot \lambda_2^{-1}(f \cdot g^{-1}) \neq 0$$

which shows that $P\sigma(T_\varphi)$ is a subgroup of Γ. If $\lambda_1 = \lambda_2$, it follows $T_\varphi(f \cdot g^{-1}) = f \cdot g^{-1}$ and, again by the one-dimensionality of the fixed space, $f \cdot g^{-1} = c \cdot \mathbf{1}$ or $f = c \cdot g$, i.e. each eigenvalue is simple. ∎

III.D Discussion

III.D.1. The "original" Poincaré theorem:
Henri Poincaré ([1890], p. 69) formulated what later on was called the recurrence theorem:

> "Théorème I. Supposons que le point P reste à distance finie, et que le volume $\int \mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3$ soit un invariant intégral; si l'on considère une région r_0 quelconque, quelqe petite que soit cette région, il y aura des trajectoires qui la traverseront une infinité de fois."

In the corollary to this theorem he mentioned some kind of probability distribution for the trajectories:

> "Corollaire. Il résulte de ce qui précède qu'il existe une infinité de trajectoires qui traversent une infinité de fois la région r_0; mais il peut en exister d'autres qui ne traversent cette région qu'un nombre fini de fois. Je me propose maintenant d'expliquer pourquoi oes dernières trajectoires peuvent être regardées oomme exceptionnelles."

III.D.2. Recurrence and the second law of thermodynamics:
As we explained in Lecture I the time evolution of physical "states" is adequately described in the language of MDS and therefore "states" are "recurrent". This (and the picture following (III.2)) seems to be in contradiction with the second law of thermodynamics which says that entropy can only increase, if it changes at all, and thus we can never come back to a state of entropy h, once we have reached a state of entropy higher than h. One explanation lies in the fact that the second law is an empirical law concerning a quantity, called entropy, that can only be determined through measurements that require *time averaging* (in the range from milliseconds to seconds). In mathematical models of "micro"-dynamics, which were the starting point of ergodic theory, such time averages should be roughly *constant* (and equal to the space mean by the ergodic hypothesis). Therefore entropy should be constant for dynamical systems (like the constant defined in Lecture XII, although at least to us it is unclear whether the two numbers, the Kolmogoroff-Sinai entropy and the physical entropy can be identified or compared in such a model). In this case there is no contradiction to Poincaré's theorem, because entropy does not really depend

on the ("micro"-)state x.

The second law of thermodynamics applies to *changes* in the underlying physical "micro"-dynamics, i.e. in the dynamical system or in the mapping φ. Such changes can occur for example if boundary conditions are changed by the experimenter or engineer; they are described on a much coarser time scale, and as a matter of fact, they can only lead in a certain direction, namely toward higher entropy.

Another way of turning this argument is the following: The thermodynamical (equilibrium) entropy is a quantity that is based on thermodynamical measurements, which always measure time averages in the range from milliseconds to seconds. In particular, such an unusual momentary state as in the picture following (III.2) cannot be measured thermodynamically, in fact the ergodic hypothesis states that we shall usually measure a time average which is close to the "space mean". Therefore a thermodynamical measurement of the number of atoms (i.e. the "pressure") in the left chamber will almost always give a result close to 500. In some branches of thermodynamics ("non-equilibrium" thermodynamics), however, a variable $e(x)$ is associated with micro states $x \in X$, which is also interpreted as the "entropy" of x, but is not constant on X. In this case Poincaré's theorem shows that the second law for this variable e cannot be strictly true, but still it is argued that a big decrease of e is very improbable. For example, we can try to capture the *momentary* state of the gas in the box, by quickly inserting a separating wall into the box at some arbitrary moment (chosen at random). Then the thermodynamical calculations of the invariant measure on the state space tell us, that we have a chance of 2^{-1000} of catching the gas in a position with all 1000 atoms in the left half of the box (low "entropy"), and a chance of 27.2% of having 495 to 505 atoms in the left half of the box (high "entropy").

III.D.3. Counterexamples:
The recurrence theorem (III.2) is not valid without the assumption of *finite* measure spaces or *measure-preserving* transformations:

(i) Take $X = \mathbb{R}$ and the Lebesgue measure m. Then the shift

$$\tau : x \mapsto x + 1$$

on X is bi-measure-preserving, but no point of $A := [0, 1)$ is recurrent to A.

(ii) The transformation

$$\varphi : x \to x^2$$

on $X = [0, 1]$ is bi-measurable, but not measure-preserving for the Lebesgue measure m. Clearly, no point of $A := [\frac{1}{2}, \frac{2}{3}]$ is recurrent to A.

III.D.4. Recurrence in random literature:
A usual typewriter has about 90 keys. If these keys are typed at random, what is the probability to type for example this book? Let us say, this book has N letters including blanks. Then the probability of typing it with N random letters is $p = 90^{-N}$. The Bernoulli shift $B(\frac{1}{90}, \dots, \frac{1}{90})$ is an MDS $(\widehat{X}, \widehat{\Sigma}, \widehat{\mu}; \tau)$ whose state space consists of sequences $(x_k)_{k \in \mathbb{Z}}$ which can be regarded as the result of infinite random typing. What is the probability, that such a sequence contains this book,

i.e. the sequence $R_1, \ldots, R_N$ of letters? From

$$\begin{aligned}
&\widehat{\mu}[\text{there exists } k \in \mathbb{Z} \text{ such that } x_{k+1} = R_1, \ldots, x_{k+N} = R_N] \\
&\quad = 1 - \widehat{\mu}[\text{for every } k \in \mathbb{Z} \text{ there exists } i \in \{1, \ldots, N\} \text{ such that } x_{k+i} \neq R_i] \\
&\quad \geqslant 1 - \prod_{k=1}^{n} \widehat{\mu}[\text{there exists } i \in \{1, \ldots, N\} \text{ such that } x_{k+i} \neq R_i] \\
&\quad = 1 - (1-p)^n \quad \text{for every } n \in \mathbb{N}
\end{aligned}$$

we conclude that this probability is 1. Now consider $A := [x_1 = R_1, \ldots, x_N = R_N]$ having $\widehat{\mu}(A) = 0$. We have just shown that for almost every $x \in \widehat{X}$ there is a number k such that $\tau^k(x) \in A$ for the shift τ. Poincaré's theorem implies that there are even infinitely many such numbers, i.e. almost every sequence contains this book infinitely often!

By Kac's theorem (Kac [1947], Petersen [1983]) and the ergodicity of $B(\frac{1}{90}, \ldots \frac{1}{90})$ the average distance between two occurrences of this book in random text is $\frac{1}{p} = 90^N$ digits. The fact that this number is very large, may help to understand the strange phenomenon depicted in (III.2)

III.D.5. Invariant sets:
The transformations $\varphi : X \to X$ which we are considering in these lectures are bijective. Therefore it is natural to call a subset $A \subseteq X$ *φ-invariant* if $\varphi(A) \subseteq A$ and $\varphi^{-1}(A) \subseteq A$, i.e. $\varphi(A) = A$. With this definition, a closed φ-invariant set $A \subseteq X$ in a TDS $(X;\varphi)$ always leads to the restricted TDS $(A;\varphi|_A)$, while $([0,1];\varphi)$, $\varphi(x) := x^2$ and $A = [0, \frac{1}{2}]$ gives an example such that $\varphi(A) \subseteq A$ but $\varphi|_A$ is not a homeomorphism of A.

For MDSs $(X, \Sigma, \mu; \varphi)$ the situation is even simpler: $\varphi(A) \subseteq A$ implies $A \subseteq \varphi^{-1}(A)$ and $\mu(A) = \mu(\varphi^{-1}(A))$ since φ is measure-preserving. Therefore $A = \varphi^{-1}(A)$ and $\varphi(A) = A$ μ-a.e..

In agreement with the definition above we define the orbit of a point $x \in X$ as $\{\varphi^k(x) : k \in \mathbb{Z}\}$. If $(X;\varphi)$ is a TDS, the smallest closed invariant set containing a point $x \in X$ is clearly the "closed orbit" $\overline{\{\varphi^k(x) : k \in \mathbb{Z}\}}$. However, the closed orbit is, in general, not a minimal set: For example consider the one point compactification of $\mathbb{Z}$

$$X := \mathbb{Z} \cup \{\infty\}$$

$$\text{and the shift} \quad \tau : \begin{cases} x \mapsto x+1 & \text{if } x \in \mathbb{Z} \\ \infty \mapsto \infty \end{cases}.$$

Then $\overline{\{\tau^k(0) : k \in \mathbb{Z}\}} = X$ is not minimal since $\tau(\infty) = \infty$.
In many cases, however, the closed orbit is minimal as can be seen in the following.

Lemma: Let $(X;\varphi)$ be a TDS, where X is a metric space (with metric d) and assume that $X = \overline{\{\varphi^s(a) : s \in \mathbb{Z}\}}$ for some $a \in X$. If for every $\varepsilon > 0$ there exists $k \in \mathbb{N}$ with

$$d(a, \varphi^{ks}a) < \varepsilon \quad \text{for all } s \in \mathbb{Z},$$

then $(X;\varphi)$ is minimal.

Proof. It suffices to show that $a \in \overline{\{\varphi^s(x) : s \in \mathbb{Z}\}}$ for every $x \in X$. Let be $x \in X$, $\varepsilon > 0$, and choose $k \in \mathbb{N}$ such that

(i) $d(a, \varphi^{ks}a) < \varepsilon$ for all $s \in \mathbb{Z}$.
Since the family of mappings $\{\varphi^0, \varphi^1, \ldots, \varphi^k\}$ is equicontinuous at x there is $\delta > 0$ such that

(ii) $d(\varphi^t x, \varphi^t y) < \varepsilon$ if $t \in \{0, \ldots, k\}$ and $d(x, y) < \delta$. The orbit of a is dense in X. Therefore, we find $r \in \mathbb{Z}$ with

(iii) $d(x, \varphi^r a) < \delta$ and by (i) a suitable $t \in \{0, \ldots, k\}$ with

(iv) $d(\varphi^{t+r}a, a) < \varepsilon$.

Combining (ii), (iii) and (iv) we conclude that

$$d(\varphi^t x, a) \leqslant d(\varphi^t x, \varphi^t(\varphi^r a)) + d(\varphi^{t+r}a, a) \leqslant 2\varepsilon.$$

■

Remark: Minimality in metric spaces is equivalently characterized by a property weaker than that given above (see Jacobs [1960], 5.1.3.).

III.D.6. Ergodicity implies "time mean equal space mean":
The physicists wanted to replace the time mean

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ \varphi^i(x)$$

of an "observable" φ in the "state" x by the space mean

$$\int_X f \, \mathrm{d}\mu \qquad \text{(see Lecture I),}$$

i.e. the above limit has to be equal the constant function $(\int_X f \, \mathrm{d}\mu) \cdot \mathbf{1}$. Obviously the time mean is a φ-invariant function, and we conclude by (III.4) that "time mean equal space mean" holds for every observable f (at least: $f \in L^p(\mu)$) if and only if (!) the dynamical system is ergodic. In this way the original problem of ergodic theory seems to be solved, but there still remains the task for the mathematician to prove the existence of the above limit (see Lecture IV and V). Even more important (and more difficult) is the problem of finding physical systems and their mathematical models, which are ergodic. The statement of Birkhoff-Koopmann [1932] "the outstanding unsolved problem in ergodic theory is the question of the truth or falsity of metrical transitivity (= ergodicity) for general Hamiltonian systems" is still valid, even if important contributions have been made for the so-called "billiard gas" by Sinai [1963] and Gallavotti-Ornstein [1974] (see Gallavotti [1975]).

III.D.7. Decomposition into ergodic components:
As indicated it is a mathematical principle to decompose an object into "irreducible" components and then to investigate these components. For an MDS this is possible (with "ergodic" for "irreducible"). In fact, such a decomposition is based on the geometrical principle of expressing a point of a (compact) convex set as a convex sum of extreme points (see books on "Choquet theory", e.g. Phelps [1966] or Alfsen [1971]), but the technical difficulties, due to the existence of null sets, are considerable, and become apparent in the following example:

Consider the MDS $(X, \mathcal{B}, m; \varphi_a)$ where $X := \{z \in \mathbb{C} : |z| \leqslant 1\}$, $\mathcal{B}$ the Borel algebra, m the Lebesgue measure $m(X) = 1$ and φ_a the rotation

$$\varphi_a(z) = a \cdot z$$

for some $a \in \mathbb{C}$ with $|a| = 1$, $a^n \neq 1$ for all $n \in \mathbb{N}$. Its ergodic "components" are the circles $X_r := \{z \in \mathbb{C} : |z| = r\}$ for $0 \leqslant r \leqslant 1$ and $(X, \mathcal{B}, m; \varphi_a)$ is "determined" by these ergodic components. For more information we refer to von Neumann [1932] or Rohlin [1966].

III.D.8. One-dimensionality of the fixed space:
Ergodicity is characterized by the one-dimensionality of the fixed space (in the appropriate function space) while minimality is not (III.4 and III.8.1). The fixed space of the induced operator T_φ in $C(X)$ is already one-dimensional if there is at least one point $x \in X$ having dense orbit $\{\varphi^n(x) : n \in \mathbb{Z}\}$ in X (see III.7, Proof). This property of a TDS, called "topological transitivity" or "topological ergodicity", is another topological analogue of ergodicity as becomes evident from the following characterizations (see Walters [1975] p. 22 and p. 117):

1. For an MDS $(X, \Sigma, \mu; \varphi)$the following are equivalent:
   a. φ is ergodic.
   b. For all $A, B \in \Sigma$, $\mu(A) \neq 0 \neq \mu(B)$, there is $k \in \mathbb{Z}$ such that $\mu(\varphi^k A \cap B) > 0$.
2. For a TDS $(X; \varphi)$, X metric, the following assertions are equivalent:
   a. φ is topologically ergodic.
   b. For all A, B open, $A \neq \varnothing \neq B$ there is $k \in \mathbb{Z}$ such that $\varphi^k A \cap B \neq \varnothing$

But even topological transitivity, although weaker than minimality, is not characterized by the fact that the fixed space is one-dimensional in $C(X)$, see (III.8).i. The reason is that T_φ in $C(X)$ lacks a certain convergence property which is automatically satisfied in $L^p(X, \Sigma, \mu)$ (see IV.7 and IV.8; for more information see IX.D.7.

III.D.9. Ergodic and minimal rotations on the n-torus:
The rotation

$$\varphi_a : z \mapsto a \cdot z$$

on the n-dimensional torus Γ^n with $a = (a_1, \ldots, a_n) \in \Gamma^n$ is is ergodic (minimal) if and only if $\{a_1, \ldots, a_n\}$ are linearly independent in the $\mathbb{Z}$-module Γ

Proof. (i) In the measure-theoretical case use the n-dimensional Fourier expansion and argue as in (III.5.i).
(ii) In the topological case we argue as in (III.8.iii) observing that for an $a = (a_1, \ldots, a_n) \in \Gamma^n$ the set $\{a^k : k \in \mathbb{Z}\}$ is dense in Γ^n iff $\{a_, \ldots, a_n\}$ is linearly independent in the $\mathbb{Z}$-module Γ (see D.8).

∎

III.D.10. Ergodic vs. minimal:
Let $(X; \varphi)$ be a TDS and μ a φ-invariant probability measure on X (see also App. S). Then $(X, \mathcal{B}, \mu; \varphi)$ is an MDS for the Borel algebra $\mathcal{B}$. In this situation, is it possible that if is ergodic but not minimal, or vice versa? The positive answer to the first part or our question is given by the Bernoulli shift, see (III.5.ii) and (III.8.iv). The construction of a dynamical system which is minimal but not ergodic is much more difficult and needs results of Lecture IV. We come back to this problem in IV.D.9.

III.D.11. Irreducible operators on Banach lattices:
Let T be a positive operator on some Banach lattice E. It is called irreducible if it leaves no non-trivial closed lattice ideal invariant. If $E = C(X)$, resp. $E = L^1(X, \sigma, \mu)$, every closed lattice ideal is of the form

$$I_A := \{f \in E : f(A) \subseteq \{0\}\}$$

where $A \subseteq X$ is closed, resp. measurable, (Schaefer [1974], p. 157). Therefore, it is not difficult to see that an induced operator T_φ on $C(X)$, resp. $L^p(X, \Sigma, \mu)$ is irreducible if and only if $(X; \varphi)$ is minimal, resp. if $(X, \Sigma, \mu; \varphi)$ ergodic. In contrast to minimal TDSs the ergodicity of an MDS $(X, \Sigma, \mu; \varphi)$ is characterized by the one-dimensionality of the T_φ-fixed space in $L^p(X, \Sigma, \mu)$, $1 \leqslant p < \infty$, (see III.4). The reason for this is the fact that the induced operators are mean ergodic on $L^p(\mu)$ but not on $C(X)$ (see Lecture IV). More generally, the following holds (see Schaefer [1974], III.8.5).

Proposition: Let T be a positive operator on a Banach lattice E and assume that T is mean ergodic with non-trivial fixed space F. The following are equivalent:

(a) T is irreducible.
(b) $F = \langle u \rangle$ and $F' = \langle \mu \rangle$ for some quasi-interior point $u \in E_+$ and a strictly positive linear form $\mu \in E'_+$.

If E is finite-dimensional, we obtain the classical concept of irreducible (= indecomposable) matrices (see IV.D.7 and Schaefer [1974], I.6).

Example: The matrix

$$\begin{pmatrix} p_0 & \cdots\cdots\cdots & p_{k-1} \\ \vdots & & \vdots \\ p_0 & \cdots\cdots\cdots & p_{k-1} \end{pmatrix}$$

of (II.6), Exercise is irreducible whereas the Bernoulli shift $B(p_0, \ldots, p_{k-1})$ is ergodic (see (III.5.ii)). This gives the impression that irreducibility is preserved under dilation (see App. **??**) at least in this example. In fact, this turns out to be true (App. **??**), and in particular in (IV.D.8) we shall show that any Markov shift is ergodic iff the corresponding matrix is irreducible. Frobenius discovered in 1912 that the point spectrum of irreducible positive matrices has nice symmetries. The same is true for operators T_φ, as shown in (III.9).

This result has been considerably generalized to irreducible positive operators on arbitrary Banach lattices. We refer to Schaefer [1974], V.5.2 for a complete treatment and quote the following theorem.

Theorem Lotz, 1968: Let T be a positive irreducible contraction on some Banach lattice E. Then $P\sigma(R) \cap \Gamma$ is a subgroup of Γ or empty, and every eigenvalue in Γ is simple.

References: Lotz [1968], Schaefer [1967/68], Schaefer [1974].

III.D.12. The origin of the word "Ergodic Theory":
In the last decades of the 19$^{\text{th}}$ century mathematicians and physicists endeavoured to explain thermodynamical phenomena by mechanical models and tried to prove the laws of thermodynamics be mechanical principles or, at least, to discover close analogies between the two. The Hungarian M.C. Szily [1872] wrote:

> "The history of the development of modern physics speaks decidedly in favour of the view that only those theories which are based on mechanical principles are capable of affording a satisfactory explanation of the phenomena."

Those efforts were undertaken particularly in connection with the second law of thermodynamics; Szily [1876] even claimed to have deduced it from the first, whereas a few years earlier he had declared:

> "What in thermodynamics we call the second proposition, is in dynamics no other than Hamilton's principle, the identical principle which has already found manifold applications in several branches of mathematical physics."

(see Szily [1872]; see also the subsequent discussion in Clausius [1872] and Szily [1873].)

In developing the Mechanical Theory of Heat three fundamentally different hypotheses were made; besides the hypothesis of the stationary or quasi-periodic motions (of R. Clausius and Szily) and the hypothesis of monocyclic systems (of H. von Helmholtz, cf. Bryan-Larmor [1892]), the latest investigations at that time concerned considerations which were based on a very large number of molecules in a gas and which established the later Kinetic Theory of Gases. This was the statistical hypothesis of L. Boltzmann, J.B. Maxwell, P.G. Tait and W. Thomson, and its fundamental theorem was the equipartition theorem of Maxwell and Boltzmann: When a system of molecules has attained a stationary state the time-average of the kinetic energy is equally distributed over the different degrees of freedom of the system. Based on this theorem there are some proofs of the second law of thermodynamics (Burbury [1876], Boltzmann [1887]), but which was the exact hypothesis for the equipartition theorem itself? In Maxwell [1879] we find the answer:

> "The only assumption which is necessary for the direct proof (of the equipartition theorem) is that the system, if left to itself in its actual state of motion, will, sooner or later, pass through every phase which is consistent with the equation of energy."

Boltzmann [1871], too, made use of a similar hypothesis:

> "Von den zuletzt entwickelten Gleichungen können wir unter einer Hypothese, deren Anwendbarkeit auf warme Körper mir nicht unwahrscheinlich scheint, direkt zum Wärmegleichgewicht mehratomiger Gasmoleküle je noch allgemeiner zum Wärmegleichgewicht eines beliebigen mit einer Gasmasse in Berührung stehenden Körpers gelangen. Die große Unregelmäßigkeit der Wärmebewegung und die Mannigfaltigkeit der Kräfte, welche von außen auf die Körper wirken, macht es wahrscheinlich, daß die Atoms derselben vermöge der Bewegung, die wir Wärme nennen, alle möglichen mit der Gleichung der lebendigen Kraft vereinbare Positionen und Geschwindigkeiten durchlaufen, daß wir also die zuletzt entwickelten Gleichungen auf die Koordinaten und die Geschwindigkeitskomponenten der Atome warmer Körper anwenden können."

Sixteen years later, Boltzmann mentioned in [1887]

> "... (Ih habe für derartige Inbegriffe von Systemen den Namen *Ergoden* vorgeschlagen.)..."

This may have induced P. and T. Ehrenfest to create the notion of "Ergodic Theory" by writing in "Begriffliche Grundlagen der statistischen Auffassung" [1911]:

> "... 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 notion "ergodic" was explained by them in a footnote:

> " ἔργον = Energie, ὁδός= Weg : Die G-Bahn geht durch alle Punkte der Energiefläche. Diese Bezeichnung gebraucht Boltzmann das erste Mai in der Arbeit [15] (1886) " (here Boltzmann [1887])

But this etymological explanation seems to be incorrect as we will see later. The hypothesis quoted above, i.e. that the gas models are ergodic systems, they called the "Ergodic Hypothesis". In the sequel they doubted the existence of ergodic systems, i.e. that their definition does not contradict itself. Actually, only few years later A. Rosenthal and M. Plancherel proved independently the impossibility of systems that are ergodic in this sense (cf. Brush [1971]). Thus, "Ergodic Theory" as a theory of ergodic systems hardly survived its definition. Nevertheless, from the explication of the "Ergodic Hypothesis" and its final negation, "Ergodic Theory" arose as a new domain of mathematical research (cf. Brush [1971], Birkhoff-Koopmann [1932].

But, P. and T. Ehrenfest were mistaken when they thought that Boltzmann used the notion "Ergodic" and "Ergodic Systems" in Boltzmann [1887] for the first time. In 1884 he had already defined the notion "Ergode" as a special type of "Monode". In his article (Boltzmann [1885]) first of all he wrote:

> "Ich möchte mir erlauben, Systeme, deren Bewegung in diesem Sinne stationär ist, als monodische Oder kürzer als Monoden zu bezeichnen. (Mit dem Namen stationär wurde von Herrn Clausius jede Bewegung bezeichnet, wobei Koordinaten und Geschwindigkeiten immer zwischen endlichen Grenzen eingeschlossen bleiben). Sie sollen dadurch charakterisiert sein, daß die in jedem Ptmkte derselben herrschende Bewegung unverändert fortdauert, also nicht Funktion der Zeit ist, solange die äußeren Kräfte unverändert bleiben, und daß auch in keinem Punkte und keiner Flc̈he derselben Masse oder lebendige Kraft oder sonst ein Agens ein- oder austritt."

In a modern language a "Monode" is a system only moving in a finite region of phase space described by a dynamic system of differential equations; a simple example is a mathematical pendulum. From Boltzmann's definition we can understand the name: μόνος means "unique", "Monode" probably comes from μονώδης which is composed of μόνο–ώδης where the suffix –ώδης means "–like".

Having specified some different kinds of "Monoden" as "Orthoden" and "Holoden", Boltzmann turned towards collections (ensembles) of systems which were all of the same nature, totally independent of each other and each, of them fulfilling a number of equations $\varphi_1 = a_1, \ldots, \varphi_k = a_k$. Of special interest to him were those collections of systems fulfilling only one equation $\varphi = a$ concerning the energy of all systems in the collection.

> "... so wollen wir den Inbegriff aller N Systeme als eine Monode bezeichnen, welche durch die Gleichungen $\varphi_1 = a_1, \ldots$ beschränkt ist ... Monoden, welche nur durch die Gleichung der lebendigen Kraft beschränkt sind, will ich als *Ergoden*, solche, welche außer dieser Gleichung auch noch durch andere beschränkt sind, als *Subergoden* bezeichnen.... Für Ergoden existiert also nur ein φ, welches gleich der für alle Systeme gleichen und während der Bewegung jedes Systems konstanten Energie eines Systems $\chi + \psi = \frac{(\phi+L)}{N}$ ist".

(Boltzmann [1885]; χ, ϕ mean the potential energy, ψ, L the kinetic energy of one system, of the collection of N systems, respectively.) The last sentence of that quotation helps us to understand the name "Ergode" in the right way: The word ἔργον = "work, energy" is used, but in a sense different from that presumed by the Ehrenfests who also did not mention Boltzmann's article [1885] in their bibliography [1911].

Boltzmann also had knowledge of "Monoden" fulfilling the "Ergodic Hypothesis" of the Ehrenfests. In the fourth paragraph of Boltzmann [1885] we read in a footnote:

> "Jedesmal, wenn jedes einzelne System der Monode im Verlaufe der Zeit alle an den verschiedenen Systemen gleichzeitig nebeneinander vorkommenden Zustände durchläuft, kann an Stelle der Monode ein einziges System gesetzt werden.... Für eine solche Monode wurde schon früher die Bezeichnung "isodisch" vorgeschlagen"

In summary an "Ergode" is a special kind of "Monode", namely one which is determined by " ἔργον" = "energy" or "work", and the word "Monode" stems from μόνος = "one" or "unique" and the suffix –ώδης = "–like" or "–full". Therefore a "Monode" is literally "one-like" i.e. atomary or indecomposable, which is just the modern meaning of ergodic. Taken literally, however, the word "Ergode" means "energy-like" or "work-full", which brings us back to our first etymological answer in Lecture I:

" difficult "!

References: Boltzmann [1885], [1887], Brush [1971], Ehrenfest [1911]

P.S. The above section originated from a source study by M. Mathieu. The Ehrenfests' explanation of the word "ergodic" is still advocated by A. LoBello:

> The etymology of the word ergodic, in: Conference on modern Analysis and Probability, New Haven 1982, Contempt.Math. 26, Amer. Math. Soc. Providence R.I., 1984, p.249.

IV. The Mean Ergodic Theorem

"Ergodic theory is the study of transformations from the point of view of ... dynamical properties connected with *asymptotic behavior*" (Walters [1975], p. 1). Here, the asymptotic behavior of a transformation φ is described by

$$\text{“}\lim_{n\to\infty}\text{”}\varphi^n$$

where it is our task first to make precise in which sense the "lim" has to be understood and second to prove its existence. Motivated by the original problem "time mean equals space mean" (see III.D.6) we investigate in this lecture the existence of the limit for $n \to \infty$ not of the powers φ^n but of the "Cesàro means"

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ \varphi^i$$

where f is an "observable" (see physicist's answer in Lecture I) contained in an appropriate function space. With a positive answer to this question - for convergence in L^2-space - ergodic theory was born as an independent mathematical discipline.

IV.1 Theorem (J. von Neumann, 1931):
Let $(X, \Sigma, \mu; \varphi)$ be and MDS and denote by T_φ the induced (unitary) operator on $L^2(X, \Sigma, \mu)$. For any $f \in L^2(\mu)$ the sequence of functions

$$f_n := \frac{1}{n}\sum_{i=0}^{n-1} T_\varphi^i f, \quad n \in \mathbb{N}$$

(norm-)converges to a T_φ-invariant function $\bar{f} \in L^2(\mu)$.

It was soon realized that only a few of the above assumptions are really necessary, while the assertion makes sense in a much more general context. Due to the importance of the concept and the elegance of the results, an axiomatic and purely functional-analytic approach seems to be the most appropriate.

IV.2 Definition:
An FDS $(E; T)$ (resp. a bounded linear operator T) is called *mean ergodic*, if the sequence

$$T_n := \frac{1}{n}\sum_{i=0}^{n-1} T^i, \quad n \in \mathbb{N}$$

converges in $\mathscr{L}(E)$ for the strong operator topology.

As above, the operators T_n will be called the *"Cesàro means"* of the powers T^i. Moreover we call $P := \lim_{n\to\infty} T_n$, if it exists, the *"projection corresponding to T"*. This language is justified by the following elementary properties of mean ergodic operators.

IV.3 Proposition:
(0) $(\mathrm{id} - T)T_n = \frac{1}{n}(\mathrm{id} - T^n)$ for every $n \in \mathbb{N}$

If T is mean ergodic with corresponding projection P, we have

(1) $TP = PT = P = P^2$.
(2) $PE = F := \{f \in E : Tf = f\}$.
(3) $P^{-1}(0) = \overline{(\mathrm{id} - T)E}$.
(4) The adjoints T_n' converge to P' in the weak*-operator topology of $\mathscr{L}(E')$ and $P'E' = F' := \{f' \in E' : T'f' = f'\}$.

(5) $(PE)'$ is (as a topological vector space) isomorphic to $P'E'$.

Proof.

(0) is obvious from the definition of T_n.

(1) Clearly, $(n+1)T_{n+1} - \mathrm{id} = nT_nT = nTT_n$ holds. Dividing by n and letting n tend to infinity we obtain $P = PT = TP$. From this we infer that $T_nP = P$ and thus $P^2 = P$.

(2) $PE \subseteq F$ follows from $TP = P$, and $F \subseteq PE$ from $P = \lim_{n\to\infty} T_n$.

(3) By the relations in (1), $(\mathrm{id} - T)E$ and (by the continuity of P) its closure is contained in $P^{-1}(0)$. Now take $f \in P^{-1}(0)$. Then

$$f = f - Pf = f - PTf = \lim_{n\to\infty}(\mathrm{id} - T_nT)f = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\mathrm{id} - T^i)f$$
$$= \lim_{n\to\infty}(\mathrm{id} - T)\frac{1}{n}\sum_{i=1}^{n} iT_if \in \overline{(\mathrm{id} - T)E}.$$

(4) By the definition of the weak* operator topology, T_n' converges to P' if $\langle T_nf, f'\rangle = \langle f, T_n'f'\rangle \to \langle f, P'f'\rangle = \langle Pf, f'\rangle$ for $f \in E$ and $f' \in E'$. This follows from the convergence of T_n to P in the strong operator topology. Together with $(PT)' = T'P' = P'$ this implies the remaining property as in (2).

(5) This statement holds for every projection on a Banach space (see B.7, Proposition).

■

Our main result contains a list of surprisingly different, but equivalent characterizations of mean ergodicity at least for operators with bounded powers.

IV.4 Theorem:
If $(E;T)$ is an FDS with $\|T^n\| \leqslant c$ for every $n \in \mathbb{N}$ the following assertions are equivalent:

(a) T is mean ergodic.

(b) T_n converges in the weak operator topology.

(c) $\{T_nf : n \in \mathbb{N}\}$ has a weak accumulation point for all $f \in E$

(d) $\overline{\mathrm{co}}\{T^if : i \in \mathbb{N}_0\}$ contains a T-fixed point for all $f \in E$.

(e) The T-fixed space F separates points of the T'-fixed space F'.

Proof. The implications (a) $\Rightarrow$ (b) $\Rightarrow$ (c) are trivial.

(c)$\Rightarrow$ (d): Take $f \in E$ and let g be a weak accumulation point of $\{T_nf : n \in \mathbb{N}\}$, i.e. $g \in \overline{\{T_nf : n > n_0\}}^{\sigma(E,E')}$ for all $n_0 \in \mathbb{N}$. Certainly, g is contained in $\overline{\mathrm{co}}\{T^if : i \in \mathbb{N}_0\}$, and we shall show that g is fixed under T: For any $n_0 \in \mathbb{N}$ we obtain

$$g - Tg = (\mathrm{id} - T)g \in (\mathrm{id} - T)\overline{\{T_nf : n > n_0\}}^{\sigma} \subseteq \overline{\{(\mathrm{id} - T)T_nf : n > n_0\}}^{\sigma}$$
$$= \overline{\{\tfrac{1}{n}(\mathrm{id} - T^n)f : n > n_0\}}^{\sigma} \subseteq \tfrac{1}{n_0}(1+c)\|f\|U,$$

where U is the closed unit ball in E – we used the fact that $(\mathrm{id} - T)$ is continuous for the weak topology and that U is weakly closed (see B.7 and B.3).

(d) $\Rightarrow$ (e) : Choose $f', g' \in F'$, $f' \neq g'$, and $f \in E$ with $\langle f, f'\rangle \neq \langle f, g'\rangle$. For all elements $f_0 \in \overline{\mathrm{co}}\{T^if : i \in \mathbb{N}_0\}$ we have $\langle f_0, f'\rangle = \langle f, f'\rangle$ and $\langle f_0, g'\rangle = \langle f, g'\rangle$ Therefore the T-fixed point $f_1 \in \overline{\mathrm{co}}\{T^if : i \in \mathbb{N}\}$ which exists by (d), satisfies

$\langle f_1, f'\rangle = \langle f, f'\rangle \neq \langle f, g'\rangle = \langle f_1, g'\rangle$, i.e. it separates f' and g'.

(e) $\Rightarrow$ (a): Consider

$$G := F \oplus \overline{(\mathrm{id} - T)E}$$

and assume that $f' \in E'$ vanishes on G. Since it vanishes on $(\mathrm{id} - T)E$ it follows immediately that $f' \in F'$. Since it also vanishes on F, which is supposed to separate F', we conclude that $f' = 0$, hence that $G = E$. But $T_n f$ converges for every $f \in F \oplus (\mathrm{id} - T)E$, and the assertion follows from the equicontinuity of $\{T_n : n \in \mathbb{N}\}$. ∎

The standard method of applying the above theorem consists in concluding mean ergodicity of an operator from the apparently "weakest" condition (IV.4.c) and the weak compactness of certain sets in certain Banach spaces. This settles the convergence problem for the means T_n as long as the operator T is defined on the right Banach space E.

IV.5 Corollary:
Let $(E; T)$ be an FDS where E is a reflexive Banach space, and assume that $\|T^n\| \leqslant c$ for all $n \in \mathbb{N}$. Then T is mean ergodic.

Proof. Bounded subsets of reflexive Banach spaces are relatively weakly compact (see B.4). Since $\{T_n f : n \in \mathbb{N}\}$ is bounded for every $f \in E$, it has a weak accumulation point. ∎

Besides matrices with bounded powers on $\mathbb{R}^n$ we have the following concrete applications:

Example 1: Let E be a Hilbert space and $T \in \mathscr{L}(E)$ be a contraction. Then T is mean ergodic and the corresponding projection P is orthogonal: By (IV.5) the Cesàro means T_n of T converge to P and the Cesaró means T_n^* of the (Hilbert space) adjoint T^* converge to a projection Q. If $(\cdot|\cdot)$ denotes the scalar product on E, we obtain from $(T_n^* f|g) \to (Qf|g)$ and $(f|T_n g) \to (f|Pg)$ for all $f, g \in E$ that $Q = P^*$. The fixed space $F = PE$ of T and the fixed space $F^* = P^*E$ of T^* are identical: Take $f \in F$. Since $\|T\| = \|T^*\| \leqslant 1$, the relation $(f|f) = (Tf|f) = (f|T^*f)$ implies $(f|f) \leqslant \|f\| \cdot \|T^*f\| \leqslant \|f\|^2 = (f|f)$, hence $T^*f = f$. The other conclusion $F^* \subseteq F$ follows by symmetry. Finally we conclude from $P = P^*P = (P^*P)^* = P^*$ that P is orthogonal.

Example 2: Let $(X, \Sigma, \mu; \varphi)$ be an MDS. The induced operator T_φ on $L^p(X, \sigma, \mu)$ for $1 < p < \infty$ is mean ergodic, and the corresponding projection P is a "conditional expectation" (see B.24):

For $f, g \in L^\infty$ and $T_\varphi f = f$ we obtain $T_\varphi(fg) = T_\varphi \cdot T_\varphi g = f \cdot T_\varphi g$. The same holds for $(T_\varphi)_n$, and therefore $P(fg) = f \cdot Pg$.

Both examples contain the case of the original von Neumann theorem (IV.1).

IV.6 Corollary:
Let $(E; T)$ be an FDS where $E = L^1(X, \Sigma, \mu)$, $\mu(X) < \infty$, and T is a positive contraction such that $T\mathbf{1} \leqslant \mathbf{1}$. Then T is mean ergodic.

Proof. The order interval $[-\mathbf{1}, \mathbf{1}] := \{f \in L^1(\mu) : -\mathbf{1} \leqslant f \leqslant \mathbf{1}\}$ is the unit ball of the dual $L^\infty(\mu)$ of $L^1(\mu)$ and therefore $\sigma(L^\infty, L^1)$-compact. The topology induced by $\sigma(L^1, L^\infty)$ in $[-\mathbf{1}, \mathbf{1}]$ is coarser than that induced by $\sigma(L^\infty, L^1)$ – since $L^\infty(\mu) \subseteq L^1(\mu)$ – but still Hausdorff. Therefore the two topologies coincide (see A.2) and

$[-\mathbf{1},\mathbf{1}]$ is weakly compact. By assumption, T and therefore the Cesàro means T_n map $[-\mathbf{1},\mathbf{1}]$ into itself, hence (IV.4.c) is satisfied for all $f \in L^\infty(\mu)$. As shown in (B.14) the same property follows for all $f \in L^1(\mu)$. ∎

Using deeper functional-analytic tools one can generalize the above corollary still further: Let T be a positive contraction on $L^1(X,\Sigma,\mu)$ and assume that the set $\{T_n u : n \in \mathbb{N}\}$ is relatively compact for some strictly positive function $u \in L^1(\mu)$. By [Schaefer, II.8.8] it follows that $\bigcup_{n\in\mathbb{N}}\{g \in l^1(\mu) : 0 \leqslant g \leqslant T_n u\}$ is also relatively weakly compact. From $0 \leqslant T_n f \leqslant T_n u$ fo $0 \leqslant f \leqslant u$, (B.14) and (IV.4.c) we conclude that T is mean ergodic (see Ito [1965], Yeadon[1980]).

Example 3: Let $(X,\Sigma,\mu;\varphi)$ be an MDS. The induced operator T_φ in $L^2(X,\Sigma,\mu)$ is mean ergodic, and the corresponding projection is a conditional expectation: The first assertion follows from (IV.6) while the second is proved as in Example 2 above.

Example 4: Let $E = L^1([0,1],\mathcal{B},m)$, m the Lebesgue measure, and $k : [0,1]^2 \to \mathbb{R}_+$ be a measurable function, such that $\int_0^1 k(x,y)\,\mathrm{d}y = 1$ for all $x \in [0,1]$. Then the kernel operator

$$T : E \to E, \quad f \mapsto Tf(x) := \int_0^1 k(x,y)f(y)\,\mathrm{d}y$$

is mean ergodic.

Even though there is still much to say about the functional-analytic properties of mean ergodic operators, we here concentrate on their ergodic properties as defined in Lecture III. A particularly satisfactory result is obtained for MDSs, since the induced operators are automatically mean ergodic on $L^p(\mu)$, $1 \leqslant p < \infty$.

IV.7 Proposition:
Let $(X,\Sigma,\mu;\varphi)$ be an MDS and $E = L^p(X,\Sigma,\mu)$, $1 \leqslant p < \infty$. Then T_φ is mean ergodic and the following properties are equivalent:

(a) φ is ergodic.

(b) The projection corresponding to T_φ has the form $P = \mathbf{1}\otimes\mathbf{1}$, i.e. $Pf = \langle f,\mathbf{1}\rangle\cdot\mathbf{1}$ for all $f \in E$

(c) $\frac{1}{n}\sum_{i=0}^{n-1}\int_X (f\circ\varphi^i)\cdot g\,\mathrm{d}\mu$ converges to $\int_X f\,\mathrm{d}\mu\cdot\int_X g\,\mathrm{d}\mu$ for all $f \in L^p(\mu)$, $g \in L^p(\mu)' = L^q(\mu)$ with $\frac{1}{p}+\frac{1}{q} = 1$.

(d) $\frac{1}{n}\sum_{i=0}^{n-1}\mu(A\cap\varphi^{-1}(B))$ converges to $\mu(A)\cdot\mu(B)$ for all $A,B \in \Sigma$.

(e) $\frac{1}{n}\sum_{i=0}^{n-1}\mu(A\cap\varphi^{-1}(A))$ converges to $\mu(A)^2$ for all $A \in \Sigma$.

Proof.
(a) ⇒ (b): Since φ is ergodic and T_φ is mean ergodic, the fixed spaces of T_φ and T'_φ are one-dimensional (III.4 and IV.4.e). Since P is a projection onto the T_φ-fixed space it must be of the form $f \mapsto Pf = \langle f,f'\rangle\mathbf{1}$ for some $f' \in E'$. But

$$\int_X f\,\mathrm{d}\mu = \langle f,\mathbf{1}\rangle = \langle f,T'_\varphi\mathbf{1}\rangle = \langle f,P'\mathbf{1}\rangle = \langle Pf,\mathbf{1}\rangle = \langle f,f'\rangle\cdot\langle\mathbf{1},\mathbf{1}\rangle = \langle f,f'\rangle$$

shows that $P = \mathbf{1}\otimes\mathbf{1}$.

(b) $\Rightarrow$ (c): Condition (c) just says that $\frac{1}{n}\sum_{i=0}^{n-1} T_\varphi^i$ converges toward $\mathbf{1}\otimes\mathbf{1}$ in the weak operator topology for the particular space $L^p(\mu)$ and its dual $L^q(\mu)$.

(c) $\Rightarrow$ (d): This follows if we take $f = \mathbf{1}_A$ and $g = \mathbf{1}_B$. The implication

(d) $\Rightarrow$ (e): is trivial.

(e) $\Rightarrow$ (a): Assume that $\varphi(A) = A \in \Sigma$. Then $\frac{1}{n}\sum_{i=0}^{n-1} \mu(A \cap \varphi^{-i}(A))$ is equal to $\mu(A)$ and converges to $\mu(A)^2$. Therefore $\mu(A)$ must be equal to 0 or 1. ∎

Remark: Further equivalences in (IV.7) are easily obtained by taking in (c) the functions f, g only from total subsets, resp. in (d) or (e) the sets A, B only from a subalgebra generating Σ.

The "automatic" mean ergodicity of T_φ in $L^p(\mu)$, $1 \leqslant p < \infty$ (by Example 2 and 5) is the reason why ergodic MDSs are characterized by the one-dimensional fixed spaces (see III.4). In fact, mean ergodicity is a rather weak property for operators on $L^p(\mu)$, $p \neq \infty$, in the sense that many operators (e.g. all contractions for $p \neq 1$ or all positive contractions satisfying $T\mathbf{1} \leqslant \mathbf{1}$ for $p = 1$) are mean ergodic.

For operators on spaces $C(X)$ the situation is quite different and mean ergodicity of $T \in \mathscr{L}(C(X))$ is a very strong property. The reason is that the sup-norm $\|\cdot\|_\infty$ is much finer than $\|\cdot\|_p$, therefore it is more difficult to identify weakly compact orbits (in order to apply IV.4.c) or the dual fixed space (in order to apply IV.4.e). Even for operators T_φ on $C(X)$ induced by a TDS one has mean ergodicity only if one makes additional assumptions, e.g. (IV.8 below or VIII.2). This non-convergence of the Cesàro means of T_φ accounts for many of the differences and additional complications in the topological counterparts to measure theoretical theorems. A first example is the characterization of minimality by one-dimensional fixed spaces.

IV.8 Proposition:
For a TDS $(X;\varphi)$ the following are equivalent:

(a) T_φ is mean ergodic in $C(X)$ and φ is minimal.
(b) There exists a unique φ-invariant probability measure, and this measure is strictly positive.

Proof. (a) $\Rightarrow$ (b): From (III.7.i) and (IV.4.e) we conclude that $\dim F = \dim F' = 1$ for the fixed spaces F in $C(X)$, resp. F' in $C(X)'$. Since T_φ is a positive operator, so is P and hence P'. Every element in $C(X)'$ is a difference of positive elements, the same is true for $F' = P'C(X)'$ and therefore F' is the subspace generated by a single probability measure called ν.

Let $0 \leqslant f \in C(X)$ with $\langle f, \nu\rangle = 0$ and define $Y := \bigcap\{[f \circ \varphi^n = 0] : n \in \mathbb{Z}\}$. Then is closed and φ-invariant, and therefore $Y = \varnothing$ or $Y = X$. If $Y = X$, then $f = 0$, whereas if $Y = \varnothing$ implies that for all $x \in X$ one has $f \circ \varphi^n(x) > 0$ for some $n \in \mathbb{Z}$. Since $\langle f, \nu\rangle = \langle F \circ \varphi^n, \nu\rangle$ for all $n \in \mathbb{N}$, this shows that $\langle f, \nu\rangle > 0$.

(b) $\Rightarrow$ (a): Let $f' \in C(X)'$ be T'_φ-invariant. Since T'_φ is positive, we obtain

$$|f'| = |T'_\varphi f'| \leqslant T'_\varphi |f'|$$

and $\langle \mathbf{1}, |f'| \rangle \leqslant \langle \mathbf{1}, T'_\varphi |f'| \rangle = \langle T_\varphi \mathbf{1}, |f'| \rangle = \langle \mathbf{1}, |f'| \rangle$. Hence $\langle \mathbf{1}, T'_\varphi |f'| - |f'| \rangle = \langle T_\varphi \mathbf{1} |f'| \rangle - \langle \mathbf{1}, |f'| \rangle = 0$, therefore $|f'|$ is T'_φ-invariant, and the dual fixed space F' is a vector lattice. Consequently every element in F' is difference of positive elements and – by assumption – F' is one-dimensional and spanned by the unique φ-invariant probability measure ν. Apply now (IV.4.e) to conclude that T_φ is mean ergodic. Again the corresponding projection is of the form $P = \mathbf{1} \otimes \nu$. Assume now that $Y \subseteq X$ is closed and φ-invariant. There exists $0 < f \in C(X)$ with $f(Y) \subseteq \{0\}$, $T_\varphi f(Y) \subseteq \{0\}$, therefore $(Pf)(Y) \subseteq \{0\}$. Hence $(\int_X f \, \mathrm{d}\nu) \mathbf{1}(Y) \subseteq \{0\}$ and Y must be empty. ∎

Example 5: The rotation φ_a induces a mean ergodic operator T_{φ_a} on $C(\Gamma)$: If $a^{n_0} = 1$ for some $n_0 \in \mathbb{N}$, the operator T_{φ_a} is periodic (i.e. $T_{\varphi_a}^{n_0} = \mathrm{id}$) and therefore mean ergodic (see IV.D.3).

In the other case, every probability measure invariant under φ_a is invariant under φ_{a^n} for all $n \in \mathbb{N}$ and therefore under all rotations. By (D.5) the normalized Lebesgue measure is the unique probability measure having this property, and the assertion follows by (IV.8.b).

The previous example may also be understood without reference to the uniqueness of Haar measure: Let G be a compact group. The mapping

$$G \to \mathscr{L}_s(C(G)) : h \mapsto T_{\varphi_h} \qquad \text{(see II.2.2)}$$

is continuous, hence the orbits – as well as their convex hulls – of any operator T_{φ_h} are relatively (norm)compact in $C(G)$. Then apply (IV.4.c) to obtain the following result.

IV.9 Proposition:
Any rotation operator on $C(G)$, G a compact group, is mean ergodic.

Exercise: The fixed space of T_{φ_g} in $C(G)$, where φ_g is the rotation by g on the compact group G, is one-dimensional if and only if $\{g^k : k \in \mathbb{Z}\}$ is dense in G.

IV.D Discussion

IV.D.0 Proposition:
Assume that $a \in \Gamma$ is not a root of unity. The induced rotation operator T_{φ_a} is mean ergodic on the Banach space $R(\Gamma)$ of all bounded Riemann integrable functions on Γ (with sup-norm), and the (normalized) Riemann integral is the unique rotation invariant normalized positive linear form on $R(\Gamma)$.

Proof. First, we consider characteristic functions χ of "segments" on Γ and show that the Cesàro means

$$T_n \chi := \frac{1}{n} \sum_{i=0}^{n-1} T_{\varphi_a}^i \chi$$

converge in sup-norm $\| \cdot \|_\infty$
For $\varepsilon > 0$ choose $f_\varepsilon, g_\varepsilon \in C(\Gamma)$ such that

$$0 \leqslant f_\varepsilon \leqslant \chi \leqslant g_\varepsilon$$

and

$$\int_\Gamma (g_\varepsilon - f_\varepsilon) \, \mathrm{d}m < \varepsilon, \quad m \text{ Lebesgue measure on } \Gamma.$$

But $T := T_{\varphi_a}$ is mean ergodic (with one-dimensional fixed space) on $C(\Gamma)$, i.e.

$$T_n g_\varepsilon \xrightarrow{\|\cdot\|_\infty} \int_\Gamma g_\varepsilon \, \mathrm{d}m \cdot \mathbf{1}$$

and
$$T_n f_\varepsilon \xrightarrow{\|\cdot\|_\infty} \int_\Gamma f_\varepsilon \, \mathrm{d}m \cdot \mathbf{1}.$$

From $T_n f_\varepsilon \leqslant T_n\chi \leqslant T_n g_\varepsilon$ we conclude that $\|\cdot\|_\infty - \lim_{n\to\infty} T_n\chi$ exists and is equal to $\int_\Gamma \chi \, \mathrm{d}m \cdot \mathbf{1}$. Now, let f be a bounded Riemann integrable function on Γ. Then for every $\varepsilon > 0$ there exist functions g_1, g_2 being linear combinations of segments such that

$$g_1 \leqslant f \leqslant g_2 \quad \text{and} \quad \int_\Gamma (g_2 - g_1) \, \mathrm{d}m < \varepsilon,$$

and an easy calculation shows that

$$\|\cdot\|_\infty - \lim_{n\to\infty} T_n f = \Big(\int_\Gamma f \, \mathrm{d}m\Big) \cdot \mathbf{1}.$$

Finally, since the fixed space of T in $R(\Gamma)$, which is equal to the fixed space under all rotations on Γ, has dimension one, the mean ergodicity implies the one-dimensionality of the dual fixed space. ∎

The preceding result is surprising, has interesting applications (see IV.D.6) and is optimal in a certain sense:

Example 6: The rotation operator T_{φ_a} induced by φ_a, $a \in \Gamma$ not a root of unity, is mean ergodic

neither on (i) $L^\infty(\Gamma, \mathcal{B}, m)$

nor on (ii) $B(\Gamma)$, the space of all bounded Borel measurable

functions on Γ endowed with the sup-norm.

Proof. (i) The rotation φ_a is ergodic on Γ, hence the fixed space of $T := T_{\varphi_a}$ in $L^1(m)$ and a fortiori in $L^\infty(m)$ has dimension one. We show that the dual fixed space F' is at least two-dimensional: Consider $A := \{a^n : n \in \mathbb{Z}\}$ and $I := \{\check{f} \in L^\infty(m) :$ there is $f \in \check{f}$ vanishing on some neighbourhood (depending on f) of $A\}$. Then I is $\neq \{0\}$, T-invariant and generates a closed (lattice or algebra) ideal J in $L^\infty(m)$. From the definition follows that $TJ \subseteq J$ and $\mathbf{1} \notin J$. Consequently, there exists $\nu \in (L^\infty(m))'$ such that $\langle \mathbf{1}, \nu\rangle = 1$, but ν vanishes on J. The same is true for $T'\nu$ and $T_n'\nu$ for all $n \in \mathbb{N}$. By the weak* compactness of the dual unit ball the sequence $\{T_n'\nu\}_{n\in\mathbb{N}}$ has a weak* accumulation point ν_0. As in (IV.4), c $\Rightarrow$ d we show that $\nu_0 \in F'$. Since $\langle \mathbf{1}, \nu_0\rangle = 1$ and $\langle f, \nu_0\rangle = 0$ for $f \in J$, we conclude $0 \neq \nu_0 \neq m$.

(ii) Take a 0-1-sequence $(c_i)_{i\in\mathbb{N}_0}$ which is not Cesàro summable, i.e.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} c_i$$

does not exist. The characteristic function χ of the set

$$\{a^n : c_n = 1\}$$

is a Borel function for which

$$T_n\chi(a)$$

does not converge, hence the functions $T_n\chi$ do not converge in $B(\Gamma)$. ∎

IV.D.1. "Mean ergodic" vs. "ergodic":
The beginner should carefully distinguish these concepts. "Ergodicity" is a mixing property of an MDS $(X, \Sigma, \mu; \varphi)$ (or a statement on the fixed space of T_φ in $L^p(X, \Sigma, \mu)$), while "mean ergodicity" is a convergence property of the Cesàro means of a linear operator on a Banach space. More systematically we agree on the following terminology: "Ergodicity" of a linear operator $T \in \mathscr{L}(E)$, E Banach space, refers to the convergence of the Cesàro means T_n with respect to the uniform, strong or weak operator topology and such operators will be called "uniformly ergodic", "strongly ergodic", resp. "weakly ergodic". For $\{T^n : n \in \mathbb{N}\}$ bounded, it follows from Theorem (IV.4) that weakly ergodic and strongly ergodic operators coincide. Therefore and in order to avoid confusion with "strongly ergodic" transformations (see IX.D.4) we choose a common and different name for such operators and called them "mean ergodic". Here, the prefix "mean" refers to the convergence in the L^2-mean in von Neumann's original ergodic theorem (IV.1). "Uniform ergodicity" is a concept much stronger than "mean ergodicity" and will be discussed in Appendix W in detail.

IV.D.2. Mean ergodic semigroups:
Strictly speaking it is not the operator T which is mean ergodic but the semigroup $\{T^n : n \in \mathbb{N}_0\}$ of all powers of T. More precisely, in the bounded case, mean ergodicity of T is equivalent by (IV.4.d) to the following property of the semigroup $\{T^m : n \in \mathbb{N}_0\}$: the closed convex hull

$$\overline{\mathrm{co}}\{T^n : n \in \mathbb{N}_0\}$$

of $\{T^n : n \in \mathbb{N}_0\}$ in $\mathscr{L}_s(E)$, which is still a semigroup, contains a zero element, i.e. contains P such that

$$SP = PS = P$$

for all $S \in \overline{\mathrm{co}}\{T^n : n \in \mathbb{N}_0\}$ (Remark: $PT = TP = P$ is sufficient!). This point of view is well suited for generalizations which shall, be carried out in Appendix Y. As an application of this method we show that every root of a mean ergodic operator is mean ergodic, too.

Theorem: Let E be a Banach space and $S \in \mathscr{L}(E)$ be a mean ergodic operator with bounded powers. Then every root of S is mean ergodic.

Proof. Assume that $S := T^k$ is mean ergodic with corresponding projection P_S. Define $P := \big(\frac{1}{k}\sum_{j=0}^{k-1} T^j\big)P_S$ and observe that $P \in \overline{\mathrm{co}}\{T^i : i \in \mathbb{N}_0\}$ and $TP := \big(\frac{1}{k}\sum_{j=0}^{k-1} T^{j+1}\big)P_S = P$, $(T^kP_S = P_S)$. Therefore, T is mean ergodic (see IV.4.d) and P is the projection corresponding to T. ∎

On the contrary, it is possible that no power of a mean ergodic operator is mean ergodic.

Example: Let $S : (x_n)_{n\in\mathbb{N}_0} \mapsto (x_{n+1})_{n\in\mathbb{N}_0}$ be the (left)shift on $\ell^\infty(\mathbb{N}_0)$ and take a 0-1-sequence $(a_n)_{n\in\mathbb{N}_0}$ which is not Cesàro summable.

For $k > 1$ we define elements $x_k \in \ell^\infty(\mathbb{N}_0)$:

$$x_k := (x_{k,n})_{n\in\mathbb{N}_0} \text{ by } \begin{cases} x_{k,n} := a_{\frac{n}{k}} & \text{for } n = ki \ (i \in \mathbb{N}_0) \\ x_{k,n} := -a_{\frac{n-1}{k}} & \text{for } n = ki + 1 \ (i \in \mathbb{N}_0) \\ x_{k,n} := 0 & \text{otherwise.} \end{cases}$$

Consider the closed S-invariant subspace E generated by $\{S^i x_k : i \in \mathbb{N}_0,\ k > 1\}$ in $\ell^\infty(\mathbb{N}_0)$ and the restriction $T := S|_E$. By construction we obtain $\|T_n x_k\| \leqslant \frac{2}{n}$ for all $k > 1$. Consequently, T is mean ergodic with corresponding projection $P = 0$. On the other hand the sequence $\left(\frac{1}{m}\sum_{i=0}^{m-1} x_{k,ki}\right)_{m\in\mathbb{N}} = \left(\frac{1}{m}\sum_{i=0}^{m-1} a_i\right)_{m\in\mathbb{N}}$ is not convergent for $k > 1$, i.e. the Cesaró means $T_m^k(x_k)$ of the powers T^{ik}, $i \in \mathbb{N}$, applied to x_k, do not converge. Therefore, no power T^k $(k > 1)$ is mean ergodic.

References: Sine [1976].

IV.D.3. Examples:

(i) A linear operator T on the Banach space $E = \mathbb{C}$ is mean ergodic if and only if $\|T\| \leqslant 1$. Express this fact in a less cumbersome way!

(ii) The following operators $T \in \mathscr{L}(E)$, E a Banach space, are mean ergodic with corresponding projection P:

(a) T periodic with $T^{n_0} = \mathrm{id}$, $n_0 \in \mathbb{N}$, implies $P = \frac{1}{n_0}\sum_{i=0}^{n_0-1} T^i$.

(b) T with spectral radius $r(T) < 1$ (e.g. $\|T\| < 1$) implies $P = 0$.

(c) T has bounded powers and maps bounded sets into relatively compact sets.

(d) $T(x_1, x_2, x_3, \dots) = (0, x_1, x_2, \dots)$ on ℓ^p, $1 < p < \infty$.

(e) $T(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots)$ on ℓ^p, $1 \leqslant p < \infty$.

(f) $Tf(x) = \int_0^x f(y)\,\mathrm{d}y$ for $f \in C([0,1])$.

(iii) The following operators are not mean ergodic:

(a) $Tf(x) = x \cdot f(x)$ on $C([0,1])$: $F = \{0\}$ but $\|T_n\| = 1$ for all $n \in \mathbb{N}$.

(b) $Tf(x) = f(x^2)$ on $C([0,1])$: $F = \langle \mathbf{1} \rangle$ but Dirac measures δ_0, δ_1 are contained in F'

(c) $T(x_1, x_2, x_3, \dots) = (0, x_1, x_2, \dots)$ on ℓ^1: $F = \{0\}$ but $\|T_n(x_k)\| = \|(x_k)\|$ for $0 \leqslant (x_k) \in \ell^1$.

(d) $T(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots)$ on ℓ^∞: $0-1$-sequence which is not Cesàro summable.

IV.D.4. Convex combinations of mean ergodic operators:

Examples of "new" mean ergodic operators can be obtained by convex combinations of mean ergodic operators. Our first lemma is due to Kakutani (see Sakai [1977], 1.6.6)

Lemma 1: Let E be a Banach space. Then the identity operator id is an extreme point of the closed unit ball in $\mathscr{L}(E)$

Proof. Take $T \in \mathscr{L}(E)$ such that $\|\mathrm{id} + T\| \leqslant 1$ and $\|\mathrm{id} - T\| \leqslant 1$. Then the same is true for the adjoints: $\|\mathrm{id}' + T'\| \leqslant 1$ and $\|\mathrm{id}' - T'\| \leqslant 1$. For $f' \in E'$ define $f_1' := (\mathrm{id}' + T')f'$. resp. $f_2' := (\mathrm{id}' - T')f'$, and conclude $f' = \frac{1}{2}(f_1' + f_2')$ and $\|f_1'\|, \|f_2'\| \leqslant \|f'\|$. A soon as f' is an extreme point of the unit ball in E' we obtain $f' = f_1' = f_2'$ and hence $T'f' = 0$. But by the Krein-Milman theorem this is sufficient to yield $T' = 0$, and hence $T = 0$. Now assume that $\mathrm{id} = \frac{1}{2}(R + S)$

for contractions $R, S \in \mathscr{L}(E)$, and define $T := \mathrm{id} - R$. This implies $\mathrm{id} - T = R$ and $\mathrm{id} + T = 2\mathrm{id} - R = S$. By the above considerations it follows that $T = 0$, i.e. $\mathrm{id} = R = S$. ■

Lemma 2: Let R, S be two commuting operators with bounded powers on a Banach space E, and consider

$$T := \alpha R + (1-\alpha)S$$

for $0 < \alpha < 1$. Then the fixed spaces $F(T)$, $F(R)$ and $F(S)$ of T, R and S are related by

$$F(T) = F(R) \cap F(S).$$

Proof. Only the inclusion $F(T) \subseteq F(R) \cap F(S)$ is not obvious. Endow E with an equivalent norm $\|x\|_1 := \sup\{\|R^n S^m x\| : n, m \in \mathbb{N}_0\}$, $x \in E$ and observe that R and S are contractive for the corresponding operator norm. From the definition of T we obtain

$$\mathrm{id}_{F(T)} = T|_{F(T)} = \alpha R|_{F(T)} + (1-\alpha)S|_{F(T)}$$

and $R|_{F(T)}, S|_{F(T)} \in \mathscr{L}(F(T))$, since R and S commute. Lemma 1 implies $R_{F(T)} = S|_{F(T)} = \mathrm{id}_{F(T)}$, i.e. $F(T) \subseteq F(R) \cap F(S)$. ■

Now we can prove the main result.

Theorem:
Let E be a Banach space and R, S two commuting operators on E with $\|R^n\|, \|S^n\| \leqslant c$ for all $n \in \mathbb{N}$. If R and S are mean ergodic, so is every convex combination

$$T := \alpha R + (1-\alpha)S, \quad 0 \leqslant \alpha \leqslant 1.$$

Proof. Let $0 < \alpha < 1$. By Lemma 2 we have $F(T) = F(R) \cap F(S)$ and $F(T') = F(R') \cap F(S')$, and by (IV.4.e) it suffices to show that $F(R) \cap F(S)$ separates $F(R') \cap F(S')$: For $f' \neq g'$ both contained in $F(R') \cap F(S')$ there is $f \in F(R)$ with $\langle f, f' \rangle \neq \langle f, g' \rangle$. Since $SF(R) \subseteq F(R)$ we have $P_S f \in F(R) \cap F(S)$ where P_S denotes the projection corresponding to S. Consequently

$$\langle P_S f, f' \rangle = \langle f, P_S' f' \rangle = \langle f, P_{S'} f' \rangle = \langle f, f' \rangle \neq \langle f, g' \rangle = \langle P_S f, g' \rangle.$$

■

The following corollaries are immediate consequences.

Corollary 1:
For T, R and S as above denote by P_R, resp. P_S the corresponding 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 2:
Let $\{R_i : 1 \leqslant i \leqslant m\}$ be a family of commuting mean ergodic operators with bounded powers. Then every convex combination $T := \sum_{i=1}^m \alpha_i R_i$ is mean ergodic.

IV.D.5. Mean ergodic operators with unbounded powers:
A careful examination of the proof of (IV.4) shows that the assumption

$$\|T^n\| \leqslant c \quad \text{for all } n \in \mathbb{N}_0,$$

may be replaced by the weaker requirements

$$\lim_{n\to\infty} \frac{1}{n}\|T^n\| = 0 \quad \text{and} \quad \|T_n\| \leqslant c \quad \text{for all } n \in \mathbb{N}.$$

The following example (Sato [1977]) demonstrates that such situations may occur. We define two sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$.

$$a_1 = 1, \quad a_n = 2\cdot 4^{n-2} \quad \text{for } n \geqslant 2$$

and
$$b_n = \sum_{i=1}^{n} a_i = \tfrac{1}{3}(2\cdot 4^{n-1}+1) \quad \text{for } n \in \mathbb{N}.$$

Endow
$$X := \{(n,i) : n \in \mathbb{N}, 1 \leqslant i \leqslant b_n\}$$

with the power set as σ-algebra Σ, and consider the measure μ defined by

$$\nu(\{(n,i)\}) := \begin{cases} 2^{1-n} & \text{if } 1 \leqslant i \leqslant a_n \\ \nu(\{(n-1, i-a_n)\}) & \text{if } a_n < i \leqslant b_n. \end{cases}$$

Observing that $\sum_{i=1}^{b_n} \nu(\{(n,i)\}) = 2^{n-1}$ we obtain a probability measure μ on Σ by

$$\mu(\{(n,i)\}) := 2\cdot 4^{-n}\cdot \nu(\{(n,i)\}).$$

The measurable (not measure-preserving!) transformation

$$\varphi : (n,i) \mapsto \begin{cases} (n, i+1) & \text{for } 1 \leqslant i < b_n \\ (n+1, 1) & \text{for } i = b_n \end{cases}$$

on X induces the desired operator $T := T_\varphi$ on $L^1(X,\Sigma,\mu)$.

First, it is not difficult to see that $\|T^k\| = 2^n$ for $k = b_n, b_n+1, \ldots, b_{n+1}-1$. This shows that $\sup\{\|T^k\| : k \in \mathbb{N}\} = \infty$ and $\lim_{k\to\infty} \frac{1}{k}\|T^k\| = 0$.

Second, for $b_n + 1 \leqslant k \leqslant b_{n+1}$ we estimate the norm of the Cesaró means

$$\|T_k\| \leqslant \frac{1}{b_n\cdot \nu(\{(m+1,1)\})} \sum_{i=1}^{b_{n+1}} \nu(\{(n+1,i)\}) = \frac{2^n}{\frac{1}{3}(2\cdot 4^{n-1}+1)\cdot 2^{-n}} \leqslant 6.$$

Finally, T is mean ergodic: With the above remark this follows from (IV.4.c) as in (IV.6).

IV.D.6. Equidistribution mod 1 (Kronecker, 1884; Weyl, 1916):
Mean ergodicity of an operator T with respect to the supremum norm in some function space is a strong and useful property. For example, if $T = T_\varphi$ for some $\varphi : X \to X$ and if $\chi = \mathbf{1}_A$ is the characteristic function of a subset $A \subseteq X$, then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} T^i\chi(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \chi(\varphi^i(x)), \quad x \in X$$

is the "mean frequency" of visits of $\varphi^n(x) \in A$. Therefore, if χ is contained in some function space on which T is mean ergodic (for $\|\cdot\|_\infty$), then this mean frequency exists (uniformly in $x \in X$). Moreover, if the corresponding projection P is one-dimensional, hence of the form $P = \mu \otimes \mathbf{1}$, the mean frequency of visits in A is equal to $\mu(A)$ for every $x \in X$.

This observations may be applied to the "irrational rotation" φ_a on Γ and to the Banach space $R(P)$ of all bounded Riemann integrable functions on Γ (see IV.D.0). Thus we obtain the following classical result on the equidistribution of sequences mod 1.

Theorem (Weyl, 1916):
Let $\xi \in [0,1]\backslash\mathbb{Q}$. The sequence $(\xi_n)_{n\in\mathbb{N}} := n\xi \mod 1$ is (uniformly) equidistributed in $[0,1]$, i.e. for every interval $[\alpha,\beta] \subseteq [0,1]$ holds

$$\lim_{n\to\infty} \frac{N(\alpha,\beta,n)}{n} = \beta - \alpha,$$

where $N(\alpha,\beta,n)$ denotes the number of elements $\xi_i \in [\alpha,\beta]$ for $1 \leqslant i \leqslant n$.

This theorem H. Weyl [1916] is the first example of number-theoretical consequences of ergodic theory. A first introduction into this circle: of ideas can be found in Jacobs [1972] or Hlawka [1979], while Furstenberg [1981] presents more and deeper results.

IV.D.7. Irreducible operators on L^p-spaces:
The equivalent statements of Proposition (IV.7) express essentially mean ergodicity and some "irreducibility" of the operator T_φ corresponding to the transformation φ. Using more operator theory, further generalizations should be possible (see also III.D.11). Here we shall generalize (IV.7) to FDSs $(E;T)$, where $E = L^p(X,\Sigma,\mu)$, $\mu(X) = 1$, $1 \leqslant p < \infty$, and $T \in \mathscr{L}(E)$ is positive satisfying $T\mathbf{1} = \mathbf{1}$ and $T'\mathbf{1} = \mathbf{1}$.

First, an operator-theoretical property naturally corresponding to "ergodicity" of a bi-measure-preserving transformation has to be defined.

Definition:
Let $(E;T)$ be an FDS as explained above. A set $A \in \Sigma$ is called T-*invariant* if $T\mathbf{1}_A(x) = 0$ for almost all $x \in X\backslash A$. The positive operator T is called irreducible if every T-invariant set has measure 0 or 1.

Remarks:
1. It is obvious that for an operator T_φ induced by an MDS $(X,\Sigma,\mu;\varphi)$ irreducibility of T_φ is equivalent to ergodicity of φ
2. If E is finite-dimensional, i.e. $X = \{x_1,\dots,x_n\}$, and T is reducible, i.e. not irreducible, then there exists a non-trivial T-invariant subset A of X. After a permutation of the points in X we may assume $A = \{x_1,\dots,x_k\}$ for $1 \leqslant k < n$. Then $T\mathbf{1}_A(x) = 0$ for all $x \in X\backslash A$ means that the matrix associated with T has the form

$$\begin{array}{cc} \begin{array}{ccc|c} & & & \!\!\!\!k \end{array} & \\ \left(\begin{array}{cc|c} \star & \star & \star \\ \star & \star & \star \\ \hline \multicolumn{2}{c|}{0} & \star \end{array}\right) & k \end{array}$$

Proposition: Let $(E;T)$ be an FDS formed by $E = L^p(X,\Sigma,\mu)$, $\mu(X) = 1$, $1 \leqslant p < \infty$, and a positive operator T satisfying $T\mathbf{1} = \mathbf{1}$ and $T'\mathbf{1} = \mathbf{1}$. Then T is mean ergodic and the following statements are equivalent:

(a) T is irreducible.
(a′) The fixed space F of T is one-dimensional, i.e. $F = \langle \mathbf{1} \rangle$.
(b) The corresponding mean ergodic projection has the form $P = \mathbf{1} \otimes \mathbf{1}$.
(c) $\langle T_n f, g\rangle$ converges to $\int_X f \, \mathrm{d}\mu \cdot \int_X g \, \mathrm{d}\mu$ for every $f \in L^p(\mu)$, $g \in L^1(\mu)$.
(d) $\langle T_n \mathbf{1}_A, \mathbf{1}_B\rangle$ converges to $\mu(A) \cdot \mu(B)$ for every $A, B \in \Sigma$.
(e) $\langle T_n \mathbf{1}_A, \mathbf{1}_A\rangle$ converges to $\mu(A)^2$ for every $A \in \Sigma$.

Proof. Observe first that the assumptions $T\mathbf{1} = \mathbf{1}$ and $T'\mathbf{1} = \mathbf{1}$ imply that T naturally induces contractions on $L^1(\mu)$, resp. $L^\infty(\mu)$. From the Riesz convexity theorem (e.g. Schaefer [1974], V.8.2) it follows that $\|T\| \leqslant 1$. Consequently, T is mean ergodic by (IV.5) or (IV.6)

(a)⇒ (a′): Assume that the T-fixed space F contains a function f which is not constant. By adding an appropriate multiple of $\mathbf{1}$ we may obtain that f assumes positive and negative values. Its absolute value satisfies

$$|f| = |Tf| \leqslant T|f| \quad \text{and} \quad \int_X |f| \, \mathrm{d}\mu = \int_X T|f| \, \mathrm{d}\mu,$$

hence $|f| \in F$ and also $0 < f^+ := \frac{1}{2}(|f| + f) \in F$ and $0 < f^- := \frac{1}{2}(|f| - f) \in F$. Analogously we conclude that for every $n \in \mathbb{N}$ the function

$$f_n^+ := \inf(n \cdot f^+, \mathbf{1}) = \frac{1}{2}(n \cdot f^+ + \mathbf{1} - |n \cdot f^+ - \mathbf{1}|)$$

is contained in F. From the positivity of T we obtain

$$\mathbf{1}_A = \sup\{f_n^+ : n \in \mathbb{N}\} \in F$$

where $A := [f^+ > 0]$. Obviously, A is a non-trivial T-invariant set.

The implications (a′) ⇒ (b) ⇒ (c) ⇒ (d) ⇒ (e) follow as in the proof of (IV.7).

(e) ⇒ (a): If A is T-invariant the hypothesis $T\mathbf{1} = \mathbf{1}$ implies $T\mathbf{1}_A \leqslant \mathbf{1}_A$ and the hypothesis $T'\mathbf{1} = \mathbf{1}$ implies that $T\mathbf{1}_A = \mathbf{1}_A$. Therefore,

$$\langle T_n \mathbf{1}_A, \mathbf{1}_A\rangle = \langle T\mathbf{1}_A, \mathbf{1}_A\rangle = \langle \mathbf{1}_A, \mathbf{1}_A\rangle = \mu(A)$$

and the condition (e) implies $\mu(A) \in \{0, 1\}$. ∎

IV.D.8. Ergodicity of the Markov shift:
As an application of (IV.7) we show that the ergodicity of the Markov shift $(\widehat{X}, \widehat{\Sigma}, \widehat{\mu}; \tau)$ (see II.6) with transition matrix $T = (a_{ij})$ and strictly positive invariant distribution $\mu = (p_0, \ldots, p_{k-1})^\top$ can be characterized by an elementary property of the $k \times k$- matrix T.

Proposition: The following are equivalent:

(a) The transition matrix T is irreducible.
(b) The Markov shift $(\widehat{X}, \widehat{\Sigma}, \widehat{\mu}; \tau)$ is ergodic.

Proof. As remarked (IV.7) ergodicity of τ is equivalent to the fact that the induced operator $\widehat{T}f := f \circ \tau$, $f \in L^1(\widehat{X}, \widehat{\Sigma}, \widehat{\mu})$, satisfies

$$\langle \widehat{T}_n \mathbf{1}_A, \mathbf{1}_B \rangle \to \widehat{\mu}(A) \cdot \widehat{\mu}(B)$$

for all $A, B \in \widehat{\Sigma}$, which are of the form

$$\begin{aligned} & A = [x_{-l} = a_{-l}, \ldots\ldots, x_l = a_l] \\ \text{and} \quad & B = [x_{-m} = b_{-l}, \ldots\ldots, x_l = b_m] \end{aligned}$$

with $a_j, b_j \in \{0, \ldots, k-1\}$.
For $n \in \mathbb{N}$ so large that $n' := n - (m + l + 1) \geqslant 0$, we obtain

$$\begin{aligned} \widehat{\mu}(\tau^n A \cap B) &= \widehat{\mu}[x_{-m} = b_{-m}, \ldots, x_m = b_m, x_{n-l} = a_{-l}, \ldots, x_{n+l} = a_l] \\ &= \sum_{c_1=0}^{k-1} \cdots \sum_{c_{n'}=0}^{k-1} \widehat{\mu}[x_{-m} = b_{-m}, \ldots, x_m = b_m, x_{m+1} = c_1, \ldots, x_{m+n'} = c_{n'}, \\ &\qquad\qquad x_{n-l} = a_{-l}, \ldots, x_{n+l} = a_l] \\ &= \sum_{c_1=0}^{k-1} \cdots \sum_{c_{n'}=0}^{k-1} \Big(p_{b_{-m}} \prod_{i=-m}^{m-1} t_{b_i b_{i+1}}\Big) \Big(t_{b_m c_1} \prod_{i=1}^{n'-1} t_{c_i c_{i+1}} t_{c_{n'} a_{-l}}\Big) \prod_{i=-l}^{l-1} t_{a_i a_{i+1}} \\ &= \widehat{\mu}(B) (T^{n-m-l})_{b_m a_{-l}} \cdot (p_{a_{-l}})^{-1} \widehat{\mu}(A). \end{aligned}$$

Thus $\lim_{n\to\infty} \langle \widehat{T}_n \mathbf{1}_A, \mathbf{1}_B \rangle = \widehat{\mu}(B) \cdot (\lim_{n\to\infty} T_n)_{b_m a_{-l}} \cdot (p_{a_{-l}})^{-1} \widehat{\mu}(A) = \widehat{\mu}(A) \cdot \widehat{\mu}(B)$, iff $(\lim_{n\to\infty} T_n)_{ij} = (\mathbf{1} \otimes \mu)_{ij} = p_j > 0$ for every $i, j \in \{0, \ldots, k-1\}$. By the assertion (b) in (IV.D.7, Proposition) the last condition is equivalent to the irreducibility of T. ∎

IV.D.9. A dynamical system which is minimal but not ergodic:
As announced in (III.D.10) we present a minimal TDS $(X; \varphi)$ such that the MDS $(X, \mathcal{B}, \mu; \varphi)$ is not ergodic for a suitable φ-invariant probability measure $\mu \in M(X)$.

Choose numbers $k_i \in \mathbb{N}$, $i \in \mathbb{N}_0$, such that

$$\begin{aligned} & (*) \qquad k_{i-1} \text{ divides } k_i \text{ for all } i \in \mathbb{N} \\ \text{and} \quad & (**) \qquad \sum_{i=1}^{\infty} \frac{k_{i-1}}{k_i} \leqslant \frac{1}{12}. \end{aligned}$$

For example we may take $k_i = 10^{(3^i)}$.
For $i \in \mathbb{N}$ define $Z_i := \{z \in \mathbb{Z} : |z - n \cdot k_i| \leqslant k_{i-1} \text{ for some } n \in \mathbb{Z}\}$ and observe that $\mathbb{Z} = \bigcup_{i \in \mathbb{N}} Z_i$, since k_i tends to infinity. Therefore

$$i(z) := \min\{j \in \mathbb{N} : z \in \mathbb{Z}_j\}$$

is well-defined for $z \in \mathbb{Z}$. Now take

$$a := (a_z)_{z \in \mathbb{Z}} \quad \text{with} \quad a_z := \begin{cases} 0 & \text{if } i(z) \text{ is even} \\ 1 & \text{if } i(z) \text{ is odd,} \end{cases}$$

and consider the shift

$$\tau : (x_z)_{z \in \mathbb{Z}} \mapsto (x_{z+1})_{z \in \mathbb{Z}}$$

on $\{0, 1\}^{\mathbb{Z}}$.

Proposition: With the above definitions and $X := \overline{\{\tau^s a : s \in \mathbb{Z}\}} \subseteq \{0,1\}^{\mathbb{Z}}$ the TDS $(X; \tau|_X)$ is minimal, and there exists a probability measure $\mu \in M(X)$ such that the MDS $(X, \mathcal{B}, \mu; \tau|_X)$ is not ergodic.

Proof. Clearly, X is T-invariant and $(X; \tau|_X)$ is a TDS. The (product) topology on $\{0,1\}^{\mathbb{Z}}$ – and on X – is induced by the metric

$$d((x_z),(y_z)) := \inf\Big\{\frac{1}{t+1} : x_z = y_z \text{for all } |z| < t\Big\}$$

The assertion is proved in several steps.

(i) Take $i \in \mathbb{N}$. By definition of the sets Z_j, $j = 1, \ldots, i$ the number $i(z)$ only depends on $z \mod k_i$ for $i(z) \leqslant i$, i.e. the finite sequence of 0's and 1's

$$a_{-i}, a_{-i+1}, \ldots, a_0, \ldots, a_{i-1}, a_i$$

reappears in $(a_z)_{z\in\mathbb{Z}}$ with constant period. Using the above metric d, the lemma in (III.D.5) shows that X is minimal

(ii) We prove that the induced operator $T := T_{\tau|_X}$, on $C(X)$ is not mean ergodic by showing that for the function $\in C(X)$ defined by

$$f((x_z)_{z\in\mathbb{Z}}) := x_1$$

the sequence $(T_n f(a))_{n\in\mathbb{N}}$ does not converge:

$$T_n f(a) = \frac{1}{n}\sum_{z=0}^{n-1} f(\tau^z a) = \frac{1}{n}\sum_{z=1}^{n} a_z,$$

and $\sum_{z=1}^{n} a_z$ is the number of those z ($1 \leqslant z \leqslant n$) for which $i(z)$ is odd. Consider $n = k_i$ and observe that the set $\{1, \ldots, k_i\} \cap Z_j$ has exactly $\frac{k_i}{k_j}(2k_{j-1}+1)$ elements for $j = 1, \ldots, i$. Now

$$\sum_{j=1}^{i} \frac{k_i}{k_j}(2k_{j-1}+1) \leqslant \sum_{j=1}^{i} \frac{3k_{j-1}k_i}{k_j} \leqslant 3k_i \cdot \frac{1}{12} = \frac{k_i}{4} \qquad \text{(use (**))},$$

i.e. $\{1, \ldots, k\} \cap \bigcup_{j=1}^{i} Z_j$ contains at most $\frac{k_i}{4}$ numbers. However $\{1, \ldots, k_i\} \subseteq Z_{i+1}$, hence

$$\Big|\{1, \ldots, k_i\} \cap (Z_{i+1} \backslash \bigcup_{j=1}^{i} Z_j)\Big| \geqslant \frac{3}{4}k_i,$$

and for all numbers in that intersection we have $i(z) = i+1$. In conclusion, one obtains

$$|T_{k_{i+1}} f(a) - T_{k_i} f(a)| \geqslant \frac{1}{2}.$$

(iii) Using (IV.8) and (App.S), Theorem 1, we conclude from (ii) taht there exist at least two different τ-invariant probability measures $\mu_1, \mu_2 \in C(X)'$. For $\mu := \frac{1}{2}(\mu_1 + \mu_2)$ the MDS $(X, \mathcal{B}, \mu; \tau|_X)$ is not ergodic by (App.S).

∎

Remark: For examples on the 2-torus see Parry [1980], and on non-metrizable subsets of the Stone-Čech compactification of $\mathbb{N}$ see Rudin [1958] and Gait-Koo [1972].

References: Ando [1968], Gait-Koo [1972], Jacobs [1960], Parry [1980], Raimi [1964], Rudin [1958].

IV.D.10. Uniquely ergodic systems and the Jewett-Krieger theorem: For an MDS $(X, \Sigma, \mu; \varphi)$ and for $f \in L^p(X, \Sigma, \mu)$, the means

$$\frac{1}{n} \sum_{i=0}^{n-1} T_\varphi^i f$$

converge with respect to the L^p-norm for $1 \leqslant p < \infty$. Concerning the convergence for L^∞-norm (i.e. sup-norm) we don't have yet a definite answer, but know that in general the sup-norm is too strong to yield mean ergodicity of T_φ on $L^\infty(\mu)$. This was shown in example 6 in Lecture IV for any ergodic rotation φ_a on the unit circle Γ. On the other hand, in this same example there exist T_φ-invariant norm-closed subalgebras $\mathscr{A}$ of $L^\infty(X, \Sigma, \mu)$ which are dense in $L^1(X, \Sigma, \mu)$ and on which T_φ becomes mean ergodic (e.g. take $\mathscr{A} = C(\Gamma)$ or even $R(\Gamma)$, see (IV.D.0)). Such a subalgebra $\mathscr{A}$ is isomorphic to a space $C(Y)$ for some compact space Y and the algebra isomorphism on $C(Y)$ corresponding to T_φ is of the form T_ψ for some homeomorphism $\psi : Y \to Y$ (use the Gelfand-Neumark theorem (C.9) and (II.D.5)). The TDS $(Y; \psi)$ is minimal, since T_ψ is mean ergodic with one-dimensional fixed space, and therefore it possesses a unique ψ-invariant, strictly positive probability measure ν (see IV.8). Such systems will be called *uniquely ergodic*, since they determine a unique ergodic MDS. On the other hand it follows from the denseness of $\mathscr{A}$ in $L^1(\Gamma, \mathcal{B}, \mu)$ that the MDS $(\Gamma, \mathcal{B}, m; \varphi_a)$ is isomorphic to $(Y, \mathcal{B}, \nu; \psi)$ (use VI.2), a fact that will be expressed by saying that the original ergodic MDS is isomorphic to some MDS that is uniquely determined by a uniquely ergodic TDS. In fact, $(\Gamma, \mathcal{B}, m; \varphi_a)$ is uniquely ergodic since $\mathscr{A}$ can be chosen to be $C(\Gamma)$, but this choice is by no means unique and $\mathscr{A} = L^\infty(\Gamma, \mathcal{B}, m)$ would not work. Therefore we pose the following interesting question! Is every ergodic MDS isomorphic to an MDS determined by a uniquely ergodic TDS? As we have explained above, this question is equivalent to the following:

Problem: Let $(X, \Sigma, \mu; \varphi)$ be an ergodic MDS. Does there always exist a T_φ-invariant closed subalgebra $\mathscr{A}$ of $L^\infty(X, \Sigma, \mu)$

(i) T_φ is mean ergodic on $\mathscr{A}$, and
(ii) $\mathscr{A}$ is dense in $L^1(X, \Sigma, \mu)$?

The subsequent answer to this problem shows that the rotation $(\Gamma, \mathcal{B}, m; \varphi_a)$ is quite typical: Isomorphic uniquely ergodic systems always exist, but the algebra $L^\infty(\mu)$ is (almost) always too large for that purpose.

Lemma: For an ergodic MDS $(X, \Sigma, \mu; \varphi)$ the following assertions are equivalent:
(a) φ is mean ergodic on $L^\infty(X, \Sigma, \mu)$.
(b) $L^\infty(X, \Sigma, \mu)$ is finite dimensional.

Proof. In view of the representation theorem in (VI.D.6) it suffices to consider operators

$$T_\psi : C(Y) \to C(Y)$$

induced by a homeomorphism on an extremally disconnected space Y. By assumption (a), T_ψ is mean ergodic with one-dimensional fixed space and strictly positive invariant linear form ν. Prom (IV.8) it follows that ψ has to be minimal, and hence $\{\psi^k(y) : k \in \mathbb{Z}\}$ is dense in Y for every $y \in Y$. The lemma in (VI.D.6) implies that $\{\psi^k(y) : k \in \mathbb{Z}\}$ and hence $\{y\}$ is not a null set for the measure corresponding to ν. Therefore, $\{y\}$ must be open and the compact space Y is discrete.

■

Having seen that T_ψ is not mean ergodic on all of $L^\infty(\mu)$ one might try to find smaller subspaces on which mean ergodicity is guaranteed.
On the other hand

$$F(T) \oplus \overline{(\mathrm{id} - T_\varphi)L^\infty}$$

is the largest subspace of $L^\infty(\mu)$ on which T_φ is mean ergodic (use **??**). Unfortunately, this subspace is "never" a subalgebra. More precisely:

IV.D.11 Proposition:
For any ergodic MDS $(X, \Sigma, \mu; \varphi)$ the following assertions are equivalent:
(a) T_φ is mean ergodic on $L^\infty(\mu)$.
(b) $L^\infty(\mu)$ is finite dimensional.
(c) $\langle \mathbf{1} \rangle \oplus \overline{(\mathrm{id} - T_\varphi)L^\infty}$ is a subalgebra of $L^\infty(\mu)$.

Proof. It suffices to show that (c) implies (a). To that purpose we assume that the Banach algebra $L^\infty(\mu)$ is represented as $C(Y)$, Y compact, and the algebra isomorphism corresponding to T_φ is of the form $T_\psi : C(Y) \to C(Y)$ for some homeomorphism $\psi : Y \to Y$ and $\psi \neq \mathrm{id}$. Denote by $\mathrm{Fix}(\psi)$ the fixed point set of ψ. Then every function $f \in \overline{(\mathrm{id} - T_\psi)C(Y)}$ vanishes on $\mathrm{Fix}(\psi)$. Take $0 \neq g \in (\mathrm{id} - T_\psi)C(Y)$. Its square g^2 is contained in the subspace on which the means of T_ψ^i converge and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} T_\psi^i g^2 = \big(\int_Y g^2 \, \mathrm{d}\nu\big)\mathbf{1}_Y$$

for the strictly positive ψ-invariant measure ν. Therefore $\mathrm{Fix}(\psi)$ must be empty. It is now a simple application of Urysohn's lemma to show that $(\mathrm{id} - T_\psi)C(Y)$ separates the points in Y. By the Stone-Weierstrass theorem we obtain that $\langle \mathbf{1} \rangle \oplus (\mathrm{id} - T_\psi)C(Y)$ is dense in $C(Y)$ and therefore that T_ψ is mean ergodic on $L^\infty(\mu)$. ■

After these rather negative results it becomes clear that our task consists in finding "large" subalgebras contained in $\langle \mathbf{1} \rangle \oplus \overline{(\mathrm{id} - T_\varphi)L^\infty(\mu)}$. This has been achieved by Jewett [1970] (in the weak mixing case) and Krieger [1972]. Theirs as well as all other available proofs rest on extremely ingenious combinatorial techniques and we regret not being able to present a functional-analytic proof of this beautiful theorem.

Theorem (Jewett-Krieger, 1970):
Let $(X, \Sigma, \mu; \varphi)$ be an ergodic MDS. There exists a T_φ-invariant closed subalgebra $\mathscr{A}$ of $L^\infty(X, \Sigma, \mu)$, dense in $L^1(X, \Sigma, \mu)$, on which T_φ is mean ergodic.

Applying an argument similar to that used in the proof of (IV.D.0) the algebra of the above theorem can be enlarged and the corresponding structure spaces become totally disconnected. In conclusion we state the following answer to the original question.

Corollary:
Every separable ergodic $(X, \Sigma, \mu; \varphi)$ is isomorphic to an MDS determined by a uniquely ergodic TDS on a totally disconnected compact metric space.

References: Bellow-Furstenberg [1979], Denker [1973], Hansel [1974], Hansel-Raoult [1973], Jewett [1970], Krieger [1972], Petersen [1983].

V. The Individual Ergodic Theorem

In $L^2(X, \Sigma, \mu)$, convergence in the quadratic mean (i.e. in L^2-norm) does not imply pointwise convergence, and therefore, von Neumann's ergodic theorem (IV.1) did not exactly answer the original question: For which observables f and for which states x does the time mean

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(\varphi^i(x)) \quad \text{exists?}$$

But very soon afterwards, and stimulated by von Neumann's result, G.D. Birkhoff came up with a beautiful and satisfactory answer.

V.1 Theorem (G.D. Birkhoff, 1931):
Let $(X, \Sigma, \mu; \varphi)$ be an MDS. For any $f \in L^2(X, \Sigma, \mu)$ and for almost every $x \in X$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(\varphi^i(x))$$

exists.

Even today the above theorem may not be obtained as easily as its norm-counterpart (IV.1). In addition, its modern generalizations are not as far reaching as the mean ergodic theorems contained in Lecture IV. This is due to the fact that for its formulation we need the concept of μ-a.e.-convergence, which is more strictly bound to the context of measure theory. For this reason we have to restrict our efforts to L^p-spaces, but proceed axiomatically as in Lecture IV.

V.2 Definition:
Let (X, Σ, μ) be a measure space and consider $E = L^p(X, \Sigma, \mu)$, $1 \leqslant p \leqslant \infty$. $\in \mathscr{L}(E)$ is called *individually ergodic* if for every $f \in E$ the Cesáro means $T_n f := \frac{1}{n} \sum_{i=0}^{n-1} T^i f$ converge μ-a.e. to some $\bar{f} \in E$.

Remark: The convergence of $T_n f$ in the above definition has to be understood in the following sense:
For every choice of functions g_n in the equivalence classes $\widetilde{T_n f}$, $n \in \mathbb{N}$, (see B.20) there exists a μ-null set N such that $g_n(x)$ converge for any $x \in X \backslash N$. Only in (V.D.6) we shall see how a.e.-convergence of sequences in $L^p(\mu)$ can be defined without referring to the values of representants.

There exist two main results generalizing Birkhoff's theorem, one for positive contractions on L^1, the other for the reflexive L^p-spaces. But in both cases the proof is guided by the following ideas: Prove first the a.e.-convergence of the Cesàro means T on some dense subspace of E (easy!). Then prove some "Maximal Ergodic Inequality" (difficult!), and – as an easy consequence – extend the a.e.-convergence to all of E.
Here we treat only the L^1-case and refer to App. V for the L^p-theorem.

V.3 Theorem (Hopf, 1954; Dunford-Schwartz, 1956):
Let (X, Σ, μ) be a probability space, $E = L^1(X, \Sigma, \mu)$ and $T \in \mathscr{L}(E)$. If T is positive, $T\mathbf{1} \leqslant \mathbf{1}$ and $T'\mathbf{1} \leqslant \mathbf{1}$, then T is individually ergodic.

Remark: The essential assumptions may also be stated as $\|T\|_\infty \leqslant 1$ and $\|T\|_1 \leqslant 1$ for the operator norms on $\mathscr{L}(L^\infty(\mu))$ and $\mathscr{L}(L^1(\mu))$. The proof of the above "individual ergodic theorem" will not be easy, but it is presented along the lines indicated above.

V.4 Lemma:
Under the assumptions of (V.3) there exists a dense subspace E_0 of $E = L^1(X, \Sigma, \mu)$ such that the sequence of functions $T_n f$ converges with respect to $\|\cdot\|_\infty$ for every $f \in E_0$.

Proof. By (IV.6), T is mean ergodic and therefore

$$L^1(\mu) = F \oplus \overline{(\mathrm{id} - T)L^1(\mu)} = F \oplus \overline{(\mathrm{id} - T)L^\infty(\mu)},$$

where F is the T-fixed space in $L^1(\mu)$. We take $E_0 := F \oplus (\mathrm{id} - T)L^\infty(\mu)$. The convergence is obvious for $f \in F$. But for $(\mathrm{id} - T)g$, $g \in L^\infty(\mu)$, we obtain, using (IV.3.0), the positivity of T and $T\mathbf{1} \leqslant \mathbf{1}$, the estimate

$$\begin{aligned}|T_n f| = |(\mathrm{id} - T)T_n g| &= \frac{1}{n}|(\mathrm{id} - T^n)g| \leqslant \frac{1}{n}(|g| + T^n|g|) \\ &\leqslant \frac{1}{n}(\|g\|_\infty \cdot \mathbf{1} + \|g\|_\infty \cdot T^n\mathbf{1}) \leqslant \frac{2}{n}\|g\|_\infty \cdot \mathbf{1}.\end{aligned}$$

■

V.5 Lemma (maximal ergodic lemma, Hopf, 1954):
Under the assumptions of (V.3) and for $f \in L^1(X, \Sigma, \mu)$, $n \in \mathbb{N}$, $\gamma \in \mathbb{R}_+$ we define

$$f_n^* := \sup\{T_k f : 1 \leqslant k \leqslant n\} \quad \text{and} \quad A_{n,\gamma}(f) := [f_n^* > \gamma].$$

Then

$$\gamma \cdot \mu(A_{n,\gamma}(f)) \leqslant \int_{A_{n,\gamma}(f)} f \,\mathrm{d}\mu \leqslant \|f\|.$$

Proof (Garsia, 1955):
We keep f, n and γ fixed and define

$$g := \sup\Big\{\sum_{i=0}^{k-1}(T^i f - \gamma) : 1 \leqslant k \leqslant n\Big\}.$$

First we observe that $A := A_{n,\gamma}(f) = [g > 0]$. Then

$$\begin{aligned}T(g^+) &\geqslant (Tg)^+, && \text{since } 0 \leqslant T \\ &\geqslant \sup\Big\{\Big(\sum_{i=0}^{k-1}(T^{i+1}f - \gamma T\mathbf{1})\Big)^+ : 1 \leqslant k \leqslant n\Big\}, && \text{analogously} \\ &\geqslant \sup\Big\{\Big(\sum_{i=0}^{k-1}(T^{i+1}f - \gamma\mathbf{1})\Big)^+ : 1 \leqslant k \leqslant n\Big\}, && \text{since } T\mathbf{1} \leqslant \mathbf{1} \\ &\geqslant \sup\Big\{\Big(\sum_{i=0}^{k-1}(T^{i+1}f - \gamma\mathbf{1})\Big)^+ : 1 \leqslant k \leqslant n-1\Big\},\end{aligned}$$

$$= \sup\Big\{\Big(\sum_{i=0}^{k}(T^i f - \gamma\mathbf{1}) - (f-\gamma\mathbf{1})\Big)^+ : 2 \leqslant k \leqslant n\Big\},$$

$$\geqslant \sup\Big\{\sum_{i=0}^{k-1}(T^i f - \gamma\mathbf{1}) - (f-\gamma\mathbf{1}) : 1 \leqslant k \leqslant n\Big\}, \qquad \geqslant g - (f-\gamma\mathbf{1}).$$

This inequality yields

$$\mathbf{1}_A \cdot (f-\gamma\mathbf{1}) \geqslant \mathbf{1}_A \cdot g - \mathbf{1}_A \cdot T(g^+) \geqslant g^+ - T(g^+).$$

Finally the hypothesis $T'\mathbf{1} \leqslant \mathbf{1}$ implies

$$\int_A (f-\gamma\mathbf{1})\,\mathrm{d}\mu = \langle \mathbf{1}_A \cdot (f-\gamma\mathbf{1}), \mathbf{1}\rangle \geqslant \langle g^+ - T(g^+), \mathbf{1}\rangle = \langle g^+, \mathbf{1}\rangle - \langle g^+, T'\mathbf{1}\rangle \geqslant 0.$$

■

Remarks:

1. $f^* := \sup\{T_k f : k \in \mathbb{N}\}$ is finite a.e., since $\mu[f^* > m] = \mu[\sup_{n\in\mathbb{N}} f_n^* > m] \leqslant \frac{\|f\|}{m}$ for every $m \in \mathbb{N}$, and therefore
$$\mu\Big(\bigcap_{m\in\mathbb{N}}[f^* > m]\Big) = 0 \quad \text{or} \quad \mu[f^* < \infty] = \mu\Big(\bigcup_{m\in\mathbb{N}}[f^* \leqslant m]\Big) = 1.$$
2. Observe that we didn't need the assumption $\mu(X) < \infty$ in (V.5). The essential condition was that T is defined on $L^\infty(\mu)$ and $L^1(\mu)$, and contractive for $\|\cdot\|_\infty$ and $\|\cdot\|_1$.

V.6. Proof of Theorem (V.3):
We take $0 \neq f \in L^1(\mu)$ and show that

$$h_f(x) := \limsup_{n,m\in\mathbb{N}} |T_n f(x) - T_m f(x)| = 0$$

for almost every $x \in X$. With the notation introduced above we have $h_f(x) \leqslant 2|f|^*(x)$ and $h_f(x) h_{f-f_0}(x)$ for every f_0 contained in the subspace E_0 of $\|\cdot\|_\infty$-convergence found in (V.4). By the maximal ergodic inequality (V.5) we obtain for $\gamma > 0$ the estimate

$$\mu[h_f > \gamma\|f-f_0\|] = \mu[h_{f-f_0} > \gamma\|f-f_0\|] \leqslant \mu[|f-f_0|^* > \tfrac{\gamma}{2}\|f-f_0\|]$$
$$\leqslant \frac{2\|f-f_0\|}{\gamma\|f-f_0\|} = \frac{2}{\gamma}.$$

For $\varepsilon > 0$ take $\gamma = \frac{1}{\varepsilon}$, choose $f_0 \in E_0$ such that $\|f-f_0\| < \varepsilon^2$, and conclude

$$\mu[h_f > \varepsilon] \leqslant 2\varepsilon.$$

This shows that $h_f = 0$ a.e.. ■

Remark: The limit function $\bar{f}(x) := \lim_{n\to\infty} T_n f(x)$ is equal to Pf where P denotes the projection corresponding to the mean ergodic operator T. Therefore $\bar{f}$ is contained in $L^1(\mu)$.

Since $L^2(X,\Sigma,\mu) \subseteq L^1(X,\Sigma,\mu)$ for finite measure spaces, the Birkhoff theorem (V.1) follows immediately from (V.3) for $T = T_\varphi$. Moreover we are able to justify why "ergodicity" is the adequate "ergodic hypothesis" (compare III.D.6).

V.7 Corollary:
For an MDS $(X, \Sigma, \mu; \varphi)$ the following assertions are equivalent:

(a) φ is ergodic.

(b) For all ("observables") $f \in L^1(X, \Sigma, \mu)$ and for almost every ("state") $x \in X$ we have

$$\text{time mean} := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(\varphi^i(x)) = \int_X f \, \mathrm{d}\mu =: \text{ space mean.}$$

Proof. By (IV.7.b) the limit function $\bar{f}$ is the constant function $(\mathbf{1}\otimes\mathbf{1})f = (\int_X f \, \mathrm{d}\mu)\mathbf{1}$. ∎

V.D Discussion

V.D.1. "Equicontinuity" for a.e.-convergence:
The reader might have expected, after having proved in (V.4) a.e.-convergence on a dense subspace to finish the proof of (V.3) by a simple extension argument. For norm convergence, i.e. for the convergence induced by the norm topology, this is possible by "equicontinuity" (see B.11). But in the present context, we make the following observation.

Lemma: In general, the a.e.-convergence of sequences in $L^1(X, \Sigma, \mu)$ is not a topological convergence, i.e. there exists no topology on $L^1(X, \Sigma, \mu)$ whose convergent sequences are the a.e.-convergent sequences.

Proof. A topological convergence has the "star"-property, i.e. a sequence converges to an element f if and only if every subsequence contains a subsequence convergent to f (see Peressini [1967], p. 45). Consider $([0,1], \mathcal{B}, m)$, m the Lebesgue measure. The sequence of characteristic functions of the intervals $[0, \frac{1}{2}]$, $[\frac{1}{2}, 1]$, $[0, \frac{1}{4}]$, $[\frac{1}{4}, \frac{1}{2}]$, $[\frac{1}{2}, \frac{3}{4}]$, $[\frac{3}{4}, 1]$, $[0, \frac{1}{8}], \ldots$ does not converge almost everywhere, while every subsequence contains an a.e.-convergent subsequence (see A.16) ∎

Consequently, the usual topological equicontinuity arguments are of no use in proving a.e.-convergence and are replaced by the maximal ergodic lemma (V.5) in the proof of the individual ergodic theorem. In a more general context this has already been investigated by Banach [1926] and the following "extension" result is known as "Banach's principle" (see Garsia [1970]).

Proposition: Let $(S_n)_{n\in\mathbb{N}}$ (be a sequence of bounded linear operators on $L^p(X, \Sigma, \mu)$, $1 \leqslant p < \infty$, and consider

$$S^* f(x) := \sup_{n\in\mathbb{N}} |S_n f(x)|$$

and

$$G := \{f \in L^p : S_n f \text{ converges } \mu\text{-a.e.}\}$$

If there exists a positive decreasing function

$$c : \mathbb{R}_+ \to \mathbb{R}$$

such that $\lim_{\gamma\to\infty} c(\gamma) = 0$ and

$$\mu[S^* f(x) > \gamma \|f\|] \leqslant c(\gamma)$$

for all $f \in L^p(\mu)$, $\gamma > 0$, then the subspace G is closed.

Proof. Replace $\frac{\|f\|}{\gamma}$ in the proof of (V.6) by $c(\gamma)$ ∎

For an abstract treatment of this problem we refer to von Weizsäcker [1974]. See also (V.D.6).

V.D.2. Mean ergodic vs. individually ergodic:
A bounded linear operator on $L^p(X, \Sigma, \mu)$ may be mean ergodic or individually ergodic, but in general no implication is valid between the two concepts.

Example 1: The (right) shift operator

$$T : (x_n) \mapsto (0, x_1, x_2, \dots)$$

on $\ell^1(\mathbb{N}) = L^1(\mathbb{N}, \Sigma, \mu)$, where $\mu(\{n\}) = 1$ for every $n \in \mathbb{N}$, is individually ergodic, but not mean ergodic (IV.D.3).

Exercise: Transfer the above example to a finite measure space.

Example 2: On $L^2([0,1], \mathcal{B}, m)$, m Lebesgue measure, there exist operators which are not individually ergodic, but contractive hence mean ergodic (see App.V.10).

But a common consequence of the mean and individual ergodic theorem may be noted: On finite measure spaces (X, Σ, μ) the L^p-convergence and the a.e.-convergence imply the μ-stochastic convergence (see App.A.16).
Therefore

$$\lim_{n\to\infty} \mu[|T_n f(x) - \bar{f}(x)| \geqslant \varepsilon] = 0$$

for every $\varepsilon > 0$, $f \in L^p$, where $\bar{f}$ denotes the limit function of the Cesàro means $T_n f$ for a mean *or* individually ergodic operator $T \in \mathscr{L}(L^p(\mu))$.

In fact, even more is true.

Theorem (Krengel [1966]):
Let (X, Σ, μ) be a finite measure space and T be a positive contraction on $L^1(\mu)$. Then the Cesàro means $T_n f$ converge stochastically for every $f \in L^1(\mu)$.

V.D.3. Strong law of large numbers (concrete example):
The strong law of large numbers "is" the individual ergodic theorem. To make this evident we have to translate it from the language of probability theory into the language of MDSs. This requires some effort and will be performed in (V.D.7). Here we content ourselves with an application of the individual ergodic theorem, i.e. the strong law of large numbers, to a concrete model. As we have seen in (II.3.ii) the Bernoulli shift $B(\frac{1}{2}, \frac{1}{2})$ is an adequate model for "coin throwing". If we take $\mathbf{1}_A$ to be the characteristic function of the rectangle

$$A = \{x = (x_n) : x_0 = 1\}$$

in $\hat{X} = \{0,1\}^{\mathbb{Z}}$, then

$$\sum_{i=0}^{n-1} \mathbf{1}_A(\tau^i x), \quad \tau \text{ the shift on } \hat{X},$$

counts the appearances of "head" in the first n performances of our "experiment" $x = (x_n)$. Since $B(\frac{1}{2}, \frac{1}{2})$ is ergodic and since $\hat{\mu}(A) = \frac{1}{2}$, the individual ergodic theorem (V.7) asserts that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_A(\tau^i x) = \frac{1}{2}$$

for a.e. $x \in \hat{X}$, i.e. the average frequency of "head" in almost every "experiment" tends to $\frac{1}{2}$.

V.D.4. Borel's theorem on normal numbers:
A number $\xi \in [0,1]$ is called normal to base 10 if in its decimal expansion

$$\xi = 0, x_1 x_2 x_3 \dots \quad , x_i \in \{0, 1, 2, \dots, 9\},$$

every digit appears asymptotically with frequency $\frac{1}{10}$.

Theorem (Borel, 1909): Almost every number in $[0,1]$ is normal.

Proof. First we observe that the decimal expansion is well defined except for a countable subset of $[0,1]$. Modulo these points we have a bijection from $[0,1]$ onto $\hat{X} := \{0, 1, \dots, 9\}^{\mathbb{N}}$ which maps the Lebesgue measure onto the product measure $\hat{\mu}$ with

$$\hat{\mu}\{(x_n) \in \hat{X} : x_1 = 0\} = \dots = \hat{\mu}\{(x_n) \in \hat{X} : x_1 = 9\} = \frac{1}{10}.$$

Consider the characteristic function χ of $\{(x_n) \in \hat{X} : x_1 = 1\}$ and the operator $T : L^1(\hat{X}, \hat{\Sigma}, \hat{\mu}) \to L^1(\hat{X}, \hat{\Sigma}, \hat{\mu})$ induced by the (left) shift

$$\tau : (x_n) \mapsto (x_{n+1}).$$

Then $\sum_{i=0}^{n-1} T^i \chi(x) = \sum_{i=0}^{n-1} \chi(\tau^i x)$ is the number of appearances of 1 in the first n digits of $x = (x_n)$. Since T is individually ergodic with one-dimensional fixed space, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} T^i \chi(x) = \int_X \chi \, \mathrm{d}\hat{\mu} = \frac{1}{10}$$

for almost every $x \in X$. The same is true for every other digit. ∎

V.D.5. Individually ergodic operators on $C(X)$:
It seems to be natural to adapt the question of a.e.-convergence of the Cesàro means $T_n f$ to other function spaces as well. Clearly, in the topological context and for the Banach space $C(X)$ the a.e.-convergence has to be replaced by pointwise convergence everywhere. But for bounded sequences $(f_n) \subseteq C(X)$ pointwise convergence to a continuous function is equivalent to weak convergence (see App.B.18), and by (IV.4.b) this "individual" ergodicity on $C(X)$ would not be different from mean ergodicity.

Proposition: For an operator $T \in \mathscr{L}(C(X))$ satisfying $\|T^n\| \leqslant c$ the following assertions are equivalent:

(a) For every $f \in C(X)$ the Cesàro means $T_n f$ converge pointwise to a function $f \in C(X)$.
(b) T is mean ergodic.

V.D.6. A.e.-convergence is order convergence:
While the mean ergodic theorem relies on the norm structure of $L^p(\mu)$ (and therefore generalizes to Banach spaces) there is strong evidence that the individual ergodic theorem is closely related to the order structure of $L^p(\mu)$. One reason – for others see App.V – becomes apparent in the following lemma.

Lemma: An order bounded sequence $(f_n) \subseteq L^p(X, \Sigma, \mu)$, $1 \leqslant p \leqslant \infty$, converges a.e. if and only if it is "order convergent", i.e.

$$\mathrm{o} - \overline{\lim_{n\to\infty}}\, f_n := \inf_{k\in\mathbb{N}} \sup_{n\geqslant k} f_n = \sup_{k\in\mathbb{N}} \inf_{n\geqslant k} =: \mathrm{o} - \underline{\lim_{n\to\infty}}\, f_n .$$

The proof is a simple measure-theoretical argument. It is important that the "functions" f in the order limit are elements of the order complete Banach lattice $L^p(\mu)$. In particular, "null sets" and "null functions" don't occur any more. Since the sequences $(T_n f)$ in the individual ergodic theorem are unbounded one needs a slightly more general concept. We decided not to discuss such a concept here since it seems to us that a *purely* vector lattice theoretical approach to the individual ergodic theorem has yet to prove its significance.

References: Ionescu Tulcea [1969], Peressini [1967], Yoshida [1940].

V.D.7. Strong law of large numbers (proof):
As indicated in (V.D.3) this fundamental theorem of probability theory can be obtained from the individual ergodic theorem by a translation of the probabilistic language into ergodic theory.

Theorem (Kolmogorov, 1933):
Let $(f_n)_{n\in\mathbb{N}_0}$ be a sequence of independent identically distributed integrable random variables. Then $\frac{1}{n}\sum_{i=0}^{n-1} f_i$ converge a.e. to the expected value $\mathbb{E}f_0$.

Explanation of the terminology: f is a *random variable* if there is a probability space $(\Omega, \mathscr{A}, \mathcal{P})$ such that $f : \Omega \to \mathbb{R}$ is measurable (for the Borel algebra $\mathcal{B}$ on $\mathbb{R}$). The probability measure $\mathcal{P} \circ f^{-1}$ is called the *distribution* of f, and for $A \in \mathcal{B}$ one usually writes

$$\mathcal{P}[f \in A] := p(f^{-1}(A)).$$

Two random variables f_i, f_j are *identically distributed* if they have the same distribution, i.e. $p[f_i \in A] = p[f_j \in B]$ for every $A \in \mathcal{B}$. A sequence (f_n) of random variables is called *independent* if for any finite set $J \subseteq \mathbb{N}$ and any sets $A_j \in \mathcal{B}$ we have

$$\mathcal{P}[f_j \in A_j \text{ for every } j \in J] := \mathcal{P}\Big(\bigcap_{j\in J} f_j^{-1}(A_j)\Big) = \prod_{j\in J} p(f_j^{-1}(A_j)) = \prod_{j\in J} \mathcal{P}[f_j \in A_j].$$

Finally, f is *integrable* $f \in L^1(\Omega, \mathscr{A}, \mathcal{P})$, and its *expected value* is

$$\mathbb{E}f := \int_\Omega f \,\mathrm{d}\mathcal{P}(\omega) = \int_\mathbb{R} t \,\mathrm{d}(\mathcal{P} \circ f^{-1})(t).$$

Proof of the Theorem. Denote by μ the distribution of (f_n), i.e.

$$\mu := \mathcal{P} \circ f_n^{-1} \quad \text{for every } n \in \mathbb{N}$$

Consider

$$\hat{X} = \mathbb{R}^\mathbb{Z}$$

with the product measure $\widehat{\mu}$ on the product σ-algebra $\widehat{\Sigma}$. With the (left) shift $\tau : \widehat{X} \to \widehat{X}$ we obtain an MDS $(\widehat{X}, \widehat{\Sigma}, \widehat{\mu}; \tau)$ which is a continuous version of the Bernoulli shift on a finite set (see II.3.iii). As in (III.5.ii) we can verify that $(\widehat{X}, \widehat{\Sigma}, \widehat{\mu}; \tau)$ is ergodic, and the individual ergodic theorem implies

$$\frac{1}{n} \sum_{i=0}^{n-1} T_\tau^i \widehat{f} \xrightarrow{\text{a.e.}} \int_{\widehat{X}} \widehat{f} \, \mathrm{d}\widehat{\mu} \quad \text{for every } \widehat{f} \in L^1(\widehat{X}, \widehat{\Sigma}, \widehat{\mu}).$$

Next, denote the projections onto the i^{th} coordinate by

$$\pi_i : \widehat{X} \to \mathbb{R},$$

i.e. $\pi_i((x_n)) = x_i$. By assumption, $\pi_0 \in L^1(\widehat{X}, \widehat{\Sigma}, wh\mu)$ and $T_\tau^i \pi_0 = \pi_i$. Therefore

$$\frac{1}{n} \sum_{i=0}^{n-1} \pi_i \xrightarrow{\text{a.e.}} \int_{\widehat{X}} \pi_0 \, \mathrm{d}\widehat{\mu} = \int_{\mathbb{R}} t \, \mathrm{d}\mu(t) = \mathbb{E} f_0.$$

In the final step we have to transfer the a.e.-convergence on $\widehat{X}$ to the a.e.-convergence on Ω. The set of all finite products $\prod_{j\in J} g_j \circ \pi_j$ with $0 \leqslant g_j \in L^1(\mathbb{R}, \mathcal{B}, \mu)$ is total in $L^1(\widehat{X}, \widehat{\Sigma}, \widehat{\mu})$ by construction of the product σ-algebra. On these elements we define a mapping Φ by

$$\Phi(\prod_{j\in J} g_j \circ \pi_j) := \prod_{j\in J} g_j \circ f_j.$$

From

$$\begin{aligned} \int_{\mathbb{R}} \Big(\prod_{j\in J} g_j \circ \pi_j \, \mathrm{d}\widehat{\mu} \Big) &= \prod_{j\in J} \Big(\int_{\mathbb{R}} g_j \, \mathrm{d}\mu \Big) = \prod_{j\in J} \Big(\int_{\Omega} g_j \circ f_j \, \mathrm{d}\mathcal{P} \Big) \\ &= \int_{\Omega} \prod_{j\in J} g_j \circ f_j \, \mathrm{d}\mathcal{P} = \int_{\Omega} \Phi\Big(\prod_{j\in J} g_j \circ f_j \Big) \, \mathrm{d}\mathcal{P} \end{aligned}$$

it follows that Φ can be extended to a linear isometry

$$\Phi : L^1(\widehat{X}, \widehat{\Sigma}, \widehat{\mu}) \to L^1(\Omega, \mathscr{A}, \mathcal{P}).$$

But, Φ is positive, hence preserves the order structure of the L^1-spaces and by (V.D.6) the a.e.-convergence. Therefore,

$$\frac{1}{n} \sum_{i=0}^{n-1} \Phi(\pi_i) = \frac{1}{n} \sum_{i=0}^{n-1} f_i$$

converges a.e. to $\int_\Omega \Phi(\pi_0) \, \mathrm{d}\mathcal{P} = \mathbb{E} f_0$. ∎

Remark: In the proof above we constructed a Markov shift corresponding to $p(x, A) = \mu(A)$, $x \in \mathbb{R}$, $A \in \mathcal{B}$.

References: Bauer [1968], Kolmogorov [1933], Lamperti [1977].

V.D.8. Ergodic theorems for non-positive operators:
The positivity of the operator is essential for the validity of the individual ergodic theorem. It is however possible to extend such theorems to operators which are dominated by positive operators. First we recall the basic definitions from Schaefer [1974].

Let E be an order complete Banach lattice. $T \in \mathscr{L}(E)$ is called regular if T is the difference of two positive linear operators. In that case,

$$|T| := \sup(T, -T)$$

exists and the space $\mathscr{L}^r(E)$ of all regular operators becomes a Banach lattice for the regular norm

$$\|T\|_r := \||T|\|.$$

If $E = L^1(\mu)$ or $E = L^\infty(\mu)$ then $L^r(E) = \mathscr{L}(E)$ and $\|\cdot\| = \|\cdot\|_r$ (Schaefer [1974], IV.1.5). This yields an immediate extension of (V.3).

Proposition 1: Let (X, Σ, μ) be a probability space, $E = L^1(X, \Sigma, \mu)$ and $T \in \mathscr{L}(E)$. If T is a contraction on $L^1(\mu)$ and on $L^\infty(\mu)$ then T is individually ergodic.

Proof. $|T|$ still satisfies the assumptions of (V.3), hence (V.4) and (V.5) are valid for $|T|$. But $\pm T \leqslant |T|$ implies the analogous assertion for T, hence T is individually ergodic. ∎

For $1 < p < \infty$, we have $\mathscr{L}^r(L^p) \neq \mathscr{L}(L^p)$ in general but by similar arguments we obtain from (App.V.8):

Proposition 2: Every regular contraction T, i.e. $\|T\|_r \leqslant 1$, on an L^p-space, $1 < p < \infty$ is individually ergodic.

References: Chacón- Krengel [1964], Gologan [1979], Krengel [1963], Sato [1977], Schaefer [1974].

V.D.9. A non-commutative individual ergodic theorem:
$L^\infty(X, \Sigma, \mu)$ is the prototype of a commutative W^*-algebra. Without the assumption of commutativity, every W^*-algebra can be represented as a weakly closed self-adjoint operator algebra on a Hilbert space (e.g. see Sakai [1971], 1.16.7). Since such algebras play an important role in modern mathematics and mathematical physics the following generalization of the Dunford-Schwartz individual ergodic theorem may be of some interest.

Theorem (Lance, 1976; Kümmerer, 1978):
Let $\mathscr{A}$ be a W^*-algebra and $T \in \mathscr{L}(\mathscr{A})$ a weak* continuous positive linear operator such that $T\mathbf{1} \leqslant \mathbf{1}$ and $T_*\mu \leqslant \mu$ for some faithful (= strictly positive) state μ in the predual $\mathscr{A}_*$. Then the Cesàro means $T_n x$ converge almost uniformly to $\bar{x} \in \mathscr{A}$ for every $x \in \mathscr{A}$, i.e. for every $\varepsilon > 0$ there exists a projection $p_\varepsilon \in \mathscr{A}$ such that $\mu(p_\varepsilon) < \varepsilon$ and $\|(T_n x - \bar{x})(\mathbf{1} - p_\varepsilon)\| \to 0$.

References: Conze-Dang Ngoc [1978], Kümmerer [1978], Lance [1976], Yeadon [1977].

VI. Isomorphism of Dynamical Systems

In an axiomatic approach to ergodic theory we should have defined isomorphism, i.e. "equality" of dynamical systems, immediately after the Definition (II.1) of the objects themselves. We preferred to wait and see what kind of properties are of interest to us. We shall now define isomorphism in such a way that these properties will be preserved. In particular, we saw that all properties of an MDS $(X, \Sigma, \mu; \varphi)$ are described by measurable sets $A \in \Sigma$ taken modulo μ-null sets (see e.g. III.1, III.3 and V.2). This suggests that the correct concept of isomorphism for MDSs should disregard null sets, and should be based on the measure algebra

$$\check{\Sigma} = \Sigma/\mathscr{N}$$

, where $\mathscr{N}$ is the σ-ideal of μ-null sets in Σ (see App.A.9).

Consequently, it is not the point to point map

$$\varphi : X \to X$$

which is our object of interest, but the algebra isomorphism

$$\check{\varphi} : \check{\Sigma} \to \check{\Sigma}$$

induced by φ and defined by

$$\check{\varphi}\check{A} := \widecheck{(\varphi^{-1}A} \quad \text{for } A \in \check{A} \in \check{\Sigma}.$$

This point of view may also be justified by the following observations:

(i) $\check{\varphi}$ is an isomorphism of the measure algebra $\check{\Sigma}$;
(ii) $(X, \Sigma, \mu; \varphi)$ is ergodic if and only if $\check{\varphi}\check{A} = \check{A}$ implies $\check{A} = \check{\emptyset}$ or $\check{A} = \check{X}$.

These considerations might motivate the following definition.

VI.1 Definition:
Two MDSs $(X, \Sigma, \mu; \varphi)$ and $(Y, T, \nu; \psi)$ are called *isomorphic* if there exists a measure-preserving isomorphism $\check{\Theta}$ from $\check{\Sigma}$ to $\check{T}$ such that the diagram

$$\begin{tikzcd} \check{\Sigma} \arrow[r, "\check{\varphi}"] \arrow[d, "\check{\Theta}"'] & \check{\Sigma} \arrow[d, "\check{\Theta}"] \\ \check{T} \arrow[r, "\check{\psi}"'] & \check{T} \end{tikzcd}$$

commutes.

While structurally simple, this definition might appear difficult to work with, since it deals with equivalence classes of measurable sets. But at least for those who are familiar with the "function" spaces $L^p(X, \Sigma, \mu)$, this causes no trouble. Indeed, the measure algebra $\check{\varphi} : \check{\Sigma} \to \check{\Sigma}$ is nothing else but the operator

$$T_\varphi : L^p(X, \Sigma, \mu) \to L^p(X, \Sigma, \mu)$$

induced by φ and restricted to the (equivalence classes of) characteristic functions, i.e.

$$T_\varphi \mathbf{1}_{=} \mathbf{1}_{\varphi^{-1}(A)} \quad \text{or} \quad T_\varphi \mathbf{1}_{\check{A}} = \mathbf{1}_{\check{\varphi}\check{a}}$$

for all $A \in \Sigma$.

Conversely, every measure-preserving measure algebra isomorphism can be uniquely extended to a linear and order isomorphism of the corresponding L^1-spaces. We therefore obtain a "linear operator version" of the above concept.

VI.2 Proposition:
Two MDSs $(X, \Sigma, \mu; \varphi)$ and $(Y, T, \nu; \psi)$ are isomorphic if and only if there exists a Banach lattice isomorphism

$$V : L^1(X, \Sigma, \mu) \to L^(Y, T, \nu)$$

with $V\mathbf{1}_X = \mathbf{1}_Y$ such that diagram

$$\begin{tikzcd} L^1(X, \Sigma, \mu) \arrow[r, "T_\varphi"] \arrow[d, "V"'] & L^1(X, \Sigma, \mu) \arrow[d, "V"] \\ L^1(Y, T, \nu) \arrow[r, "T_\psi"'] & L^1(Y, T, \nu) \end{tikzcd}$$

commutes.

Proof. The (equivalence classes of) characteristic functions χ are characterized by

$$\chi \wedge (\mathbf{1} - \chi) = 0.$$

Therefore, an isometric lattice isomorphism V maps the characteristic functions on X onto the characteristic functions on Y and thereby induces a measure-preserving isomorphism

$$\breve{\Theta} : \breve{\Sigma} \to \breve{T}$$

. Conversely, every measure-preserving algebra isomorphism

$$\breve{\Theta} : \breve{\Sigma} \to \breve{T}$$

induces an isometry preserving the lattice operations from the sub-lattice of all characteristic functions contained in $L^1(X, \Sigma, \mu)$ onto the sublattice of all characteristic functions in $L^1(Y, T, \nu)$. This isometry extends uniquely to a lattice isomorphism

$$V : L^1(X, \Sigma, \mu) \to L^1(Y, T, \nu).$$

Since $\breve{\Theta}$ determines V, and $\breve{\varphi}$, resp. $\breve{\psi}$, determine T_φ, resp. T_ψ, (and conversely) the commutativity of one diagram implies the commutativity of the other. ∎

Remarks:

1. The isometric lattice isomorphism $V : L^1(X, \Sigma, \mu) \to L^1(Y, T, \nu)$ in (VI.2) may be restricted to the corresponding L^p)-spaces, $1 \leqslant p <\leqslant \infty$ (use the Riesz convexity theorem, see Schaefer [1974], V.8.2). These restrictions are still isometric lattice isomorphisms for which the corresponding L^p-diagram commutes.
2. The proposition above (as II.D.6 and V.D.6) shows that the order structure of L^p and the positivity of T_φ is decisive in ergodic theory. Therefore, many ergodic-theoretical problems can be treated in the framework of Banach lattices (see Schaefer [1974], ch. III).

In the topological case the appropriate definition of isomorphism is quite evident.

VI.3 Definition:
Two TDSs $(X; \varphi)$ and $(Y; \psi)$ are called *isomorphic* if there exists a homeomorphism

$$\Theta : X \to Y$$

such that the diagram

$$\begin{tikzcd} X \arrow[r, "\varphi"] \arrow[d, "\Theta"'] & X \arrow[d, "\Theta"] \\ Y \arrow[r, "\psi"'] & Y \end{tikzcd}$$

commutes.

Note that by considering the Banach lattice (or Banach algebra) $C(X)$ one obtains an operator-theoretical version analogous to (VI.2).

VI.4 Remark Hilbert space isomorphism:
For historical reasons and because of the spectral properties (III.4.b) and (IX.4) one occasionally considers a concept of isomorphism for MDSs ("spectral isomorphism"), which is defined in analogy to (VI.2), but only requires the map

$$V : L^2(X, \Sigma, \mu) \to L^2(Y, T, \nu)$$

to be a Hilbert space isomorphism.

By Remark 1 following (VI.2) this concept is weaker than (VI.1). One can therefore lose "ergodic properties" which are not "spectral properties" in passing from one MDS to another which is spectrally isomorphic to the first. A trivial example is furnished by $([0,1], \mathcal{B}, m; \mathrm{id})$ with Lebesgue measure m and $(\mathbb{N}, \mathscr{P}(\mathbb{N}), \nu; \mathrm{id})$ with $\nu(\{n\}) := 2^{-n}$. These two MDSs are spectrally isomorphic but not isomorphic. The reason is that $L^2([0,1], \mathcal{B}, m)$ is – as a Hilbert space – isomorphic to $\ell^2(\mathbb{N})$ but $(\mathbb{N}, \mathscr{P}(\mathbb{N}), \nu; \mathrm{id})$, unlike $([0,1], \mathcal{B}, m; \mathrm{id})$, has minimal invariant sets with non-zero measure.

More important examples are the Bernoulli shifts $B(p_0, \ldots, p_{k-1})$ which are apectrally isomoprhic (see VII.D.5) but necessarily isomorphic (**??**).
This again indicates that Hilbert spaces are insufficient for the purposes of ergodic theory.

VI.5 Remark point isomorphism:
For practical reasons and in analogy to Definition (II.1), which uses point to point maps φ, another concept of isomorphism for MDSs is usually considered. It is defined analogously to (VI.1) but the measure-preserving algebra isomorphism

$$\breve{\Theta} : \breve{\Sigma} \to \breve{T}$$

is replaced by a bi-measure-preserving map $\Theta : X \to Y$ such that the diagram

$$\begin{tikzcd} X \arrow[r, "\varphi"] \arrow[d, "\Theta"'] & X \arrow[d, "\Theta"] \\ Y \arrow[r, "\psi"'] & Y \end{tikzcd}$$

commutes.

This point isomorphism is stronger than isomorphism since Θ induces an algebra isomorphism

$$\breve{\Theta} : \breve{T} \to \breve{\Sigma}$$

by $\breve{\Theta}\breve{B} := \widetilde{(\theta^{-1}B}$ for $B \in T$. In fact, there exist MDSs which are isomorphic but not pointwise isomorphic:

Take $(X, \Sigma, \mu; \varphi)$ with $X = \{x\}$, $\Sigma = \mathscr{P}(X)$, $\mu(X) = 1$, $\varphi = \text{id}$ and $(Y, T, \nu; \psi)$ with $Y = \{x, y\}$, $T = \{\varnothing, Y\}$, $\nu(Y) = 1$, $\psi = \text{id}$.

Nevertheless, most isomorphisms appearing in the applications and in concrete examples are point to point maps and not only measure algebra isomorphisms. For this reason we defined the concept of an MDS using point maps $\varphi : X \to Y$, and therefore one might prefer the concept of "point isomorphism".

The following classical result shows however that the distinction between isomorphic and point isomorphic (but not between isomorphic and spectrally isomorphic) is rather artificial. Consequently, we shall use the term isomorphism synonymously for algebra isomorphisms and point isomorphisms.

VI.6 Theorem von Neumann, 1932:
Two MDSs on compact metric probability spaces are isomorphic if and only if they are point isomorphic.

Proof. On compact metric probability spaces every measure-preserving measure algebra isomorphism is induced by a bi-measure-preserving point map (see **??**). Then the commutativity of the diagram in (VI.1) implies the commutativity of the corresponding diagram (VI.6) for point to point maps. ∎

VI.7. The isomorphism problem
is one of the central mathematical problems in modern ergodic theory. It consists in deciding whether two given MDSs (or TDSs) are isomorphic. This is easy if you succeed in constructing an isomorphism. If you don't succeed – even after great efforts – you cannot conclude on "non-isomorphism". The adequate mathematical principle for proving non-isomorphism of two MDSs is the following:

Consider isomorphism invariants of MDSs, i.e. properties of MDSs, which are preserved under isomorphisms. As soon as you find an isomorphism invariant distinguishing the two systems they can't be isomorphic. But even it is not impossible to construct an isomorphism between two MDSs (i.e. if they are isomorphic), such a construction might be extremely difficult. On the other hand, it might be easier to calculate the values of all "known" isomorphism invariants. Such a system of isomorphism invariants is called *complete* if two systems are isomorphic as soon as all of these invariants coincide. To find such a complete system of invariants for all MDSs is the dream of many ergodic theorists. Only for certain subclasses of MDSs this has been achieved (see Lecture **??** and (**??**).

VI.D Discussion

VI.D.1.

VI.D.2.

VI.D.3.

VI.D.4.

VI.D.5.

VI.D.6.

VII. Compact Operator Semigroups

Having investigated the asymptotic behavior of the Cesàro means

$$T_n := \frac{1}{n}\sum_{i=0}^{n-1} T^i$$

and having found convergence in many cases, we are now interested in the behavior of the powers

$$T^n$$

of $T(= T_\varphi)$ themselves. The problems and methods are functional-analytic, and for a better understanding of the occurring phenomena the theory of compact operator semigroups – initiated by Glicksberg-de Leeuw [1959] and Jacobs [1956] – seems to be the appropriate framework.

Therefore, in this lecture we present a brief introduction to this field, restricting ourselves to cases which will be applied to measure-theoretical and topological dynamical systems.

In the following, a semigroup S is a set with an associative multiplication

$$(t,s) \mapsto t \cdot s.$$

However such objects become interesting (for us) only if they are endowed with some additional topological structure.

VII.1 Definition:
A semigroup S is called a *semitopological semigroup* if S is a topological space such that the multiplication is separately continuous on $S \times S$. *Compact semigroups* are semitopological semigroups which are compact.

Remark: This terminology is consistent with that of App.D, since every compact (semitopological) group has jointly continuous multiplication (see VII.D.6) and therefore is a compact topological group.

For a theory applicable to operators on Banach spaces, it is important to assume that the multiplication is only separately continuous (see B.16). But this is still enough to yield an interesting structure theorem for compact semigroups. We present this result in the commutative case and recall first that an ideal in a commutative semigroup S is a nonempty subset J such that $SJ := \{st : s \in S\} \subseteq J$.

VII.2 Theorem:
Every commutative compact semigroup S contains a unique minimal ideal K, and K is a compact group.

Proof. Choose closed ideals $J_1, \dots, J_n$ in S. Since

$$\varnothing \neq J_1 J_2 \dots J_n \subseteq \bigcap_{i=1}^{n} J_i,$$

we conclude that the family of closed ideals in S has the finite intersection property, and therefore the ideal

$$K := \bigcap \big\{ J : \ J \text{ is a closed ideal} \big\}$$

is non-empty by the compactness of S. By the separate continuity of the multiplication, the principal ideal $Ss = sS$ generated by $s \in S$ is closed. This shows that K is contained in every ideal of S. Next we show that K is a group: $sK = K$ for every $s \in S$ since K is minimal. Hence there exists $q \in K$ such that $sq = s$. Moreover for any $r \in K$ there exists $r' \in K$ such that $r's = r$. This implies

$$rq = r'sq = r's = r,$$

i.e. q is a unit in K. Again from $sK = K$ we infer the existence of $t(= s^{-1})$ such that $st = q$. Finally, we have to show that the multiplication on a compact semigroup which is algebraically a group is already jointly continuous. As remarked above, this is a consequence of a famous theorem of Ellis (see VII.D.6). ∎

By the above theorem, in every compact commutative semigroup S we have a unique idempotent q, namely the unit of K, such that

$$K = qS$$

is an ideal in S and a compact group with unit q. Now we will apply this abstract result to semigroups generated by certain operators on Banach spaces. The situations which occurred in (IV.5) and (IV.6) are the main applications we have in mind.

VII.3 Lemma:
Let $(E; T)$ be an FDS satisfying

$$(*) \qquad \{T^n f : n \in \mathbb{N}\} \text{ is relatively weakly compact for every } f \in E.$$

Denote by $\mathscr{S} := \overline{\{T^n : n \in \mathbb{N}\}}$ the closure of $\{T^n : n \in \mathbb{N}\}$ in $\mathscr{L}(E)$ with respect to the weak operator topology. Then $\mathscr{S}$ and its closed convex hull $\overline{\text{co}}(\mathscr{S})$ are commutative compact semigroups.

Proof. Multiplication is separately continuous for the weak operator topology (see App.B.16), hence $\{T^n : n \in \mathbb{N}\}$ is a commutative semitopological semigroup in $\mathscr{L}(E)$. It is remarkable that separate continuity is sufficient to prove that its closure is still a semigroup and even commutative. We show the second assertion while the proof of the first is left to the reader. From the separate continuity it follows that operators in $\mathscr{S}$ commute with operators in $\{T^n : n \in \mathbb{N}\}$. Now take $0 \neq R_1, R_2 \in \mathscr{S}$, $f \in E$, $f' \in E'$ and $\varepsilon > 0$. Then there exists $R \in \{T^n : n \in \mathbb{N}\}$ such that

$$\begin{aligned} |\langle (R_2 - R)f, R_1 f' \rangle| &\leqslant \frac{\varepsilon}{2} \qquad \text{and} \\ |\langle (R_2 - R)R_1 f, f' \rangle| &\leqslant \frac{\varepsilon}{2}. \end{aligned}$$

Therefore we have

$$\begin{aligned} |\langle (R_1R_2 - R_2R_1)f, f' \rangle| &= |\langle (R_1R_2 - R_1R + RR_1 - R_2R_1)f, f' \rangle| \\ &\leqslant |\langle (R_1(R_2 - R)f, f' \rangle| + |\langle (R - R_2)R_1 f, f' \rangle| \leqslant \varepsilon, \end{aligned}$$

which implies $R_1R_2 = R_2R_1$.

Finally, the condition $(*)$ implies that $\mathscr{S}$ is compact in $\mathscr{L}_w(E)$ (see App.B.14).

Since the closed convex hull of a weakly compact set in E is still weakly compact (see App.B.6), and since the convex hull $\text{co}(\mathscr{S})$ is a commutative semigroup, the same arguments as above apply to $\overline{\text{co}}(\mathscr{S})$. ∎

Now we apply (VII.2) to the semigroups $\mathscr{S}$ and $\overline{\text{co}}(\mathscr{S})$. Thereby the semigroup $\overline{\text{co}}(\mathscr{S})$ leads to the already known results of Lecture IV.

VII.4 Proposition:
Let $(E;T)$ be an FDS satisfying $(*)$. Then T is mean ergodic with corresponding projection P, and $\{P\}$ is the minimal ideal of the compact semigroup $\overline{\text{co}}\{T^n : n \in \mathbb{N}_0\}$.

In particular, $\quad E = F \oplus F_0$

where $\quad F := PE = \{f \in E : Tf = f\}$

and $\quad F_0 := P^{-1}(0) = \overline{(\text{id} - T)E} = \big\{f \in E : 0 \in \overline{\text{co}}\{T^n f : n \in \mathbb{N}_0\}\big\}$.

Proof. The mean ergodicity of T follows from (IV.4.c), and $TP = PT = P$ (see IV.3.1) shows that $\{P\}$ is the minimal ideal in $\overline{\text{co}}\{T^n : n \in \mathbb{N}_0\}$. The remaining statements have already been proved in (IV.3) except the last identity which follows from (IV.4.d). ∎

Analogous reasoning applied to the semigroup

$$\mathscr{S} := \overline{\{T^n : n \in \mathbb{N}_0\}} \subseteq \mathscr{L}_w(E)$$

yields another splitting of E into T-invariant subspaces. The main point in the following theorem is the fact that we are again able to characterize these subspaces.

VII.5 Theorem:
Let $(E;T)$ be an FDS satisfying $(*)$. Then there exists a projection

$$Q \in \mathscr{S} := \overline{\{T^n : n \in \mathbb{N}_0\}}$$

such that $\quad \mathscr{K} := Q\mathscr{S}$

is the minimal ideal of $\mathscr{S}$ and a compact group with unit Q.

In particular, $\quad E = G \oplus G_0$

where $\quad G := QE = \overline{\text{lin}}\Big\{f \in E : Tf = \lambda f \text{ for some } \lambda \in \mathbb{C},\ |\lambda| = 1\Big\}$

and $\quad GF_0 := Q^{-1}(0) = \Big\{f \in E : 0 \in \overline{\{T^n f : n \in \mathbb{N}_0\}}^{\sigma(E,E')}\Big\}$.

Proof. (VII.2) and (VII.3) imply the first part of the theorem, while the splitting $E = G \oplus G_0 = QE \oplus Q^{-1}(0)$ is obvious since Q is a projection.

The characterizations of $Q^{-1}(0)$ and QE are given in three steps:

1. We show that $Q^{-1}(0) = \{f \in E : 0 \in \overline{\{T^n f : n \in \mathbb{N}_0\}}^{\sigma}\}$. Since for every $f \in E$ map $S \mapsto Sf$ is continuous from $\mathscr{L}_w(E)$ into E_σ and since Q is contained in $\mathscr{S}$, we see that $Qf = 0$ implies $0 \in \overline{\{T^n f : n \in \mathbb{N}_0\}}$. Conversely, if $0 \in \overline{\{T^n f : n \in \mathbb{N}_0\}}$, there exists an operator R in the compact semigroup $\mathscr{S}$ such that $Rf = 0$. A fortiori
$$QRf = 0 \quad \text{and} \quad Qf = R'QRf = 0$$
where R' is the inverse of QR in the group $\mathscr{K} = Q\mathscr{S}$.

2. Next we prove that

$$QE \subseteq H := \overline{\text{lin}}\{f \in E : Tf = \lambda f \text{ for some } |\lambda| = 1\}.$$

Denote by $\widehat{\mathscr{K}}$ the character group of $\mathscr{K}$ and define for every character $\gamma \in \widehat{\mathscr{K}}$ the operator P_γ

$$P_\gamma(f) := \int_{\mathscr{K}} \overline{\gamma(S)} Sf \, \mathrm{d}m(S), \quad f \in E.$$

Here, m is the normalized Haar measure on $\mathscr{K}$, and the integral is understood in the weak topology on E, i.e.

$$\langle P_\gamma(f), f' \rangle := \int_{\mathscr{K}} \overline{\gamma(S)} \langle Sf, f' \rangle \, \mathrm{d}m(S), \quad \text{for every } f' \in E'.$$

$P_\gamma(f)$ is an element of the bi-dual E'' contained in $\overline{\text{co}}\{\overline{\gamma(S)} \cdot Sf : S \in \mathscr{K}\}$. However by Krein's theorem (App.B.6) this set is $\sigma(E, E')$-compact and hence contained in E. Therefore P_γ is a well-defined bounded linear operator on E. Now take $R \in \mathscr{K}$ and observe that

$$\begin{aligned} RP_\gamma(f) &= R\Big(\int_{\mathscr{K}} \overline{\gamma(S)} Sf \, \mathrm{d}m(S)\Big) = \int_{\mathscr{K}} \overline{\gamma(S)} RSf \, \mathrm{d}m(S) \\ &= \gamma(R) \int_{\mathscr{K}} \overline{\gamma(RS)} RSf \, \mathrm{d}m(RS) = \gamma(R) P_\gamma(f) \quad \text{for every } f \in \mathrm{E} \end{aligned}$$

i.e., $\qquad RP_\gamma = P_\gamma R = \gamma(R) P_\gamma.$

For $R := TQ$ we obtain $TP_\gamma = TQP_\gamma = \gamma(TQ)P_\gamma$ and therefore $P_\gamma(H) \subseteq H$. The assertion is proved if we show that $QE \subseteq \overline{\text{lin}} \bigcup \{P_\gamma E : \gamma \in \widehat{\mathscr{K}}\}$ or equivalently that $\{P_\gamma E : \gamma \in \widehat{\mathscr{K}}\}$ is total in QE.

Take $f' \in E'$ vanishing on the above set, i.e., such that $\int_{\mathscr{K}} \overline{\gamma(S)} \langle Sf, f' \rangle \, \mathrm{d}m(S) = 0$ for all $\gamma \in \widehat{\mathscr{K}}$ and all $f \in E$. Since the mapping $S \mapsto \langle Sf, f' \rangle$ is continuous, and since the characters form a complete orthonormal basis in $L^2(\mathscr{K}, m)$ (see App.D.7) this implies that $< Sf, f' >= 0$ for all $S \in \mathscr{K}$. In particular, taking $S = Q$ we conclude that f' vanishes on QE.

3. Finally, we show that $H \subseteq QE$. This inclusion is proved if Q, the unit of $\mathscr{K}$ is the identity operator on H. Every eigenvector of T is also an eigenvector of T^n and hence an eigenvector of $R \in \mathscr{S}$. Now take $\varepsilon > 0$ and a finite set

$$\mathcal{F} := \{f_1, \dots, f_n\}$$

of normalized eigenvectors of T (and R) with

$$Rf_i = \lambda_i f_i, \quad |\lambda_i| = 1, \; 1 \leqslant i \leqslant n.$$

By the compactness of the torus Γ we find $m \in \mathbb{N}$ such that

$$\begin{aligned} |1 - \lambda_i^m| &\leqslant \varepsilon \quad \text{and consequently} \\ \|R^m f_i - f_i\| &\leqslant \varepsilon \quad \text{simultaneously for } i = 1, \dots, n. \end{aligned}$$

This proves that the set

$$A_{\mathcal{F},\varepsilon} := \big\{R \in \mathscr{K} : \|Rf - f\| \leqslant \varepsilon \text{ for } f \in \mathcal{F}\big\}$$

is non-empty and closed. By the compactness of $\mathscr{K}$ we conclude that $\bigcap\limits_{\mathcal{F},\varepsilon} A_{\mathcal{F},\varepsilon} \neq \emptyset$, i.e. $\mathscr{K}$ contains an element which is the identity operator on H. Since Q is the unit of $\mathscr{K}$ it must be the identity on H.

■

The minimal ideal $\mathscr{K}$ of $\mathscr{S}$ in the above theorem may be identified with a group of operators on $H = \overline{\text{lin}}\{f \in E : Tf = \lambda f \text{ for some } |\lambda| = 1\}$ which is compact in the weak operator topology and has unit $Q = \text{id}_H$. Moreover, the weak and strong topologies coincide on the one-dimensional orbits $\mathscr{S} f$ for every eigenvector f. Therefore the group $\mathscr{K}$ is even compact for the strong operator topology. Operators for which $H = E$ (and therefore $Q = \text{id}_E$ and $\mathscr{S} = \mathscr{K}$) are of particular importance and will be called "operators with discrete spectrum". The following is an easy consequence of these considerations.

VII.6 Corollary:
For an FDS $(E; T)$ with $\|T^n\| \leqslant c$ the following properties are equivalent:

(a) T has discrete spectrum, i.e. the eigenvectors corresponding to the unimodular eigenvalues of T are total in E.
(b) $\mathscr{S} = \overline{\{T^n : n \in \mathbb{N}_0\}} \subseteq \mathscr{L}_w(E)$ is a compact group with unit id_E.
(c) $\mathscr{S} = \overline{\{T^n : n \in \mathbb{N}_0\}} \subseteq \mathscr{L}_s(E)$ is a compact group with unit id_E.

The following example is simple, but very instructive and should help to avoid pitfalls.

VII.7 Example:
Take the Hilbert $\ell^2(\mathbb{Z})$ and the shift

$$T : (x_z) \to (x_{z+1}).$$

Then $\{T^n : n \in \mathbb{Z}\}$ is a group, its closure in $\mathscr{L}_w(\ell^2(\mathbb{Z}))$ is a compact semigroup with minimal ideal $\mathscr{K} = \{0\}$.

VII.8. Programmatic remark:
The semigroups in

$$\mathscr{S} := \overline{\{T_\varphi^n : n \in \mathbb{N}_0\}}$$

in $\mathscr{L}_w(L^p(X, \Sigma, \mu))$, $1 \leqslant p < \infty$, appearing in (measure-theoretical) ergodic theory are compact and therefore yield projections P (as in VII.4) and Q (as in VII.5) such that

$$\text{id} \geqslant Q \geqslant P \geqslant \mathbf{1} \otimes \mathbf{1},$$

where the order relation for projections is defined by the inclusion of the range spaces. While we have seen in (IV.7) that "ergodicity" is characterized by $P = \mathbf{1} \otimes \mathbf{1}$ we will study in the subsequent lectures the following "extreme" cases:

$$\begin{array}{ll} \text{Lecture VIII:} & \text{id} = Q > P = \mathbf{1} \otimes \mathbf{1}, \\ \text{Lecture IX:} & \text{id} > Q = P = \mathbf{1} \otimes \mathbf{1}. \end{array}$$

VII.D Discussion

VII.D.1. Semitopological semigroups:
One might expect that semigroups S – if topologized – should have jointly continuous multiplication, i.e.,

$$(t,s) \mapsto t \cdot s$$

should be continuous from $S \times S$ into S. In fact, there exists a rich theory for such objects (see Hofmann-Mostert [1966]), but the weaker requirement of separately continuous multiplication still yields interesting results as (VII.2) (see Berglund-Hofmann [1967]) and occurs in non-trivial examples:

The one point compactification $S = \mathbb{Z} \cup \{\infty\}$ of $(\mathbb{Z}, +)$ is a semitopological semigroup if $a + \infty = \infty + a = \infty$ for every $a \in S$. But the addition is not jointly continuous since

$$0 = \lim_{n\to\infty}(n + (-n)) \neq \lim_{n\to\infty} n + \lim_{n\to\infty}(-n) = \infty.$$

Obviously, the minimal ideal is $K = \{\infty\}$.

VII.D.2. Weak vs. strong operator topology on $\mathscr{L}(E)$:
In ergodic theory it is the semigroup $\{T^n : n \in \mathbb{N}_0\}$ – $T \in \mathscr{L}(E)$ and E a Banach space – which is of interest. In most cases this semigroup is algebraically isomorphic to the semigroup $\mathbb{N}_0$. But since our interest is in the asymptotic behavior of the powers T^n, we need some topology on $\mathscr{L}(E)$. If we choose the norm topology or the strong operator topology, and if $\|T^n\| \leqslant c$, then $\{T^n : n \in \mathbb{N}_0\}$ and $\overline{\{T^n : n \in \mathbb{N}_0\}}$ become topological semigroups with jointly continuous multiplication. Unfortunately, these topologies are too fine to yield convergence in many cases. In contrast, if we take the weak operator topology, then $\overline{\{T^n : n \in \mathbb{N}_0\}}$ has only separately continuous multiplication, but in many cases (see IV.5, IV.6 and VII.3) it is compact, and convergence of T^n or some subsequence will be obtained. The following example illustrates these remarks:

Take $E = \ell^2(\mathbb{Z})$ and T the shift as in (VII.7). Then T^n does not converge with respect to the strong operator topology (Proof: If $T^n f$ converges, its limit must be a T-fixed vector, hence equal to 0, but $\|f\| = \|T^n f\|$.), but for the weak operator topology we have $\lim_{n\to\infty} T^n = 0$. The fact that the multiplication is not jointly continuous for the weak operator topology may be seen from

$$0 = \lim_{n\to\infty} T^n \cdot \lim_{n\to\infty} T^{-n} \neq \lim_{n\to\infty}(T^n \cdot T^{-n}) = \mathrm{id}.$$

VII.D.3. Monothetic semigroups:
The semitopological semigroup

$$\mathscr{S} = \overline{\{T^n : n \in \mathbb{N}_0\}} \subseteq \mathscr{L}_w(E)$$

generated by some FDS $(E;T)$ contains an element whose powers are dense in $\mathscr{S}$. Such an element is called *generating*, and the semigroup is called *monothetic*. We mention the following examples of monothetic semigroups:

(i) The set $S := \{2^{-n} : n \in \mathbb{N}\}$ and its closure $\overline{S} = \{2^{-n} : n \in \mathbb{N}\} \cup \{0\}$, endowed with topology and multiplication induced by $\mathbb{R}$, are the simplest monothetic semigroups.

(ii) The unit circle Γ is a (compact) monothetic group, and every $a \in \Gamma$ which is not a root of unity is generating (see III.8.iii).

(iii) The n-torus Γ^n, $n \in \mathbb{N}$ is a (compact) monothetic group, and $a = (a_1, \dots, a_n) \in \Gamma^n$ is generating iff $\{a_1, \dots, a_n\}$ is linearly independent in the $\mathbb{Z}$-module (see App.D.8).

(iv) $S := \Gamma \cup \{\frac{n+1}{n} e^{ni} : n \in \mathbb{N}\}$, $i^2 = -1$, is compact monothetic semigroup for the topology induced by $\mathbb{C}$, the canonical multiplication on Γ,

$$\frac{n+1}{n} e^{in} \cdot \frac{m+1}{m} e^{mi} := \frac{n+m+1}{n+m} e^{(n+m)i} \quad \text{for } n, m \in \mathbb{N}$$

and
$$\frac{n+1}{n} e^{in} \cdot \gamma = \gamma \cdot \frac{n+1}{n} e^{in} := \gamma \cdot e^{in} \quad n \in \mathbb{N},\ \gamma \in \Gamma.$$

The element $2e^i$ is generating (compare Hofmann-Mostert [1966], p. 72).

VII.D.4. Compact semigroups generated by operators on $L^p(X, \Sigma, \mu)$:
The operators $T_\varphi : L^p X, \Sigma, \mu) \to L^p(X, \Sigma, \mu)$ appearing in the ergodic theory of MDS's $(X, \Sigma, \mu; \varphi)$ generate compact semigroups which will be discussed now in more generality. To that purpose, consider a probability space (X, Σ, μ) and a positive operator

$$T : L^1(X, \Sigma, \mu) \to L^1(X, \Sigma, \mu)$$

satisfying $T\mathbf{1} \leqslant \mathbf{1}$ and $T'\mathbf{1} \leqslant \mathbf{1}$. By the Riesz convexity theorem (see Schaefer [1974], V.8.2) T leaves invariant every $L^p(\mu)$, $1 \leqslant p \leqslant \infty$, and the restrictions

$$T_p : L^p(X, \Sigma, \mu) \to L^p(X, \Sigma, \mu)$$

are contractive for $1 \leqslant p \leqslant \infty$. The semigroups

$$\mathscr{S}_p := \overline{\{T_p^n : n \in \mathbb{N}_0\}}$$

in $\mathscr{L}_w(E)$ are compact for $1 \leqslant p < \infty$: if $1 < p < \infty$, argue as in (IV.5); if $p = 1$, as in (IV.6). Moreover, it follows from the denseness of $L^\infty(\mu)$ in $L^p(\mu)$ that all these semigroups are algebraically isomorphic, and that all these weak operator topologies coincide (use App.A.2). Therefore the compact semigroups generated by T in $L^p(\mu)$ for $1 \leqslant p < \infty$ will be denoted by $\mathscr{S}$.
If $L^1(\mu)$ is separable we can find a sequence $\{\chi_n : n \in \mathbb{N}\}$ of characteristic functions which is total in $L^1(\mu)$. The seminorms

$$p_{n,m} := |\langle R\chi_n, \chi_m \rangle|, \quad R \in \mathscr{L}(L^1(\mu)),$$

induce a Hausdorff topology on $\mathscr{S}$ weaker than the weak operator topology. Since $\mathscr{S}$ is compact, both topologies coincide, and therefore $\mathscr{S}$ is a compact metrizable semigroup.

VII.D.5. Operators with discrete spectrum:
Clearly, the identity on any Banach space has discrete spectrum. More interesting examples follow:

(i) Consider $E = C(\Gamma)$ and $T := T_{\varphi_a}$ for some rotation

$$\varphi_a : z \mapsto a \cdot z.$$

The functions $f_n : z \mapsto z^n$ are eigenfunctions of T for every $n \in \mathbb{Z}$ and are total in $C(\Gamma)$ by the Stone-Weierstrass theorem. Therefore, T has discrete spectrum in $C(\Gamma)$.

(ii) The operator T_{φ_a} induced on $L^p(\Gamma, \mathcal{B}, m)$, $1 \leqslant p < \infty$, has discrete spectrum since it has the same eigenfunctions as the operator in (i) and since $C(\Gamma)$ is dense in $L^p(\mu)$ for $1 \leqslant p < \infty$.

(iii) Analogous assertions are valid for all operators induced by any rotation on a compact Abelian group (choose the characters as eigenfunctions), and we will see in Lecture VIII in which sense this situation is typical for ergodic theory.

(iv) There exist operators having discrete spectrum but unbounded powers: For $n > 2$ endow $E_n := \mathbb{C}^n$ with the norm

$$\|(x_1, \ldots, x_n)\| := \max\{(n+1-i)^{-1}|x_i| : 1 \leqslant i \leqslant n\}$$

and consider the rotation operators

$$S_{(n)} : E_n \to E_n : \ (x_1, \ldots, x_n) \mapsto (x_n, x_1, \ldots, x_{n-1}).$$

Every $S_{(n)}$, $n \geqslant 2$, has discrete spectrum in E_n. An easy calculation shows that $\||S_{(n)}\| \leqslant 2$ and $\sup\{\|S_{(i)}^{n-1}\| : i \geqslant 2\} \leqslant \|S_{(n)}^{n-1}\| = n$ for all $n \geqslant 2$. Now, take the ℓ^1-direct sum $E := \bigoplus_{n \geqslant 2} E_n$ and $T := \bigoplus_{n \geqslant 2} S_{(n)}$. Clearly $\|T^i\| = i+1$ for every $i \in \mathbb{N}$, but T has discrete spectrum in E.

VII.D.6. Semitopological vs. topological groups (the Ellis Theorem): In the remark following Definition (VII.1) we stated that a semitopological group which is compact is a topological group. Usually this fact is derived from a deep theorem of Ellis [1957] but the proof of the property we needed in Lecture VII is actually quite easy – at least for metrizable groups.

Proposition: Let G be a group, $\mathscr{O}$ a metrizable, compact Hausdorff topology on G such that the mapping

$$(g, h) \mapsto gh : \ G \times G \to G$$

is separately continuous. Then $(G, \mathscr{O})$ is a topological group.

Proof. Suppose that the multiplication is not continuous at $(s, t) \in G \times G$. Then there exists $\varepsilon > 0$ such that for every neighbourhood U of s and V of t

$$\varepsilon \leqslant d(st, s_U t_V)$$

for some suitable $(s_U, t_V) \in U \times V$, and $d(\cdot, \cdot)$ a metric on G generating $\mathscr{O}$. Since multiplication is separately continuous there exists a neighbourhood U_0 of s and V_0 of t, such that

$$d(st, s't) \leqslant \frac{\varepsilon}{4} \quad \text{for every } s' \in U_0,$$

and

$$d(s_{U_0} t, s_{U_0} t') \leqslant \frac{\varepsilon}{4} \quad \text{for every } t' \in V_0.$$

From this we obtain the contradiction

$$\varepsilon \leqslant d(st, s_{U_0} t_{V_0} \leqslant d(st, s_{U_0} t) + d(s_{U_0} t, s_{U_0} t_{V_0}) \leqslant \frac{\varepsilon}{2}.$$

Therefore the multiplication is jointly continuous on G.

It remains to prove that the mapping $g \mapsto g^{-1}$ is continuous on G. Take, $g \in G$ and choose a sequence $(g_n)_{n \in \mathbb{N}}$ contained in G such that $\lim_{n \to \infty} g_n = g$. Since $(G, \mathscr{O})$ is compact and metrizable, the sequence (g_n^{-1}) has a convergent subsequence in G. Thus we may assume that $\lim_{n \to \infty} g_n^{-1} = h$ for some $h \in G$. From the joint continuity of the multiplication we obtain $1 = gh = hg$, thus $h = g^{-1}$, which proves the assertion. ∎

VIII. Dynamical Systems with Discrete Spectrum

As announced in (VI.7), in this lecture we tackle and solve the isomorphism problem at least for a subclass of MDSs: If $(X, \Sigma, \mu; \varphi)$ is ergodic and has "discrete spectrum", then the eigenvalues of T_φ are a complete system of invariants.

Before proving this statement let us say a few words about the hypothesis we are going to make throughout this lecture. In particular, we have to prepare ourselves to apply the results on semigroups of Lecture VII to the present ergodic-theoretical situation.

Let $(X, \Sigma, \mu; \varphi)$ be an ergodic MDS. As usual we consider the induced operator $T := T_\varphi \in \mathscr{L}(L^p(\mu))$, $1 \leqslant p < \infty$, and also the compact abelian semigroup

$$\mathscr{S} := \overline{\{T^n : n \in \mathbb{N}_0\}} \subseteq \mathscr{L}_w(L^p(\mu)) \quad \text{see VII.D.4}$$

Since φ is ergodic, the corresponding mean ergodic projection P is of the form

$$P = \mathbf{1} \times \mathbf{1} \in \overline{\text{co}}\mathscr{S} \quad \text{(see IV.7).}$$

Since $\mathscr{S}$ is compact, there exists another projection

$$Q \in \mathscr{S}$$

such that $Q\mathscr{S}$ is a compact group (see VII.5). In contrast to Lecture IX we require here that Q is much "larger" than P or more precisely

$$Q = \text{id}$$

, i.e. $\mathscr{S}$ is a compact group in $\mathscr{L}_w(L^p(\mu))$ – or $\mathscr{L}_s(L^p(\mu))$, see VII.6 having the operator id as unit. In other words, we assume that $(X, \Sigma, \mu; \varphi)$ is ergodic and has discrete spectrum, i.e. T_φ has discrete spectrum in $L^p(X, \Sigma, \mu)$, $1 \leqslant p < \infty$. Under these assumptions we seek a complete system of isomorphism invariants.

It is helpful to start with the analogous problem for TDSs. We therefore assume that $(X; \varphi)$ is a minimal TDS, and that T_φ has discrete spectrum in $C_{\mathbb{C}}(X)$. The following example shows that such systems appear quite frequently and are of some importance.

VIII.1 Example:
Let G be a compact group. If G is monothetic with generating element $g \in G$ (i.e. $\{g^n : n \in \mathbb{N}_0\}$ is dense in G, see VII.D.3) then the rotation Rot$g := \varphi_g$ is minimal.

Moreover, every character $\gamma \in \widehat{G}$ is an eigenfunction of T_{φ_g} because

$$T_{\varphi_g}\gamma(h) = \gamma(gh) = \gamma(g) \cdot \gamma(h)$$

for every $h \in G$. Since the product of two characters is still a character and since the characters separate points of G (see App.D.7) it follows from the Stone-Weierstrass theorem that T_{φ_g} has discrete spectrum in $C(G)$.

Conversely, the following theorem shows that the example above is typical.

VIII.2 Theorem:
Let $(X; \varphi)$ be a minimal TDS such that T_φ has discrete spectrum in $C(X)$. Then it is isomorphic to a rotation on a compact monothetic group.

Proof. From (VII.6) it follows that the induced operator $T := T_\varphi$ in $C(X)$ generates a compact group

$$\mathscr{G} := \overline{\{T^n : n \in \mathbb{N}_0\}} \subseteq \mathscr{L}_s(C(X)).$$

We shall show that $(X;\varphi)$ is isomorphic to $(\mathscr{G};\mathrm{Rot}T)$. The operator T is a Banach algebra isomorphism of $C(X)$. Since $\mathscr{G}$ is a group, the same is true for every $S \in \mathscr{G}$. Therefore there exist homeomorphisms

$$\varphi_S : X \to X$$

such that $$Sf = f \circ \varphi_S \text{ for every } S \in \mathscr{G},\ f \in C(X)$$

and $$\varphi_{S_1 S_2} = \varphi_{S_1} \circ \varphi_{S_2} \text{ for } S_1, S_2 \in \mathscr{G} \text{ (see II.D.5).}$$

Choose $x_0 \in X$ and define

$$\Theta : \mathscr{G} \to X \text{ by } \Theta(S) := \varphi_S(x_0) \text{ for } S \in \mathscr{G}.$$

This map yields the isomorphism between $(\mathscr{G};\mathrm{Rot}T)$ and $(X;\varphi)$:

1. Θ is continuous: If the net $(S_\alpha)_{\alpha\in A}$ converges to S in the strong operator topology, then
$$f(\Theta(S_\alpha)) = S_\alpha f(x_0) \text{ converges to } Sf(x_0) = f(\Theta(S))$$
for every $f \in C(X)$. But this implies that $(\Theta(S_\alpha))_{\alpha\in A}$ converges to $\Theta(S)$ in X.
2. Θ is surjective: $\Theta(\mathscr{S})$ is a closed subset of X which is φ-invariant. From the minimality it follows that $\Theta(\mathscr{S}) = X$.
3. Θ is injective: If $\Theta(S_1) = \Theta(S_2)$, for $S_1, S_2 \in \mathscr{G}$, we conclude that $\varphi_{S_1}(x_0) = \varphi_{S_2}(x_0)$ or $\varphi_{S_2^{-1}S_1}(x_0) = x_0$ and $\varphi_{S_2^{-1}S_1}(\varphi^n(x_0)) = \varphi^n(\varphi_{S_2^{-1}S_1}(x_0)) = \varphi^n(x_0)$ for all $n \in \mathbb{N}$. Again from minimality of φ it follows that $\{\varphi^n(x_0) : n \in \mathbb{N}\}$ is dense in X, an therefore that $\varphi_{S_2^{-1}S_1} = \mathrm{id}_X$ or $S_2 = S_1$
4. The diagram
$$\begin{tikzcd} X \arrow[r, "\varphi"] \arrow[d, "\Theta"'] & X \arrow[d, "\Theta"] \\ \mathscr{G} \arrow[r, "\mathrm{Rot}T"'] & \mathscr{G} \end{tikzcd}$$
commutes:
For $S \in \mathscr{G}$ we obtain $\varphi(\Theta(S)) = \varphi(\varphi_S(x_0)) = \Theta(TS)$.

■

As an application of this representation theorem we can solve the isomorphism problem for minimal TDSs with discrete spectrum.

VIII.3 Corollary:

(i) For minimal TDSs with discrete spectrum the point spectrum of the induced operator is a subgroup of the unit circle Γ, and as such a complete isomorphism invariant.

(ii) Let Γ_0 be an arbitrary subgroup of Γ and endow Γ_0 with the discrete topology. The rotation on the compact group

$$G := \widehat{\Gamma_0}$$

by the character $\mathrm{id} : \lambda \mapsto \lambda$ on Γ_0 is (up to isomorphism) the unique minimal TDS with discrete spectrum having Γ_0 as point spectrum.

Proof. (i) In (III.9) we proved that for a minimal TDS $(X;\varphi)$ the point spectrum $P\sigma(T_\varphi)$ of the induced operator T_φ is a subgroup of Γ. Now consider two minimal TDSs $(X_1;\varphi_1)$ and $(X_2;\varphi_2)$ having discrete spectrum such that $P\sigma(T_{\varphi_1}) = P\sigma(T_{\varphi_2})$. By (VIII.2) $(X_1;\varphi_1)$ is isomorphic to a rotation by a generating element on a compact group $(G_1;\varphi_a)$, and analogously $X_2;\varphi_2) \simeq (G_2;\varphi_b)$. The next step is to show that the character group $\widehat{G_1}$ is isomorphic to $P\sigma(T_{\varphi_a})$:

Every $\gamma \in \widehat{G_1}$ is a continuous eigenfunction of T_{φ_a} with corresponding eigenvalue $\gamma(a)$. It is easy to see that

$$\Theta : \gamma \mapsto \gamma(a)$$

defines a group homomorphism from $\widehat{G_1}$ into $P\sigma(T_{\varphi_a})$. Furthermore, Θ is injective since $\gamma_1(a) = \gamma_2(a)$ implies that $\gamma_1(a^n) = \gamma_2(a^n)$ for every $n \in \mathbb{Z}$, hence $\gamma_1 = \gamma_2$ for (continuous) characters γ_1, γ_2. The map Θ is surjective since to every eigenvalue $\lambda \in P\sigma(T_{\varphi_a})$ there corresponds a unique eigenfunction $f \in C(G_1)$ normalized by $f(a) = \lambda$ (see III.9). By induction we obtain

$$f(a^n) = T_{\varphi_a} f(a^n) = \lambda f(a^n) = \lambda^{n+1}$$

for all $n \in \mathbb{N}$, and by continuity we conclude that f is a character on G_1 with $\Theta(f) = \lambda$.

Therefore, $\widehat{G_1}$ is isomorphic to $P\sigma(T_{\varphi_a}) = P\sigma(T_{\varphi_1})$, and analogously $\widehat{G_2} \simeq P\sigma(T_{\varphi_b}) = P\sigma(T_{\varphi_1})$. From $P\sigma(T_{\varphi_1}) = P\sigma(T_{\varphi_2})$ and Pontrjagin's duality theorem (App.D.6) we conclude $G_1 \simeq G_2$.

Finally, identifying G_1 and G_2 we have to prove that $(G_1,\varphi_a) = (G_1;\varphi_b)$ where a and b are two generating elements in G_1 such that $P\sigma(T_{\varphi_a}) = P\sigma(T_{\varphi_b})$.

For $\lambda \in P\sigma(T_{\varphi_a})$ there exist unique eigenfunction f_λ for T_{φ_a}, resp. g_λ for T_{φ_b}, normalized by $f_\lambda(a) = \lambda$, resp. $g_\lambda(a) = \lambda$. The mapping $f_\lambda \to g_\lambda$, $\lambda \in P\sigma(T_{\varphi_a})$ has a unique extension to a Banach algebra isomorphism V on $C(G_1)$. Clearly $V \circ T_{\varphi_a} = T_{\varphi_b} \circ V$, and therefore $(G_1;\varphi_a) \simeq (G_2;\varphi_b)$ by (IV.3).

(ii) By (i) it remains to show that $P\sigma(T_{\varphi_{\mathrm{id}}}) = \Gamma_0$. But this follows from (App.D.6):

$$P\sigma(T_{\varphi_{\mathrm{id}}}) \simeq \hat{G} = \hat{\hat{\Gamma}}_0 = \Gamma_0.$$

∎

We have seen that the classification of minimal TDSs with discrete spectrum reduces to the classification of compact monothetic groups. The tori Γ^n, $n \in \mathbb{N}$, yield the standard examples (see **??**). In the second part of this lecture we return to measure-theoretical ergodic theory, and we can use (VIII.2) in order to obtain a solution of the analogous problem for MDSs.

VIII.4 Theorem Halmos-von Neumann, 1942:
Let $(X,\Sigma,\mu;\varphi)$ be an ergodic MDS such that T_φ has discrete spectrum in $L^p(X,\Sigma,\mu)$, $1 \leqslant p < \infty$. Then it is isomorphic to a rotation on a compact monothetic group endowed with the normalized Haar measure.

Proof. If $f \in L^p(\mu)$ is an eigenfunction of $T := T\varphi$, for an eigenvalue λ, $|\lambda| = 1$, we conclude that

$$T|f| = |Tf| = |\lambda||f| = |f| = c \cdot \mathbf{1}$$

since φ is ergodic (see III.4). Therefore, the linear span of

$$\{f \in L^p(\mu) : Tf = \lambda f \text{ for some } |\lambda| = 1\}$$

is a conjugation-invariant subalgebra of $L^\infty(\mu)$, and its closure in $L^\infty(\mu)$ denoted by $\mathscr{A}$, is a commutative C^*-algebra with unit. By the Gelfand-Neumark theorem (App.C.9) there exists an isomorphism

$$j : \mathscr{A} \to C(Y)$$

for some compact space Y.
The restriction of T_φ to $\mathscr{A}$ is an algebra isomorphism on $\mathscr{A}$. Therefore, its isomorphic image $j \circ T_\varphi \circ j$ inf on $C(Y)$ is induced by some homeomorphism $\psi : Y \to Y$. Next we show that $(Y; \psi)$ is a minimal TDS with discrete spectrum: T_ψ has discrete spectrum in $C(Y)$ as T_φ has in $\mathscr{A}$. Therefore, T_ψ is mean ergodic by (VII.6) and (IV.4.c). Thus the fixed space of T_φ in $\mathscr{A}$, and therefore of T_ψ in $C(Y)$ is one-dimensional. Since (the restriction of) μ is a strictly positive, T_φ invariant linear form on $\mathscr{A}$, we obtain a strictly positive, ψ-invariant probability measure $\tilde{\mu}$ on Y. Hence the minimality of $(Y; \psi)$ follows from (IV.4.e) and (IV.8).
Now we can apply Theorem (VIII.2) to the TDS (Y, ψ) and obtain a homeomorphism

$$\Theta : G \to Y,$$

where G is a compact monothetic group with generating element a, making commutative the following diagram:

$$\begin{tikzcd}
\mathscr{A} \arrow[r, "T_\varphi"] \arrow[d, "j"'] & \mathscr{A} \arrow[d, "j"'] \\
C(Y) \arrow[r, "T_\psi"] \arrow[d, "T_\Theta"'] & C(Y) \arrow[d, "T_\Theta"] \\
C(G) \arrow[r, "\mathrm{Rot}a"'] & C(G)
\end{tikzcd}$$

where $(\mathrm{Rot}a)f(g) := f(ag)$ for $f \in C(G)$. But $\mathscr{A}$, $C(Y)$ and $C(G)$ are dense subspaces in $L^p(X, \Sigma, \mu)$, $L^p(Y, \tilde{\mu})$ and $L^p(G, m)$ respectively, where m is the Haar measure on G. From the construction above it follows that $j'\tilde{\mu} = \mu$. Since m is the unique probability measure invariant under Rot$a$, we also conclude $T'_\Theta m \tilde{\mu}$. Therefore we can extend j and T_Θ continuously to positive isometries (hence lattice isomorphisms, see App.C.4) on the corresponding L^p-spaces. Obviously, the same can be done for T_φ, T_ψ and Rot$a$. Finally, we obtain an analogous diagram for the L^p-spaces, which proves the isomorphism of $(X, \Sigma, \mu; \varphi)$ and $G, \mathcal{B}, m; \mathrm{Rot}a)$ by (VI.2). ∎

As in the topological case we deduce from the above theorem that ergodic MDSs with discrete spectrum are completely determined by their point spectrum.

VIII.5 Corollary:

(i) For ergodic MDSs with discrete spectrum the point spectrum of the induced operator is a subgroup of Γ and as such a complete isomorphism invariant.

(ii) Let Γ_0 be an arbitrary subgroup of Γ and endow Γ_0 with the discrete topology. The rotation on the compact group

$$G := \widehat{\Gamma_0}$$

with normalized Haar measure m by the character $\mathrm{id} : \lambda \mapsto \lambda$ on Γ_0 is (up to isomorphism) the unique ergodic MDS with discrete spectrum having point spectrum Γ_0.

VIII.D Discussion

IX. Mixing

Now we return to the investigation of "mixing properties" of dynamical systems, and the following experiment might serve as an introduction to the subsequent problems and results: two glasses are taken, one filled with red wine, the other with water, and one of the following procedures is performed once a minute.

A. The glasses are interchanged.
B. Nothing is done.
C. Simultaneously, a spoonful of the liquid in the right glass is added to the left glass and vice versa.

Intuitively, the process A is not really mixing because it does not approach any invariant "state", B is not mixing either because it stays in an invariant "state" which is not the equidistribution of water and wine, while C is indeed mixing. However, if in A the glasses are changed very rapidly it will appear to us, as if A were mixing, too.

It is our task to find correct mathematical models of the mixing procedures described above, i.e. we are looking for dynamical systems which are *converging* (in some sense) toward an *"equidistribution"*. The adequate framework will be that of MDSs (compare IV.8 and the remark proceeding it). More precisely, we take an MDS $(X, \Sigma, \mu; \varphi)$. The operator $T := T_\varphi$ induced on $L^p(X, \Sigma, \mu)$, $1 \leqslant p < \infty$, generates a compact semigroup

$$\mathscr{S} := \overline{\{T^n : n \in \mathbb{N}_0\}}$$

in $\mathscr{L}(L^p(\mu))$ for the weak operator topology. Moreover, if we assume $L^p(\mu)$ to be separable, this semigroup is metrizable (see VII.D.4).

The above experiments lead to the following mathematical questions:

convergence: under which conditions and in which sense do the powers T^n converge as $n \to \infty$?

If convergence of T^n holds in any reasonable topology then $P := \lim_{n\to\infty} T^n$ is projection onto the T-fixed space in $L^p(\mu)$. Therefore, the second property describing "mixing" may be expressed as follows.

equidistribution: under which conditions does the T-fixed space contain only the constant functions ?

One answer to these questions – in analogy to the case of the fast version of A – has already been given in Lecture IV, but will be repeated here.

IX.1 Theorem:
An MDS $(X, \Sigma, \mu; \varphi)$ is ergodic if and only if one of the following equivalent properties is satisfied:

(a) $T_n \to \mathbf{1} \otimes \mathbf{1}$ in the weak operator topology.

(b) $\langle T_n f, g\rangle \to \left(\int f \,\mathrm{d}\mu\right)\left(\int g \,\mathrm{d}\mu\right)$ for all $f, g \in L^\infty(X, \Sigma, \mu)$.

(c) $\displaystyle\frac{1}{n}\sum_{i=0}^{n-1} \mu(\varphi^{-i}A \cap B) \to \mu(A)\cdot\mu(B)$ for all $A, B \in \Sigma$.

(d) 1 is simple eigenvalue of T.

Proof. See (III.4) and (IV.7) including the remark. ∎

The really mixing case C is described by the (weak operator) convergence of the powers of T toward the projection $\mathbf{1}\otimes\mathbf{1}$. In analogy to the theorem above we obtain the following result.

IX.2 Theorem:
For an MDS $(X, \Sigma, \mu; \varphi)$ the following are equivalent.

(a) $T^n \to \mathbf{1}\otimes\mathbf{1}$ in the weak operator topology.
(b) $\langle T^n f, g\rangle \to \big(\int f \,\mathrm{d}\mu\big)\big(\int g \,\mathrm{d}\mu\big)$ for all $f, g \in L^\infty(X, \Sigma, \mu)$.
(c) $\mu(\varphi^{-n} A \cap B) \to \mu(A)\cdot\mu(B)$ for all $A, B \in \Sigma$

IX.3 Definition:
An MDS $(X, \Sigma, \mu; \varphi)$, resp. the transformation φ, satisfying one of the equivalent properties of (IX.2) is called *strongly mixing*.

Even if this concept perfectly describes the mixing-procedure C which seems to be the only one of some practical interest, we shall introduce one more concept:

Comparing the equivalences of (IX.1) and (IX.2) one observes that there is lacking a (simple) spectral characterization of strongly mixing. Obviously, the existence of an eigenvalue $\lambda \neq 1$, $|\lambda| = 1$, of T excludes the convergence of the powers T^n. Therefore, we may take this non-existence of non-trivial eigenvalues as the defining property of another type of mixing which possibly might coincide with strong mixing.

IX.4 Definition:
An MDS $(X, \Sigma, \mu; \varphi)$, resp. the transformation φ, is called *weakly mixing* if 1 is a simple and the unique eigenvalue of T in $L^p(X, \Sigma, \mu)$.

The results of Lecture VII applied to the compact semigroup

$$\mathscr{S} := \overline{\{T^n : n \in \mathbb{N}\}}^{\sigma}$$

will clarify the structural significance of this definition:
Let P be the projection corresponding to the mean ergodic operator T, i.e. $\{P\}$ is the minimal ideal of $\overline{\mathrm{co}}\mathscr{S}$, and denote by $Q \in \mathscr{S}$ the projection generating the minimal ideal

$$\mathscr{K} = Q\mathscr{S}$$

of $\mathscr{S}$. The fact that 1 is a simple eigenvalue of T corresponds to the fact that $P = \mathbf{1}\otimes\mathbf{1}$, see (IV.7), hence

$$\mathbf{1}\otimes\mathbf{1} \in \overline{\mathrm{co}}\mathscr{S}.$$

In (VII.5) we proved that Q is a projection onto the subspace spanned by all unimodular eigenvectors, hence

$$QE = PE = \langle \mathbf{1}\rangle.$$

From $Q \in \mathscr{S}$ it follows as in (IV.7) that

$$Q = P = \mathbf{1}\otimes\mathbf{1},$$

or equivalently

$$\{\mathbf{1}\otimes\mathbf{1}\} = \mathscr{K}$$

is the minimal ideal in $\mathscr{S}$. Briefly, weakly mixing systems are those for which the mean ergodic projection is already contained in $\mathscr{S}$ and is of the form $\mathbf{1} \otimes \mathbf{1}$. The following theorem shows in which way weak mixing lies between ergodicity (IX.1) and strong mixing (IX.2).

IX.5 Theorem:
Let $(X, \Sigma, \mu; \varphi)$ be an MDS. If $E := L^p(X, \Sigma, \mu)$, $1 \leqslant p < \infty$ is separable, the following assertions are equivalent:

(a) $T^{n_i} \to \mathbf{1} \otimes \mathbf{1}$ for the weak operator topology and for some subsequence $\{n_i\} \subseteq \mathbb{N}$.

(a′) $T^{n_i} \to \mathbf{1} \otimes \mathbf{1}$ for the weak operator topology and for some subsequence $\{n_i\} \subseteq \mathbb{N}$ having density 1.

(a″) $\frac{1}{n} \sum\limits_{i=0}^{n-1} |\langle T^i f, g\rangle - \langle f, \mathbf{1}\rangle \cdot \langle \mathbf{1}, g\rangle| \to 0$ for all $f \in E$, $g \in E'$.

(b) $\langle T^{n_i} f, g\rangle \to \left(\int f \,\mathrm{d}\mu\right) \cdot \left(\int g \,\mathrm{d}\mu\right)$ for all $f, g \in L^\infty(X, \Sigma, \mu)$ and for some subsequence $\{n_i\} \subseteq \mathbb{N}$.

(c) $\mu(\varphi^{-n_i} A \cap B) \to \mu(A) \cdot \mu(B)$ for all $A, B \in \Sigma$ and for some subsequence $\{n_i\} \subseteq \mathbb{N}$.

(d) φ is weakly mixing.

(e) $\varphi \otimes \varphi$ is ergodic.

(f) $\varphi \otimes \varphi$ is weakly mixing.

IX.6 Remarks:

1. A subsequence $\{n_i\} \subseteq \mathbb{N}$ has density 1 if
$$\lim_{k\to\infty} \frac{1}{k}\big|\{n_i\} \cap \{1, 2, \ldots, k\}\big| = 1 \qquad \text{(see App.E.1).}$$
2. The definition $\varphi \otimes \varphi : (x, y) \mapsto (\varphi(x), \varphi(y)$ makes $(X \times X, \Sigma \otimes \Sigma, \mu \otimes \mu; \varphi \otimes \varphi)$ an MDS.
3. (a) and (a′) are formally weaker than (IX.2.a), while (a″) (called "strong Cesàro convergence") is formally stronger than (IX.1.a).
4. "Primed" versions of (b) and (c) analogous to (a) are easily deduced.
5. Further equivalences are easily obtained by taking in (b) the functions f, g only from a subset of $L^\infty(\mu)$ which is total in $L^1(\mu)$, resp. in (c) the sets A, B only from a subalgebra generating Σ.

Proof. The general considerations above imply that (d) is equivalent to $\mathbf{1} \otimes \mathbf{1} \in \mathscr{S} = \overline{\{T^n : n \in \mathbb{N}\}}$. But by (VII.D.4), $\mathscr{S}$ is metrizable for the weak operator topology, hence there even exists a subsequence in $\{T^n : n \in \mathbb{N}\}$ converging to $\mathbf{1} \otimes \mathbf{1}$, which shows the equivalence of (a) and (d).

(a) $\Rightarrow$ (a′): We recall again that $\mathscr{S}$ is a commutative compact semigroup containing $\mathbf{1} \otimes \mathbf{1}$ as a zero, i.e. $R \cdot (\mathbf{1} \otimes \mathbf{1}) = \mathbf{1} \otimes \mathbf{1}$ for all $R \in \mathscr{S}$. Define the operator
$$\tilde{T} : C(\mathscr{S}) \to C(\mathscr{S})$$

induced by the rotation by T on $\mathscr{S}$, i.e.

$$\tilde{T}\tilde{f}(R) = \tilde{f}(TR) \quad \text{for } R \in \mathscr{S},\ \tilde{f} \in C(\mathscr{S}).$$

First, we show that this operator is mean ergodic with projection $\tilde{P}$ defined as

$$\tilde{P}\tilde{f}(R) = \tilde{f}(\mathbf{1} \otimes \mathbf{1}) \quad \text{for } R \in \mathscr{S},\ \tilde{f} \in C(\mathscr{S}).$$

Since multiplication by T is (uniformly) continuous on $\mathscr{S}$, the mapping from $\mathscr{S}$ into $\mathscr{L}(C(\mathscr{S}))$ which associates to every $R \in \mathscr{S}$ its rotation operator $\tilde{R}$ is well defined. Consider a sequence $(S_k)_{k\in\mathbb{N}}$ in $\mathscr{S}$ converging to S. Then $\tilde{S}_k\tilde{f}(R) = \tilde{f}(S_kR)$ converges to $\tilde{f}(SR) = \tilde{S}\tilde{f}(R)$ for all $R \in \mathscr{S}$, $\tilde{f} \in C(\mathscr{S})$. But the pointwise convergence and the boundedness of $\tilde{S}_k\tilde{f}$ imply weak convergence (see App.B.18), hence $\tilde{S}_k \to \tilde{S}$ in $\mathscr{L}_w(C(\mathscr{S}))$, and the mapping $S \mapsto \tilde{S}$ is continuous from $\mathscr{S}$ into $\mathscr{L}_w(C(\mathscr{S}))$. Therefore, from $T^{n_i} \to \mathbf{1} \otimes \mathbf{1}$ we obtain $\tilde{T}^{n_i} \to \widetilde{\mathbf{1} \otimes \mathbf{1}} = \tilde{P} \in \mathscr{L}_w(C(\mathscr{S}))$. Applying (IV.4.d) we conclude that the Cesàro means of $\tilde{T}^n$ converge strongly to $\tilde{P}$. Take now $f \in E$, $g \in E'$ and define a continuous function $\tilde{f} \in C(\mathscr{S})$ by

$$\tilde{f}(R) := |\langle Rf, h\rangle - \langle f, \mathbf{1}\rangle \cdot \langle \mathbf{1}, g\rangle|.$$

Obviously, we have $\tilde{P}\tilde{f}(T) = \tilde{f}(\mathbf{1} \otimes \mathbf{1}) = 0$. Therefore

$$0 = \lim_{n\to\infty} \tilde{T}_n\tilde{f}(T) \lim_{n\to\infty} = \frac{1}{n}\sum_{i=0}^{n-1} \left|\langle T^i f, g\rangle - \langle f, \mathbf{1}\rangle \cdot \langle \mathbf{1}, g\rangle\right|.$$

(a″) ⇒ (a): Since $\mathscr{S}$ is metrizable and compact for the topology induced from $\mathscr{L}_w(E)$, there exist countably many $f_k \in E$, $g_l \in E'$ such that the seminorms

$$p_{k,l}(R) := |\langle Rf_k, g_l\rangle|$$

define the topology on $\mathscr{S}$. By the assumption (a″) and by (App.E.2) for every pair (k, l) we obtain a subsequence

$$\{n_i\}^{k,l} \subseteq \mathbb{N}$$

with density 1, such that

$$\langle T^{n_i} f, g\rangle \to \langle f_k, \mathbf{1}\rangle \cdot \langle \mathbf{1}, g_l\rangle.$$

By (App.E.3) we can find a new subsequence, still having density 1, such that the concergence is valid simultaneously for all f_k and g_l. As usual, we apply (App.B.15) to obtain weak operator convergence.

(a′) ⇒ (a) is clear.

The equivalences (a) ⇔ (b) ⇔ (c) follow if we observe that the topologies we are considering in (b) and (c) are Hausdorff and weaker than the weak operator topology for which $\mathscr{S}$ is compact. Therefore, these topologies coincide on $\mathscr{S}$.

(c) ⇒ (f): Take $A, A', B, B' \in \Sigma$. For a suitable but fixed subsequence $(n_i) \subseteq \mathbb{N}$ $\mu(\varphi^{-n_i}A \cap B)$, resp. $\mu(\varphi^{-n_i}A' \cap B')$ converges to $\mu(A) \cdot \mu(B)$, resp. $\mu(A') \cdot \mu(B')$, as $n_i \to \infty$. This implies that

$$(\mu \otimes \mu)(((\varphi \otimes \varphi)^{-n_i} A \times A') \cap (B \times B')) = \mu(\varphi^{-n_i}A \cap B) \cdot \mu(\varphi^{-n_i}A' \cap B')$$

converges to $\mu(A) \cdot \mu(B) \cdot \mu(A') \cdot \mu(B') = (\mu \otimes \mu)(A \times A') \cdot (\mu \otimes \mu)(B \otimes B')$. Since the same assertion holds for disjoint unions of sets of the form $A \times A'$ we obtain the desired convergence for all sets in a dense subalgebra of $\Sigma \otimes \Sigma$. Using an argument as in the above proof of (a) ⇔ (b) ⇔ (c) we conclude that the MDS

$(X \times X, \Sigma \otimes \Sigma, \mu \otimes \mu; \varphi \otimes \varphi)$ satisfies a convergence property as (c), hence it is weakly mixing.

(f) $\Rightarrow$ (e) is clear.

(e) $\Rightarrow$ (d): Assume that $T_\varphi f = \lambda f$, $|\lambda| = 1$, for $0 \neq f \in L^1(\mu)$. Then we have $T_\varphi \bar{f} = \bar{\lambda}\bar{f}$ and, for the function $f \otimes \bar{f} : (x, y) \mapsto f(x) \cdot \bar{f}(y)$, $(x, y) \in X \times X$, we obtain $T_{\varphi\otimes\varphi}(f \otimes \bar{f}) = \lambda f \otimes \bar{\lambda}\bar{f} = |\lambda|^2(f \otimes \bar{f}) = f \otimes \bar{f}$. But 1 is a simple eigenvalue of $T_{\varphi\otimes\varphi}$ with eigenvector $\mathbf{1}_X \otimes \mathbf{1}_X$. Therefore we conclude $f = c\mathbf{1}_X$ and $\lambda = 1$ i.e. φ is weakly mixing. ∎

IX.7 Example: While it is easy to find MDSs which are ergodic but not weakly mixing (e.g. the rotation φ_a, $a^n \neq 1$ for all $n \in \mathbb{N}$, on the circle Γ has all powers of a as eigenvalues of T_{φ_a}), it remained open for a long time whether weak mixing implies strong mixing. That this is not the case will be shown in the next lecture.

The Bernoulli shift $B(p_0, \ldots, p_{k-1})$is strongly mixing as can be seen in proving (IX.2.c) for the rectangles, analogously to (III.5.ii).

IX.D Discussion

IX.D.1. Mathematical models of mixing procedures:
We consider the apparatus described at the beginning of this lecture. Our mathematical model is based on the assumption that two liquids contained in the same glass will mix rapidly whereas the transfer of liquid from one glass into the other is controlled by the experimenter. This leads to the following model:

Let $(X, \Sigma, \mu; \varphi)$ be a strongly mixing MDS. Take $X' := X \times \{0, 1\}$, Σ' the obvious σ-algebra on X' and μ' defined by $\mu'(A' \times \{1\}) = \mu'(A \times \{0\}) = \frac{1}{2}\mu(A)$ for $A \in \Sigma$. We obtain MDS $(X', \Sigma', \mu'; \varphi')$ by

A. $$\varphi'(x, j) := (\varphi(x), 1 - j)$$

B. $$\varphi'(x, j) := (\varphi(x), j)$$

C. $$\varphi'(x, j) := \begin{cases} (\varphi(x), j) & \text{for } x \in X \backslash S \\ (\varphi(x), 1 - j) & \text{for } x \in S \end{cases}$$

Exercise: Show that C is strongly mixing, B is not ergodic, but the powers of $T_{\varphi'}$ converge, and A is ergodic, but the powers of $T_{\varphi'}$ do not converge.

IX.D.2. Further equivalences to strong mixing:
To (IX.2) we can add the following equivalences:

(e) $(T^n f|f) \to (f|\mathbf{1})^2$ for all $f \in L^\infty(X, \Sigma, \mu)$, where $(\cdot|\cdot)$ denotes the scalar product in $L^2(X, \Sigma, \mu)$.

(f) $\frac{1}{n}\sum\limits_{i=0}^{n-1} T^{k_i} \to \mathbf{1} \times \mathbf{1}$ in the weak operator topology for every subsequence $(k_i) \subseteq \mathbb{N}$.

Proof. (d) $\Rightarrow$ (a): By (App.B.15) it suffices to show that $\langle T^n f, g\rangle$ converge to $\langle f, \mathbf{1}\rangle \cdot \langle \mathbf{1}, g\rangle$ for all g in a total subset of $L^2(\mu)$ and $f \in L^\infty(\mu)$. To that purpose we consider the closed T-invariant subspace

$$E_0 := \overline{\text{lin}}\{\mathbf{1}, f, Tf, T^2f, \dot{\}} \subseteq L^2(\mu).$$

The assertion is trivial for $g \in E_0^\perp$ and follows from the assumption for $g = T^n f$.

(b) $\Leftrightarrow$ (e): It is elementary to see that a sequence of real or complex numbers converges if and only if every subsequence is convergent in the Cesàro sense. ■

Certainly, the equivalence of (b) and (e) remains valid under much more general circumstances. But for operators induced by an MDS the *weak* operator convergence of $\frac{1}{n}\sum_{i=0}^{n-1} T^{k_i}$ as in (e) is equivalent to the *strong* operator convergence of these averages. 'This surprising result will be discussed in (IX.D.5).

IX.D.3. Strong operator convergence of T^n:

IX.D.4. Weak mixing implies "strong ergodicity":

IX.D.5. Weak convergence implies strong convergence of averages:

IX.D.6. Weak mixing in Banach spaces:

IX.D.7. Mixing in C(X):

X. Category Theorems and Concrete Examples

The construction and investigation of concrete dynamical systems with different ergodic-theoretical behaviour is an important and difficult task. In this lecture we will show that there exist weakly mixing MDS's which are not strongly mixing. But, following the historical development, we present an explicit construction of such an example only after having proved its existence by categorical considerations with regard to the set of all bi-measure-preserving transformations.

In the following we always take $(X, \mathcal{B}, m)$ to be the probability space $X = [0, 1]$ with Borel algebra $\mathcal{B}$ and Lebesgue measure m. In order to describe the set of all m-preserving transformations on X we first distinguish some very important classes.

X.1 Definition:
Let $(X, \mathcal{B}, m; \varphi)$ be an MDS.

(i) A point $x \in X$ is called *periodic* (*with period* $n_0 \in \mathbb{N}$) if $\varphi^{n_0} x = x$ and ($\varphi^n(x) \neq x$ for $n = 1, \ldots, n_0 - 1$).

(ii) The transformation φ is *periodic* (*with period* $n \in \mathbb{N}_0$) if $\varphi^{n_0} = \mathrm{id}$ (and $\varphi^n \neq \mathrm{id}$ for $n = 1, \ldots, n_0 - 1$).

(iii) The transformation φ is *antiperiodic* if the set of periodic points in X is a m-null set.

Remarks:

1. If the transformation is periodic, so is every point, but not conversely since the set of all periods may be unbounded.
2. The set $A_n := \{x \in X : x \text{ has period } n\}$ is measurable: Consider a "separating base" $\{B_k \in \mathcal{B} : k \in \mathbb{N}\}$, i.e. a sequence which generates $\mathcal{B}$ and separates the points of X (see A.13 and **??**). Then we obtain
$$\{x \in X : \varphi^n x = x\} = \bigcap_{k \in \mathbb{N}} (B_k \cap \varphi^n B_k) \cup ((X \backslash B_k) \cap \varphi^n (X \backslash B_k))$$
for every $n \in \mathbb{N}$, and therefore we conclude that $A_n \in \mathcal{B}$.
3. An arbitrary transformation φ may be decomposed into periodic and antiperiodic parts:
As above take A_n to be the set of all points in X with period n and $A_{\mathrm{ap}} := X \backslash \bigcup_{n \in \mathbb{N}} A_n$. Then X is the disjoint union of the φ-invariant sets A_n, $n \in \mathbb{N}$ and A_{ap}. The restriction of φ to A_n is periodic with period n and φ is antiperiodic on A_{ap}.
4. An ergodic transformation on $([0, 1], \mathcal{B}, m)$ is antiperiodic. This is an immediate consequence of the following important lemma.

X.2 Lemma Rohlin's lemma: Consider an MDS $(X, \mathcal{B}, m\varphi)$.

(i) If every point $x \in X$ has period n then there exists $A \in \mathcal{B}$ such that $A, \varphi A$, $\varphi^2 A, \ldots, \varphi^{n-1} A$ are pairwise disjoint and $m(A) = \frac{1}{n}$.

(ii) If φ is antiperiodic then for every $n \in \mathbb{N}$ and $\varepsilon > 0$ there exists $A \in \mathcal{B}$ such that$A, \varphi A, \varphi^2 A, \ldots, \varphi^{n-1} A$ are pairwise disjointand $m\big(\bigcup_{k=0}^{n-1} \varphi^k A) > (1 - \varepsilon)$.

Proof. (i) If $n > 1$ there exists a measurable set C_n such that $m(C_1 \triangle \varphi C_1) > 0$ (use the existence of a separating base) and therefore $m(C_1 \backslash \varphi C_1) = m(C_1) - m(c_1 \cap$

$\varphi C_1) = m(\varphi C_1) - m(\varphi C_1 \cap C_1) = m(\varphi C_1 \cap C_1) = m(\varphi C_1 \backslash C_1) > 0$. Certainly, $B_1 := C_1 \backslash \cap C_1$ is disjoint from φB_1. If $n > 2$ there exists $C_2 \subseteq B_1$ $m(C_2 \triangle \varphi^2 C_2) > 0$. For $B_2 := C_2 \backslash \varphi^2 C_2$ we have $m(B_2) > 0$, and the sets $B_2, \varphi B_2, \varphi^2 B_2$ are pairwise disjoint. Proceeding in this way we obtain B_{n-1} such that $m(B_{n-1}) > 0$ and $B_{n-1}, \varphi B_{n-1}, \ldots, \varphi^{n-1} B_{n-1}$ are pairwise disjoint.
Consider the measure algebra $\breve{\mathcal{B}}$ and the equivalence classes $\breve{B} \in \breve{\mathcal{B}}$ of sets $B \in \mathcal{B}$ such that $B, \varphi B, \ldots, \varphi^{n-1} B$ are pairwise disjoint. Since $\breve{\mathcal{B}}$ is a complete Boolean algebra (see A.9) an application of Zorn's lemma yields $\breve{A} \in \breve{\mathcal{B}}$ which is maximal such that $A, \varphi A, \ldots, \varphi^{n-1} A$ are pairwise disjoint for some $A \in \breve{A}$. If we assume $m(A) < \frac{1}{n}$ we can apply the above construction to the φ-invariant set $X \backslash \bigcup_{i=0}^{n-1} \varphi^i A$ and obtain contradiction to the maximality of A. Therefore, $\mu(A) = \frac{1}{n}$, and the assertion is proved.

(ii) We may take $\varepsilon = \frac{1}{p}$ for some $p \in \mathbb{N}$. For $r := n[$ and as in the proof of (i) we construct $B \in \mathcal{B}$ such that $B, \varphi B, \ldots, \varphi^{n-1} B$ are pairwise disjoint and such that B is maximal relative to this property. For $1 \leqslant k \leqslant r$ define

$$B_k : +\{x \in \varphi^{r-1} B : \varphi^k x \in B \text{ and } \varphi^j x \notin B \text{ for } 1 \leqslant j < k\}.$$

These sets are pairwise disjoint, and the same holds for $B_k, \varphi B_k, \ldots, \varphi^k B_k$ for any $k = 1, \ldots, r$. Therefore, the maximality of B implies

$$(*) \qquad m\Big(\varphi^{r-1} B \backslash \bigcup_{k=1}^{r} B_k\Big) = 0.$$

Moreover, the sets

$$\begin{aligned}
&\varphi B_2 \\
&\varphi B_3, \varphi^2 B_3 \\
&\varphi B_4, \varphi^2 B_4, \varphi^3 B_4 \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
&\varphi B_r, \varphi^2 B_r, \varphi^3 B_r, \ldots\ldots\ldots \varphi^{r-1} B_r
\end{aligned}$$

are disjoint from any $\varphi^k B$ for $0 \leqslant k \leqslant r-1$, since

$$\varphi^i B_j \cap \varphi^k B = \varphi^i (B_j \cap \varphi^{k-i} B) \subseteq \varphi^i (\varphi^{r-1} B \cap \varphi^{k-i}) = \varnothing$$

if $0 < i < j \leqslant r$ and $i \leqslant k$ (resp. $\varphi^i B_j \cap \varphi^k B = \varphi^k (\varphi^{i-k} B_j \cap B) = \varnothing$ if $k < i$).
Finally, they are pairwise disjoint as can be seen considering sets contained in the same, resp. in different columns. In particular, we find that $\varphi B_1, \varphi^2 B_2, \varphi^3 B_3, \ldots, \varphi^r B_r$ are pairwise disjoint subsets of B. Therefore, by (**??**) we obtain

$$m\Big(\bigcup_{k=1}^{r} \varphi^k B_k\Big) = m\Big(\bigcup_{k=1}^{r} B_k\Big) = m(\varphi^{r-1} B) = m(B).$$

Now, consider

$$B^* := \bigcup_{k=0}^{r-1} \varphi^k B \cup \bigcup_{1 \leqslant i \leqslant j \leqslant n} \varphi^i B_j,$$

which is φ-invariant modulo m-null sets. Since B is maximal and φ is antiperiodic it follows that

$$B^* = X.$$

Finally, we obtain the desired set:

$$A := \bigcup_{k=0}^{p-1} \varphi^{kn} B \cup \bigcup_{k=0}^{p-2} \bigcup_{j=(k+1)n+1}^{r} \varphi^{kn+1} B_j.$$

Obviously, $A, \varphi A, \ldots, \varphi^{n-1}A$ are pairwise disjoint, and $\bigcup\limits_{i=0}^{n-1} \varphi^i A$ contains every $\varphi^k B$, $0 \leqslant k \leqslant r-1$. From $B^* = X$ it follows that $X \setminus \bigcup_{i=0}^{n-1} \varphi^i A$ is contained $\bigcup_{k=0}^{n-1} \bigcup_{0<i<j\leqslant n} \varphi^{kn+1} B_{kn+j}$. Therefore, we conclude that

$$m\Big(X \setminus \bigcup_{i=0}^{n-1} \varphi^i A\Big) \leqslant n \cdot m(B) \leqslant \tfrac{n}{r} = \varepsilon.$$

∎

The lemma above will be used to show that the periodic transformations occur frequently in the set of all bi-measure-preserving transformations on X. To that purpose we denote by $\tilde{\mathscr{G}}$ the group of all bi-measure-preserving bijections on $(X, \mathcal{B}, m)$. Here we identify transformations which coincide m-almost everywhere.
The set $\mathscr{G} := \{T_\varphi : \varphi \in \tilde{\mathscr{G}}\}$ of all induced operators

$$T_\varphi : L^1(X, \mathcal{B}, m) \to L^(X, \mathcal{B}, m)$$

is a group in $\mathscr{L}(L^1(X, \mathcal{B}, m))$. The following lemma shows that the map $\varphi \mapsto T_\varphi$ from $\tilde{\mathscr{G}}$ onto $\mathscr{G}$ is a group isomorphism.

X.3 Lemma:
If $\varphi \in \tilde{\mathscr{G}}$ and $m\{x \in X : \varphi(x) \neq x\} > 0$, then $T_\varphi \neq \mathrm{id}$.

Proof. The assumption $m\{x \in X : \varphi(x) \neq x\} > 0$ implies that at least one of the measurable sets A_n, $n \geqslant 2$, or A_{ap} defined in Remrarks 2, 3 following (X.1) has non-zero measure. By (X.2) we obtain a measurable set A such that $m(A) > 0$ and $A \cap \varphi(A) = \varnothing$. This yields

$$A \cap \varphi^{-1}(A) = \varnothing \quad \text{and} \quad T_\varphi \mathbf{1}_A = \mathbf{1}_{\varphi^{-1}(A)} \neq \mathbf{1}_A.$$

∎

On $\mathscr{G}$ we consider the topology which is induced by the strong operator topology on $\mathscr{L}(L^1(m))$. This topology coincides on $\mathscr{G}$ with the topology of pointwise convergence on all characteristic functions $\mathbf{1}_{B_k}$, $k \in \mathbb{N}$, where $\{B_k : k \in \mathbb{N}\}$ generates $\mathcal{B}$ (use B.11), and will be transferred to $\tilde{\mathscr{G}}$. In particular, T_{φ_i} converges to T_φ (resp. φ_i converges to φ) if and only if $m(\varphi_i(A) \triangle \varphi(A)) \to 0$ for every $A \in \mathcal{B}$. Since the multiplication on bounded subsets of $\mathscr{L}(L^1(m))$ is continuous for the strong operator topology, $\mathscr{G}$ (and $\tilde{\mathscr{G}}$) is a topological group which is metrizable. In (**??**) we shall see that $\mathscr{G}$ is complete, hence $\mathscr{G}$ and $\tilde{\mathscr{G}}$ are complete metric spaces, and Baire's category theorem is applicable (see A.6).

X.4 Proposition:
For every $n \in \mathbb{N}$ the set of all periodic transformations on $(X, \mathcal{B}, m)$ with period larger than n is dense in $\tilde{\mathscr{G}}$.

Proof. Consider $\varphi \in \tilde{\mathscr{G}}$, $\varepsilon > 0$ and characteristic functions $\chi_1, \ldots, \chi_m \in L^1(m)$. We shall construct $\psi \in \tilde{\mathscr{G}}$ with period larger than n such that

$$\|T_\varphi \chi_i - T_\psi \chi_i\| \leqslant 3\varepsilon \quad \text{for } i = 1, \ldots, m.$$

To that aim we decompose X as in (X.1), Remark 3, into antiperiodic part A_{ap} and periodic parts A_j, $j \in \mathbb{N}$. Then choose $l \in \mathbb{N}$ such that $m(\bigcup_{j\geqslant l} A_j) < \frac{\varepsilon}{2}$. Defining $B := A_1 \cup \cdots \cup A_l$ we observe that $\varphi|_B$ is periodic with period at most equal to $l!$. In the next step, we choose $k \in \mathbb{N}$ such that k is a multiple of $l!$ and larger than $\max\{n, \frac{2}{\varepsilon}\}$. Now, apply (X.2.ii) and find a measurable set $C \subseteq A_{\mathrm{ap}}$. such that $C, \varphi C, \ldots, \varphi^{k-1}C$ are pairwise disjoint and

$$\tfrac{1}{k}(1-\tfrac{\varepsilon}{2})\cdot m(A_{\mathrm{ap}}) \leqslant m(C) \leqslant \tfrac{1}{k}m(A_{\mathrm{ap}}).$$

The transformation $\psi \in \tilde{\mathscr{G}}$ defined as

$$\psi(x) := \begin{cases} \varphi(x) & \text{for } x \in B \cup C \cup \varphi C \cup \cdots \cup \varphi^{k-1}C \\ \varphi^{1-k}(x) & \text{for } x \in \varphi^{k-1}C \\ x & \text{for all other } x \in X, \end{cases}$$

is periodic with period $k > n$. But, ψ coincides with φ outside of a set R with measure

$$m(R) \leqslant \tfrac{1}{k}m(A_{\mathrm{ap}}) + \tfrac{\varepsilon}{2}m(A_{\mathrm{ap}}) + \tfrac{\varepsilon}{2} \leqslant 3\cdot\tfrac{\varepsilon}{2}.$$

Therefore, we conclude $\|T_\varphi\chi_i - T\psi\chi_i\| \leqslant 2\cdot m(R) \leqslant 3\varepsilon$ for $i = 1, \ldots, m$. ■

X.5 Theorem Rohlin, 1948:
The set $\tilde{\mathscr{S}}$ of all strongly mixing transformations on $(X, \mathcal{B}, m)$ is of first category in $\tilde{\mathscr{G}}$.

Proof. Proofs Let $A := [0, \frac{1}{2}] \subseteq X$. For every $k \in \mathbb{N}$,

$$\tilde{\mathscr{M}}_k := \{\varphi \in \tilde{\mathscr{G}} : |m(A \cap \varphi^k A)| - \tfrac{1}{4}| \leqslant \tfrac{1}{5}\}$$

is closed. If $\varphi \in \tilde{\mathscr{G}}$ is strongly mixing, we

$$\lim_{k\to\infty} m(A \cap \varphi^k A) = m(A)^2 = \frac{1}{4} \quad \text{(by IX.2)},$$

hence $\varphi \in \tilde{\mathscr{M}}_k$ for all $k \geqslant k_0$, or

$$\tilde{\mathscr{S}} \subset \bigcup_{n\in\mathbb{N}} \tilde{\mathscr{N}}_n \quad \text{for} \quad \tilde{\mathscr{N}}_n := \bigcap_{k\geqslant n} \tilde{\mathscr{M}}_k.$$

Since $\tilde{\mathscr{N}}_n$ is closed it remains to show that $\tilde{\mathscr{G}}\backslash\tilde{\mathscr{N}}_n$ is dense in $\tilde{\mathscr{G}}$. If φ is periodic, say $\varphi^k = \mathrm{id}$, then

$$m(A \cap \varphi^k A) - \frac{1}{4} = \frac{1}{4}, \quad \text{hence } \varphi \in \tilde{\mathscr{G}}\backslash\tilde{\mathscr{M}}_k.$$

Therefore $\bigcup\limits_{k\geqslant n}\{\varphi \in \tilde{\mathscr{G}} : \varphi^k = \mathrm{id}\} \subseteq \tilde{\mathscr{G}}\backslash \bigcap\limits_{k\geqslant n} \tilde{\mathscr{M}}_k = \tilde{\mathscr{G}}\backslash\tilde{\mathscr{N}}_k$, and the assertion follows from (X.4). ■

X.6 Proposition:
The set $\tilde{\mathscr{W}}$ of all weakly mixing transformations on $(X, \mathcal{B}, m)$ is dense in $\tilde{\mathscr{G}}$.

For the somewhat technical proof using "dyadic permutations" of $[0, 1]$ we refer to Halmos [1956], p. 65, or Jacobs [1960], p.126, but we draw the following beautiful conclusion.

X.7 Theorem Halmos, 1944:
The set $\tilde{\mathscr{W}}$ of all weakly transformations on $(X, \mathcal{B}, m)$is of second category in $\tilde{\mathscr{G}}$.

Proof. Since $\tilde{\mathscr{G}}$ is a complete metric space, Baire's category theorem (see A.6) asserts that $\tilde{\mathscr{G}}$ is of second category. Therefore and by (X.7) it is enough to show that $\tilde{\mathscr{W}}$ is the intersection of a sequence of open sets. We prove this assertion for the (induced) operator sets $\mathscr{W} := \{T_\varphi \in \mathscr{L}(L^1(m)) : \varphi \in \tilde{\mathscr{W}}\}$. Let $\{f_i\}_{i\in\mathbb{N}}$ be a subset of $L^\infty(m)$ which is dense in$L^1(m)$. Define

$$\mathscr{W}_{ijkn} := \left\{T_\varphi \in \mathscr{G} : |\langle T^n f_i, f_j\rangle - \langle f_i, \mathbf{1}\rangle| < \tfrac{1}{k}\right\} \quad \text{for } i,j,k,n \in \mathbb{N}.$$

By (**??**) the sets $\mathscr{W}_{ijkn}$ and therefore $\mathscr{W}_{ijk} := \bigcup_{n\in\mathbb{N}} \mathscr{W}_{ijkn}$ are open. We shall show that $\mathscr{W} = \bigcap_{i,j,k} \mathscr{W}_{ijk}$. The inclusion $\mathscr{W} \subseteq \bigcap_{i,j,k} \mathscr{W}_{ijk}$ is obvious by (IX.5.a). On the other hand, if φ is not weakly mixing, then there exists a non-constant eigenvector $h \in L^1(m)$ of T_φ with unimodular eigenvalue λ. It is possible to choose h with $\|h\| = 1$ and $\langle h, \mathbf{1}\rangle = 0$. Now, choose $k \in \mathbb{N}$ such that $\|h - f_k\| \leqslant \frac{1}{10}$. We obtain

$$\begin{aligned} &\left|\langle T_\varphi^n f_k, f_k\rangle - \langle f_k, \mathbf{1}\rangle \cdot \langle f_k, \mathbf{1}\rangle\right| = \\ &\left|\langle T_\varphi^n (f_k - h), (f_k - h)\rangle - \langle (f_k - h), \mathbf{1}\rangle \cdot \langle (f_k - h), \mathbf{1}\rangle + \langle T^n h, h\rangle\right| \geqslant \tfrac{1}{2} \end{aligned}$$

for every $n \in \mathbb{N}$. This yields $T_\varphi \notin \mathscr{W}_{kk2}$, and the theorem is proved. ∎

Combining (X.5) and (**??**) we conclude that there exist weakly mixing transformations on $(X, \mathcal{B}, m)$ which are not strongly mixing. But, even if "most" transformations are of this type no explicit example was known before Chacon and Kakutani in 1965 presented the first concrete construction. Later on, Chacon and others developed a method of constructing MDS's enjoying very different properties ("stacking method"); We shall use this method in its simplest form in order to obtain a weakly mixing MDS which is not strongly mixing. The basic concepts of the construction are set down in the following definition.

X.8 Definition:

(i) A *column* $C := (I_j)_{j=,1\ldots,q}$ of *height* q is a q-tuple of disjoint intervals $I_j = [a_j, b_j) \subseteq [0,1)$ of equal length.

(ii) With a column C there is associated a piecewise linear mapping

$$\varphi_C : \bigcup_{j=1}^{q-1} I_j \to \bigcup_{j=2}^{q} I_j \quad \text{defined by}$$
$$\varphi_C(x) := (x - a_j) + a_{j+1} \quad \text{for } x \in I_j.$$

Remark: A column is represented diagrammatically as follows:

Therefore the mapping φ_C moves a point $x \in I_j$, $j \leqslant q-1$ vertically upwards to $\varphi_C(x) \in I_{j+1}$.

The main part in the construction of the desired MDS $(X, \mathcal{B}, m; \varphi)$ consists in the definition of a sequence $C(n) = (I_j(n))_{j=1,\ldots,q(n)}$ of columns. Then we use the associated mappings $\varphi_n := \varphi_{C(n)}$ to define φ on X.

Take $C(0) := ([0, \frac{1}{2}))$ and denote the remainder by $R(0) := [\frac{1}{2}, 1)$. Cut $C(0)$ and $R(0)$ "in half" and let

$$C(1) := ([0, \tfrac{1}{4}), [\tfrac{1}{4}, \tfrac{1}{2}), [\tfrac{1}{2}, \tfrac{3}{4})) \quad \text{and} \quad R(1) = [\tfrac{3}{4}, 1).$$

In this way we proceed! More precisely, from $I_j(n) = [a_j(n), b_j(n)) \in C(n)$ we produce

$$I'_j(n) := \Big[a_j(n), \tfrac{a_j(n)+b_j(n)}{2}\Big)$$

and

$$I''_j(n) := \Big[\tfrac{a_j(n)+b_j(n)}{2}, b_j(n)\Big),$$

and from $R(n)$ we produce

$$R'(n) := \Big[b_{q(n)}(n), \tfrac{b_{q(n)}(n)+1}{2}\Big)$$

and

$$R''(n) := \Big[\tfrac{b_{q(n)}(n)+1}{2}, 1\Big).$$

Then we define

$$C(n+1) := (I'_1(n), \ldots, I'_{q(n)}(n), I''_1(n), \ldots, I''_{q(n)}(n), R'(n))$$

and $\quad R(n+1) := R''(n))$.

This procedure can be illustrated as follows:
The objects defined above possess the following properties:

1. $m(R(n)) = 2^{-n+1}$ converges to zero as n tends to infinity.
2. Every interval $I_j(n) \in C(n)$ is a union of intervals in $C(n+1)$.
3. The σ-algebra $\sigma\big(\bigcup_{n=k}^{\infty} \bigcup_{j=1}^{q(n)} I_j(n)\big)$, $k \in \mathbb{N}$, is equal to the Borel algebra.
4. The mapping φ_{n+1} is an extension of φ_n.
5. For every $x \in [0,1)$ there exists n such that
$$\varphi(x) := \varphi_n(x), \quad n \geqslant n_0,$$
is defined.

Now: $(X, \mathcal{B}, m; \varphi)$ is an MDS if we take φ as the mapping just defined $X = [0,1)$.

X.9 Theorem:
The MDS $(X, \mathcal{B}, m; \varphi)$ is weakly but not strongly mixing.

Proof. (i) $(X, \mathcal{B}, m; \varphi)$ is not strongly mixing: Take $A := I_1(1) = [0, \frac{1}{4})$. By (1) above A is a union of intervals in $C(n)$, and by definition of φ it follows $m(\varphi^{-q(n)}(I_j(n)) \cap I_j(n)) \geqslant \frac{1}{2} m(I_j(n))$. Therefore

$$m(\varphi^{-q(n)}(A) \cap A) \geqslant \frac{1}{2} m(A) = \frac{1}{8} \quad \text{for every } n \in \mathbb{N}.$$

But if φ were strongly mixing, then $m(\varphi^{-q(n)}(A) \cap A)$ would converge to $(m(A))^2 = \frac{1}{16}$ (see IX.2 and IX.3).

(ii) The weak mixing of $(X, \mathcal{B}, m; \varphi)$ is proved three steps.

1) For $n \in \mathbb{N}$ and $A \in \mathcal{B}$ choose $L_{n,A} \subseteq \{1, 2, \ldots, q(n)\}$ such that $m(A \triangle \bigcup_{j \in L_{n,A}} I_j(n))$ is minimal and define

$$A(n) := \bigcup_{j \in L_{n,A}} I_j(n).$$

By property 2) above and by (A.11) $m(A \triangle A(n))$ converges to zero as $n \to \infty$. Now, $m(A(n)) = |L_{n,A}| m(I_1(n)) = q(n)^{-1} \cdot |L_{n,A}| \cdot (1 - m(R(n))$ implies that $\lim_{n\to\infty} q(n)^{-1} \cdot |L_{n,A}| = \lim_{n\to\infty} m(A(n)) = m(A)$ by property (0).

2) $(X, \mathcal{B}, \mu; \varphi)$ is ergodic: Assume $\varphi(A) = A \in \mathcal{B}$. This implies for any $j = 1, \ldots, q(n)$ that

$$m(I_j(n) \cap A) = m(\varphi^{j-1}(I_1(n) \cap A)) = m(I_1(n) \cap A) = q(n)^{-1}(m(A) - m(R(n) \cap A)$$

and therefore

$$m(A(n) \cap A) = q(n)^{-1} \cdot |L_{n,A}| \cdot (m(A) - m(R(n) \cap A).$$

The following calculation

$$\begin{aligned} 0 &= \lim_{n\to\infty} m(A(n) \triangle A) = \lim_{n\to\infty} (m(A(n)) + m(A) - 2m(A(n) \cap A)) \\ &= \lim_{n\to\infty} (m(A(n)) + m(A) - 2q(n)^{-1} |L_{n,A}| (m(A) - m(R(n) \cap A))) \\ &= m(A) + m(A) - 2m(A) \cdot m(A) \\ &= 2m(A)(1 - m(A)) \end{aligned}$$

proves that $m(A) = 0$ or $m(A) = 1$, i.e. φ is ergodic.

3)Finally, it remains to show that 1 is the only eigenvalue of the induced operator $T_v a$ (see IX.4): Assume $T_\varphi f = \lambda f$, $0 \neq f \in L^\infty(m)$, and take $0 < \varepsilon < \frac{1}{8}$. By Lusin's theorem (see A.15) there exists a closed set $D \subseteq [0, 1)$ of positive measure on which f is uniformly continuous, so that there is $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$ for $x, y \in D$. Choosing n large enough we obtain a set $L \subseteq \{1, \ldots, q(n)\}$ such that $D' := \bigcup_{i \in L} I_i(n)$ satisfies $D \subseteq D'$ and $m(D' \backslash D) < \varepsilon \cdot m(D) \leqslant \varepsilon m(D')$ and $m(I_i(n)) < \delta$ for $i \in L$. Now, define $I := I_j(n) \cap D$, where $m(I_j(n) \backslash D) < \varepsilon \cdot m(I_j(n))$ for a suitable $j \in L$. From the definition of φ it follows that

$$m(\varphi^{q(n)}(I_j(n)) \cap I_j(n)) \geqslant \tfrac{1}{2} m(I_j(n))$$

and

$$m(\varphi^{q(n)+1}(I_j(n)) \cap I_j(n)) \geqslant \tfrac{1}{4} m(I_j(n)).$$

Therefore, we conclude that

$$\begin{aligned} m(\varphi^{q(n)}(I) \cap I) &= m(\varphi^{q(n)}(I_j(n)) \cap I) - m(\varphi^{q(n)}(I_j(n) \backslash D) \cap I) \\ &\geqslant m(\varphi^{q(n)}(I_j(n)) \cap I_j(n)) - m(\varphi^{q(n)}(I_j(n((\cap(I_j(n) \backslash D)) - \varepsilon \cdot m(I_j(n)) \\ &\geqslant \tfrac{1}{2} m(I_j(n)) - 2\varepsilon \cdot m(I_j(n)) > 0 \end{aligned}$$

and analogously

$$m(\varphi^{q(n)+1}(I) \cap I) \geqslant \tfrac{1}{4} m(I_j(n)) - 2 \cdot \varepsilon \cdot m(I_j(n)) > 0.$$

If $x = \varphi 6q(n)(y) \in \varphi^{q(n)}(I) \cap I$ we obtain

$$f(x) = f(\varphi^{q(n)}(y)) = \lambda^{q(n)} f(y) \quad \text{and} \quad |f(x) - f(y)| < \varepsilon.$$

If $x' = \varphi 6q(n) + 1(y') \in \varphi^{q(n)+1}(I) \cap I$ we obtain

$$f(x') = f(\varphi^{q(n)+1}(y')) = \lambda^{q(n)+1} f(y') \quad \text{and} \quad |f(x') - f(y')| < \varepsilon.$$

Finally,

$$\lambda = \frac{\lambda^{q(n)+1}}{\lambda^{q(n)}} = \frac{f(x')}{f(y')} \cdot \frac{f(y)}{f(x)}$$

implies

$$|\lambda - 1| \leqslant \left| \frac{f(x')}{f(y')} \cdot \left(\frac{f(y)}{f(x)} - 1 \right) \right| + \left| \frac{f(x')}{f(y')} - 1 \right| \leqslant 2\varepsilon$$

which proves that 1 is the only eigenvalue of T_φ. ∎

X.D Discussion

XI. Information of Covers

XII. Entropy of Dynamical Systems

XIII. Uniform Entropy and Comparison of Entropies

Appendix A. Some Topology and Measure Theory

(i) Topology

The concept of a topological space is so fundamental in modern mathematics that we don't feel obliged to recall its definitions or basic properties. Therefore we refer to Dugundji 1966 for everything concerning topology, nevertheless we shall briefly quote some results on compact and metric spaces which we use frequently.

A.1. Compactness:
A topological space $(X, \mathscr{O})$, $\mathscr{O}$ the family of open sets in X, is called *compact* if it is Hausdorff and if every open cover of X has a finite subcover. The second property is equivalent to the *finite intersection property*: every family of closed subsets of X, every finite subfamily of which has non-empty intersection, has itself non-empty intersection.

A.2. The continuous image of a compact space is compact if it is Hausdorff. Moreover, if X is compact, a mapping $\varphi : X \to X$ is already a homeomorphism if it is continuous and bijective. If X is compact for some topology $\mathscr{O}$ and if $\mathscr{O}'$ is another topology on X, coarser than $\mathscr{O}$ but still Hausdorff, then $\mathscr{O} = \mathscr{O}'$.

A.3. Product spaces:
Let $(X_\alpha)_{\alpha \in A}$ a non-empty family of non-empty topological spaces. The product $X := \prod_{\alpha \in A} X_\alpha$ becomes a topological space if we construct a topology on X starting with the base of open rectangles, i.e. with sets of the form $\{x = (x_\alpha)_{\alpha \in A} : x_{\alpha_i} \in O_{\alpha_i}$ for $i = 1, \ldots, n\}$ for $\alpha_1, \ldots, \alpha_n \in A$, $n \in \mathbb{N}$ and O_{α_i} open in X_{α_i}. Then Tychonov's theorem asserts that for this topology, X is compact if and only if each X_α, $\alpha \in A$ is compact.

A.4. Urysohn's lemma:
Let X be compact and A, B disjoint closed subsets of X. Then there exists a continuous function $f : X \to [0, 1]$ with $f(A) \subseteq \{0\}$ and $f(B) \subseteq \{1\}$.

A.5. Lebesgue's covering lemma: If (X, d) is a compact metric space and α is is a finite open cover of X, then there exists a $\delta > 0$ such that every set $A \subseteq X$ with diameter diam $(A) < \delta$ is contained in some element of α.

A.6. Category: A subset A of a topological space X is called *nowhere dense* if the closure of A, denoted by $\overline{A}$, has empty interior: $\overset{\circ}{\overline{A}} = \varnothing$. A is called of *first category* in X if A is the union of countably many nowhere dense subsets of X. A is called of *second category* in X if it is not of first category. Now let X be a compact or a complete metric space. Then Baire's category theorem states that every non-empty open set is of second category.

(ii) Measure theory

Somewhat less elementary but even more important for ergodic theory is the concept of an abstract measure space. We shall use the standard approach to measure- and integration theory and refer to Bauer [1972] and Halmos [1950]. The advanced reader is also directed to Jacobs [l978]. Although we again assume that the reader is familiar with the basic results, we present a list of more or less known definitions and results.

A.7. Measure spaces and null sets:
A triple (X, Σ, μ) is a *measure space* if X is a set, Σ σ-algebra of subsets of X and μ a measure on Σ, i.e.

$$\mu : \Sigma \to \mathbb{R}_+ \cup \{\infty\}$$

σ-additive and and $\mu(\varnothing) = 0$.

If $\mu(X) < \infty$ (resp. $\mu(X) = 1$), X, Σ, μ is called a *finite* measure space (resp. a *probability space*); it is called *σ-finite*, if $X = \bigcup_{n\in\mathbb{N}} A_n$ with $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$.

A set $N \subseteq \Sigma$ is a *μ-null set* if $\mu(N) = 0$.

Properties, implications, conclusions etc. are valid "μ-almost everywhere" or for "almost all $x \in X$" if they are valid for all $x \in X\backslash N$ where N is some μ-null set. If no confusion seems possible we sometimes write "... is valid for all x" meaning "... is valid for almost all $x \in X$".

A.8. Equivalent measures:
Let (X, Σ, μ) be a σ-finite measure space and ν another measure on Σ. ν is called *absolutely continuous* with respect to μ if every μ-null set is ν-null set. ν is equivalent to μ iff ν is absolutely continuous with respect to μ and conversely. The measures which are absolutely continuous with respect to μ can be characterized by the Radon-Nikodỳm theorem (see Halmos [1950], §31).

A.9. The measure algebra:
In a measure space (X, Σ, μ) the μ-null sets form a σ-ideal $\mathscr{N}$. The Boolean algebra

$$\check{\Sigma} := \Sigma/\mathscr{N}$$

is called the corresponding *measure algebra*. We remark that $\check{\Sigma}$ is isomorphic to the algebra of characteristic functions in $L^\infty(X, \Sigma, \mu)$ (see App.B.20) and therefore is a *complete Boolean algebra*.

For two subsets A, B of X,

$$A \triangle B := (A \cup B)\backslash(A \cap B) = (A\backslash B) \cup (B\backslash A)$$

denotes the symmetric difference of A and B, and

$$d(A, B) := \mu(A \triangle B)$$

defines a semi-metric on X vanishing on Σ the elements of $\mathscr{N}$ (if $\mu(X) < \infty$). Therefore we obtain a metric on $\check{\Sigma}$ still denoted by d.

A.10 Proposition: The measure algebra $(\check{Z}, d)$ of a finite measure space X, Σ, μ is a complete metric space.

Proof. It suffices to show that (Σ, d) is complete. For a Cauchy sequence $(A_n)_{n\in\mathbb{N}} \subseteq \Sigma$, choose a subsequence $(A_{n_i})_{i\in\mathbb{N}}$ such that $d(A_k, A_l) < 2^{-i}$ for $k, l > n_i$. Then $A := \bigcap_{m=1}^\infty \bigcup_{j=m}^\infty A_{n_j}$ is the limit of (A_n). Indeed, with $B_m := \bigcup_{j=m}^\infty A_{n_j}$ we have

$$d(B_m, A_{n_m}) \leqslant \sum_{j=m}^\infty \mu(A_{n_{j+1}} \backslash A_{n_j}) \leqslant \sum_{j=m}^\infty 2^{-j} = 2 \cdot 2^{-m}$$

and

$$d(A,B_m) \leqslant \sum_{j=m}^{\infty} \mu(B_j \backslash B_{j+1}) \leqslant \sum_{j=m}^{\infty} \big(d(B_j, A_{n_j}) + d(A_{n_j}, A_{n_{j+1}}) + d(A_{n_{j+1}}, B_{j+1})\big)$$

$$\leqslant \sum_{j=m}^{\infty} \big(2 \cdot 2^{-j} + 2^{-j} + 2 \cdot 2^{-(j+1)}\big) \leqslant 8 \cdot 2^{-m}.$$

Therefore

$$d(A, A_k) \leqslant d(A, B_m) + d(B_m, A_{n_m}) + d(A_{n_m}, A_k) \leqslant 11 \cdot 2^{-m}$$

for $k \geqslant n_m$. ∎

A.11. For a subset $\breve{W}$ of $\check{\Sigma}$ we denote by $a(\breve{W})$ the *Boolean algebra generated* by $\breve{W}$, by $\sigma(\breve{W})$ the *Boolean σ-algebra generated* by $\breve{W}$.
$\check{\Sigma}$ is called countably generated, if there exists a countable subset $\breve{W} \subseteq \check{\Sigma}$ such that $\sigma(\breve{W}) = \check{\Sigma}$.
The metric d relates $a(\breve{W})$ and $\sigma(\breve{W})$. More precisely, using an argument as in (A.10) one can prove that in a finite measure space

$$\sigma(\breve{W}) = \overline{a(\breve{W})}^{d} \quad \text{for every } \breve{W} \subseteq \check{\Sigma}$$

.

A.12. The Borel algebra:
In many applications a set X bears a topological structure and a measure space structure simultaneously. In particular, if X is a compact space, we *always* take the σ-algebra $\mathscr{B}$ generated by the open sets, called the *Borel algebra* on X. The elements of $\mathcal{B}$ are called Borel sets, and a measure defined on $\mathcal{B}$ is a Borel measure. Further, we only consider *regular Borel measures*: here, μ is called regular if for every $A \in \mathcal{B}$ and $\varepsilon > 0$ there is a compact set $K \subseteq A$ and an open set $U \supseteq A$ such that $\mu(A \backslash K) < \varepsilon$ and $\mu(U \backslash A) < \varepsilon$.

A.13 Example:

Let $X = [0,1[$ be endowed with the usual topology. Then the Borel algebra $\mathcal{B}$ is generated by the set of all dyadic intervals

$$\mathscr{D} := \big\{[k \cdot 2^{-i}, (k+1) \cdot 2^{-i}] : i \in \mathbb{N}; k = 0, \ldots, 2^i - 1\big\}.$$

$\mathscr{D}$ is called a *separating base* because it generates $\mathcal{B}$ and for any $x, y \in X$, $x \neq y$, there is $D \in \mathscr{D}$ such that $x \in D$ and $y \notin D$, or $x \notin D$ and $y \in D$.

A.14. Measurable mappings:
Consider two measure spaces (X, Σ, μ) and (Y, T, ν). A mapping $\varphi : X \to Y$ is called *measurable*, if $\varphi^{-1}(A) \in \Sigma$ for every $A \in T$, and called *measure-preserving*, if, in addition, $\mu(\varphi^{-1}(A)) = \nu(A)$ for all $A \in T$ for all $A \in T$ (abbreviated: $\mu \circ \varphi^{-1} = \nu$).

For real-valued measurable functions f and g on (X, Σ, μ), where $\mathbb{R}$ is endowed with the Borel algebra, we use the following notation:

$$[f \in B] := f^{-1}(B) \quad \text{for } B \in \mathcal{B},$$
$$[f = g] = \{x \in X : f(x) = g(x)\},$$
$$[f \leqslant g] := \{x \in X : f(x) \leqslant g(x)\}.$$

Finally,

$$\mathbf{1}_A : x \mapsto \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \quad \text{denotes the characteristic}$$

function of $A \subseteq X$. If $A = X$, we often write $\mathbf{1}$ instead of $\mathbf{1}_X$.

A.15. Continuous vs. measurable functions:
Let X be compact, $\mathcal{B}$ the Borel algebra on X and μ a regular Borel measure. Clearly, every continuous function $f : X \to \mathbb{C}$ is measurable for the corresponding Borel algebras. On the other hand there is a partial converse:

Theorem Lusin: Let $f : X \to \mathbb{C}$ be measurable and $\varepsilon > 0$. Then there exists a compact set $A \subseteq X$ such that $\mu(X \backslash A) < \varepsilon$ and f is continuous on A.

Proof (Feldman [1981]): Let $\{U_j\}_{j\in\mathbb{N}}$ be a countable base of open subsets of $\mathbb{C}$. Let V_j be open such that $f^{-1}(U_j) \subseteq V_+j$ and $\mu(VV_j \backslash f^{-1}(U_j)) < \frac{\varepsilon}{2}2^{-j}$. If we take $B := \bigcup_{j=1}^{\infty}(V_j \backslash f^{-1}(U_j))$, we obtain $\mu(B) < \frac{\varepsilon}{2}$, and we show that $g := f|_{B^c}$ is continuous. To this end observe that

$$\begin{aligned} V_j \cap B^c &= V_j \cap (V_j \backslash f^{-1}(U_j))^c \cap B^c = V_j \cap (V_j^c \cup f^{-1}(U_j)) \cap B^c \\ &= V_j \cap f^{-1}(U_j) \cap B^c = f^{-1}(U_j) \cap B^c = g^{-1}(U_j). \end{aligned}$$

Since any open subset U of $\mathbb{C}$ can be written as $U = \bigcup_{j\in M} U_j$, we have $G^{-1}(U) = \bigcup_{j\in M} g^{-1}(U_j) = \bigcup_{j\in M} V_j \cap B^c$, which is open in B^c. Now we choose a compact set $A \subseteq B^c$ with $\mu(B^c \backslash A) < \frac{\varepsilon}{2}$, and conclude that f is continuous on A and that $\mu(X \backslash A) = \mu(B) + \mu(B^c \backslash A) < \varepsilon$. ∎

A.16. Convergence of integrable functions:
Let (X, Σ, μ) be a finite measure space and $1 \leqslant p < \infty$. A measurable (real) function f on X is called *p-integrable*, if $\int |f|^p \, \mathrm{d}\mu < \infty$ (see Bauer [1972], 2.6.3).

For sequences $(f_n)_{n\in\mathbb{N}}$ of p-integrable functions we have three important types of convergence:

1. $(f_n)_{n\in\mathbb{N}}$ converges to f *μ-almost everywhere* if
$$\lim_{n\to\infty} (f_n(x) - f(x)) = 0 \quad \text{for almost all } x \in X.$$
2. $(f_n)_{n\in\mathbb{N}}$ converges to f *in the p-norm* if
$$\lim_{n\to\infty} \int |f_n - f|^p \, \mathrm{d}\mu = 0 \quad \text{see (B.20).}$$
3. $(f_n)_{n\in\mathbb{N}}$ converges to f *μ-stochastically* if
$$\lim_{n\to\infty} \mu[|f_n - f| \geqslant \varepsilon] = 0 \quad \text{for every } \varepsilon > 0.$$

Proposition: Let $(f_n)_{n\in\mathbb{N}}$ be p-integrable functions and f be measurable.

(i) If $f_n \to f$ μ-almost everywhere or in the p-norm, then $f_n \to f$ μ-stochastically (see Bauer [1972], 2.11.3 and 2.11.4).

(ii) If $(f_n)_{n\in\mathbb{N}}$ converges to f in the p-norm, then there exists a subsequence (f_{n_k}) converging to f μ-a.e. (see Bauer [1972], 2.7.5).

(iii) If $(f_n)_{n\in\mathbb{N}}$ converges to f μ-a.e. and if there is a p-integrable function g such that $|f_n(x)| \leqslant g(x)$ μ-a.e., then $f_n \to f$ in the p-norm and f is p-integrable (Lebesgue's dominated convergence theorem, see Bauer [1972], 2.7.4).

Simple examples show that in general no other implications are valid.

A.17. Product spaces:
Given a countable family $(X_\alpha, \Sigma_\alpha, \mu_\alpha)_{\alpha\in A}$ of probability spaces, we can consider the cartesian product $X = \prod_{\alpha\in A}$ and the so-called product σ-algebra $\Sigma = \bigotimes_{\alpha\in A}\Sigma_\alpha$ which is generated by the set of all *measurable rectangles*, i.e. sets of the form

$$R_{\alpha_1,\ldots,\alpha_n}(A_{\alpha_1},\ldots,A_{\alpha_n}) := \big\{x = (x_\alpha)_{\alpha\in A} : x_{\alpha_i} \in A_{\alpha_i} \text{ for } i = 1,\ldots,n\big\}$$

for $\alpha_1,\ldots,\alpha_n \in A$, $n \in \mathbb{N}$, $A_{\alpha_i} \in \Sigma_{\alpha_i}$.

The well known extension theorem of Hahn-Kolmogorov implies that there exists a unique probability measure $\mu : \bigotimes_{\alpha\in A}\mu_\alpha$ on Σ such that

$$\mu(R_{\alpha_1,\ldots,\alpha_n}(A_{\alpha_1},\ldots,A_{\alpha_n})) = \prod_{i=1}^{n}\mu_{\alpha_i}(A_{\alpha_i})$$

for every measurable rectangle (see Halmos [1950], §383 Theorem B).

Then X, Σ, μ is called the *product (measure) space* defined by $(X_\alpha, \Sigma_\alpha, \mu_\alpha)_{\alpha\in A}$

Finally, we mention an extension theorem dealing with a different situation (see also Ash [1972], Theorem 5.11.2).

Theorem: Let $(X_n)_{n\in\mathbb{Z}}$ be a sequence of compact spaces, $\mathcal{B}_n$ the Borel algebra on X_n. Further, we denote by Σ the product σ-algebra on $X = \prod_{n\in\mathbb{Z}} X_n$, by $\mathscr{F}_m$ the set of all measurable sets in X whose elements depend only on the coordinates $-m,\ldots,0,\ldots,m$. Finally we put $\mathscr{F} = \bigcup_{m\in\mathbb{N}}\mathscr{F}_m$. If μ is a function on $\mathscr{F}$ such that it is a regular probability measure on $\mathscr{F}_m$ for each $m \in \mathbb{N}$, then μ has a unique extension to a probability measure on Σ.

Remark: Let $\varphi_n : X \to Y_n := \prod_{-n}^{n} X_i$; $(x_j)_{j\in\mathbb{Z}} \mapsto (x_{-n},\ldots,x_n)$. Then we assume above that $\nu_n(A) := \mu(\varphi_n^{-1}(A))$, A measurable in Y_n, defines a regular Borel probability measure on Y_n for every $n \in \mathbb{N}$.

Proof. The set function μ has to be extended from $\mathscr{F}$ to $\sigma(\mathscr{F}) = \Sigma$. By the classical Carathèodory extension theorem (see Bauer [1972], 1.5) it suffices to show that $\lim_{i\to\infty}\mu(C_i) = 0$ for any decreasing sequence $(C_i)_{i\in\mathbb{N}}$ of sets in $\mathscr{F}$ satisfying $\bigcap_{i\in\mathbb{N}} C_i = \varnothing$. Assume that $\mu(C_i) \geqslant \varepsilon$ for all $i \in \mathbb{N}$ and some $\varepsilon > 0$. For each C_i there is an $n \in \mathbb{N}$ such that $C_i \in \mathscr{F}_n$ and $A_i \subseteq Y_n$ with $C_i = \varphi_n^{-1}(A_i)$. Let B_i a closed subset of A_i such that $\nu_n(A_i \backslash B_i) \leqslant \frac{\varepsilon}{2}\cdot 2^{-i}$. Then $D_i := \varphi_n^{-1}(B_i)$ is compact in X and $\mu(C_i\backslash D_i) \leqslant \frac{\varepsilon}{2}\cdot 2^{-i}$. Now the sets $G_k := \bigcap_{i=1}^{k} D_i$ form a decreasing sequence of compact subsets of X, and we have

$$\begin{aligned} G_k \subseteq C_k \text{ and } \mu(G_k) &= \mu(C_k) - \mu(C_k\backslash G_k) = \mu(C_k) - \mu\Big(\bigcup_{i=1}^{k}(C_i\backslash D_i)\Big) \\ &\geqslant \mu(C_k) - \sum_{i=1}^{k}\mu(C_i\backslash D_i) \geqslant \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}. \end{aligned}$$

Hence $G_k \neq \varnothing$ and therefore $\bigcap_{i\in\mathbb{N}} C_i$, which contains $\bigcap_{i\in\mathbb{N}} G_i$, is non-empty, a contradiction. ∎

Appendix B. Some Functional Analysis

As indicated in the introduction, the present lectures on ergodic theory require some familiarity with functional-analytic concepts and with functional-analytic thinking. In particular, properties of Banach spaces E, their duals E' and the bounded linear operators on E and E' play a central role. It is impossible to introduce the newcomer into this world of Banach spaces in a short appendix. Nevertheless, in a short "tour d'horizon" we put together some more or less standard definitions, arguments and examples – not as an introduction into functional analysis but as a reminder of things you (should) already know or as a reference of results we use throughout the book. Our standard source is Schaefer [1971].

B.1. Banach spaces:
Let E be a real or complex *Banach space* with norm $\|\cdot\|$ and closed unit ball $U := \{f \in E : \|f\| \leqslant 1\}$. We associate to E its dual E' consisting of all continuous linear functionals on E. Usually, E' will be endowed with the dual norm

$$\|f'\| := \sup\{|\langle f, f'\rangle| : \|f\| \leqslant 1\}$$

where $\langle \cdot, \cdot \rangle$ denotes the canonical bilinear form

$$(f, f') \mapsto \langle f, f'\rangle := f'(f) \quad \text{on } E \times E'.$$

B.2. Weak topologies:
The topology on E of pointwise convergence on E' is called the weak topology and will be denoted by $\sigma(E, E')$. Analogously, one defines on E' the topology of pointwise convergence on E, called the weak* topology and denoted by $\sigma(E', E)$. These topologies are weaker than the corresponding strong (= norm) topologies, and we need the following properties.

B.3. While in general not every strongly closed subset of a Banach space E is weakly closed, it is true that the strong and weak closure coincide for convex sets (Schaefer [1971], II.9.2, Corollary 2).

B.4. Theorem Alaoglu-Bourbaki:
The dual unit ball $U^\circ := \{f' \in E' : \|f'\| \leqslant 1\}$ in E' is weak* compact (Schaefer [1971], IV.5.2).

From this one deduces: A Banach space E is reflexive (i.e. the canonical injection from E into the bidual E'' is surjective) if and only if its unit ball is weakly compact (Schaefer [1971], IV.5.6).

B.5. Theorem of Krein-Milman
Every weak* compact, convex subset of E' is the closed, convex hull of its set of extreme points (Schaefer [1971], II.10.4).

B.6. Theorem of Krein:
The closed, convex hull of a weakly compact set is still weakly compact (Schaefer [1971], IV. 11.4).

B.7. Bounded operators:
Let T be a bounded (=continuous) linear operator on the Banach space E. Then T is called a *contraction* if $\|Tf\| \leqslant \|f\|$, and an *isometry* if $\|Tf\| = \|f\|$ for all $f \in E$. We remark that every bounded linear operator T on E is automatically continuous for the weak topology on E (Schaefer [1971], III.1.1). For $f \in E$ and $f' \in E'$ we define the corresponding *one-dimensional operator*

$$f' \otimes f \quad \text{by} \quad (f' \otimes f)(g) := \langle g, f' \rangle \cdot f$$

for all $g \in E$. Moreover we call a bounded linear operator P on E a *projection* if $P^2 = P$. In that case we have $P^2 = P$.

Proposition: For a projection P on a Banach space E the dual of PE is (as a topological vector space) isomorphic to the closed subspace $P'E'$ of E'.

Proof. The linear map $\Phi : E' \to (PE)'$ defined by $\Phi f' := f|_{PE}$ is surjective by the Hahn-Banach theorem. Therefore $(PE)'$ is isomorphic to $E'/\ker \Phi$. From $\ker \Phi = {P'}^{-1}(0)$ and $E' = P'E' \oplus {P'}^{-1}(0)$ we obtain $(PE)' \simeq E'/{P'}^{-1}(0) \simeq P'E'$. ∎

B.8. The space $\mathscr{L}(E)$ of all bounded linear operators on E becomes a Banach space if endowed with the *operator norm*

$$\|T\| := \sup\{\|Tf\| : \|f\| \leqslant 1\}.$$

But other topologies on $\mathscr{L}(E)$ will be used as well. We write $\mathscr{L}_s(E)$ if we endow $\mathscr{L}(E)$ with the *strong operator topology* i.e. with the topology of simple (= pointwise) convergence on E with respect to the norm topology. Therefore, a net $\{T_\alpha\}$ converges to T in the strong operator topology iff $T_\alpha \xrightarrow{\|\cdot\|} Tf$ for all $f \in E$. Observe that the strong operator topology is the topology on $\mathscr{L}(E)$ induced from the product topology on $(E, \|\cdot\|)^E$.

The *weak operator topology* on $\mathscr{L}(E)$ – write $\mathscr{L}_w(E)$ – is the topology of simple convergence on E with respect to $\sigma(E, E')$. Therefore,

$$\begin{aligned} &T_\alpha \text{ converges to } T \text{ in the weak operator topology} \\ \text{iff} \quad &\langle T_\alpha f, f' \rangle \to \langle Tf, f' \rangle \quad \text{for all } f \in E,\ f' \in E'. \end{aligned}$$

Again, this topology is the topology on $\mathscr{L}(E)$ inherited from the product topology on $(E, \sigma(E, E'))^E$.

B.9. Bounded subsets of $\mathscr{L}(E)$:
For $M \subseteq \mathscr{L}(E)$ the following are equivalent:
(a) M is bounded for the weak operator topology.
(b) M is bounded for the strong operator topology.
(c) M is uniformly bounded, i.e. $\sup\{\|T\| : T \in M\} < \infty$.
(d) M is equicontinuous for $\|\cdot\|$.

Proof. See Schaefer [1971], III.4.1, Corollary, and III.4.2 for (b) ⇔ (c) ⇔ (d); for (a) ⇔ (b) observe that the duals $\mathscr{L}_s(E)$ and $\mathscr{L}_w(E)$ are identical (Schaefer [1971], IV.4.3, Corollary 4). Consequently, the bounded subsets agree (Schaefer [1971], IV.3.2, Corollary 2). ∎

B.10. If M is a bounded subset of $\mathscr{L}(E)$, then the closure of M as subset of the product$(E, \|\cdot\|)^E$ is still contained in $\mathscr{L}(E)$ (Schaefer [1971], III.4.3).

B.11. On bounded subsets M of $\mathscr{L}(E)$, the topology of pointwise convergence on a total subset A of E coincides with the strong operator topology. Here we call A "total" if its linear hull is dense in E (Schaefer [1971], III.4.5).

The advantage of the strong, resp. weak, operator topology versus the norm topology on $\mathscr{L}(E)$ is that more subsets of $\mathscr{L}(E)$ become compact. Therefore, the following assertions (B.12)–(B.15) are of great importance.

B.12 Proposition:
For $M \subseteq \mathscr{L}(E)$, $g \in E$, we define the orbit $Mg : \{Tg : T \in M\} \subseteq E$, and the

subspaces $\qquad G_s := \{f \in E : Mf \text{ is relatively } \|\cdot\|\text{-compact}\}$

and $\qquad G_\sigma := \{f \in E : Mf \text{ is relatively } \sigma(E, E')\text{-compact}\}.$

If M is bounded, then G_s and G_σ are $\|\cdot\|$-closed in E

Proof. The assertion for G_s follows by a standard diagonal procedure. The argument for G_σ is more complicated: Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in G_σ converging to $f \in E$. By the theorem of Eberlein (Schaefer [1971], IV.11.2) it suffices to show that every sequence $(T_k f)_{k\in\mathbb{N}}$, $T_k \in M$ has a subsequence which converges weakly. Since $f_1 \in G_\sigma$ there is a subsequence $(T_{k_{i_1}} f_1)$ weakly converging to some $g_1 \in E$. Since $f_2 \in G_\sigma$, there exists a subsequence such that $(T_{k_{i_2}} f_2)$ such weakly converges to g_2, and so on. Applying a diagonal procedure we find a subsequence $(T_{k_i})_{i\in\mathbb{N}}$ of $(T_k)_{k\in\mathbb{N}}$ such that $T_{k_i} f_n \overset{i\to\infty}{\longrightarrow} g_n \in E$ weakly for every $n \in \mathbb{N}$. From

$$\begin{aligned}\|g_n - g_m\| &= \sup\bigl\{\langle g_n - g_m, f'\rangle : \|f'\| \leqslant 1\bigr\}\\ &= \sup\bigl\{\lim_{i\to\infty} |\langle T_{k_i} f_n - T_{k_i} f_m, f'\rangle| : \|f'\| \leqslant 1\bigr\}\\ &\leqslant \|T_{k_i}\| \cdot \|f_n - f_m\|\end{aligned}$$

it follows that $(g_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, and therefore converges to some $g \in E$. A standard 3ε-argument shows $T_{k_i} f \overset{i\to\infty}{\longrightarrow} g$ for $\sigma(E, E')$. ■

B.13 Proposition:
For a bounded subset $M \subseteq \mathscr{L}(E)$ the following are equivalent:

(a) M is relatively compact for the strong operator topology.
(b) Mf is relatively compact in E for every $f \in E$.
(c) Mf is relatively compact for every f in a total subset of E.

Proof. (a) $\Rightarrow$ (b) follows by the continuity of the mapping $T \mapsto Tf$ from $\mathscr{L}_s(E)$ into E.

(b) $\Leftrightarrow$ (c) follows from (B.12), and (c) $\Rightarrow$ (a) is a consequence of (A.3) and (B.10). ■

B.14 Proposition:
For a bounded subset $M \subseteq \mathscr{L}(E)$ the following are equivalent:

(a) M is relatively compact for the weak operator topology.
(b) Mf is relatively weakly compact for every $f \in E$.
(c) Mf is relatively weakly compact for every f in a total subset of E.

The proof follows as in (B.13).

B.15 Proposition:
Let $M \subseteq \mathscr{L}(E)$ be compact and choose a total subset $A \subseteq E$ and a $\sigma(E', E)$-total subset $A \subseteq E'$. Then the weak operator topology on M coincides with the topology of pointwise convergence on A and A'. In particular, M is metrizable if E is separable and E' is $\sigma(E', E)$-separable ("separable" means that there exists a countable dense set).

Proof. The semi-norms

$$P_{f,f'}(T) := |\langle Tf, f' \rangle|, \quad T \in M, f \in A, f' \in A'$$

define a Hausdorff topology on M coarser than the weak operator topology. Since M is compact, both topologies coincide (see A.2). ∎

B.16. Continuity of the multiplication in $\mathscr{L}(E)$:
In Lecture VII the *multiplication*

$$(S, T) \mapsto S \circ T$$

in $\mathscr{L}(E)$ plays an important role. Therefore, we state its continuity properties: The multiplication is jointly continuous on $\mathscr{L}(E)$ for the norm topology. In general, it is only separately continuous for the strong or the weak operator topology. However, it is jointly continuous on bounded subsets of $\mathscr{L}_s(E)$ (see Schaefer [1971], p. 183).

B.17. Spectral theory:
Let E be a complex Banach space and $T \in \mathscr{L}(E)$. The *resolvent set* $\rho(T)$ consists of all complex numbers λ for which the *resolvent* $R(\lambda, T) := (\lambda - T)^{-1}$ exists. The mapping $\lambda \mapsto R(\lambda, T)$ is holomorphic on $\rho(T)$. The *spectrum* $\sigma(T) := \mathbb{C}\backslash\rho(T)$ is a non-empty compact subset of $\mathbb{C}$, and two subsets of $\sigma(T)$ are of special interest: the *point spectrum*

$$P\sigma(T) := \{\lambda \in \sigma(T) : (\lambda - T) \text{ is not injective}\}$$

and the *approximate point spectrum*

$$A\sigma(T) := \{\lambda \imath \sigma(T) : (\lambda - T)f_n \to 0 \text{ for some normalized sequence } (f_n)\}.$$

A complex number λ is called an (approximate) eigenvalue if $\lambda \in P\sigma(A)$ (resp. $\lambda \in A\sigma(T)$), and $F_\lambda := \{f \in E : (\lambda - T) = 0\}$ is the eigenspace corresponding to the eigenvalue λ; λ is a simple eigenvalue if $\dim F_\lambda = 1$.

The real number $r(T) := \sup\{|\lambda| : \lambda \in \sigma(T)\}$ is called the *spectral radius* of T, and may be computed from the formula $r(T) = \lim_{n\to\infty} \left(\|T^n\|\right)^{\frac{1}{n}}$.

If $|\lambda| > r(T)$ the resolvent can be expressed by the Neumann series

$$R(\lambda, T) = \sum_{n=0}^{\infty} \lambda^{(n+1)} T^n.$$

For more information we refer to Schaefer [1971], App. 1 and Reed-Simon [1972].

B.18. The spaces $C(X)$ and their duals $M(X)$:
Let X be a compact space. The space $C(X)$ of all real (resp. complex) valued continuous functions on X becomes a Banach space if endowed with the norm

$$\|f\| := \sup\{|f(x)| : x \in X\}, \quad f \in C(X).$$

The dual of $C(X)$, denoted $M(X)$, is called the space of Radon measures on X. By the theorem of Riesz (Bauer [1972], 7.5) $M(X)$ is (isomorphic to) the set of all regular real-(resp. complex-)valued Borel measures on X (see A.12).

The Dirac measures δ_x, $x \in X$, defined by $\langle \delta_x, f\rangle := f(x)$ for all $f \in C(X)$, are elements of $M(X)$, and we obtain from Lebesgue's dominated convergence theorem (see A.16) the following:

If $f_n, f \in C(X)$ with $\|f_n\| \leqslant c$ for all $n \in \mathbb{N}$, then f_n converges to f for $\sigma(C(X), M(X))$ if and only if $\langle f_n, \delta_x\rangle \to \langle f, \delta_x\rangle$ for all $x \in X$.

B.19. Sequence spaces:
Let D be a set and take $1 \leqslant p < \infty$. The sequence space $\ell^p(D)$ is defined by

$$\ell^p(D) := \Big\{(x_d)_{d\in D} : \sum_{d\in D} |x_d|^p < \infty\Big\}$$

where x_D are real (or complex) numbers.

Analogously, we define

$$\ell^\infty(D) := \Big\{(x_d)_{d\in D} : \sup_{d\in D} |x_d| < \infty\Big\}.$$

The vector space $\ell^p(D)$, resp. $\ell^\infty(D)$, becomes a Banach space if endowed with the norm

$$\|(x_d)_{d\in D}\| := \Big(\sum_{d\in D} |x_d|^p\Big)^{1/p},$$

resp.

$$\|(x_d)_{d\in D}\| := \sup_{d\in D} |x_d|.$$

In our lectures, D equals $\mathbb{N}$, $\mathbb{N}_0$ or $\mathbb{Z}$. Instead of $\ell^p(D)$ we write ℓ^p if no confusion is possible.

B.20. The $L^p(X, \Sigma, \mu)$:
Let (X, Σ, μ) be a measure space and take $1 \leqslant p < \infty$. By $\mathscr{L}(X, \Sigma, \mu)$ we denote the vector space of all real- or complex-valued measurable functions on X with $\int_X |f|^p \,\mathrm{d}\mu < \infty$. Then

$$\|f\|_p := \Big(\int_X |f|^p \,\mathrm{d}\mu\Big)^{1/p}$$

is a semi-norm on $\mathscr{L}^p(X, \Sigma, \mu)$,

$$N_\mu := \big\{f \in \mathscr{L}^p(X, \Sigma, \mu) : \|f\|_p = 0\big\}$$

and is a closed subspace. The quotient space

$$L^p(X, \Sigma, \mu) = L^p(\mu) := \mathscr{L}^p(X, \Sigma, \mu)/N_\mu$$

endowed with the quotient norm is a Banach space. Analogously, one denotes by $\mathscr{L}^\infty(X, \Sigma, \mu)$ the vector space of μ-essentially bounded measurable functions on X. Again,

$$\|f\|_\infty := \big\{c \in \mathbb{R}^+ : \mu[|f| > c] = 0\big\}$$

yields a semi-norm on $\mathscr{L}^\infty(X, \Sigma, \mu)$ and the subspace

$$N_\mu := \left\{f \in \mathscr{L}^\infty(X, \Sigma, \mu) : \|f\|_\infty = 0\right\}$$

is closed. The quotient space

$$L^\infty(X, \Sigma, \mu) = L^\infty(\mu) := \mathscr{L}^\infty(X, \Sigma, \mu)/N_\mu$$

is a Banach space.

Even if the elements of $L^p(X, \Sigma, \mu)$ are equivalence classes of functions it generally causes no confusion if we calculate with the function $f \in \mathscr{L}^p(X, \Sigma, \mu)$ instead of its equivalence class $\check{f} \in L^p(X, \Sigma, \mu)$ (see II.D.4).

In addition, most operators used in ergodic theory are initially defined on the spaces $\mathscr{L}^p(X, \Sigma, \mu)$. However, if they leave invariant N_μ, we can and shall consider the induced operators on $L^p(X, \Sigma, \mu)$

B.21. For $1 \leqslant p < \infty$ the Banach space $L^p(X, \Sigma, \mu)$ is separable if and only if the measure algebra $\check{\Sigma}$ is separable.

B.22. If the measure space (X, Σ, μ) is finite, then

$$L^\infty(\mu) \subseteq L^{p_2}(\mu) \subseteq L^{p_1}(\mu) \subseteq L^1(\mu)$$

for $1 \leqslant p_1 \leqslant p_2 \leqslant \infty$.

B.23. Let (X, Σ, μ) be σ-finite. Then the dual of $L^p(X, \Sigma, \mu)$, $1 < p < \infty$ is isomorphic to $L^q(X, \Sigma, \mu)$ where $\frac{1}{p} + \frac{1}{q} = 1$, and the canonical bilinear form is given by

$$\langle f, g\rangle = \int f \cdot g \, \mathrm{d}\mu \quad \text{for } f \in L^p(\mu),\ g \in L^1(\mu).$$

Analogously, the dual of $L^1(\mu)$ is isomorphic to $L^\infty(\mu)$.

B.24. Conditional expectation:
Given a measure space (X, Σ, μ) and a sub-σ-algebra $\Sigma_0 \subseteq \Sigma$, we denote by J the canonical injection from $L^p(X, \Sigma_0, \mu)$ into $L^p(X, \Sigma, \mu)$ for $1 \leqslant p \leqslant \infty$. J is contractive and positive (see C.4). Its (pre-)adjoint

$$P : L^q(X, \Sigma, \mu) \to L^q(X, \Sigma_0, \mu)$$

is a positive contractive projection satisfying

$$P(f \cdot g) = g \cdot P(f) \quad \text{for } f \in L^q(X, \Sigma, \mu),\ g \in L^\infty(X, \Sigma_0, \mu).$$

Proof. P is positive and contractive since J enjoys the same properties. The above identity follows from

$$\langle P(fg), h\rangle = \langle fg, Jh\rangle = \int fgh \, \mathrm{d}\mu = \langle f, J(gh)\rangle\rangle = \langle (Pf)g, h\rangle$$

for all (real) $h \in L^p(X, \Sigma_0, \mu)$. ∎

We call P the *conditional expectation operator* corresponding to Σ_0. For its probabilistic interpretation see Ash [1972], Ch. 6.

B.25. Direct sums:
Let E_i, $i \in \mathbb{N}$, be Banach spaces with corresponding norms $\|\cdot\|_i$, and let $1 \leqslant p < \infty$. The ℓ^p-*direct sum* of $(E_i)_{i\in\mathbb{N}}$ is defined by

$$E := \bigoplus_p E_i := \Big\{(x_i)_{i\in\mathbb{N}} : x_i \in E_i \text{ for all } i \in \mathbb{N} \text{ and } \textstyle\sum_{i\in\mathbb{N}} \|x_i\|_i^p < \infty\Big\}.$$

E is a Banach space under the norm

$$\|(x_i)_{i\in\mathbb{N}}\| := \Big(\sum_{i\in\mathbb{N}} \|x_i\|_i^p\Big)^{1/p}$$

Given $S_i \in \mathscr{L}(E_i)$ with $\sup_{i\in\mathbb{N}} \|S_i\| < \infty$, then

$$\bigoplus S_i : (x_i)_{i\in\mathbb{N}} \mapsto (S_i x_i)_{i\in\mathbb{N}}$$

is a bounded linear operator on E with $\|\bigoplus S_i\| = \sup\{\|S_i\| : i \in \mathbb{N}\}$. Analogously one defines the ℓ^∞-direct sum $\bigoplus_\infty E_i$.

Appendix C. Remarks on Banach Lattices and Commutative Banach Algebras

(i) Banach lattices

A large part of ergodic theory, as presented in our lectures, takes place in the concrete function spaces as introduced in (B.18)–(B.20). But these spaces bear more structure than simply that of a Banach space. Above all it seems to us to be the order structure of these function spaces and the positivity of the operators under consideration which is decisive for ergodic theory. For the abstract theory of Banach lattices and positive operators we refer to the monograph of H.H. Schaefer [1974] where many of the methods we apply in concrete cases are developed. Again, for the readers convenience we collect some of the fundamental examples, definitions and results.

C.1. Order structure on function spaces:
Let E be one of the real function spaces $C(X)$ or$L^p(X, \Sigma, \mu)$, $1 \leqslant p < \infty$. Then we can transfer the order structure of $\mathbb{R}$ to E in the following way:

For $f, g \in E$ we call f *positive*, denoted $f \geqslant 0$, if $f(x) \geqslant 0$ for all $x \in X$, and define

$f \vee g$, the *supremum* of f and g, by $(f \vee g)(x) := \sup\{f(x), g(x)\}$ for all $x \in X$

$f \wedge g$, the *infimum* of f and g, by $(f \wedge g)(x) := \inf\{f(x), g(x)\}$, for all $x \in X$

$|f|$, the absolute value of f, by $|f|(x) := |f(x)|$ for all $x \in X$.

The new functions $f \vee g$, $f \wedge g$ and $|f|$ again are elements of E.

Remark that for $E = L^p(X, \Sigma, \mu)$ the above definitions make sense either by considering representatives of the equivalence classes or by performing the operations for μ-almost all $x \in X$.

Using the positive cone $E_+ := \{f \in E : f \geqslant 0\}$ we define an order relation on E by $f \geqslant g$ if $(g - f) \in E_+$. Then E becomes an ordered vector space which is a lattice for $\vee$ and $\wedge$.
Moreover, the norm of E is compatible with the lattice structure in the sense that $0 \leqslant f \leqslant g$ implies $\|f\| \leqslant \|g\|$, and $\||f|\| = \|f\|$ for every $f \in E$.

If we consider a *complex* function space E then the order relation "$\leqslant$" is defined only on the real part E_r consisting of all real valued functions in E. But the absolute value $|f|$ makes sense for all $f \in E$, and $\||f|\| = \|f\|$ holds.

C.2. A *Banach lattice* E is a real Banach space endowed with a vector ordering "$\leqslant$" making it into a vector lattice (i.e. $|f| = f \vee (-f)$ exists for every $f \in E$ and satisfying the compatibility condition:

$$|f| \leqslant g \quad \text{implies} \quad \|f\| \leqslant \|g\| \quad \text{for all } fg \in E.$$

Complex Banach lattices can be defined in a canonical way analogous to the complex function spaces in (C.1) (see Schaefer [1974], Ch.II,§11).

C.3. Let E be a Banach lattice. A subset A of E is called *order bounded* if A is contained in some *order interval* $[g, h] := \{f \in E : g \leqslant f \leqslant h\}$ for $g, h \in E$. The Banach lattice E is order complete if for every order bounded subset A the supremum $\sup A$ exists. Examples of order complete Banach lattices are the spaces $L^p(\mu)$, $1 \leqslant p \leqslant \infty$, while $C([0, 1])$ is not order complete.

C.4. Positive operators:
Let E, F be (real or complex) Banach lattices and $T : E \to F$ a continuous linear operator. T is *positive* if $TE_+ \subseteq F_+$, or equivalently, if $T|f| \geqslant |Tf|$ for all $f \in E$.

The morphisms for the vector lattice structure, called *lattice homomorphisms*, satisfy the stronger condition $T|f| = |Tf|$ for every $f \in E$.

If the norm on E is strictly monotone (i.e. $0 \leqslant f < g$ implies $\|f|| < \|g\|$); e.g. $E = L^p(\mu)$ for $1 \leqslant p < \infty$) then every positive isometry T on E is a lattice homomorphism. In fact, in that case $|Tf| \leqslant T|f|$ and $\||Tf|\| = \|Tf\| = \|f\| = \||f|\| = \|T|f|\|$ imply $|Tf| = T|f|$.

Finally, T is called *order continuous* (countably order continuous) if $\inf_{\alpha \ni A} Tx_\alpha = 0$ for every downward directed net (sequence) $(x_\alpha)_{\alpha \in A}$ with $\inf_{\alpha \ni A} x_\alpha = 0$.

C.5. Examples of positive operators are provided by positive matrices and integral operators with positive kernel (see Schaefer [1974], Ch. IV, §8).

Further, the multiplication operator

$$M_g : C(X) \to C(X) \quad (\text{resp. } L^p(X, \Sigma, \mu) \to L^p(X, \Sigma, \mu))$$

is a lattice homomorphism for every $0 \leqslant g \in C(X)$ (resp. $0 \leqslant g \in L^\infty(X, \Sigma, \mu)$).
The operators

$$T_\varphi : f \mapsto f \circ \varphi$$

induced in $C(X)$ or $L^p(X, \Sigma, \mu)$, $1 \leqslant p \leqslant \infty$, by suitable transformations

$$\varphi : X \to X$$

are even lattice homomorphisms (see II.4).

(ii) Commutative Banach algebras

While certainly order and positivity are more important for ergodic theory, in some places we use the multiplicative structure of certain function spaces.

C.6. Algebra structure on function spaces:
Let E be one of the complex function spaces $C(X)$ or $L^\infty(X, \Sigma, \mu)$. Then the multiplicative structure of $\mathbb{R}$ can be transferred to E: for $f, g \in E$ we define

$f \cdot g$, the *product* of f and g, by $(f - g)(x) := f(x) \cdot g(x)$ for all $x \in X$,

f^*, the *adjoint* of f, by $f^*(x) := \overline{f(x)}$ for all $x \in X$ where "—" denotes the complex conjugation.
The function R$\mathbf{1}$, defined by $\mathbf{1}(x) := 1$ for all $x \in X$, is the neutral element of the above commutative multiplication. The operation "$*$" is an *involution.*

C.7. A C^**-algebra* $\mathscr{A}$ is a complex Banach space and an algebra with involution $*$ satisfying

$$\|f \cdot f^*\| = \|f\|^2$$

for all $f \in \mathscr{A}$.
For our purposes we may restrict our attention to commutative C^*-algebras. As shown in (C.6) the function spaces $C(X)$ and $L^\infty(X, \Sigma, \mu)$ are commutative C^*-algebras. Another example is the sequence space ℓ^∞.

C.8. Multiplicative operators:
Let $\mathscr{A}_1$ and $\mathscr{A}_2$ be two C^*-algebras. The morphisms

$$T : \mathscr{A}_1 \to \mathscr{A}_2$$

corresponding to the C^*-algebra structure of $\mathscr{A}_1$ and $\mathscr{A}_2$ are continuous linear operators satisfying

$$T(f \cdot g) = (Tf) \cdot (Tg)$$

and

$$T(f^*) = (Tf)^*$$

for all $f, g \in \mathscr{A}$.

Let $\mathscr{A} = C(X)$, resp. $L^\infty(X, \Sigma, \mu)$. If $\varphi : X \to X$ is a continuous, resp. measurable, transformation, the induced operator

$$T_\varphi : f \mapsto f \circ \varphi$$

is a *multiplicative* operator on $\mathscr{A}$ satisfying $T_\varphi \mathbf{1} = \mathbf{1}$ and $T_\varphi f^* = (T_\varphi f)^a st$ (see II.4).

C.9. Representation theorem of Gelfand-Neumark:
Every commutative C^*-algebra $\mathscr{A}$ with unit is isomorphic to a space $C(X)$. Here X may be identified with the set of all non-zero multiplicative linear forms on $\mathscr{A}$, endowed with the weak* topology (see Sakai [1971], 1.2.1).

We remark that for $\mathscr{A} = \ell^\infty(\mathbb{N})$ the space X is homeomorphic to the Stone-Čech compactification $\beta\mathbb{N}$ of $\mathbb{N}$ (see Schaefer [1974], p. 106), and for $\mathscr{A} = L^\infty(Y, \Sigma, \mu)$, X may be identified with the Stone representation space of the measure algebra $\check{\Sigma}$ (see VI.D.6).

Appendix D. Remarks on Compact Commutative Groups

Important examples in ergodic theory are obtained by rotations on compact groups, in particular on the tori Γ^n. In our Lectures VII and VIII we use some facts about compact groups and character theory of locally compact abelian groups. Therefore, we mention the basic definitions and main results and refer to Hewitt-Ross [1979] for more information.

D.1. Topological groups:
A group $(G, \cdot)$ is called a *topological group* if it is a topological space and the mappings

$$(g, h) \mapsto g \cdot h \qquad \text{on } G \times G$$

and

$$g \mapsto g^{-1} \qquad \text{on } G$$

are continuous. A topological group is a *compact group* if G is compact. An isomorphism of topological groups is a group isomorphism which simultaneously is a homeomorphism.

D.2. The Haar measure:
Let G be a compact group. Then there exists a unique (right and left) invariant probability measure m on G, i.e. $= R_g' m = L_g' m$ for all $g \in G$ where R_g denotes the right rotation $R_g f(x) := f(xg)$, $x \in G$, $f \in C(G)$, and L_g the left rotation on $C(G)$.
m is called the *normalized Haar measure* on G.
The existence of Haar measure on compact groups can be proved using mean ergodic theory (e.g. (**??**.1) or Schaefer [1977], III.7.9, Corollary 1). For a more general and elementary proof see Hewitt-Ross [1979] 15.5–15.13.

D.3. Character group:
Let G be a locally compact abelian group. A continuous group homomorphism χ from G into the unit circle Γ is called a character of G. The set of all characters of G is called the *character group* or *dual group* of G, denoted by $\hat{G}$. Endowed with the pointwise multiplication and the compact-open topology G becomes a topological group which is commutative and locally compact (see Hewitt-Ross [1979], 23.15).

D.4 Proposition:
If G is a compact abelian group then $\hat{G}$ is discrete; and if $\hat{G}$ is a discrete abelian group, G is compact (see Hewitt-Ross [1979], 23.17).

D.5 Example: Let $\Gamma := \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle with multiplication and topology induced by $\mathbb{C}$. Then Γ is a compact group. Moreover, each character of Γ is of the form

$$z \mapsto z^n$$

for some $n \in \mathbb{Z}$, and therefore $\hat{\Gamma}$ is isomorphic to $\mathbb{Z}$. Finally, the normalized Haar measure is the normalized one-dimensional Lebesgue measure m on Γ.

D.6. Pontrjagin's duality theorem:
Let G be a locally compact abelian group, and denote by $\hat{\hat{G}}$ the dual group of $\hat{G}$.

$\hat{\hat{G}}$ is naturally isomorphic to G, where the isomorphism

$$\Phi : G \to \hat{\hat{G}}$$

is given by
$$g \mapsto \hat{g} \quad \text{with } \hat{g}(\chi) := \chi(g)$$

for all $\chi \in \hat{G}$ (see Hewitt-Ross [1979], 24.8).
In particular, this theorem asserts that a locally compact abelian group is uniquely determined by its dual.

D.7 Corollary:
The characters of a compact abelian group G form an orthonormal basis for $L^2(G, \mathcal{B}, m)$, $\mathcal{B}$ the Borel algebra and m the normalized Haar measure on G.

Proof. First, we prove the orthogonality by showing that $\int \chi(g)\,\mathrm{d}m(g) = 0$ for $\chi \neq 1$. Choose $h \in G$ with $\chi(h) \neq 1$. Then we have

$$\int \chi(g)\,\mathrm{d}m(g) = \int \chi(hg)\,\mathrm{d}m(g) = \chi(h) \int \chi(g)\,\mathrm{d}m(g)$$

and hence
$$\int \chi(g)\,\mathrm{d}m(g) = 0.$$

Clearly, every character is a normalized function in $L^2(G, \mathcal{B}, m)$. Let $g, h \in G$, and observe by (D.6) that there is a $\chi \in \hat{G}$ such that $\chi(g) \neq \chi(h)$, i.e. the characters separate the points of G. Therefore, the Stone-Weierstrass theorem implies that the algebra $\mathscr{A}$ generated by G, i.e. the vector space generated by $\hat{G}$, is dense in $C(G)$, and thus in $L^2(G, \mathcal{B}, m)$. ∎

We conclude this appendix with Kronecker's theorem which is useful for investigating rotations on the torus Γ^n. For elementary proofs see (III.8.iii) for $n = 1$ and Katznelson [1976], Ch. VI, 9.1 for general $n \in \mathbb{N}$. Our abstract proof follows Hewitt-Ross [1979], using duality theory.

D.8. Kronecker's theorem:
Let $a := (a_1, \ldots, a_n) \in \Gamma^n$ be such that $\{a_, \ldots, a_n\}$ linearly independent in the $\mathbb{Z}$-module Γ, i.e. $1 = a_1^{z_1} \ldots a_n^{z_n}$, $z_i \in \mathbb{Z}$ implies $z_i = 0$ for $i = 1, \ldots, n$. Then the subgroup $\{a^z : z \in \mathbb{Z}\}$ is dense in Γ^n.

Proof. Endow $\hat{\mathbb{Z}} = \Gamma$ with the discrete topology and form the dual group $\hat{\hat{\mathbb{Z}}}_d = \widehat{\Gamma_d}$. $\widehat{\Gamma_d}$ is a compact subgroup of the product Γ^Γ – note that here the compact-open topology on $\widehat{\Gamma_d}$ is the topology induced from the product Γ^Γ.
We consider the continuous monomorphism

$$\Phi : \mathbb{Z} \to \hat{\hat{\mathbb{Z}}}_d$$
$$z \mapsto \Phi(z) \quad \text{defined by } \Phi(z)(\gamma) := \gamma^z \text{ for all } \gamma \in \Gamma = \hat{\mathbb{Z}}.$$

Then the duality theorem yields that $\Phi(\mathbb{Z})$ is dense in $\hat{\hat{\mathbb{Z}}}_d$.
Now let $b := (b_1, \ldots, b_n) \in \Gamma^n$ and $\varepsilon > 0$. Since $\{a_1, \ldots, a_n\}$ is linearly independent in the $\mathbb{Z}$-module Γ there exists a $\mathbb{Z}$-linear mapping

$$\chi \in \widehat{\Gamma_d} \quad \text{with } \chi(a_i) = b_i \text{ for } i = 1, \ldots, n.$$

By definition of the product topology on Γ^Γ and by denseness of $\Phi(\mathbb{Z})$ in $\widehat{\Gamma_d}$ we obtain $z \in \mathbb{Z}$ such that

$$\left|a_i^z - b_i\right| = \left|\Phi(z)(a_i) - \chi(a_i)\right| < \varepsilon,$$

for $i = 1, \ldots, n$. ∎

Appendix E. Some Analytic Lemmas

Here, we prove some analytic lemmas which we use in the present lectures but don't prove there in order not to interrupt the main line of the arguments. First, we recall two definitions.

E.1 Definition:

1. A sequence $(x_n)_{n\in\mathbb{N}}$ of real (or complex) numbers is called Cesàro-summable if $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=0}^{n-1} x_i$ exists.
2. Let $(n_i)_{i\in\mathbb{N}}$ be a subsequence of $\mathbb{N}_0$. Then $(n_i)_{i\in\mathbb{N}}$ has *density* $s \in [0,1]$, denoted by $d((n_i)_{i\in\mathbb{N}}) = s$, if

$$\lim_{k\to\infty} \frac{1}{k}\big|\{n_i : i \in \mathbb{N}\} \cap \{0,1,\dots,k-1\}\big| = s$$

where $|\cdot|$ denotes the cardinality.

E.2 Lemma:
For $(x_n)_{n\in\mathbb{N}_0}$ the following conditions are equivalent:

(i) $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=0}^{n-1} |x_i| = 0$.

(ii) There exists a subsequence N of $\mathbb{N}_0$ with $d(N) = 1$ such that $\lim\limits_{\substack{n\in N\\ n\to\infty}} x_n = 0$.

Proof. We define $N_k := \{0,1,\dots,k-1\}$.
(i) $\Rightarrow$ (ii): Let $J_k := \{n \in \mathbb{N}_0 : |x_n| \geqslant \frac{1}{k}\}$, $k > 0$, and observe that $J_1 \subseteq J_2 \subseteq \cdots$. Since $\frac{1}{n}\sum\limits_{i=0}^{n-1} |x_i| \geqslant \frac{1}{n}\cdot\frac{1}{k}|J_k \cap N_n|$, each J_k has density 0. Therefore, we can choose integers $0 = n_0 < n_1 < n_2 < \cdots$ such that

$$\frac{1}{n}|J_{k+1} \cap N_n| < \frac{1}{k+1} \qquad \text{for } n \geqslant n_k.$$

Define $J := \bigcup_{k\in\mathbb{N}} (J_{k+1} \cap (N_{n_{k+1}} \backslash N_{n_k}))$ and show $d(J) = 0$.
Let $n_k \leqslant n < n_{k+1}$. Then, we obtain

$$J \cap N_n = (J \cap N_{n_k}) \cup (J \cap (N_n \backslash N_{n_k})) \subseteq (J_k \cap N_{n_k}) \cup (J_{k+1} \cap N_n),$$

and conclude that

$$\frac{1}{n}\big|J \cap N_n\big| \leqslant \frac{1}{k} + \frac{1}{k+1}.$$

If n tends to infinity, the same is true for k, and hence, J has density 0. Obviously, the sequence $N := \mathbb{N}\backslash J$ has the desired properties.

(ii) $\Rightarrow$ (i): Let $\varepsilon > 0$ and $c := \sup\{|x_n| : n \in \mathbb{N}_0\}$. Because of (ii) and $d(\mathbb{N}\backslash N) = 0$ there exists $n_\varepsilon \in \mathbb{N}$ such that $n \geqslant n_\varepsilon$ implies $|x_n| < \varepsilon$ for $n \in N$ and $\frac{1}{n}|(\mathbb{N}\backslash N) \cap N_n| <$

ε. If $n \geqslant n_\varepsilon$we conclude that

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{n-1}|x_i| &= \frac{1}{n}\sum_{i\in(\mathbb{N}\backslash N)\cap N_n}|x_i| + \frac{1}{n}\sum_{i\in N\cap N_n}|x_i|\\ &\leqslant \frac{c}{n}|(\mathbb{N}\backslash N)\cap N_n| + \varepsilon\\ &\leqslant (c+1)\cdot\varepsilon.\end{aligned}$$

■

E.3 Lemma:
Take a sequence $(z_n)_{n\in\mathbb{N}}$ of complex numbers such that

$$\sum_{n=1}^{\infty} n|z_{n+1}-z_n|^2 < \infty.$$

If $\lim\limits_{n\to\infty}\dfrac{1}{n}\sum\limits_{i=1}^{n} z_i = 0$, then $\lim_{n\to\infty} z_n = 0$.

Proof. Define $c_n := \sum_{k=n}^{\infty} k|z_{k+1}-z_k|^2$. Then

$$\max\{|z_{n+k}-z_n| : 1 \leqslant k \leqslant n-2\} \leqslant \sum_{k=n}^{2n-3}|z_{k+1}-z_k| \leqslant \Big(\sum_{k=n}^{2n-3}|z_{k+1}-z_k|^2(n-2)\Big)^{1/2} \leqslant c_n$$

and $$|z_n| = \Big|b_{n-1} - 2b_{2n-2} + \frac{1}{n-1}\sum_{k=1}^{n-2}(z_{n+k}-z_n)\Big| \quad \text{for } b_n := \frac{1}{n}\sum_{i=1}^{n} z_i.$$

■

E.4 Lemma:
Let N_i, $i = 1, 2, \ldots$ be a subsequence of $\mathbb{N}_0$ with density $d(N_i) = 1$. Then there exists a subsequence N of $\mathbb{N}_0$ such that $d(N) = 1$ and $N\backslash N_i$ is finite for every $i \in \mathbb{N}$.

Proof. There exists an increasing sequence $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ such that

$$1 - 2^{-i} \leqslant \frac{1}{k}\big|N_i \cap \{0,\ldots,k-1\}\big| \quad \text{for all } k \geqslant k_i.$$

If we define $N := \bigcap_{i\in\mathbb{N}} N_i \cup \{0,\ldots,k_i-1\}$, then N has the desired properties. ■

E.5 Lemma:
If $(x_n)_{n\in\mathbb{N}}$ is a sequence of psoitive reals satisfying $x_{n+m} \leqslant x_n + x_m$ for all $n, m \in \mathbb{N}$, then $\lim_{n\to\infty}\frac{x_n}{n}$ exists and equals $\inf_{n\in\mathbb{N}}\frac{x_n}{n}$.

Proof. Fix $n > 0$, and for $j > 0$ write $j = kn + m$ where $k \in \mathbb{N}_0$ and $0 \leqslant m < n$. Then

$$\frac{x_j}{j} = \frac{x_{kn+m}}{kn+m} \leqslant \frac{x_{kn}}{kn} + \frac{x_m}{kn} \leqslant \frac{kx_n}{kn} + \frac{x_m}{kn} = \frac{x_n}{n} + \frac{x_m}{kn}.$$

If $j \to \infty$ then $k \to \infty$, too, and we obtain

$$\limsup_{j\to\infty}\frac{x_j}{j} \leqslant \frac{x_n}{n}, \quad \text{and even} \quad \limsup_{j\to\infty}\frac{x_j}{j} \leqslant \inf_{n\in\mathbb{N}}\frac{x_n}{n},$$

On the other $\inf_{n\in\mathbb{N}}\frac{x_n}{n} \leqslant \liminf_{n\to\infty}\frac{x_n}{n}$, and the lemma is proved. ■

Appendix S. Invariant Measures

If $(X;\varphi)$ is a TDS it is important to know whether there exists a probability measure ν on X which is invariant under φ. Such an invariant measure allows the application of the measure-theoretical results in the topological context. It is even more important to obtain a φ-invariant measure on X which is equivalent to a particular probability measure (e.g. to the Lebesgue measure). The following two results show that the answer to the first question is always positive while the second property is equivalent to the mean ergodicity of some induced linear operator.

S.1 Theorem (Krylov-Bogoliubov, 1937):
Let X be compact and $\varphi : X \to X$ continuous. There exists a probability measure $\nu \in C(X)'$ which is φ-invariant.

Proof. Consider the induced operator $T := T_\varphi$ on $C(X)$. Its adjoint T' leaves invariant the weak*-compact set $\mathscr{P}$ of all probability measures in $M(X)$. If $\nu_0 \in \mathscr{P}$, then the sequence $\{T_n'\nu_0 : n \in \mathbb{N}\}$ has a weak$*$-accumulation point ν. It is easy to see (use IV.3.0) that $T'\nu = \nu$, i.e. ν is φ-invariant. ∎

As a consequence we observe that every TDS $(X;\varphi)$ may be converted into an MDS $(X,\mathcal{B},\mu;\varphi)$ where $\mathcal{B}$ is the Borel algebra and μ some φ-invariant probability measure. Moreover, the set $\mathscr{P}_\varphi$ of all φ-invariant measures in $\mathscr{P}$ is a convex $\sigma(C(X)',C(X))$-compact subset of $C(X)'$. Therefore, the Krein-Milman theorem yields many extreme points of $\mathscr{P}_\varphi$ called “ergodic measures”. The reason for that nomenclature lies in the following characterization.

S.2 Corollary:
Let $(X;\varphi)$ be a TDS. μ is an extreme point of $\mathscr{P}_\varphi$ if and only if $(X,\mathcal{B},\mu;\varphi)$ is an ergodic MDS.

Proof. If $(X,\mathcal{B},\mu;\varphi)$ is not ergodic there exists $A \in \mathcal{B}$, $0 < \mu(A) < 1$, such that $\varphi(A) = A$ and $\varphi(X\backslash A) = X\backslash A$. Define two different measures

$$\mu_1(B) := \frac{\mu(B \cap A)}{\mu(A)}$$
$$\mu_2(B) := \frac{\mu(B \cap (X\backslash A))}{\mu(X\backslash A)} \quad \text{for } B \in \mathcal{B}.$$

Clearly, $\mu = \mu(A)\cdot\mu_1 + (1-\mu(A))\cdot\mu_2$, and μ not an extreme point of $\mathscr{P}_\varphi$.

On the other hand, assume $(X,\mathcal{B},\mu;\varphi)$ to be ergodic. If $\mu = \frac{1}{2}(\mu_1+\mu_2)$ for $\mu_1,\mu_2 \in \mathscr{P}_\varphi$, then $\mu_1 \leqslant 2\mu$ and hence $\mu_1 \in L^1(\mu)' = L^\infty(\mu)$. But the fixed space of T_φ' in $L^\infty(\mu)$ contains μ and $\mu_1$and is one-dimensional by (IV.6), (IV.4.e) and (III.4). Therefore we conclude $\mu = \mu_1$, i.e. μ must be an extreme point of $\mathscr{P}_\varphi$. ∎

The question, whether there exist φ-invariant probability measures equivalent to some distinguished measure, is more difficult and will be converted into a “mean ergodic” problem.

S.3 Theorem:
Let μ be a strictly positive probability measure on some compact space X and let $\varphi : X \to X$ be Borel measurable and non-singular with respect to μ (i.e. $\mu(A) = 0$ implies $\mu(\varphi^{-1}(A)) = 0$ for $A \in \mathcal{B}$). The following conditions are equivalent:

(a) There exists a φ-invariant probability measure ν on X which is equivalent to μ.
(b) For the induced operator $T := T_\varphi$ on $L^\infty(X, \mathcal{B}, \mu)$ the Cesàro means T converge in the $\sigma(L^\infty, L^1)$-operator topology to some strictly positive projection $P \in \mathscr{L}(L^\infty(\mu))$, i.e. $Pf > 0$ for $0 < f \in L^\infty$.
(c) The pre-adjoint T' of $T = T_\varphi$ is mean ergodic on $L^1(\mu)$ and $T'u = u$ for some strictly positive $u \in L^1(\mu)$.

Proof. The assumptions on φ imply that $T = T_\varphi$ is a well-defined positive contraction on $L^\infty(\mu)$ having a pre-adjoint T' on $L^1(m)$ (see Schaefer [1974], III.9, Example 1).

(a) $\Rightarrow$ (c): By the Radon-Nikodym theorem the φ-invariant probability measure ν equivalent to μ corresponds to a normalized strictly positive T-invariant function $u \in L^1(\mu)$. But for such functions the order interval

$$[-u, u] := \{f \in L^1(\mu) : -u \leqslant f \leqslant u\}$$

is weakly compact and total in $L^1(\mu)$. Therefore $Tu = u$ implies the mean ergodicity of T as in (IV.6).

(c) implies (b) by a simple argument using duality theory.

(b) $\Rightarrow$ (a): The projection $P : L^\infty(\mu) \to L^\infty(\mu)$ satisfies $PT = TP = P$ and maps $L^\infty(\mu)$ onto the T-fixed space. Consider

$$\nu_0 := \mu \circ P$$

which is a strictly positive φ-invariant linear form on $L^\infty(\mu)$. Since the dual of $L^\infty(\mu)$ decomposes into the band $L^1(\mu)$ and its orthogonal band we may take ν as the band component of ν_0 in $L^1(\mu)$.

By Ando [1968], Lemma 1, ν is still strictly positive and hence defines a measure equivalent to μ. Moreover, $T'\nu$ is contained in $L^1(\mu)$ and dominated by ν_0, hence $T'\nu \leqslant \nu$. From $T\mathbf{1} = \mathbf{1}$ we conclude $T'\nu = \nu$ and that ν is φ invariant. Normalization of ν yields the desired probability measure. ∎

These abstract results are not only elegant and satisfying from a theoretical standpoint, they can also help to solve rather concrete problems:

Let $\varphi : [0,1] \to [0,1]$ be a transformation which is *piecewise* C^2, i.e. there is a finite partition of $[0,1]$ in intervals A_i such that φ can be extended continuously from the interior $\mathring{A}_i$ to the closure $\overline{A_i}$ and the resulting function φ_i is twice continuously differentiable on $\overline{A_i}$. Moreover we assume that the derivatives $\dot{\varphi}_i$ do not vanish on $\mathring{A}_i$, φ_i is increasing or decreasing.

In this case, φ is measurable and non-singular with respect to the Lebesgue measure m, and

$$Tf := f \circ \varphi$$

defines a positive contraction on $L^\infty([0,1], \mathcal{B}, m)$ satisfying $T\mathbf{1} = \mathbf{1}$ and having a pre-adjoint T' on $L^1(m)$.
As a consequence of this theorem, one concludes that φ possesses an invariant probability measure which is absolutely continuous with respect to m iff $\dim F(T') \geqslant 1$. In particular, this follows if T' is mean ergodic.

To find out under which conditions on φ this holds, we observe that the pre-adjoint T' can be written as

$$T'f(x) = \sum_i f \circ \varphi_i^{-1}(x)\sigma_i(x)\mathbf{1}_{B_i}(x),$$

where $B_i = \varphi_i(\overline{A_i})$ and σ_i is the absolute value of the derivative of φ_i^{-1}.

In fact: For every $x \in (0,1)$,

$$\int_0^x T'f\,\mathrm{d}m = \int_0^1 f \cdot \mathbf{1}_{(0,x)} \circ \varphi\,\mathrm{d}m = \int_{\varphi^{-1}(0,x)} f\,\mathrm{d}m.$$

Thus $T'f$ is the derivative $\dot{g}$ of the function $g(x) = \int_{\varphi^{-1}(0,x)} f\,\mathrm{d}m$.

If φ is piecewise C^2, we can calculate this derivative and obtain the above formula. Recall that the *variation* $v(f)$ of a function $f : [a,b] \to \mathbb{R}$ is defined as

$$v(f) := \sup_{n\in\mathbb{N}}\Big\{\sum_{j=1}^n |f(t_j) - f(t_{j-1})| : a = t_0 < t_1 \cdots < t_n = b\Big\}.$$

With this concept and using some elementary analysis, one proves that

$$(*) \qquad v(f \cdot g) \leqslant v(f)\|g\|_\infty + \int_a^b |f \cdot \dot{g}|\,\mathrm{d}m$$

if f is piecewise continuous and g continuously differentiable.

After these preparations we present the main result.

S.4 Proposition:
Let $\varphi : [0,1] \to [0,1]$ be piecewise C^2 such that

$$s := \inf\big\{|\dot{\varphi}(t)| : t \in (0,1) \text{ and } \varphi \text{ differentiable at } t\big\} > 1.$$

Then there exists a φ-invariant probability measure on $[0,1]$ which is absolutely continuous with respect to the Lebesgue measure m.

Proof. By (S.3) we have to show that the pre-adjoint T'_φ of T_φ is mean ergodic on $L^1(m)$. The first part of the proof is of a technical nature. Choose $n \in \mathbb{N}$ such that $s^n > 2$ and consider the map

$$\Phi := \varphi^n$$

which again is piecewise C^2. Clearly,

$$\inf\{|\dot{\Phi}(t)| : t \in (0,1) \text{ and } \Phi \text{ differentiable at } t\} \geqslant s^n > 2.$$

Now we estimate the variation $v(T'_\varphi f)$ for any piecewise continuous function $f : [0,1] \to \mathbb{R}$. To this purpose we need some constants determined by the function Φ. Take the partition of $[0,1]$ into intervals A_i corresponding to φ and write

$$T'_\Phi f(x) = \sum_{i=1}^m f \circ \Phi_i^{-1}(x)\sigma_i(x)\mathbf{1}_{B_i}(x)$$

where $B_i = \Phi_i(\overline{A_i})$ and $\sigma_i(x) = |(\dot{\Phi}_i^{-1})(x)|$.

1. For σ_i we have $\sigma_i(x) \leqslant s^{-n} \leqslant \frac{1}{2}$ for every $x \in B_i$.
2. Put $k := \max\big\{|\dot{\sigma}_i(x)| : x \in \overline{B_i};\ i = 1,\dots,m\big\} \cdot \max\big\{|\dot{\Phi}_i(x)| : x \in \overline{A_i};\ i = 1,\dots,m\big\}$.

3. For the interval $\overline{A_i} = [a_{i-1}, a_i]$ we estimate

$$\begin{aligned}|f(a_{i-1})| + |f(a_i)| &\leqslant 2\inf\{|f(x)| : x \in A_i\} + v(f|_{A_i}) \\ &\leqslant \frac{2}{m(A_i)} \int_{a_i} |f| \,\mathrm{d}m + v(f|_{A_i}) \\ &\leqslant 2h \int_{A_i} |f| \,\mathrm{d}m + +v(f|_{A_i})\end{aligned}$$

for $h := \max\{\frac{1}{m(A_i)} : i = 1, \ldots, m\}$.

Now, we can calculate:

$$\begin{aligned}v(T'_\Phi f) &\leqslant \sum_{i=1}^{m} v\big(f \circ \Phi_i^{-1}(x)\sigma_i(x) \cdot \mathbf{1}_{B_i}(x)\big) \\ &\leqslant \sum_{i=1}^{m} \Big(\|\sigma_i\|_\infty \cdot v\big(f \circ \Phi_i^{-1}(x) \cdot \mathbf{1}_{B_i}(x)\big) + \int_{B_i} \big|f \circ \Phi_i^{-1} \cdot \dot{\sigma}_i\big| \,\mathrm{d}m \Big) \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{(by inequality (}*\text{) above)} \\ &\leqslant \sum_{i=1}^{m} \Big(s^{-n}(|f(a_{i-1})| + |f(a_i)| + v(f|_{A_i})) + k \int_{B_i} |f \circ \Phi_i^{-1}| \cdot \sigma_i \,\mathrm{d}m \Big)\end{aligned}$$

$$\text{(since } \max\{|\dot{\Phi}_i(x)| : x \in \overline{A_i}; i = 1, \ldots, m\} = \min\{\sigma_i(x) : x \in B_i; i = 1, \ldots, m\}\text{)}$$

$$\begin{aligned}&\leqslant \sum_{i=1}^{m} \Big(s^{-n}(2h \int_{a_i} |f| \,\mathrm{d}m + 2v(f|_{A_i})) + k \int_{A_i} |f| \,\mathrm{d}m \Big) \\ &\leqslant (h + k)\|f\|_1 + 2s^{-n} v(f).\end{aligned}$$

Observing that $v(\mathbf{1}) = 0$ and ${T'_\Phi}^r \mathbf{1}$ is again piecewise continuous, we obtain by induction

$$v({T'_\Phi}^r \mathbf{1}) \leqslant (h + k) \sum_{i=0}^{r-1} (2s^{-n})^i \leqslant \frac{h + k}{1 - 2s^{-n}} \quad \text{for every } r \in \mathbb{N},$$

and therefore

$$\|{T'_\Phi}^r \mathbf{1}\|_\infty \leqslant \|{T'_\Phi}^r \mathbf{1}\|_1 + v({T'_\Phi}^r \mathbf{1}) \leqslant 1 + \frac{h + k}{1 - 2s^{-n}},$$

i.e. ${T'_\Phi}^r \mathbf{1} \leqslant M \cdot \mathbf{1}$ for $r \in \mathbb{N}$ and some $M > 0$. For the final conclusion the abstract mean ergodic theorem (IV.6) implies that T'_Φ is mean ergodic. Since $T'_\Phi = {T'_\varphi}^n$, the same is true for T'_φ by (IV.D.2). ∎

In conclusion, we present some examples showing the range of the above proposition.

S.5 Examples:

1. The transformation

$$\varphi(t) := \begin{cases} 2t & \text{for } 0 \leqslant t \leqslant \frac{1}{2} \\ 2 - 2t & \text{for } \frac{1}{2} \leqslant t \leqslant 1 \end{cases}$$

satisfies the assumptions of our proposition and has a φ-invariant measure. In fact, m itself is invariant.

2. For
$$\varphi(t) := \begin{cases} \frac{t}{1-t} & \text{for } 0 \leqslant t \leqslant \frac{1}{2} \\ 2t-1 & \text{for } \frac{1}{2} \leqslant t \leqslant 1 \end{cases}$$
The assumption $|\dot{\varphi}(t)| > 1$ is violated at $t = 0$. In fact, there is no φ-invariant and with respect to m absolutely continuous measure on $[0, 1]$, since ${T'_\varphi}^n f$ converges to 0 in measure for $f \in L^1(m)$ (see Lasota-Yorke [1973]).
3. For $\varphi(t) := 4t \cdot (1-t)$ is strongly violated, nevertheless there is a φ-invariant measure: Indeed, the equation $\int_{[0,x]} f \,\mathrm{d}m = \int_{\varphi^{-1}[0,x]} f \,\mathrm{d}m$ together with the plausible assumption that $f(t) = f(1-t)$ leads to
$$F(x) := \int_0^x f(t)\,\mathrm{d}t = 2 \cdot \int_0^{\frac{1}{2}-\frac{1}{2}\sqrt{1-x}} f(t)\,\mathrm{d}t = 2 \cdot F(\tfrac{1}{2} - \tfrac{1}{2}\sqrt{1-x}).$$
By substituting $x = \sin^2 \xi$ we obtain
$$F(\sin^2 \xi) = 2F(\tfrac{1}{2} - \tfrac{1}{2}\cos \xi) = 2F(\sin^2 \tfrac{\xi}{2})$$
which shows that $F(x) = \arcsin \sqrt{x}$ is a solution. Thus the function
$$f(x) = \frac{1}{2\sqrt{x(1-x)}}$$
yields a φ-invariant measure $f \cdot m$ on $[0, 1]$
4. Finally, $\varphi(t) := 2(t - 2^{-i}$ for $2^{-i} < t \leqslant 2^{1-i}$, $i \in \mathbb{N}$, has $\dot{\varphi}_i(t) = 2$, but infinitely many discontinuities. Again there exists no φ-invariant measure since ${T'_\varphi}^n f$ converges to zero in measure for $f \in L^1(m)$.

References: Ando [1968], Bowen [1979], Brunel [1970], Hajian-Ito [1967], Lasota [1980], Lasota-Yorke [1973], Neveu [1967], Oxtoby [1952], Pianigiani [1979], Takahashi [1971].

Appendix U. Dilation of Positive Operators

Appendix V. Akcoglu's Individual Ergodic Theorem

V.1.

V.2.

V.3.

V.4.

V.5.

V.6.

V.7.

V.8.

V.9.

V.10.

Appendix Y. Mean Ergodic Operator Semigroups

Y.1.

Y.2.

Y.3.

Y.4.

Y.5.

Y.6.

Y.7.

Y.8.

Y.9.

Y.10.

References

1. R. L. Adler, A. G. Konheim, and M. H. McAndrew, *Topological entropy*, Trans. Amer. Math. Soc. **114** (1965), 309–319. MR MR0175106 (30 #5291)
2. R. L. Adler, P. Shields, and M. Smorodinsky, *Irreducible Markov shifts*, Ann. Math. Statist. **43** (1972), 1027–1029. MR MR0307348 (46 #6468)
3. Roy L. Adler, *Ergodic and mixing properties of infinite memory channels*, Proc. Amer. Math. Soc. **12** (1961), 924–930. MR MR0136471 (24 #B2504)
4. Roy L. Adler and Benjamin Weiss, *Similarity of automorphisms of the torus*, Memoirs of the American Mathematical Society, No. 98, American Mathematical Society, Providence, R.I., 1970. MR MR0257315 (41 #1966)
5. M. Akcoglu and L. Sucheston, *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. MR MR0326433 (48 #4777)
6. M. A. Akcoglu, *Pointwise ergodic theorems*, Trans. Amer. Math. Soc. **125** (1966), 296–309. MR MR0230876 (37 #6434)
7. ———, *A pointwise ergodic theorem in L_p-spaces*, Canad. J. Math. **27** (1975), no. 5, 1075–1082. MR MR0396901 (53 #761)
8. M. A. Akcoglu and A. Brunel, *Contractions on L_1-spaces*, Trans. Amer. Math. Soc. **155** (1971), 315–325. MR MR0276439 (43 #2186)
9. M. A. Akcoglu, J. P. Huneke, and H. Rost, *A counter example to the Blum Hanson theorem in general spaces*, Pacific J. Math. **50** (1974), 305–308. MR MR0350455 (50 #2947)
10. M. A. Akcoglu and L. Sucheston, *Weak convergence of positive contractions implies strong convergence of averages*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **32** (1975), 139–145. MR MR0377559 (51 #13730)
11. ———, *Dilations of positive contractions on L_p spaces*, Canad. Math. Bull. **20** (1977), no. 3, 285–292. MR MR0458230 (56 #16433)
12. ———, *A ratio ergodic theorem for superadditive processes*, Z. Wahrsch. Verw. Gebiete **44** (1978), no. 4, 269–278. MR MR509202 (80h:28021)
13. Mustafa A. Akcoglu, *Positive contractions of L_1-spaces*, Math. Z. **143** (1975), 5–13. MR MR0379795 (52 #700)
14. ———, *Pointwise ergodic theorems in L_p spaces*, Ergodic theory (Proc. Conf., Math. Forschungsinst., Oberwolfach, 1978), Lecture Notes in Math., vol. 729, Springer, Berlin, 1979, pp. 13–15. MR MR550405 (82b:47004)
15. Mustafa A. Akcoglu and P. Ekkehard Kopp, *Construction of dilations of positive L_p-contractions*, Math. Z. **155** (1977), no. 2, 119–127. MR MR0448125 (56 #6435)
16. Mustafa A. Akcoglu and Louis Sucheston, *On convergence of iterates of positive contractions in L_p spaces*, J. Approximation Theory **13** (1975), 348–362, Collection of articles dedicated to G. G. Lorentz on the occasion of his sixty-fifth birthday, IV. MR MR0367151 (51 #3393)
17. L. Alaoglu and G. Birkhoff, *General ergodic theorems*, Ann. of Math. (2) **41** (1940), 293–309. MR MR0002026 (1,339a)
18. Erik M. Alfsen, *Compact convex sets and boundary integrals*, Springer-Verlag, New York, 1971, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 57. MR MR0445271 (56 #3615)
19. Steven Alpern, *New proofs that weak mixing is generic*, Invent. Math. **32** (1976), no. 3, 263–278. MR MR0402005 (53 #5828)
20. T. Andô, *Invariante Masse positiver Kontraktionen in $C(X)$*, Studia Math. **31** (1968), 173–187. MR MR0239046 (39 #405)
21. Hirotada Anzai, *Ergodic skew product transformations on the torus*, Osaka Math. J. **3** (1951), 83–99. MR MR0040594 (12,719d)
22. ———, *On an example of a measure preserving transformation which is not conjugate to its inverse*, Proc. Japan Acad. **27** (1951), 517–522. MR MR0047742 (13,924c)
23. Kazuo Anzai, *On a mean ergodic theorem for several operators*, Keio Engrg. Rep. **30** (1977), no. 7, 59–68. MR MR0487512 (58 #7137)
24. V. I. Arnol′d, *Gewöhnliche Differentialgleichungen*, Springer-Verlag, Berlin, 1980, Translated from the Russian by Brigitte Mai. MR MR569932 (81e:34001)

25. V. I. Arnol′d and A. Avez, *Ergodic problems of classical mechanics*, Translated from the French by A. Avez, W. A. Benjamin, Inc., New York-Amsterdam, 1968. MR MR0232910 (38 #1233)
26. Robert B. Ash, *Real analysis and probability*, Academic Press, New York, 1972, Probability and Mathematical Statistics, No. 11. MR MR0435320 (55 #8280)
27. ______, *Information theory*, Dover Publications Inc., New York, 1990, Corrected reprint of the 1965 original. MR MR1088248 (91k:94001)
28. Robert Atalla and Robert Sine, *Random Markov mappings*, Math. Ann. **221** (1976), no. 3, 195–199. MR MR0410420 (53 #14170)
29. Robert E. Atalla, *Markov operators and quasi-Stonian spaces*, Proc. Amer. Math. Soc. **82** (1981), no. 4, 613–618. MR MR614888 (83f:47009)
30. Wojciech Bartoszek, *Norm residuality of ergodic operators*, Bull. Acad. Polon. Sci. Sér. Sci. Math. **29** (1981), no. 3-4, 165–167. MR MR638759 (83a:47009)
31. Heinz Bauer, *Kennzeichnung kompakter Simplexe mit abgeschlossener Extremalpunktmenge*, Arch. Math. **14** (1963), 415–421. MR MR0164053 (29 #1352)
32. ______, *Probability theory and elements of measure theory*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1981, Second edition of the translation by R. B. Burckel from the third German edition, Probability and Mathematical Statistics. MR MR636091 (82k:60001)
33. A. Bellow and H. Furstenberg, *An application of number theory to ergodic theory and the construction of uniquely ergodic models*, Israel J. Math. **33** (1979), no. 3-4, 231–240 (1980), A collection of invited papers on ergodic theory. MR MR571532 (81f:28013)
34. Alexandra Bellow, *An L^p-inequality with application to ergodic theory*, Houston J. Math. **1** (1975), no. 1, 153–159. MR MR0385070 (52 #5940)
35. ______, *A problem in L^p-spaces*, Measure theory (Proc. Conf., Oberwolfach, 1975), Springer, Berlin, 1976, pp. 381–388. Lecture Notes in Math., Vol. 541. MR MR0450510 (56 #8804)
36. J. F. Berglund and K. H. Hofmann, *Compact semitopological semigroups and weakly almost periodic functions*, Lecture Notes in Mathematics, No. 42, Springer-Verlag, Berlin, 1967. MR MR0223483 (36 #6531)
37. J. F. Berglund, H. D. Junghenn, and P. Milnes, *Compact right topological semigroups and generalizations of almost periodicity*, Lecture Notes in Mathematics, vol. 663, Springer, Berlin, 1978. MR MR513591 (80c:22003)
38. Patrick Billingsley, *Ergodic theory and information*, John Wiley & Sons Inc., New York, 1965. MR MR0192027 (33 #254)
39. G. D. Birkhoff, *What is the ergodic theorem?*, Amer. Math. Monthly **49** (1942), 222–226. MR MR0006619 (4,15b)
40. J. R. Blum and D. L. Hanson, *On the mean ergodic theorem for subsequences*, Bull. Amer. Math. Soc. **66** (1960), 308–311. MR MR0118803 (22 #9572)
41. ______, *On the isomorphism problem for Bernoulli schemes*, Bull. Amer. Math. Soc. **69** (1963), 221–223. MR MR0143862 (26 #1412)
42. Heikki Bonsdorff, *Characterizations of uniform recurrence for general Markov chains*, Ann. Acad. Sci. Fenn. Ser. A I Math. Dissertationes (1980), no. 32, 49. MR MR599981 (82b:60086)
43. Rufus Bowen, *Entropy for group endomorphisms and homogeneous spaces*, Trans. Amer. Math. Soc. **153** (1971), 401–414. MR MR0274707 (43 #469)
44. ______, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Springer-Verlag, Berlin, 1975, Lecture Notes in Mathematics, Vol. 470. MR MR0442989 (56 #1364)
45. ______, *Invariant measures for Markov maps of the interval*, Comm. Math. Phys. **69** (1979), no. 1, 1–17, With an afterword by Roy L. Adler and additional comments by Caroline Series. MR MR547523 (81e:28010)
46. Ola Bratteli and Derek W. Robinson, *Operator algebras and quantum statistical mechanics. Vol. 1*, Springer-Verlag, New York, 1979, C^*- and W^*-algebras, algebras, symmetry groups, decomposition of states, Texts and Monographs in Physics. MR MR545651 (81a:46070)
47. Leo Breiman, *The individual ergodic theorem of information theory*, Ann. Math. Statist. **28** (1957), 809–811. MR MR0092710 (19,1148g)
48. ______, *On achieving channel capacity in finite-memory channels*, Illinois J. Math. **4** (1960), 246–252. MR MR0118574 (22 #9347)

49. A. L. Brown and A. Page, *Elements of functional analysis*, Van Nostrand Reinhold Company, London-New York-Toronto, Ont., 1970, The New University Mathematics Series. MR MR0358266 (50 #10732)
50. James R. Brown, *Approximation theorems for Markov operators*, Pacific J. Math. **16** (1966), 13–23. MR MR0192552 (33 #777)
51. ———, *Ergodic theory and topological dynamics*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976, Pure and Applied Mathematics, No. 70. MR MR0492177 (58 #11323)
52. A. Brunel, *New conditions for existence of invariant measures in ergodic theory.*, Contributions to Ergodic Theory and Probability (Proc. Conf., Ohio State Univ., Columbus, Ohio, 1970), Springer, Berlin, 1970, pp. 7–17. MR MR0268355 (42 #3253)
53. A. Brunel and D. Revuz, *Quelques applications probabilistes de la quasi-compacité*, Ann. Inst. H. Poincaré Sect. B (N.S.) **10** (1974), 301–337. MR MR0373008 (51 #9210)
54. Antoine Brunel, *Sur un lemme ergodique voisin du lemme de E. Hopf, et sur une de ses applications*, C. R. Acad. Sci. Paris **256** (1963), 5481–5484. MR MR0152633 (27 #2608)
55. Stephen G. Brush, *Milestones in mathematical physics. Proof of the impossibility of ergodic systems: the 1913 papers of Rosenthal and Plancherel*, Transport Theory Statist. Phys. **1** (1971), no. 4, 287–298. MR MR0409089 (53 #12851)
56. R. B. Burckel, *Weakly almost periodic functions on semigroups*, Gordon and Breach Science Publishers, New York, 1970. MR MR0263963 (41 #8562)
57. D. L. Burkholder, *Semi-Gaussian subspaces*, Trans. Amer. Math. Soc. **104** (1962), 123–131. MR MR0138986 (25 #2426)