# 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*

# Contents

## Appendices S — Z

# Preface

As a mathematical discipline, ergodic theory is only about fifty years old and its origin in statistical mechanics is still visible. On the other hand, it has grown rapidly and developped extensive ramifications into many well-established fields of modern mathematics. Mathematicians with quite different backgrounds and different aims have therefore become interested in ergodic theory. For a long time the best approach for a beginner was through P.R. Halmos' "Lectures on Ergodic Theory" published in 1956. Since then ergodic theory has made considerable progress, of which Peter Walters (1982) has recently given a fairly comprehensive account.

In the present book we develop the most important and basic results of modern ergodic theory within the more comprehensive framework of functional analysis. Methods of functional analysis often make it possible to formulate more general results, which elucidate structural similarities of problems arising in different branches of ergodic theory. The 13 Lectures together with the Discussions (and the Introductory Appendices if necessary) should provide a compact introduction into modern ergodic theory for the newcomer. The book will be, however, much easier to comprehend for students with a solid background in functional analysis.

These Lectures are an introduction to ergodic theory, not a monograph on the subject. Therefore, they contain, first of all, those definitions and theorems constituting the core of today's ergodic theory. In addition, one might expect some brief excursions into more specialized topics, especially where applications of more general functional analytic concepts are concerned. We hope to fulfill these promises, even if we sometimes (e.g. in Lecture VII on the theory of compact operator semigroups) seem to go beyond the preassigned limits. The reason is our appreciation of functional-analytic methods and our belief that large parts of ergodic theory, in particular the ergodic and mixing theorems, can be simplified and unified by these methods. Another consequence of this belief is our joint treatment of topological and measure-theoretical ergodic theory. We emphasize the methods common to both cases, and we hope that even our basic Lectures may thus occasionally be interesting to a specialist (e.g. Lecture IV., VIII. or XI.). Certainly, some important topics are missing: we do not give a proof of Ornstein's isomorphism theorem for Bernoulli shifts, for example, and do not touch on Sinai's work on the ergodicity of the billiard flow. In fact, the vast theory of dynamical systems on differentiable manifolds is completely absent.

The present book originated from lectures delivered by R.N. at the University of Tübingen (1977/78), at the Scuola Normale Superiore at Pisa (1978) and at a summer school at Cortona (1979). We have tried to give this book an organization which reflects the typical process of learning in advanced mathematics courses:

The core of the book consists of the 13 Lectures.

As a prerequisite to these Lectures, the reader should be familiar with topology, measure theory and functional analysis. When some of this knowledge is missing, it can be looked up in the Introductory Appendices.

We hope that the Discussion section following each Lecture helps the student to a better understanding of the concepts and problems that are sometimes only touched upon in the lecture itself. The Discussions also point to some relations of the lectures to other areas of mathematics.

For the specialist or the advanced student the Supplementary Appendices present in a concise fashion some more special topics, each of which could be the subject of a seminar.

## 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: "elementyra 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 ?"

<u>Naive answer</u>:

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.

<u>First etymological answer</u>:

εργοδης = d i f f i c u l t .

<u>Historical answer</u>:

| | | | | | |
|---|---|---|---|---|---|
| 1880 | - | Boltzmann, Maxwell | - | ergodic hypothesis |
| 1900 | - | Poincaré | - | recurrence theorem |
| 1931 | - | v. 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 branche 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: $\varphi$ preserves the (normalized) Lebesgue measure $\mu$ 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 $x \in X$.<br>Problem: Find $\lim f(\varphi^n(x))$! | |
| 1st objection: Time change is much faster than our observations. | Modified problem: Find the time mean $M_t f(x) := \lim \frac{1}{n} \sum_{i=0}^{n-1} f(\varphi^i(x))$ ! | |
| 2nd objection: In practice, it is impossible to determine the state 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 $M_t f(x) = \int f \, d\mu$ = space mean, which is independent of the state x.<br>"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. u. T. Ehrenfest 1911 ,p.30).

"Correct" etymological answer:

ἔργον = energy

-ώδης = -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 ἔργον (Arbeit) und ὁδός (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 näherer 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 théorie cinétique des gaz, où l'expérience suggère und tendence à l'"uniformité": 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 homogène."

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 autors).

Elementary mathematical answer:

Let $X$ be a set, $\varphi : X \to X$ a mapping. The induced operator $T_\varphi$ maps functions $f : X \to \mathbb{R}$ into $T_\varphi f := f \circ \varphi$. Ergodic theory investigates the asymptotic behavior of $\varphi^n$ and $T_\varphi^n$ 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 $\mu$-integrable functions on X.

## II. Dynamical Systems

Many of the answers presented in lecture I indicate that ergodic theory deals with pairs $(X;\varphi)$ where $X$ is a set whose points represent the "states" of a physical system while $\varphi$ 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 $\varphi$ 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 $(X,\Sigma,\mu)$ 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 a 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 sence: There exists a subset $X_o$ of $X$ with $\mu(X_o) = 1$ such that the restriction $\varphi_o : X_o \to X_o$ of $\varphi$ is bijective, and both $\varphi_o$ and its inverse are measurable and measure-preserving for the induced $\sigma$-algebra $\Sigma_o := \{A \cap X_o : A \in \Sigma\}$.

2. If $\varphi$ is bi-measure-preserving with respect to $\mu$, we call $\mu$ a $\varphi$-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, $\mathcal{B}$ its Borel algebra, and $m$ the normalized Lebesgue measure on $\Gamma$.
Choose $a \in \Gamma$ and define

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

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

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

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

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

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.

Tomorrow, the point $(x_n) = (\ldots\ldots x_{-1}, x_o, x_1, x_2, \ldots\ldots)$

will be $(x_{n+1}) = (\ldots x_{-1}, x_o, x_1, x_2, \ldots\ldots\ldots)$ ,

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

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

$(\hat{X}; \tau)$ is a TDS, and $\tau$ is called the (left)<u>shift</u> 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 $\sigma$-algebra $\hat{\Sigma}$ on $\hat{X}$, and that $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ is an MDS.

On $\hat{X}$ there are many $\tau$-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 $(\{0,1\}, \mathfrak{P}\{0,1\}, p)^{\mathbb{Z}}$ (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 $(X, \Sigma, p)$ be a probability space, where $X := \{0, \dots, k-1\}$, $k > 1$, is finite, $\Sigma$ the power set of $X$ and $p = (p_0, \dots, p_{k-1})$ a probability measure on $X$.
Take $\hat{X} := X^{\mathbb{Z}}$, the product $\sigma$-algebra $\hat{\Sigma}$ on $\hat{X}$, the product measure $\hat{\mu}$ and the shift $\tau$ on $X$. Then we obtain an MDS $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$, called the <u>Bernoulli shift</u> with distribution $p$ and denoted by <u>$B(p_0, \dots, p_{k-1})$</u>.

II. 4 <u>Induced operators</u>:

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 complexvalued) continuous functions on $X$ (see B.18).

Define the "induced operator"

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

It is easy to see that $T_\varphi$ is an isometric linear operator on $C(X)$, and hence $(C(X); T_\varphi)$ is an FDS.
Moreover, we observe that $T_\varphi$ 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 $T_\varphi$ is a $^*$-algebra isomorphism (see C.8).

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

Define

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

or more precisely: $T_\varphi \check{f} := \widecheck{f \circ \varphi}$ where $\check{f}$ denotes the equivalence class in $L^p(X, \Sigma, \mu)$ corresponding to the function $f$.
Again, the "induced operator" $T_\varphi$ is an isometric (resp. unitary) linear operator on $L^p(X, \Sigma, \mu)$ (resp. on $L^2(\mu)$) since $\varphi$ is measure-preserving, and hence $(L^p(X, \Sigma, \mu); T_\varphi)$ is an FDS. As above, $T_\varphi$ 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 $T_\varphi$ 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 VI.S.3). 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 $\mathbb{R}^k = C(\{0,\dots,k-1\})$ and $T$ is any $k\times k$-matrix. We single out a particular case or special interest in probability theory: Let $T$ be stochastic, i.e.

$T = (a_{ij})$ such that $0 \le a_{ij}$ and $\sum_{j=0}^{k-1} a_{ij} = 1$ for $i = 0, 1,\dots, k-1$.

Then $(E;T)$ is an FDS and $T\mathbb{1} = \mathbb{1}$ where $\mathbb{1} = (1,\dots,1)$.

The matrix $T$ has the following interpretation in probability theory. We consider $X = \{0,1,\dots,k-1\}$ as the "state space" of a certain system, and $T$ as a description of time evolution of the states in the following sense: $a_{ij}$ 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 $(X;\varphi)$).

II. 6 Markov shifts:

Let
$$T : \mathbb{R}^k \to \mathbb{R}^k$$
be a stochastic matrix $(a_{ij})$ as in (II.5).

Let $\mu = \begin{pmatrix} p_0 \\ \vdots \\ p_{k-1} \end{pmatrix}$ be an invariant probability vector, i.e.

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

and $\mu$ 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 $\mu$ 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

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

with the product $\sigma$-algebra $\hat{\Sigma}$.

For $0 \le l \le k-1$ $pr[x_o=l]$ denotes the probability that $x \in \hat{X}$ is in the state $l$ at time 0. We define

$$pr[x_o=l] := p_l ,$$
$$pr[x_o=l,\ x_1=m] := p_l\, a_{lm} ,$$
$$pr[x_o=l_o,\ x_1=l_1,\ \dots,\ x_t=l_t] := p_{l_o}\, a_{l_o l_1}\, a_{l_1 l_2} \cdots a_{l_{t-1} l_t} .$$

Moreover, since $\mu$ is invariant,

$$pr[x_1=l] = \sum_{i=0}^{k-1} pr[x_o=i,\ x_1=l] = \sum_{i=0}^{k-1} p_i a_{il} = p_l = pr[x_o=l] ,$$

$pr[x_t=l] = p_l = pr[x_o=l]$, and finally

(*) $pr[x_s=l_o,\ x_{s+1}=l_1,\ \dots,\ x_{s+t}=l_t] = p_{l_o}a_{l_o l_1}a_{l_1 l_2}\cdots a_{l_{t-1}l_t} =$
$pr[x_o=l_o,\ x_1=l_1,\dots,\ x_t=l_t]$ for any choice of $s\in\mathbb{Z}$, $t\in\mathbb{N}_o$ and $l_o,\ \dots,\ l_t \in \{0,\dots,k-1\}$.

The equation (*) gives a probability measure on each algebra $\mathfrak{F}_m := \{A\in\hat{\Sigma} : A = \bigcap_{i=-m}^{m} [x_i \in A_i],\ A_i \subset X\}$. By (A.17) this determines exactly one probability measure $\hat{\mu}$ on the product $\sigma$-algebra $\hat{\Sigma}$ on $\hat{X}$. This measure $\hat{\mu}$ is obviously invariant unter 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 $\mu$ 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 $\{0,1,\dots,k-1\}$, whereas in (II.6) the shift $\tau$ is defined on the state space $\{0,1,\dots,k-1\}^{\mathbb{Z}}$. We have refined (i.e. enlarged) the state space of (II.5) to make the model "deterministic".

An analogous construction can be earried 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 $B(p_o,\dots,p_{k-1})$ 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

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

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

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 $T_\varphi$ 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 $T_\varphi$ 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 $T_\varphi$ (like any positive linear operator) is uniquely determined by its restriction to this real part.

II. D.4 Null sets in $(X, \Sigma, \mu)$:

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 $x \in X \setminus A$, $\mu(A) = 0$.

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

are induced by a homeomorphism

$$\varphi : X \longrightarrow X$$

in the sense that $T = T_\varphi$ ?

A complete answer is given as follows.

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

(i) T is a lattice isomorphism satisfying $T \mathbb{1} = \mathbb{1}$.

(ii) T is an algebra isomorphism.

(iii) $T = T_\varphi$ for a (unique) homeomorphism $\varphi$ 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 $\varphi$ 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 <u>Which FDSs are MDSs</u> ?

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

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

are induced by a bi-measure-preserving transformation

$$\varphi : X \longrightarrow X,$$

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 <u>Discrete vs. continuous time</u>:

Applying $\varphi$ (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 $\varphi$ means advancing in time with a discrete time scale in steps of $\Delta t$. Intuitively it is more realistic to consider a continuous time scale, and in our mathematical model the transformation $\varphi$ and the group homomorphism

$$n \longmapsto \varphi^n$$

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

$$t \longmapsto \varphi_t$$

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

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

in the discrete model is replaced by

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

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 behaviour of the system as t tends to infinity.

II. D.8 <u>From a differential equation to a dynamical system</u>:

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

$$(*) \qquad \dot{x} = \frac{dx}{dt} = f(x)$$

(or in coordinates: $\dot{x}_i = f_i(x_1,\ldots,x_n)$, $i = 1,\ldots,n$).

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

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

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

<u>Examples</u>:

(i) Let $\Gamma^2 = \mathbb{R}^2/\mathbb{Z}^2$ be the 2-dimensional torus and let

$$\dot{x} = 1$$
$$\dot{y} = \alpha$$

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

$$\varphi_t\left(\binom{x}{y}\right) = \binom{(x+t) \mod 1}{(y+\alpha t) \mod 1} .$$

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

$$\dot{x} = F\left(\binom{x}{y}\right)$$
$$\dot{y} = \alpha\cdot F\left(\binom{x}{y}\right)$$

where $F$ is a $C^1$-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 nonautonomous 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 $\varphi_t$-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 $q = (q_1,\dots,q_n)$ (coordinates) and $p = (p_1,\dots,p_n)$ (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

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

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

$$\frac{\partial}{\partial q}\left(\frac{\partial H}{\partial p}\right) + \frac{\partial}{\partial p}\left(-\frac{\partial H}{\partial q}\right) = 0 .$$

Therefore, the measure $dq_1 \dots dq_n \cdot dp_1 \dots dp_n$ 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

$$\frac{dH}{dt} = \frac{\partial H}{\partial q}\dot{q} + \frac{\partial H}{\partial p}\dot{p} = \frac{\partial H}{\partial q}\frac{\partial H}{\partial p} + \frac{\partial H}{\partial p}\left(-\frac{\partial H}{\partial q}\right) = 0$$

i.e. H is a first integral of (✱✱) (conservation of energy!).
This means that $X_E := \{(q,p) \in \mathbb{R}^{2n} : H(p,q) = E\}$ 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-dimendional Lebesgue measure from the 2-dimendional Lebesgue measure. This induced measure is $(\varphi_t)$-invariant and finite, and we obtain "continuous measure-theoretical dynamical systems".

<u>Example</u> (linear harmonic oscillator) :
Let $X = \mathbb{R}^2$ and let $\binom{q}{p}$ 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

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

The system (* *) becomes

$$\dot{q} = p$$
$$\dot{p} = -q$$

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

$$\varphi_t\left(\binom{q}{p}\right) = \begin{pmatrix} \sqrt{p^2+q^2}\ \sin(t+\beta) \\ \sqrt{p^2+q^2}\ \cos(t+\beta) \end{pmatrix}$$

where $\beta \in [0,2\pi)$ is defined by $\sqrt{p^2+q^2} \cdot \sin\beta = q$ and $\sqrt{p^2+q^2}\cos\beta = p$.
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 $\left\{\binom{0}{0}\right\}$.
For $E > 0$ the $(\varphi_t)_{t \in \mathbb{R}}$-invariant manifold

$X_E := \left\{\binom{q}{p} \in \mathbb{R}^2 : H(p,q) = E\right\}$ 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.9 <u>Dilating an FDS to an MDS:</u>

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

$$\begin{array}{ccc} L^1(X,\Sigma,\mu) & \xrightarrow{\quad T^n \quad} & L^1(X,\Sigma,\mu) \\ J\big\downarrow & & \big\uparrow Q \\ L^1(\hat{X},\hat{\Sigma},\hat{\mu}) & \xrightarrow[\quad \hat{T}^n_{\varphi} \quad]{} & L^1(\hat{X},\hat{\Sigma},\hat{\mu}) \end{array}$$

commutes for all n = 0, 1, 2, .... ?

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 <u>lattice dilation</u> of $(L^1(X,\Sigma,\mu);T)$ if - in the diagram above - J is an isometric lattice homomorphism (with $J\,\mathbb{1} = \hat{\mathbb{1}}$), and Q is a positive contraction. From these requirements it follows that T has to be positive with $T\,\mathbb{1} = \mathbb{1}$ and $T'\,\mathbb{1} = \mathbb{1}$. 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, $A_{rec} := A \cap (\varphi^{-1}A \cup \varphi^{-2}A \cup \ldots.)$ is the set of all points of A which are recurrent to A.
If $B := A \cup \varphi^{-1}A \cup \varphi^{-2}A \cup \ldots$ we obtain $\varphi^{-1}B \subset B$ and $A \setminus A_{rec} = B \setminus \varphi^{-1}B$.
Since $\varphi$ is measure-preserving and $\mu$ finite, we conclude

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

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

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

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

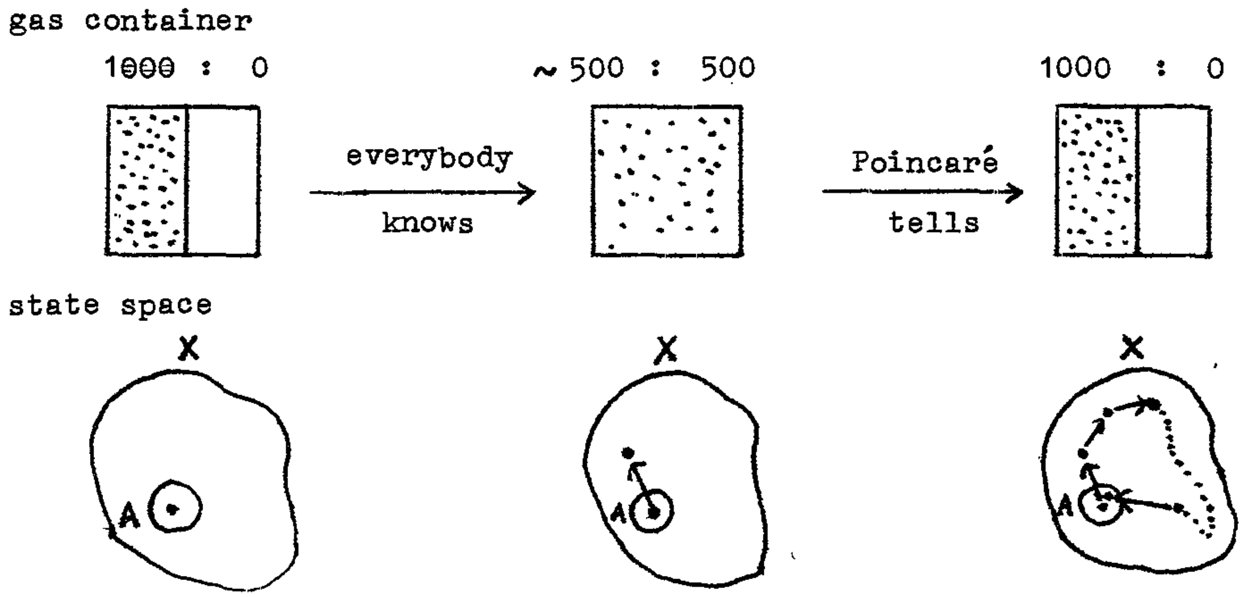
We explained in the physicist's answer in Lecture I that the dynamics can be described by the MDS $(X, \Sigma, \mu; \varphi)$ on the state space

$X := \{$coordinates of the possible locations and impulses of the 1000 molecules in the box$\} \subset \mathbb{R}^{6000}$ .

As the set A to which recurrence is expected we choose

$A := \{x \in X$: 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.




"Ergodic theory is the study of transformations from the point of view of <u>mixing properties</u>" (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 v.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 <u>Definition</u>:

An MDS $(X, \Sigma, \mu; \varphi)$ is called <u>ergodic</u> if there are no non-trivial $\varphi$-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 fayt 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 $T_\varphi : L^p(\mu) \longrightarrow L^p(\mu)$.

III. 4 <u>Proposition:</u> 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 $T_\varphi$ is one-dimensional, or: 1 is a simple eigenvalue of $T_\varphi$ in $L^p(\mu)$ for $1 \leq p \leq \infty$ .

Proof:

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

(b) $\Rightarrow$ (a): If $A \in \Sigma$ , $0 < \mu(A) < 1$, is $\varphi$-invariant, then $1\!\!1_A \in F$ and $\dim F \geq 2$.

(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 $\varphi$ -invariant, and hence trivial.

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

∎

III. 5 <u>Examples</u>:

(i) The <u>rotation</u> $(\Gamma, \mathcal{B}, m; \varphi_a)$ is ergodic iff $a \in \Gamma$ is not a root of unity:

If $a^{n_o} = 1$ for some $n_o \in \mathbb{N}$, then $1\!\!1$ and $f: z \to z^{n_o}$ are in F, and so

$\varphi_a$ 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 $f_n(z) = z^n$ form an orthonormal basis in $L^2(m)$ we obtain

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

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

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

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

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 $\varphi$-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 $\varphi$-invariant closed sets $A \subset X$, i.e. $\varphi(A) = A$, A closed, implies $A = \emptyset$ 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 $\varphi$-invariant, closed subset Y of X such that $(Y;\varphi|_Y)$ 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 $\varphi$-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 = const..

(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 <u>Examples</u>:

(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 $([0,1], \mathcal{B}, m; id)$ there exists no "minimal" invariant subset with positive measure.

(iii) The <u>rotation</u> $(\Gamma;\varphi_a)$ is minimal iff $a\in\Gamma$ is not a root of unity: If $a^{n_o} = 1$ for some $n_o\in\mathbb{N}$, then $\{z\in\Gamma: z^{n_o} = 1\}$ is closed and $\varphi_a$-invariant. For the other implication, we show that the orbit of every point in $\Gamma$ is dense. To do this we need only prove that $\{1, a, a^2, \dots\}$ is dense in $\Gamma$. Choose $\varepsilon > 0$. Since by assumption $a^{n_1} \neq a^{n_2}$ for $n_1 \neq n_2$, there exists $1 < k\in\mathbb{N}$ such that $|a^1 - a^k| < \varepsilon$. But $|a^1 - a^k| = |1 - a^{k-1}| = |a^{(k-1)n} - a^{(k-1)(n+1)}| < \varepsilon$ for all $n\in\mathbb{N}$. Since the set of "segments" $\{(a^{(k-1)n}, a^{(k-1)(n+1)}) : n\in\mathbb{N}\}$ covers $\Gamma$, we proved that there is at least one power of a in every $\varepsilon$-segment of $\Gamma$.

(iv) The <u>shift</u> $\tau$ on $\{0,1,\dots, k-1\}^{\mathbb{Z}}$ is <u>not</u> 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 $T_\varphi$ (see III.4 and III.7.i). In particular, it can be expected that the set $P\sigma(T_\varphi)$ of all eigenvalues of $T_\varphi$ 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 <u>Proposition</u>: 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 $T_\varphi$ on $C(X)$ (resp. $L^p(X,\Sigma,\mu)$, $1 \le p \le \infty$) is a subgroup of $\Gamma$, and each eigenvalue is simple.

Proof:

Since $T_\varphi$ is a bijective isometry the spectrum of $T_\varphi$ is contained in $\Gamma$.

Let $T_\varphi f = \lambda f$, $\|f\| = 1 = |\lambda|$. Since $T_\varphi$ is a lattice homomorphism we conclude

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

and hence $|f| = \mathbb{1}$ 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 $T_\varphi$ is also an algebra homomorphism 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 $\Gamma$.

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, $f\cdot g^{-1} = c\cdot\mathbb{1}$ or $f = c\cdot g$, i.e. each eigenvalue is simple. ∎

## III. D <u>Discussion</u>

### III. D.1 <u>The "original" Poincaré theorem:</u>

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 $\int dx_1 dx_2 dx_3$ 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 ces dernières trajectoires peuvent être regardées comme 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 $\varphi$. 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 <u>momentary</u> 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 <u>Counterexamples</u>: The recurrence theorem (III.2) is not valid without the assumption of <u>finite</u> measure spaces or <u>measure-preserving</u> 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

$$\varphi : x \mapsto x^2$$

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 <u>Recurrence in random literature</u>:

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 $(\hat{X},\hat{\Sigma},\hat{\mu};\tau)$ 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 $R_1,\dots,R_N$ of letters?

From

$$\hat{\mu}[\text{there exists } k\in\mathbb{Z} \text{ such that } x_{k+1}=R_1,\dots,x_{k+N}=R_N]$$
$$= 1-\hat{\mu}[\text{for every } k\in\mathbb{Z} \text{ there exists } i\in\{1,\dots,N\} \text{ such that } x_{k+i}\neq R_i]$$
$$\geq 1-\prod_{k=1}^{n}\hat{\mu}[\text{there exists } i\in\{1,\dots,N\} \text{ such that } x_{k+i}\neq R_i]$$
$$= 1-(1-p)^n \quad \text{for every } n\in\mathbb{N}$$

we conclude that this probability is 1.

Now consider $A := [x_1=R_1,\dots,x_N=R_N]$ having $\hat{\mu}(A) = p>0$. We have just shown that for almost every $x\in\hat{X}$ there is a number k such that $\tau^k(x)\in A$ for the shift $\tau$. 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 $B(\frac{1}{90},\dots,\frac{1}{90})$ the average distance between two occurrences of this book in random text is $1/p = 90^N$ 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 $A\subset X$ <u>$\varphi$-invariant</u> if $\varphi(A)\subset A$ a n d $\varphi^{-1}(A)\subset A$, i.e. $\varphi(A)=A$.

With this definition, a closed $\varphi$-invariant set $A\subset X$ 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 $\varphi(A)\subset A$ but $\varphi|_A$ is not a homeomorphism of A.

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

In agreement with the definition above we define the <u>orbit</u> 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 $\mathbb{Z}$

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

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

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

Since the family of mappings $\{\varphi^0, \varphi^1, \dots, \varphi^k\}$ is equicontinuous at x there is $\delta > 0$ such that

(ii) $\quad d(\varphi^t x, \varphi^t y) < \varepsilon$

if $t \in \{0,\dots,k\}$ and $d(x,y) < \delta$ .

The orbit of a is dense in X. Therefore, we find $r \in \mathbb{Z}$ with

(iii) $\quad d(x, \varphi^r a) < \delta$

and by (i) a suitable $t \in \{0,\dots,k\}$ with

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

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

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

∎

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" f in the "state" x by the space mean

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

i.e. the above limit has to be equal the constant function $(\int_X f \, d\mu) \cdot 1\!\!1$. Obviously the time mean is a $\varphi$-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 <u>Decomposition into ergodic components</u>:
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 $X := \{z \in \mathbb{C} : |z| \leq 1\}$, $\mathcal{B}$ the Borel algebra, $m$ the Lebesgue measure with $m(X) = 1$ and $\varphi_a$ the rotation

$$\varphi_a(z) := az$$

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 $0 \leq r \leq 1$ and $(X, \mathcal{B}, m; \varphi_a)$ is "determined" by these ergodic components. For more information we refer to v.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.i).

The fixed space of the induced operator $T_\varphi$ 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. $\varphi$ 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. $\varphi$ is topologically ergodic.
   b. For all A, B open, $A \neq \emptyset \neq B$ there is $k \in \mathbb{Z}$ such that $\varphi^k A \cap B \neq \emptyset$.

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 $T_\varphi$ in $C(X)$ lacks a certain convergence property which is automatically satisfied in $L^p(X, \Sigma, \mu)$, $1 \leq p < \infty$, (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

$$\varphi_a : z \mapsto az$$

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

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 $a = (a_1, \ldots, a_n) \in \Gamma^n$ the set $\{a^z: z \in \mathbb{Z}\}$ is dense in $\Gamma^n$ iff $\{a_1, \ldots, a_n\}$ is linearly independent in the $\mathbb{Z}$-module $\Gamma$ (see D.8). ∎

III. D.10 Ergodic vs. minimal:

Let $(X; \varphi)$ be a TDS and $\mu$ a $\varphi$-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 $\varphi$ 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 mor 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

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

where $A \subset X$ is closed, resp. measurable, ( Schaefer [1974],p.157).

Therefore, it is not difficult to see that an induced operator $T_\varphi$ on $C(X)$, resp. $L^1(X, \Sigma, \mu)$, is irreducible if and only if $(X; \varphi)$ is minimal, resp. if $(X, \Sigma, \mu; \varphi)$ is ergodic.

In contrast to minimal TDSs the ergodicity of an MDS $(X, \Sigma, \mu; \varphi)$ is characterized by the one-dimensionality of the $T_\varphi$-fixed space in $L^p(X, \Sigma, \mu)$, $1 \leq p < \infty$, (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 classicel concept of irreducible (= indecomposable) matrices (see IV.D.7 and Schaefer [1974],I.6).

Example: The matrix

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

of (II.6), Exercise is irreducible whereas the Bernoulli shift $B(p_o,\ldots,p_{k-1})$ is ergodic (see III.5.ii). This gives the impression that irreducibility is preserved under dilation (see App.U) at least in this example. In fact, this turns out to be true (App.U), 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 $T_\varphi$ 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.

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

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

III. D.12 <u>The origin of the word "Ergodic Theory"</u>:
In the last decades of the 19th 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 particularily 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.C. 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, ja 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 Atome 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]:

"... (Ich habe für derartige Inbegriffe von Systemen den Namen <u>Ergoden</u> vorgeschlagen.)..."

This may have induced P.and T.Ehrenfest to create the notion of "Ergodic Theory" by writing in "Begriffliche Grundlagen der atatistischen 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 <u>durch jeden Phasenpunkt hindurch</u>, 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 Mal 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 "<u>Ergode</u>" as a <u>special type of "Monode"</u>. 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 <u>Monoden</u> 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 Punkte 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 Fläche 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 $\varphi_1 = a_1, \dots,$ $\varphi_k = a_k$. 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 $\varphi_1 = a_1, \dots$ 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 $\varphi$, 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]; $\chi$, $\Phi$ mean the potential energy, $\psi$, $L$ 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 Ehrenfest 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 <u>asymptotic behavior</u>" (Walters [1975], p.1). Here, the asymptotic behavior of a transformation $\varphi$ 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 equal space mean" (see III.D.6) we investigate in this lecture the existence of the limit for $n \to \infty$ not of the powers $\varphi^n$ but of the "Cesaro 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 <u>Theorem</u> (J.v.Neumann, 1931):
Let $(X,\Sigma,\mu;\varphi)$ be an MDS and denote by $T_\varphi$ the induced (unitary) operator on $L^2(X,\Sigma,\mu)$. For any $f\in L^2(\mu)$ the sequence of functions

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

(norm-)converges to a $T_\varphi$ -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 <u>Definition</u>:
An FDS (E;T) (resp. a bounded linear operator T) is called <u>mean ergodic</u>, if the sequence

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

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

As above, the operators $T_n$ will be called the "Cesaro 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) $(Id - T)\, T_n = \frac{1}{n}(Id - T^n)$ for every $n \in \mathbb{N}$.

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

(1) $T\,P = P\,T = P = P^2$ .

(2) $PE = F := \{f \in E: Tf = f\}$ .

(3) $P^{-1}(0) = \overline{(Id - T)E}$ .

(4) The adjoints $T_n'$ converge to $P'$ in the weak* operator topology of $\mathcal{L}(E')$ 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, $(n+1)T_{n+1} - Id = nT_nT = nT\,T_n$ 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) $PE \subset F$ follows from $TP = P$, and $F \subset PE$ from $P = \lim_{n\to\infty} T_n$ .

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

$$f = f - Pf = f - PTf = \lim_{n\to\infty} (Id - T_nTf) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (Id - T^i)f$$

$$= \lim_{n\to\infty} (Id - T)\frac{1}{n} \sum_{i=1}^{n} iT_if \in \overline{(Id - T)E} \ .$$

(4) By the definition of the weak* operator topology, $T_n'$ converges to $P'$ if

$\langle T_nf, f' \rangle = \langle f, T_n'f' \rangle \rightarrow \langle f, P'f' \rangle = \langle Pf, f' \rangle$ 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 $\|T^n\| \leq c$ 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) $\{T_n f : n \in \mathbb{N}\}$ has a weak accumulation point for all $f \in E$.

(d) $\overline{\text{co}}\{T^i f : i \in \mathbb{N}_0\}$ 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 $\{T_n f : n \in \mathbb{N}\}$, i.e. $g \in \overline{\{T_n f : n > n_0\}}^{\sigma(E,E')}$ for all $n_0 \in \mathbb{N}$. Certainly, $g$ is contained in $\overline{\text{co}}\{T^i f : i \in \mathbb{N}_0\}$, and we shall show that $g$ is fixed under $T$: For any $n_0 \in \mathbb{N}$ we obtain
$g - Tg = (Id - T)g \in (Id - T)\ \overline{\{T_n f : n > n_0\}}^{\sigma} \subset \overline{\{(Id - T)T_n f : n > n_0\}}^{\sigma}$
$= \overline{\{\frac{1}{n}\ (Id - T^n)f : n > n_0\}}^{\sigma} \subset \frac{1}{n_0}(1 + c)\|f\|\ U$, where $U$ is the closed unit ball in $E$ - we used the fact that $(Id - T)$ 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 $\langle f, g'\rangle \neq \langle f, f'\rangle$. For all elements $f_0 \in \overline{\text{co}}\{T^i f : i \in \mathbb{N}_0\}$ 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 $f_1 \in \overline{\text{co}}\{T^i f : i \in \mathbb{N}\}$, 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

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

and assume that $f' \in E'$ vanishes on $G$. Since it vanishes on $(Id - T)E$

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 $f \in F \oplus (Id - T)E$, 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 $\|T^n\| \leq c$ 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 $T \in \mathcal{L}(E)$ be a contraction. Then T is mean ergodic and the corresponding projection P is orthogonal: By (IV.5) the Cesaro means $T_n$ of T converge to P and the Cesaro 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 $(T_n^*|g) \to (Qf|g)$ 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 $\|T\| = \|T^*\| \leq 1$, the relation $(f|f) = (Tf|f) = (f|T^*f)$ implies $(f|f) \leq \|f\| \cdot \|T^*f\| \leq \|f\|^2 = (f|f)$, hence $T^*f = f$. The other conclusion $F^* \subset F$ 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 $T_\varphi$ 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 $T_\varphi(fg) = T_\varphi f \cdot T_\varphi g = f \cdot T_\varphi g$. The same holds for $(T_\varphi)_n$, and therefore $P(fg) = f \cdot Pg$.

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

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

Proof:
The order interval $[-\mathbb{1}, \mathbb{1}] := \{f \in L^1(\mu) : -\mathbb{1} \leq f \leq \mathbb{1}\}$ 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 $\sigma(L^1, L^\infty)$ on $[-\mathbb{1}, \mathbb{1}]$ is coarser than that induced by $\sigma(L^\infty, L^1)$ - since $L^\infty(\mu) \subset L^1(\mu)$ - but still Hausdorff. Therefore the two topologies coincide (see A.2) and $[-\mathbb{1}, \mathbb{1}]$ is weakly compact. By assumption, $T$ and therefore the Cesaro means $T_n$ map $[-\mathbb{1}, \mathbb{1}]$ 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 weakly compact for some strictly positive function $u \in L^1(\mu)$. By [S], (II.8.8) it follows that $\bigcup_{n \in \mathbb{N}} \{g \in L^1(\mu) : 0 \leq g \leq T_n u\}$ is also relatively weakly compact. From $0 \leq T_n f \leq T_n u$ for $0 \leq f \leq u$, (B.14) and (IV.4.c) we conclude that $T$ is mean ergodic (see Ito [1965], Yeadon [1980]).

<u>Example 3</u>: Let $(X, \Sigma, \mu; \varphi)$ be an MDS. The induced operator $T_\varphi$ in $L^1(X, \Sigma, \mu)$ 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.

<u>Example 4:</u> 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 $\int_0^1 k(x,y)\,dy = 1$ for all $x \in [0,1]$. Then the kernel operator

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

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 $L^p(\mu)$, $1 \leq p < \infty$ .

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

(a) $\varphi$ is ergodic.

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

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

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

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

Proof:

(a) $\Rightarrow$ (b): Since $\varphi$ is ergodic and $T_\varphi$ is mean ergodic, the fixed spaces of $T_\varphi$ and $T_\varphi'$ are one-dimensional (III.4 and IV.4.e). Since P is a projection onto the $T_\varphi$-fixed space it must be of the form $f \mapsto Pf = \langle f, f' \rangle \, 1\!\!1$ for some $f' \in E'$. But
$\int_X f \, d\mu = \langle f, 1\!\!1 \rangle = \langle f, T_\varphi' 1\!\!1 \rangle = \langle f, P' 1\!\!1 \rangle = \langle Pf, 1\!\!1 \rangle = \langle f, f' \rangle \langle 1\!\!1, 1\!\!1 \rangle = \langle f, f' \rangle$
shows that $P = 1\!\!1 \otimes 1\!\!1$ .

(b) $\Rightarrow$ (c): Condition (c) just says that $\frac{1}{n} \sum_{i=0}^{n-1} T_\varphi^i$ converges toward $1\!\!1 \otimes 1\!\!1$ 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 $f = 1\!\!1_A$ and $g = 1\!\!1_B$. 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 $\Sigma$ .

The "automatic" mean ergodicity of $T_\varphi$ in $L^p(\mu)$, $1 \leq p < \infty$, (by Example 2 and 3) 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 $T 1\!\!1 \leq 1\!\!1$ for $p = 1$) are mean ergodic.

For operators on spaces $C(X)$ the situation is quite different and mean ergodicity of $T \in \mathcal{L}(C(X))$ 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 $T_\varphi$ 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 Cesaro means of $T_\varphi$ 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) $T_\varphi$ is mean ergodic in $C(X)$ and $\varphi$ is minimal.

(b) There exists a unique $\varphi$-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 $T_\varphi$ 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 $\nu$.

Define $Y := \bigcap \{[f = 0] : f > 0, \langle f, \nu \rangle = 0\}$. Then $Y \neq X$ is closed and $\varphi$-invariant, and therefore $Y = \emptyset$. This implies $\langle f, \nu \rangle > 0$ for every $0 < f \in C(X)$.

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

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

and $\langle 1\!\!1 , |f'| \rangle \leq \langle 1\!\!1 , T'_\varphi |f'| \rangle = \langle T_\varphi 1\!\!1 , |f'| \rangle = \langle 1\!\!1 , |f'| \rangle$.
Hence $\langle 1\!\!1 , T'_\varphi |f'| - |f'| \rangle = \langle T_\varphi 1\!\!1 , |f'| \rangle - \langle 1\!\!1 , |f'| \rangle = 0$, therefore $|f'|$ is $T'_\varphi$ -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 $\varphi$ -invariant probability measure $\nu$. Apply now (IV.4.e) to conclude that $T_\varphi$ is mean ergodic. Again the corresponding projection is of the form $P = \nu \otimes 1\!\!1$. Assume now that $Y \subsetneqq X$ is closed and $\varphi$ -invariant. There exists $0 < f \in C(X)$ with $f(Y) \subseteq \{0\}$, $T_\varphi f(Y) \subseteq \{0\}$ and therefore $Pf(Y) \subseteq \{0\}$. Hence $(\int_X f \, d\nu) \, 1\!\!1 (Y) \subseteq \{0\}$ and Y must be empty. ■

<u>Example 5</u>: The rotation $\varphi_a$ induces a mean ergodic operator $T_{\varphi_a}$ on $C(\Gamma)$: If $a^{n_0} = 1$ for some $n_0 \in \mathbb{N}$, the operator $T_{\varphi_a}$ is periodic (i.e. $T_{\varphi_a}^{n_0} = Id$) and therefore mean ergodic (see IV.D.3).
In the other case, every probability measure invariant under $\varphi_a$ is invariant under $\varphi_{a^n}$ 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

$$G \rightarrow \mathcal{L}_s(C(G)) : h \mapsto T_{\varphi_h} \qquad \text{(see II.2.ii)}$$

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

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

Exercise: The fixed space of $T_{\varphi_g}$ in C(G), where $\varphi_g$ is the rotation by g on the compact group G, is one-dimensional if and only if $\{g^z : z \in \mathbb{Z}\}$ 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 $T_{\varphi_a}$ is mean ergodic on the Banach space $R(\Gamma)$ of all bounded Riemann integrable functions on $\Gamma$ (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 $\chi$ of "segments" on $\Gamma$ and show that the Cesaro means

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

converge in sup-norm $\|\cdot\|_\infty$.

For $\varepsilon > 0$ choose $f_\varepsilon, g_\varepsilon \in C(\Gamma)$ such that

$$0 \le f_\varepsilon \le \chi \le g_\varepsilon$$

and $\int_\Gamma (g_\varepsilon - f_\varepsilon)\, dm < \varepsilon$, $m$ Lebesgue measure on $\Gamma$.

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

$$T_n g \xrightarrow{\|\ \|_\infty} \int_\Gamma g_\varepsilon\, dm \cdot 1\!\!1$$

and

$$T_n f \xrightarrow{\|\ \|_\infty} \int_\Gamma f_\varepsilon\, dm \cdot 1\!\!1 .$$

From $T_n f_\varepsilon \le T_n \chi \le T_n g_\varepsilon$ we conclude that $\|\ \|_\infty - \lim_{n\to\infty} T_n \chi$ exists and is equal to $\int_\Gamma \chi\, dm$.

Now, let f be a bounded Riemann integrable function on $\Gamma$. Then for every $\varepsilon > 0$ there exist functions $g_1, g_2$ being linear combinations of segments such that

$$g_1 \le f \le g_2 \quad \text{and} \quad \int_\Gamma (g_2 - g_1)\, dm < \varepsilon ,$$

and an easy calculation shows that

$$\|\cdot\|_\infty - \lim_{n\to\infty} T_n f = \left(\int_\Gamma f\, dm\right) \cdot 1\!\!1 .$$

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

The preceeding result is surprising, has interesting applications (see IV.D.5) and is optimal in a certain sense:

Example 6: The rotation operator $T_{\varphi_a}$ induced by $\varphi_a$, $a \in \Gamma$ 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 $\Gamma$ endowed with the sup-norm.

Proof:

(i) The rotation $\varphi_a$ is ergodic on $\Gamma$, 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 a 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 $TJ \subset J$ and $\mathbb{1} \notin J$. Consequently, there exists $\nu \in (L^\infty(m))'$ such that $\langle \mathbb{1}, \nu \rangle = 1$, but $\nu$ 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 $\nu_0$. As in (IV.4, c $\Rightarrow$ d) we show that $\nu_0 \in F'$. Since $\langle \mathbb{1}, \nu_0 \rangle = 1$ 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 Cesaro summable, i.e.

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

does not exist. The characteristic function $\chi$ 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 $T_\varphi$ in $L^p(X, \Sigma, \mu)$), while "mean ergodicity" is a convergence property of the Cesaro means of a linear operator on a Banach space. More systematically we agree on the following terminology: "Ergodicity" of a linear operator $T \in \mathcal{L}(E)$, E Banach space, refers to

the convergence of the Cesaro 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 $\{T^n: n \in \mathbb{N}_0\}$ 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 v.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 $\{T^n: n \in \mathbb{N}\}$ 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 $\{T^n: n \in \mathbb{N}_0\}$ : the closed convex hull

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

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

$$SP = PS = P$$

for all $S \in \overline{co}\,\{T^n: n \in \mathbb{N}_0\}$ (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 $S \in \mathcal{L}(E)$ 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 $P := (\frac{1}{k} \sum_{j=0}^{k-1} T^j)\, P_s$ and observe that $P \in \overline{co}\,\{T^i: i \in \mathbb{N}_0\}$ and $TP = (\frac{1}{k} \sum_{j=0}^{k-1} T^{j+1})\, P_s = P$ (use $T^k P_S = P_S$). Therefore, T is mean ergodic (see IV.4.d) and P is the projection corresponding to T.

On the contrary, it is possible that now 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 $l^\infty(\mathbb{N}_0)$ and take a 0 - 1-sequence $(a_n)_{n\in\mathbb{N}_0}$ which is not Cesaro summable.
For $k > 1$ we define elements $x_k \in l^\infty(\mathbb{N}_0)$:

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

Consider the closed S-invariant subspace E generated by $\{S^i x_k : i\in\mathbb{N}_0, k>1\}$ in $l^\infty(\mathbb{N}_0)$ and the restriction $T := S|_E$ .

By construction we obtain $\| T_n x_k \| \leq \frac{2}{n}$ for all $k > 1$. Consequently, T is mean ergodic with corresponding projection P = 0. On the other hand the sequence $(\frac{1}{m}\sum_{i=0}^{m-1} x_{k,ki})_{m\in\mathbb{N}} = (\frac{1}{m}\sum_{i=0}^{m-1} a_i)_{m\in\mathbb{N}}$ is not convergent for $k>1$, i.e. the Cesaro means $(T^k)_m(x_k)$ 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.

Reference: 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 $\|T\| \leq 1$. Express this fact in a less cumbersome way!

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

(a) T periodic with $T^{n_0} = Id$, $n_0 \in \mathbb{N}$, 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 weak 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 $l^p$, $1 \le p < \infty$.

(f) $Tf(x) = \int_0^x f(y)\,dy$ 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 $\| T_n 1\!\!1 \| = 1$ for all $n \in \mathbb{N}$.

(b) $Tf(x) = f(x^2)$ on $C([0,1])$ : $F = \langle 1\!\!1 \rangle$ but the Dirac measures $\delta_0$, $\delta_1$ 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 $0 \le (x_k) \in l^1$.

(d) $T(x_1,x_2,x_3,\dots) = (x_2,x_3,x_4,\dots)$ on $l^\infty$ : use a 0-1-sequence which is not Cesaro 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 [1971], 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 $\mathcal{L}(E)$.

Proof:
Take $T \in \mathcal{L}(E)$ such that $\| Id + T \| \le 1$ and $\| Id - T \| \le 1$. Then the same is true for the adjoints: $\| Id' + T' \| \le 1$ and $\| Id' - T' \| \le 1$.
For $f' \in E'$ define $f_1' := (Id' + T')f'$, resp. $f_2' := (Id' - T')f'$, and conclude $f' = \frac{1}{2}(f_1' + f_2')$ and $\| f_1' \|$, $\| f_2' \| \le \| f' \|$. As 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 $Id = \frac{1}{2}(R + S)$ for contractions $R, S \in \mathcal{L}(E)$, and define $T := Id - R$.
This implies $Id - T = R$ and $Id + T = 2\,Id - R = S$. By the above considerations it follows that $T = 0$, i.e. $Id = R = S$. ∎

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 $F(T) \subset F(R) \cap F(S)$ 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

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

and $R|_{F(T)}, S|_{F(T)} \in \mathcal{L}(F(T))$, since R and S commute.
Lemma 1 implies $R|_{F(T)} = S|_{F(T)} = Id_{F(T)}$, i.e. $F(T) \subset F(R) \cap F(S)$.

∎

Now we can prove the main result.

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

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

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 $SF(R) \subset F(R)$ 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 denoted 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 $\{ R_i : 1 \leq i \leq m \}$ 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

$$\|T^n\| \leq c \qquad \text{for all } n \in \mathbb{N}.$$

may be replaced by the weaker requirements

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

The following example (Sato [1977]) demonstrates that such situations may occur.

We define two sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ by

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

and

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

Endow

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

with the power set as $\sigma$-algebra $\Sigma$, and consider the measure $\nu$ defined by

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

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

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

The measurable (not measure-preserving!) transformation

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

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 $k = b_n, b_n+1, \dots, b_{n+1}-1$. This shows that $\sup\{\|T^k\| : k \in \mathbb{N}\} = \infty$ and $\lim_{k\to\infty} \frac{1}{k}\|T^k\| = 0$.

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

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

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 $\chi = 1\!\!1_A$ is the characteristic function of a subset $A \subset X$, 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 $\chi$ 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 $P = \mu \otimes 1\!\!1$, 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" $\varphi_a$ on $\Gamma$ and to the Banach space $R(\Gamma)$ of all bounded Riemann integrable functions on $\Gamma$ (see IV.D.0). Thus we obtain the following classical result on the equidistribution of sequences mod 1.

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

This theorem of 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 [1972b] 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 $T_\varphi$ corresponding to the transformation $\varphi$. 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 $E = L^p(X,\Sigma,\mu)$, $\mu(X) = 1$, $1 \le p < \infty$, and $T \in \mathcal{L}(E)$ is positive satisfying $T1\!\!1 = 1\!\!1$ and $T' 1\!\!1 = 1\!\!1$.

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

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

Remarks:

1. It is obvious that for an operator $T_\varphi$ induced by an MDS $(X, \Sigma, \mu; \varphi)$ irreducibility of $T_\varphi$ is equivalent to ergodicity of $\varphi$.
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 $1 \le k < n$. Then $T\mathbb{1}_A(x) = 0$ for $x \in X \setminus A$ means that the matrix associated with $T$ has the form

$$\overset{\hphantom{xx}k}{\left(\begin{array}{c:c} & \\ \hdashline 0 & \end{array}\right)}k \quad .$$

<u>Proposition</u>:
Let $(E;T)$ be an FDS formed by $E = L^p(X, \Sigma, \mu)$, $\mu(X) = 1$, $1 \le p < \infty$, and a positive operator $T$ satisfying $T\mathbb{1} = \mathbb{1}$ and $T'\mathbb{1} = \mathbb{1}$. 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. $F = \langle \mathbb{1} \rangle$.

(b) The corresponding mean ergodic projection has the form $P = \mathbb{1} \otimes \mathbb{1}$.

(c) $\langle T_n f, g \rangle$ converges to $\int_X f\, d\mu \cdot \int_X g\, d\mu$ for every $f \in L^p(\mu)$, $g \in L^q(\mu)$.

(d) $\langle T_n \mathbb{1}_A, \mathbb{1}_B \rangle$ converges to $\mu(A) \cdot \mu(B)$ for every $A, B \in \Sigma$.

(e) $\langle T_n \mathbb{1}_A, \mathbb{1}_A \rangle$ converges to $\mu(A)^2$ for every $A \in \Sigma$.

Proof:
Observe first that the assumptions $T\mathbb{1} = \mathbb{1}$ and $T'\mathbb{1} = \mathbb{1}$ 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 $\|T\| \le 1$. Consequently, $T$ is mean ergodic by (IV.5) or (IV.6).

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

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

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

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

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

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

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

The implications (a') $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) follow as in the proof of (IV.7).

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

$$< T_n \mathbb{1}_A, \mathbb{1}_A > \; = \; < T\mathbb{1}_A, \mathbb{1}_A > \; = \; < \mathbb{1}_A, \mathbb{1}_A > \; = \; \mu(A)$$

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 $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ (see II.6) with transition matrix $T = (a_{ij})$ and strictly positive invariant distribution $\mu = \begin{pmatrix} p_0 \\ \vdots \\ p_{k-1} \end{pmatrix}$ can be characterized by an elementary property of the kxk - matrix T.

Proposition: The following are equivalent:

(a) The transition matrix T is irreducible.

(b) The Markov shift $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ is ergodic.

Proof:

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

$$< \hat{T}_n \mathbb{1}_A, \mathbb{1}_B > \; \longrightarrow \; \hat{\mu}(A) \cdot \hat{\mu}(B)$$

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

$A = [x_{-\ell} = a_{-\ell}, \ldots\ldots, x_\ell = a_\ell]$
and $B = [x_{-m} = b_{-m}, \ldots\ldots, x_m = b_m]$
with $a_i, b_j \in \{0, \ldots, k-1\}$.
For $n \in \mathbb{N}$ so large that $n' := n - (m + \ell + 1) \geq 0$, we obtain

$$\hat{\mu}(\tau^n A \cap B) = \hat{\mu}[x_{-m} = b_{-m}, \ldots, x_m = b_m, x_{n-\ell} = a_{-\ell}, \ldots, x_{n+\ell} = a_\ell]$$

$$= \sum_{c_1=0}^{k-1} \cdots \sum_{c_n=0}^{k-1} \hat{\mu}[x_{-m}=b_{-m}, \ldots, x_m=b_m, x_{m+1}=c_1, \ldots, x_{m+n}=c_n, x_{n-\ell}=a_{-\ell}, \ldots, x_{n+\ell}=a_\ell]$$

$$= \sum_{c_1=0}^{k-1} \cdots \sum_{c_n=0}^{k-1} (p_{b_{-m}} \prod_{i=-m}^{m-1} t_{b_i b_{i+1}})\,(t_{b_m c_1} \prod_{i=1}^{n'-1} t_{c_i c_{i+1}}\; t_{c_n a_{-\ell}})\,(\prod_{i=-\ell}^{\ell-1} t_{a_i a_{i+1}})$$

$$= \hat{\mu}(B)\,(T^{n-m-\ell})_{b_m a_{-\ell}} \cdot (p_{a_{-\ell}})^{-1}\,\hat{\mu}(A).$$

Thus $\lim_{n\to\infty} \langle \hat{T}_n 1\!\!1_A, 1\!\!1_B \rangle = \hat{\mu}(B)\cdot(\lim_{n\to\infty} T_n)_{b_m a_{-\ell}} \cdot (p_{a_{-\ell}})^{-1} \cdot \hat{\mu}(A) = \hat{\mu}(B)\cdot\hat{\mu}(A)$,

iff $(\lim_{n\to\infty} T_n)_{ij} = (1\!\!1 \otimes 1\!\!1)_{ij} = p_j > 0$ 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 $\varphi$-invariant probability measure $\mu \in M(X)$.

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

$$(*) \quad k_{i-1} \text{ divides } k_i \text{ for all } i \in \mathbb{N}$$

and

$$(**) \quad \sum_{i=1}^{\infty} \frac{k_{i-1}}{k_i} \leq \frac{1}{12}.$$

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

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

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

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

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 $X := \overline{\{\tau^s a : s \in \mathbb{Z}\}} \subset \{0, 1\}^{\mathbb{Z}}$, 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 $\tau$-invariant and $(X;\tau|_X)$ is a TDS. The (product) topology on $\{0, 1\}^{\mathbb{Z}}$ - and on $X$ - is induced by the metric

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

The assertion is proved in several steps.

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

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

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 $f \in C(X)$ 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 $z$ ($1 \leq z \leq n$) for which $i(z)$ is odd. Consider $n = k_i$ and observe that the set $\{1,\dots, k_i\} \cap Z_j$ has exactly $\frac{k_i}{k_j}(2k_{j-1} + 1)$ elements for $j = 1,\dots,i$.

Now

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

i.e. $\{1,\dots, k_i\} \cap \bigcup_{j=1}^{i} Z_j$ contains at most $\frac{k_i}{4}$ numbers.

However $\{1,\dots, k_i\} \subset Z_{i+1}$, hence

$$\left|\{1,\dots, k_i\} \cap \left(Z_{i+1} \setminus \bigcup_{j=1}^{i} Z_j\right)\right| \geq \frac{3}{4} k_i,$$

and for all numbers in that intersection we have $i(z) = i + 1$.

In conclusion, one obtains

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

(iii) Using (IV.8) and (App.S), Theorem 1, we conclude from (ii) that there exist at least two different $\tau$-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

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

converge with respect to the $L^p$-norm for $1 \leq p < \infty$. Concerning the convergence for $L^\infty$-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 $T_\varphi$ on $L^\infty(\mu)$. This was shown in example 6 in Lecture IV for any ergodic rotation $\varphi_a$ on the unit circle $\Gamma$. On the other hand, in this same example there exist $T_\varphi$-invariant norm-closed subalgebras $\mathcal{A}$ of $L^\infty(\Gamma, \mathcal{B}, m)$ which are dense in $L^1(\Gamma, \mathcal{B}, m)$ and on which $T_\varphi$ becomes mean ergodic (e.g. take $\mathcal{A} = C(\Gamma)$ or even $R(\Gamma)$, see (IV.D.0). Such a subalgebra $\mathcal{A}$ is isomorphic to a space $C(Y)$ for some compact space $Y$ and the algebra isomorphism on $C(Y)$ corresponding to $T_\varphi$ is of the form $T_\psi$ 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 $T_\psi$ is mean ergodic with one-dimensional fixed space, and therefore it possesses a unique $\psi$-invariant, strictly positive probability measure $\nu$ (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 $\mathcal{A}$ in $L^1(\Gamma, \mathcal{B}, m)$

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 $\mathcal{A}$ can be chosen to be $C(\Gamma)$, but this choice is by no means unique and $\mathcal{A} = L^\infty(\Gamma, \mathcal{B}, m)$ 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 $T_\varphi$-invariant closed subalgebra $\mathcal{A}$ of $L^\infty(X, \Sigma, \mu)$ such that

(i) $T_\varphi$ is mean ergodic on $\mathcal{A}$, and

(ii) $\mathcal{A}$ 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) $T_\varphi$ 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), $T_\psi$ is mean ergodic with one-dimensional fixed space and strictly positive invariant linear form $\nu$. From (IV.8) it follows that $\psi$ has to be minimal, and hence $\{\psi^z(y) : z \in \mathbb{Z}\}$ is dense in Y for every $y \in Y$. The lemma in (VI.D.6) implies that $\{\psi^z(y) : z \in \mathbb{Z}\}$ and hence $\{y\}$ is not a null set for the measure corresponding to $\nu$. Therefore, $\{y\}$ must be open and the compact space Y is discrete. ∎

Having seen that $T_\varphi$ 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

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

is the largest subspace of $L^\infty(\mu)$ on which $T_\varphi$ is mean ergodic (use IV.3.0). Unfortunately, this subspace is "never" a subalgebra. More precisely:

Proposition: For any ergodic MDS $(X,\Sigma,\mu;\varphi)$ the following assertions are equivalent:

(a) $T_\varphi$ is mean ergodic on $L^\infty(\mu)$.

(b) $L^\infty(\mu)$ is finite-dimensional.

(c) $\langle 1\!\!1 \rangle \oplus \overline{(Id-T_\varphi)L^\infty(\mu)}$ 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 $T_\varphi$ is of the form $T_\psi : C(Y) \to C(Y)$ for some homeomorphism $\psi : Y \to Y$ and $\psi \neq id_Y$. Denote by Fix $(\psi)$ the fixed point set of $\psi$. Then every function $f \in \overline{(Id-T_\psi)C(Y)}$ vanishes on Fix $(\psi)$. Take $0 \neq g \in (Id-T_\psi)C(Y)$. Its square $g^2$ is contained in the subspace on which the means of $T_\psi^i$ converge and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} T_\psi{}^i g^2 = \left(\int_Y g^2\, d\nu\right) 1\!\!1_Y$$

for the strictly positive $\psi$-invariant measure $\nu$. Therefore Fix $(\psi)$ must be empty. It is now a simple application of Urysohn's lemma to show that $(Id-T_\psi)C(Y)$ separates the points in $Y$. By the Stone-Weierstrass theorem we obtain that $\langle 1\!\!1 \rangle \oplus (Id-T_\psi)C(Y)$ is dense in $C(Y)$ and therefore that $T_\varphi$ 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 $\langle 1\!\!1 \rangle \oplus \overline{(Id-T_\varphi)L^\infty(\mu)}$. 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 $T_\varphi$-invariant closed subalgebra $\mathcal{A}$ of $L^\infty(X,\Sigma,\mu)$, dense in $L^1(X,\Sigma,\mu)$, on which $T_\varphi$ is mean ergodic.

Applying an argument similar to that used in the proof of (IV.D.O) 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 MDS $(X, \Sigma, \mu; \varphi)$ 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, v.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

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

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

V. 1 <u>Theorem</u> (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 $\mu$-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 <u>Definition:</u>
Let $(X,\Sigma,\mu)$ be a measure space and consider $E = L^p(X,\Sigma,\mu), 1 \leq p \leq \infty$. $T \in \mathcal{L}(E)$ is called <u>individually ergodic</u> if for every $f \in E$ the Cesaro means $T_n f := \frac{1}{n}\sum_{i=0}^{n-1} T^i f$ converge $\mu$-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 $\check{T_n f}$, $n \in \mathbb{N}$, (see B.20) there exists a $\mu$-null set N such that $g_n(x)$ converge for any $x \in X \setminus N$. Only in (V.D.6) we shall see how a.e.-convergence of sequences in $L^p(\mu)$ can be defined without refering 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 Cesaro means $T_n$ 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 <u>Theorem</u> (Hopf, 1954; Dunford-Schwartz, 1956 ):

Let $(X, \Sigma, \mu)$ be a probability space, $E = L^1(X, \Sigma, \mu)$ and $T \in \mathcal{L}(E)$. If T is positive, $T\mathbb{1} \leq \mathbb{1}$ and $T'\mathbb{1} \leq \mathbb{1}$, then T is individually ergodic.

Remark:

The essential assumptions may also be stated as $\|T\|_\infty \leq 1$ and $\|T\|_1 \leq 1$ for the operator norms on $\mathcal{L}(L^\infty(\mu))$ and $\mathcal{L}(L^1(\mu))$.

The proof of the above "individual ergodic theorem" will not be easy, but it is presented along the lines indicated above.

V. 4 <u>Lemma</u>:

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

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

where F is the T-fixed space in $L^1(\mu)$.

We take $E_0 := F \oplus (Id - T) L^\infty(\mu)$. The convergence is obvious for $f \in F$. But for $f = (Id - T)g$, $g \in L^\infty(\mu)$, we obtain, using (IV.3.0), the positivity of T and $T1\!\!1 \leq 1\!\!1$, the estimate

$$|T_n f| = |(Id - T)T_n g| = \frac{1}{n}|(Id - T^n)g| \leq \frac{1}{n}(|g| + T^n|g|)$$
$$\leq \frac{1}{n}(\|g\|_\infty \cdot 1\!\!1 + \|g\|_\infty \cdot T^n 1\!\!1) \leq \frac{2}{n}\|g\|_\infty \cdot 1\!\!1 . \qquad \blacksquare$$

V. 5 <u>Lemma</u> (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

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

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

Proof (Garsia, 1965):

We keep f,n and $\gamma$ fixed and define

$$g := \sup\left\{\sum_{i=0}^{k-1}(T^i f - \gamma) \;:\; 1 \leq k \leq n\right\}.$$

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

$$\begin{aligned}
T(g^+) &\geq (Tg)^+ , \text{ since } 0 \leq T, \\
&\geq \sup\left\{\left(\sum_{i=0}^{k-1}(T^{i+1}f - \gamma T1\!\!1)\right)^+ : 1 \leq k \leq n\right\}, \text{ analogously}, \\
&\geq \sup\left\{\left(\sum_{i=0}^{k-1}(T^{i+1}f - \gamma 1\!\!1)\right)^+ : 1 \leq k \leq n\right\}, \text{ since } T1\!\!1 \leq 1\!\!1, \\
&\geq \sup\left\{\left(\sum_{i=0}^{k-1}(T^{i+1}f - \gamma 1\!\!1)\right)^+ : 1 \leq k \leq n-1\right\} \\
&= \sup\left\{\left(\sum_{i=0}^{k-1}(T^i f - \gamma 1\!\!1) - (f - \gamma 1\!\!1)\right)^+ : 2 \leq k \leq n\right\} \\
&\geq \sup\left\{\sum_{i=0}^{k-1}(T^i f - \gamma 1\!\!1) - (f - \gamma 1\!\!1) : 1 \leq k \leq n\right\} \\
&\geq g - (f - \gamma 1\!\!1).
\end{aligned}$$

This inequality yields

$$1\!\!1_A \cdot (f - \gamma 1\!\!1) \geq 1\!\!1_A \cdot g - 1\!\!1_A \cdot T(g^+) \geq g^+ - T(g^+) .$$

Finally the hypothesis $T' 1\!\!1 \leq 1\!\!1$ implies

$$\int_A (f - \gamma 1\!\!1) d\mu = \langle 1\!\!1_A \cdot (f - \gamma 1\!\!1), 1\!\!1 \rangle \geq \langle g^+ - T(g^+), 1\!\!1 \rangle = \langle g^+, 1\!\!1 \rangle - \langle g^+, T' 1\!\!1 \rangle \geq 0.$$

$\blacksquare$

Remarks:

1. $f^* := \sup\{T_k f : k \in \mathbb{N}\}$ is finite a.e., since $\mu[f^* > m] = \mu[\sup_{n\in\mathbb{N}} f_n^* > m]$
$= \sup_{n\in\mathbb{N}} \mu[f_n^* > m] \le \frac{\|f\|}{m}$ for every $m \in \mathbb{N}$, and therefore
$\mu(\bigcap_{m\in\mathbb{N}} [f^* > m]) = 0$ or $\mu[f^* < \infty] = \mu(\bigcup_{m\in\mathbb{N}} [f^* \le m]) = 1$.
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 $\|\;\|_1$.

V. 6 Proof of Theorem (V.3):

We take $0 \ne f \in L^1(\mu)$ and show that

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

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

$$\mu[h_f > \gamma\|f - f_o\|] = \mu[h_{f-f_o} > \gamma\|f - f_o\|] \le \mu[|(f - f_o)^*| > \frac{\gamma}{2}\|f - f_o\|]$$

$$\le \frac{2\|f - f_o\|}{\gamma\|f - f_o\|} = \frac{2}{\gamma}.$$

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

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

This shows that $h_f = 0$ a.e. ∎

Remark:

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

Since $L^2(X,\Sigma,\mu) \subset L^1(X,\Sigma,\mu)$ 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) $\varphi$ is ergodic.

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

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

Proof:

By (IV.7.b) the limit function $\overline{f}$ is the constant function $(1\!\!1 \otimes 1\!\!1)f = (\int_X f \, d\mu) \cdot 1\!\!1$. ∎

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 $L^p(X, \Sigma, \mu)$ is not a topological convergence, i.e. there exists no topology on $L^p(\mu)$ 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 $L^p(X,\Sigma,\mu)$, $1\leq p<\infty$, and consider

$$S^*f(x) := \sup_{n\in\mathbb{N}} |S_nf(x)| \quad \text{and}$$
$$G := \{f\in L^p : S_nf \text{ converges a.e.}\} .$$

If there exists a positive decreasing function

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

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

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

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 v.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

$$T : (x_n) \to (0, x_1, x_2, \ldots)$$

on $l^1(\mathbb{N}) = L^1(\mathbb{N},\Sigma,\mu)$ 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 $L^2([0, 1],\mathfrak{B},m)$, $m$ 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 $(X, \Sigma, \mu)$ the $L^P$-convergence and the a.e.-convergence imply the $\mu$-stochastic convergence (see App.A.16).
Therefore

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

for every $\varepsilon > 0$, $f \in L^P$, where $\bar{f}$ denotes the limit function of the Cesaro means $T_n f$ for a mean <u>or</u> individually ergodic operator $T \in \mathcal{L}(L^P(\mu))$.
In fact, even more is true.

<u>Theorem</u> (Krengel [1966]):
Let $(X, \Sigma, \mu)$ be a finite measure space and T be a positive contraction on $L^1(\mu)$. Then the Cesaro means $T_n f$ converge $\mu$-stochastically for every $f \in L^1(\mu)$.

V. D.3 <u>Strong law of large numbers (concrete example)</u>:
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". In we take $1\!\!1_A$ to be the characteristic function of the rectangle

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

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

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

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

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

for a.e. $x \in X$, 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

$$\xi = 0.x_1x_2x_3\ldots , \quad x_i \in \{0, 1, \ldots, 9\} ,$$

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 $\hat{X} := \{0, 1, \ldots, 9\}^{\mathbb{N}}$ which maps the Lebesgue measure onto the product measure $\hat{\mu}$ with

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

Consider the characteristic function $\chi$ 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

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

for almost every $x \in \hat{X}$. 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 Cesaro 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 $(f_n) \subset C(X)$ 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.

<u>Proposition:</u> For an operator $T \in \mathcal{L}(C(X))$ satisfying $\|T^n\| \leq c$ the following assertions are equivalent:

(a) For every $f \in C(X)$ the Cesaro means $T_n f$ converge pointwise to a function $\bar{f} \in C(X)$.

(b) T is mean ergodic.

V. D.6 <u>A.e.-convergence is order convergence</u>:

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.

<u>Lemma:</u>

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

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

The proof is a simple measure-theoretical argument. It is important that the "functions" $f_n$ 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 <u>purely</u> vector lattice theoretical approach to the individual ergodic theorem has yet to prove its significance.

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

V. D.7 <u>Strong law of large numbers (proof)</u>:

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 $(f_n)_{n \in \mathbb{N}_o}$ be a sequence of independent identically distributed integrable random variables. Then $\frac{1}{n}\sum_{i=o}^{n-1} f_i$ converge a.e. to the expected value $Ef_o$ .

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

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

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

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

Finally, f is integrable if $f \in L^1(\Omega, \mathfrak{A}, p)$, and its expected value is

$$Ef := \int_{\Omega} f(\omega)\ dp(\omega) = \int_{\mathbb{R}} t\ d(p \circ f^{-1})(t) .$$

Proof of the theorem:
Denote by $\mu$ the distribution of $f_n$, i.e.

$$\mu := p \circ f_o^{-1} = p \circ f_n^{-1} \quad \text{for every } n \in \mathbb{N}. \quad \text{Consider}$$

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

with the product measure $\hat{\mu}$ on the product $\sigma$-algebra $\hat{\Sigma}$ .
With the (left) shift

$$\tau : \ \hat{X} \to \hat{X}$$

we obtain an MDS $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ 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 $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ is ergodic, and the individual ergodic theorem implies

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

Next, denote the projections onto the i-th coordinate by

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

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

$$\frac{1}{n}\sum_{i=0}^{n-1} \pi_i \xrightarrow{\text{a.e.}} \int_{\hat{X}} \pi_o \, d\hat{\mu} = \int_{\mathbb{R}} t \, d\mu(t) = Ef_o .$$

In the final step we have to transfer the a.e.-convergence on $\hat{X}$ to the a.e.-convergence on $\Omega$.

The set of all finite products $\prod_{j\in J} g_j \circ \pi_j$ with $0 \le g_j \in L^1(\mathbb{R}, \mathcal{B}, \mu)$ is total in $L^1(\hat{X}, \hat{\Sigma}, \hat{\mu})$ by construction of the product $\sigma$-algebra. On these elements we define a mapping $\Phi$ by

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

From $\int_{\hat{X}} \Big(\prod_{j\in J} g_o \circ \pi_j\Big) d\hat{\mu} = \prod_{j\in J}\Big(\int_{\mathbb{R}} g_j d\mu\Big) = \prod_{j\in J}\Big(\int_{\Omega} g_j \circ f_j dp\Big)$

$= \int_{\Omega}\Big(\prod_{j\in J} g_j \circ f_j\Big) dp = \int_{\Omega} \Phi\Big(\prod_{j\in J} g_j \circ f_j\Big) dp$ it follows that $\Phi$ can be extended to a linear isometry

$$\Phi : L^1(\hat{X}, \hat{\Sigma}, \hat{\mu}) \longrightarrow L^1(\Omega, \mathcal{A}, p) .$$

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

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

converge a.e. to $\int_{\Omega} \Phi(\pi_o) \, dp = Ef_o$ . ∎

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 <u>Ergodic theorems for non-positive operators</u>:
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. $T \in \mathcal{L}(E)$ is called <u>regular</u> if T is the difference of two positive linear operators. In that case,

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

exists and the space $\mathcal{L}^r(E)$ 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 $\mathcal{L}^r(E) = \mathcal{L}(E)$ and $\|\cdot\| = \|\cdot\|_r$ (Schaefer 1974, IV.1.5). This yields an immediate extension of (V.3).

<u>Proposition 1</u>: Let $(X, \Sigma, \mu)$ be a probability space, $E = L^1(X, \Sigma, \mu)$ and $T \in \mathcal{L}(E)$.

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 $\pm T \leq |T|$ implies the analogous assertions for T, hence T is individually ergodic. ∎

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

<u>Proposition 2</u>: Every regular contraction T, i.e. $\|T\|_r \leq 1$, 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 <u>A non-commutative individual ergodic theorem</u>:

$L^\infty(X, \Sigma, \mu)$ 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.

<u>Theorem</u> (Lance, 1976; Kümmerer, 1978):

Let $\mathfrak{A}$ be a W*-algebra and $T \in \mathcal{L}(\mathfrak{A})$ a weak* continuous positive linear operator such that $T\mathbb{1} \leq \mathbb{1}$ and $T_* \mu \leq \mu$ for some faithful

( = strictly positive) state $\mu$ in the predual $\mathfrak{A}_*$.
Then the Cesaro means $T_n x$ converge almost uniformly to $\bar{x} \in \mathfrak{A}$ for every $x \in \mathfrak{A}$ , i.e. for every $\varepsilon > 0$ there exists a projection $p_\varepsilon \in \mathfrak{A}$ such that $\mu(p_\varepsilon) < \varepsilon$ and $\|(T_n x - \bar{x})(1\!\!1 - p_\varepsilon)\| \to 0$.

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 $\mu$-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

$$\check{\Sigma} := \Sigma/\mathcal{N} ,$$

where $\mathcal{N}$ is the $\sigma$-ideal of $\mu$-null sets in $\Sigma$ (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 $\varphi$ and defined by

$$\check{\varphi}\check{A} := (\varphi^{-1}A)^{\vee} \qquad \text{for } A \in \check{A} \in \check{\Sigma} .$$

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 <u>Definition</u>:

Two MDSs $(X, \Sigma, \mu; \varphi)$ and $(Y, \mathcal{T}, \nu; \psi)$ are called <u>isomorphic</u> if there exists a measure-preserving isomorphism $\check{\Theta}$ from $\check{\Sigma}$ onto $\check{\mathcal{T}}$ such that the diagram

$$\begin{array}{ccc} \check{\Sigma} & \xrightarrow{\check{\varphi}} & \check{\Sigma} \\ \check{\Theta} \downarrow & & \downarrow \check{\Theta} \\ \check{\mathcal{T}} & \xrightarrow[\check{\psi}]{} & \check{\mathcal{T}} \end{array}$$

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 isomorphism $\check{\varphi} : \check{\Sigma} \longrightarrow \check{\Sigma}$ is nothing else but the operator

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

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

$$T_\varphi \mathbb{1}_A = \mathbb{1}_{\varphi^{-1}(A)} \qquad \text{or} \qquad \check{T}_\varphi \mathbb{1}_{\check{A}} = \mathbb{1}_{\check{\varphi}\check{A}}$$

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 <u>Proposition</u>:

Two MDSs $(X, \Sigma, \mu; \varphi)$ and $(Y, \mathcal{T}, \nu; \psi)$ are isomorphic if and only if there exists a Banach lattice isomorphism

$$V : L^1(X, \Sigma, \mu) \longrightarrow L^1(Y, \mathcal{T}, \nu)$$

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

the diagram

$$\begin{array}{ccc} L^1(X, \Sigma, \mu) & \xrightarrow{T_\varphi} & L^1(X, \Sigma, \mu) \\ V \downarrow & & \downarrow V \\ L^1(Y, \mathcal{T}, \nu) & \xrightarrow[T_\psi]{} & L^1(Y, \mathcal{T}, \nu) \end{array}$$

commutes.

Proof:

The (equivalence classes of) characteristic functions $\chi$ are characterized by

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

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

$$\check{\theta} : \check{\Sigma} \longrightarrow \check{\mathcal{T}} \quad .$$

Conversely, every measure-preserving algebra isomorphism

$$\check{\theta} : \check{\Sigma} \to \check{T}$$

induces an isometry preserving the lattice operations from the sublattice 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

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

Since $\check{\theta}$ determines V, and $\check{\varphi}$, resp. $\check{\psi}$, determine $T_{\varphi}$, resp. $T_{\psi}$, (and conversely) the commutativity of one diagram implies the commutativity of the other. ∎

Remarks:

1. The isometric lattice isomorphism $V : L^1(X, \Sigma, \mu) \longrightarrow L^1(Y, T, \nu)$ in (VI.2) may be restricted to the corresponding $L^p$-spaces, $1 \leq p \leq \infty$ (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 $T_{\varphi}$ 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 <u>Definition</u>:

Two TDSs $(X; \varphi)$ and $(Y; \psi)$ are called <u>isomorphic</u> if there exists a homeomorphism $\theta : X \to Y$ such that the diagram

$$\begin{array}{ccc} X & \xrightarrow{\ \varphi\ } & X \\ \theta \downarrow & & \downarrow \theta \\ Y & \xrightarrow[\ \psi\ ]{} & Y \end{array}$$

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 <u>Remark</u> (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

$$V : L^2(X, \Sigma, \mu) \longrightarrow L^2(Y, \mathcal{T}, \nu)$$

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 $([0, 1], \mathfrak{B}, m; \mathrm{id})$ with Lebesgue measure $m$ and $(\mathbb{N}, \mathfrak{P}(\mathbb{N}), \nu; \mathrm{id})$, with $\nu(n) := 2^{-n}$. These two MDSs are spectrally isomorphic but not isomorphic. The reason is that $L^2([0, 1], \mathfrak{B}, m)$ is - as a Hilbert space - isomorphic to $l^2(\mathbb{N})$ but $(\mathbb{N}, \mathfrak{P}(\mathbb{N}), \nu; \mathrm{id})$, unlike $([0, 1], \mathfrak{B}, m; \mathrm{id})$, has minimal invariant sets with non-zero measure.

More important examples are the Bernoulli shifts $B(p_0, \dots, p_{k-1})$ which are all spectrally isomorphic (see VI.D.5) but not necessarily isomorphic (XII.D.1).
This again indicates that Hilbert spaces are insufficient for the purposes of ergodic theory.

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

$$\check{\Theta} : \check{\Sigma} \longrightarrow \check{\mathcal{T}}$$

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

$$\begin{array}{ccc} X & \xrightarrow{\ \varphi\ } & X \\ \Theta \downarrow & & \downarrow \Theta \\ Y & \xrightarrow[\ \psi\ ]{} & Y \end{array}$$

commutes.
This point isomorphism is stronger than isomorphism since $\Theta$ induces an algebra isomorphism

$$\check{\check{\Theta}} : \check{\mathcal{T}} \longrightarrow \check{\Sigma}$$

by $\check{\check{\Theta}}\check{B} := \widecheck{(\Theta^{-1}B)}$ for $B \in \mathcal{T}$. In fact, there exist MDSs which are

isomorphic but not pointwise isomorphic:
Take $(X, \Sigma, \mu; \varphi)$ with $X = \{x\}$, $\Sigma = \mathfrak{P}(X)$, $\mu(X) = 1$, $\varphi = \text{id}$ and $(Y, \mathcal{T}, \nu; \psi)$ with $Y = \{x, y\}$, $\mathcal{T} = \{\emptyset, Y\}$, $\nu(Y) = 1$, $\psi = \text{id}$.
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 $\varphi : X \to X$, 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 (v.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 VI.D.1).
Then the commutativity of the diagram in (VI.1) implies the commutativity of the corresponding diagram (VI.5) 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 <u>complete</u> 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 VIII and (XII.D.1)).

## VI. D <u>Discussion</u>

### VI. D.1 <u>Point vs. algebra isomorphism</u>:

The adequate isomorphisms in measure-theoretical ergodic theory are given by algebra isomorphisms; in most but not all cases these algebra isomorphisms are induced by point to point isomorphisms:
Take $(X, \Sigma, \mu; \varphi)$ as in (VII.5) and $(Y, \mathcal{T}, \nu; \psi)$ with $Y = \mathbb{R}$,

$$\mathcal{T} = \{ A \subset \mathbb{R}: A \text{ or } \mathbb{R} \setminus A \text{ is at most countable} \},$$

$$\nu(A) = \begin{cases} 0 & \text{if } A \text{ is at most countable} \\ 1 & \text{if } \mathbb{R} \setminus A \text{ is at most countable} \end{cases}, \text{ and } \psi = \text{id}.$$

Again, the two systems are algebra isomorphic, but there is no point isomorphism inducing this isomorphism.

The difference between the two isomorphism concepts and the difficulty in passing from one to the other can be easily explained to a functional-analyst familiar with the Gelfand - Neumark representation theorem and its consequence (II.D.5):
Let $(X, \Sigma, \mu)$ and $(Y, \mathcal{T}, \nu)$ be two (measure algebra) isomorphic probability spaces and denote by

$$V : L^\infty(X, \Sigma, \mu) \longrightarrow L^\infty(Y, \mathcal{T}, \nu)$$

the corresponding Banach lattice (resp. Banach algebra) isomorphism (see VI.2). Next we assume that $\mu$ is a regular Borel measure on the compact space X, so that there is a canonical Banach algebra morphism

$$i : C(X) \longrightarrow L^\infty(X, \Sigma, \mu) .$$

Finally we assume that $(Y, \mathcal{T}, \nu)$ possesses a <u>lifting</u>, i.e. there exists a Banach algebra morphism

$$L : L^\infty(Y, \mathcal{T}, \nu) \longrightarrow M^\infty(Y, \mathcal{T}, \nu) ,$$

where $M^\infty(Y,\mathcal{T},\nu)$ consists of all bounded functions in $\mathcal{L}^\infty(Y,\mathcal{T},\nu)$, which is a right inverse to the quotient map from $M^\infty(Y,\mathcal{T},\nu)$ into $L^\infty(Y,\mathcal{T},\nu)$.
The composition

$$j := L \circ V \circ i$$

is an algebra morphism from $C(X)$ into $M^\infty(Y,\mathcal{T},\nu)$, and its adjoint maps the set of multiplicative linear forms $\delta_y$, $y \in Y$, into the set of all Dirac measures $\delta_x$, $x \in X$. In this way we obtain a transformation $\theta : Y \to X$ which is measurable and induces $V$.

It is not our intention to prove these statements (see for example Ionescu Tulcea - Ionescu Tulcea [1969], ch.X), but we want to point out that the main ingredient implying coincidence of point and measure algebra isomorphism is the existence of a lifting. For most measure spaces (e.g. Borel measures on metric spaces) such liftings do exist. This justifies the tacit assumption in ergodic theory that point isomorphism equals algebra isomorphism.

VI. D.2 <u>The baker's transformation is a Bernoulli shift</u>:
As announced in (II.3) we show that the baker's transformation

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

on $([0,\ 1]^2, \mathcal{B}, m)$ is isomorphic to the Bernoulli shift $B(\frac{1}{2},\frac{1}{2})$ ($= (\hat{X},\hat{\Sigma},\hat{\mu};\tau)$ where $\hat{X} = [0,\ 1]^{\mathbb{Z}}$ and $\hat{\mu}$ the product measure defined by $(\frac{1}{2},\frac{1}{2})$ ):
Define the $\hat{\mu}$-null set $X$ in $\hat{X}$ by

$$X := \bigcup_{n\in\mathbb{N}} \{(x_k) \in \hat{X} : x_k = x_n \text{ for all } k \geq n\} \cup \bigcup_{m\in\mathbb{N}} \{(x_k) \in \hat{X} : x_{-k} = x_m \text{ for all } k \geq m\},$$

and analogously the $m$-null set $Y \subset [0,\ 1]^2$ by

$$Y := \{n\cdot 2^{-k} : k \in \mathbb{N};\ n = 0,\dots,2^k\} \times [0,1] \cup [0,\ 1] \times \{n\cdot 2^{-k} : k \in \mathbb{N}; n=0,\dots,2^k\}.$$

Now, the mapping

$$\theta : \hat{X} \setminus X \to [0,\ 1]^2 \setminus Y \quad \text{defined by}$$

$$(x_k)_{k\in\mathbb{Z}} \mapsto \left( \sum_{k=0}^{\infty} 2^{-k-1} x_k\ ,\ \sum_{k=1}^{\infty} 2^{-k} x_{-k} \right)$$

is bijective and bi-measure-preserving.
Finally, it is not difficult to check that $\theta \circ \tau = \varphi \circ \theta$ , and therefore the two dynamical systems are isomorphic.

VI. D.3 Examples of isomorphism invariants:

In most cases, isomorphism invariants are real or complex numbers. For MDSs the so-called "spectral invariants" are the most obvious. These are isomorphism invariants which are even preserved by Hilbert space isomorphisms as in (IV.4). In particular we mention the set of all eigenvalues, the point spectrum

$$P\,\sigma(T_\varphi)$$

of the induced operator $T_\varphi$ in $L^2(X,\Sigma,\mu)$. Similarily, the whole spectrum $\sigma(T_\varphi)$ is an isomorphism invariant. Unfortunately, this invariant is not very useful:

Proposition: Let $(X,\Sigma,\mu;\varphi)$ be an ergodic MDS such that $L^2(X,\Sigma,\mu)$ is not finite dimensional. Then $\sigma(T_\varphi) = \Gamma$.

Proof:

In this proof we restrict ourselves to the case $X = [0, 1]$, $\Sigma = \mathcal{B}$ the Borel algebra and $m$ the Lebesgue measure on X. Since $T_\varphi$ is unitary it suffices to thow that $\Gamma \subset \sigma(T_\varphi)$. Take $\lambda \in \Gamma$. By (X.D.4) for every $n \in \mathbb{N}$ there exists a measurable set $A_n$ such that $A_n, \varphi A_n, \ldots, \varphi^{n-1}A_n$ are pairwise disjoint, and such that $m(A_n) \geq \frac{1}{n} - \frac{1}{n^2}$.

Define $L^2$-functions

$$g_n := \frac{1}{n \cdot m(A_n)} \sum_{i=0}^{n-1} \lambda^{n-i} T_\varphi^i \mathbb{1}_{A_n}$$

and

$$f_n(x) := \begin{cases} \lambda^{-(n-i)} & \text{for } x \in \varphi^i A_n \ (i = 0, 1, \ldots, n-1) \\ 0 & \text{otherwise.} \end{cases}$$

An easy calculation shows, that $\langle g_n, f_n \rangle = 1$ for every $n \in \mathbb{N}$, and from $\| f_n \| \leq 1$ we obtain $\| g_n \| \geq 1$.

Now,

$$\begin{aligned} \| T_\varphi g_n - \lambda g_n \| &= \frac{1}{n \cdot m(A_n)} \Big\| \sum_{i=0}^{n-1} (\lambda^{n-i} T_\varphi^{i+1} - \lambda^{n-i+1} T_\varphi^i) \mathbb{1}_{A_n} \Big\| \\ &= \frac{1}{n \cdot m(A_n)} \| (\lambda T_\varphi^n \mathbb{1}_{A_n} - \lambda^{n+1} \mathbb{1}_{A_n}) \| \\ &\leq \frac{1}{n \cdot m(A_n)} \cdot 2 \cdot m(A_n)^{\frac{1}{2}} = \frac{2}{n \cdot m(A_n)^{\frac{1}{2}}} \longrightarrow 0, \end{aligned}$$

which shows $\lambda \in \sigma(T_\varphi)$. ∎

Remark:

For arbitrary MDSs the proposition follows from Schaefer [1974], V.4.4 and Schaefer-Wolff-Arendt [1978], 2.2 .

Other well known spectral invariants concern the asymptotic behavior of $T^n$ (e.g. mixing, see Lecture IX). But it took quite a long time until Kolmogorov in 1956 introduced an invariant, the "entropy", which is not a spectral invariant. See Lectures XI - XIII.

VI. D.4 Every MDS is spectrally isomorphic to its inverse:

If $(X,\Sigma,\mu;\varphi)$ is an MDS, then

$$T_\varphi : L^2(X,\Sigma,\mu) \longrightarrow L^2(X,\Sigma,\mu)$$

is a unitary operator, and its spectrum $\sigma(T_\varphi)$ is either a finite union of finite subgroups of $\Gamma$ or equal $\Gamma$ (compare VI.D.3).

More precisely, if $\lambda \in \Gamma$ is an eigenvalue of $T_\varphi$ with eigenfunction $f \in L^2(\mu)$, i.e.

if $\qquad T_\varphi f = \lambda \cdot f$ ,

then $\qquad T_\varphi \bar{f} = \bar{\lambda} \cdot \bar{f}$

and $\qquad T_\varphi f^{(n)} = (T_\varphi f)^{(n)} = \lambda^n \cdot f^{(n)}$ , $n \in \mathbb{N}$,

where $f^{(0)} := |f|$, $f^{(n+1)} := \operatorname{sign} f \cdot f^{(n)}$, and $(\operatorname{sign} f)(x) := \frac{f(x)}{|f(x)|}$ if $f(x) \neq 0$ and $(\operatorname{sign} f)(x) = 0$ if $f(x) = 0$ .

This shows that the point spectrum of $T_\varphi$ is invariant under complex conjugation and cyclic (i.e. $\lambda \in P\sigma(T_\varphi)$ implies $\bar{\lambda} \in P\sigma(T_\varphi)$ and $\lambda \in P\sigma(T_\varphi)$ implies $\lambda^n \in P\sigma(T_\varphi)$ for $n \in \mathbb{N}$). By analogous arguments the same holds for the approximate point spectrum $A\sigma(T_\varphi)$.

In particular, we observe that $T_\varphi$ and its inverse $T_\varphi^{-1}$ $(= T_\varphi^*)$ possess the same eigenvalues and the same spectrum. Therefore we are not able to distinguish between $(X,\Sigma,\mu;\varphi)$ and $(X,\Sigma,\mu;\varphi^{-1})$ by using these spectral invariants. Actually, even more is true.

Proposition 1: $(X,\Sigma,\mu;\varphi)$ and $(X,\Sigma,\mu;\varphi^{-1})$ are spectrally isomorphic.

Proof:

On $L^2(X,\Sigma,\mu)$ we consider the conjugate linear involution

$$I : f \to \bar{f}$$

and observe that $T_\varphi \circ I = I \circ T_\varphi$ .

Next we apply the spectral theorem (see Reed-Simon [1972]) and find an $L^2$-space H and a unitary multiplication operator $M_u : g \mapsto ug$ on H such that the two FDSs

$$(L^2(X,\Sigma,\mu);T_\varphi) \quad \text{and} \quad (H;M_u)$$

are isomorphic. This isomorphism may be given by the unitary operator $V : L^2(X,\Sigma,\mu) \to H$. On H, the complex conjugation again is conjugate linear and will be denoted by J. Since the adjoint of $M_u$ is the multiplication operator $M_{\bar u}$ we obtain

$$J \circ M_u = M_u^* \circ J \ .$$

All these relations are collected in the following commuting diagram:

$$\begin{array}{ccccccccc}
L^2(\mu) & \xrightarrow{\ I\ } & L^2(\mu) & \xrightarrow{\ V\ } & H & \xrightarrow{\ J\ } & H & \xrightarrow{\ V^*\ } & L^2(\mu) \\
T_\varphi \downarrow & & T_\varphi \downarrow & & M_u \downarrow & & \downarrow M_u^* & & \downarrow T_\varphi^* = T_{\varphi^{-1}} \\
L^2(\mu) & \xrightarrow[\ I\ ]{} & L^2(\mu) & \xrightarrow[\ V\ ]{} & H & \xrightarrow[\ J\ ]{} & H & \xrightarrow[\ V^*\ ]{} & L^2(\mu)
\end{array}$$

We obtain the conclusion that

$$W := V^* \circ J \circ V \circ I$$

is a unitary operator yielding a Hilbert space isomorphism between $(L^2(\mu);\ T_\varphi)$ and $(L^2(\mu);\ T_{\varphi^{-1}})$. ■

In general, a unitary operator $U \in \mathcal{L}(H)$, H a Hilbert space, is not isomorphic to its inverse: Take $X := \{z \in \mathbb{C} : |z| = 1,\ 0 \leq \arg\ z \leq \pi\}$,

$$H := L^2(X,\mathcal{B},m) \quad \text{and} \quad Uf(z) := z \cdot f(z).$$

Since $\sigma(U) = X$ but $\sigma(U^*) = \{\bar z : z \in X\}$, U and $U^*$ cannot be isomorphic.

On the other hand, one sees from (VI.4) and (VI.5.2) that spectral isomorphism is a rather weak isomorphism for MDSs. In particular, the spectral isomorphism of $(X,\Sigma,\mu;\varphi)$ and $(X,\Sigma,\mu;\varphi^{-1})$ does not imply its isomorphism.

Therefore one must look for other isomorphism invariants by which one might be able to distinguish between $(X,\Sigma,\mu;\varphi)$ and its inverse. Unfortunately, the other important invariant besides the spectral invariants does not help.

<u>Proposition 2:</u> $(X,\Sigma,\mu;\varphi)$ and $(X,\Sigma,\mu;\varphi^{-1})$ have the same entropy:

$$h_\mu(X;\varphi) = h_\mu(X;\varphi^{-1}) \ .$$

This is proved in (XII.4.v) and leaves us with two possibilities: either $(X,\Sigma,\mu;\varphi)$ and its inverse are always isomorphic or entropy and spectral invariants together are not a complete system of isomorphism invariants for MDSs.

The examples in Anzai [1951 a,b] show that the latter case holds (see also XIII.D.5).

References: Anzai [1951 a,b], Ornstein [1973].

VI. D.5 Bernoulli shifts have countable Lebesgue spectrum:
If the MDS $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ is a Bernoulli shift $B(p_0, \ldots, p_{k-1})$ then we are able to describe the induced operator

$$T_\tau : L^2(\hat{X}, \hat{\Sigma}, \hat{\mu}) \longrightarrow L^2(\hat{X}, \hat{\Sigma}, \hat{\mu})$$

in terms of an adequately chosen Hilbert space basis. To this end we consider the k-dimensional space

$$E := L^2(\{0, \ldots, k-1\}, (p_0, \ldots, p_{k-1})),$$

and choose a complete orthonormal system $f_0, \ldots, f_{k-1} \in E$ with $f_0 = 1\!\!1$.
The projections

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

induce isometric injections

$$P_j : E \longrightarrow L^2(\hat{X}, \hat{\Sigma}, \hat{\mu})$$
$$f \longmapsto f \circ \pi_j$$

satisfying $\quad P_j f_0 = 1\!\!1_{\hat{X}}$.

It is not difficult to show that the countably many functions

$$g := \prod_{j \in \mathbb{Z}} P_j f_{n_j} \quad ,$$

where $n_j = 0$ except for finitely many indices, form a complete orthonormal system in $L^2(\hat{X}, \hat{\Sigma}, \hat{\mu})$.
Moreover

$$T_\tau g = T_\tau \left( \prod_{j \in \mathbb{Z}} P_j f_{n_j} \right) = \prod_{j \in \mathbb{Z}} P_{j+1} f_{n_j} \; ,$$

which shows that the basis in $L^2(\hat{X}, \hat{\Sigma}, \hat{\mu})$ may be written in the form

$$\{ h_0, T_\tau^i h_m : i \in \mathbb{Z}, m \in \mathbb{N} \}$$

where $T_\tau h_0 = h_0 = 1\!\!1_{\hat{X}}$.

The property of an MDS of having an $L^2$-basis as above is expressed by saying that it has <u>countable Lebesgue spectrum</u>. Clearly, for any two MDSs with countable Lebesgue spectrum the FDSs induced on the corresponding $L^2$-spaces are isomorphic. In other words, we have the following result.

Proposition:

(i) Any Bernoulli shift has countable Lebesgue spectrum.

(ii) Any two Bernoulli shifts are spectrally isomorphic.

The statement (ii) indicates that spectral invariants are very incomplete. Compare (XII.D.1).

Exercise: Show that an MDS with countable Lebesgue spectrum is strongly mixing.

References: Helson-Parry [1978], Rohlin [1967], Walters [1975].

VI. D.6 From an MDS to a TDS - the Stone space of the measure algebra:

Since every TDS possesses at least one invariant measure (see App.S.1) it can be viewed as an MDS. We now associate to every MDS a suitable TDS. More precisely: Given an MDS $(X,\Sigma,\mu;\varphi)$ we construct an isomorphic MDS $(X_1,\Sigma_1,\mu_1;\varphi_1)$ such that $(X_1;\varphi_1)$ is a TDS.

First we investigate the "static" situation and consider $L^\infty(X,\Sigma,\mu)$ as a commutative $C^*$-algebra (resp. as an AM-space with unit). From the Gelfand-Neumark theorem we know that there exists a compact space $\check{X}$ such that $L^\infty(X,\Sigma,\mu)$ is isomorphic to $C(\check{X})$ (see App.C.9), The adjoint of this isomorphism transports the positive linear form $\mu$ on $L^\infty(X,\Sigma,\mu)$ to a Radon measure $\check{\mu} \in C(\check{X})'$ which corresponds to a regular Borel measure, still denoted by $\check{\mu}$, on the Borel algebra $\mathcal{B}$ of $\check{X}$. Since the isomorphism from $L^\infty(X,\Sigma,\mu)$ to $C(\check{X})$ extends to an isometric isomorphism from $L^1(X,\Sigma,\mu)$ to $L^1(\check{X},\mathcal{B},\check{\mu})$, we conclude by (VI.2) that the probability spaces $(X,\Sigma,\mu)$ and $(\check{X},\mathcal{B},\check{\mu})$ are isomorphic.

Another characterization of $\check{X}$ avoiding the use of the Gelfand-Neumark theorem is possible:

Consider the measure algebra $\check{\Sigma} := \Sigma/\mathcal{N}$ (see App.A.9) which is a Boolean algebra. By the Stone representation theorem (see Halmos [1974], p.78) there exists a unique compact space $X_{\check{\Sigma}}$ such that $\check{\Sigma}$ is isomorphic to the Boolean algebra of all open-closed subsets of $X_{\check{\Sigma}}$. The compact space $X_{\check{\Sigma}}$ is called the Stone space of $\check{\Sigma}$, and one can prove that

$$\check{X} \text{ and } X_{\check{\Sigma}} \text{ are homeomorphic.}$$

For a more detailed analysis of $\check{X}$ we return to the functional analytic approach via Gelfand-Neumark and recall that $L^\infty(X,\Sigma,\mu)$ is isomorphic to $C(\check{X})$.

Lemma:

(i) $\check{X}$ is extremally disconnected, i.e. the closure of every open set is open.

(ii) The $\check{\mu}$-null sets in $\check{X}$ are precisely the nowhere dense subsets of $\check{X}$.

(iii) $L^\infty(\check{X}, \mathcal{B}, \check{\mu})$ is canonically isomorphic to $C(\check{X})$, i.e. every equivalence class $\check{f} \in L^\infty(\check{X}, \mathcal{B}, \check{\mu})$ contains exactly one continuous function $f \in C(\check{X})$.

Proof:

(i) $L^\infty(X, \Sigma, \mu)$ and hence $C(\check{X})$ is an order complete Banach lattice, and the assertion follows from Schaefer [1974], II.7.7.

(ii) This surprising coincidence is due to the existence of the strictly positive and order continuous linear form $\mu$ on $L^\infty(X, \Sigma, \mu)$, resp. $\check{\mu}$ on $C(\check{X})$:

Let N be a closed subset of $\check{X}$ having empty interior. Since $\check{X}$ is extremally disconnected, we obtain

$$0 = \inf \{ 1\!\!1_{\mathcal{O}} : N \subset \mathcal{O} \text{ open-closed} \}$$

and therefore $\check{\mu}(\mathcal{O}) \downarrow \check{\mu}(N) = 0$.

On the other hand assume $\check{\mu}(B) = 0$ for some Borel set $B \subset \check{X}$. Since $\check{\mu}$ is a regular Borel measure, there exists a decreasing sequence $(U_n)$ of open sets containing B and such that $\lim_{n\to\infty} \check{\mu}(U_n) = 0$. But from the order continuity of $\check{\mu}$ on $C(\check{X})$ it follows that $\check{\mu}(U_n) = \check{\mu}(\overline{U}_n)$, where $\overline{U}_n$ is open and closed. Obviously $B_o := \bigcap_{n\in\mathbb{N}} \overline{U}_n$ contains B and is closed. It remains to observe that B is nowhere dense:

Assume $\emptyset \neq \mathcal{O} \subset B$ for some open-closed set $\mathcal{O}$. The strict positivity of $\check{\mu}$ implies $0 < \check{\mu}(\mathcal{O}) \leq \check{\mu}(B) \leq \check{\mu}(U_n)$ which is a contradiction.

(iii) For characteristic functions $1\!\!1_B$, $B \in \mathcal{B}$, this has been proved in (ii). But this already implies the assertion since these functions are total in $L^\infty(\check{X}, \mathcal{B}, \check{\mu})$. ∎

Remark:

The surprising fact expressed by the previous lemma is the complete coincidence of measure-theoretical and topological concepts in the Stone space $\check{X}$ of a probability space $(X, \Sigma, \mu)$. For instance, (ii) asserts that a set is topologically small (=nowhere dense) in $\check{X}$ if and only if it is measure-theoretically small (= $\check{\mu}$-null set). This is far from true for arbitrary Borel measures on arbitrary compact spaces.

After these preparations it is quite easy to obtain the desired isomorphism for dynamical systems.

Theorem:

Let $(X, \Sigma, \mu; \varphi)$ be an MDS and $(\check{X}, \mathcal{B}, \check{\mu})$ the probability space constructed above. There exists a homeomorphism $\psi$ on $\check{X}$ such that $(\check{X}, \mathcal{B}, \check{\mu}; \psi)$ is isomorphic to $(X, \Sigma, \mu; \varphi)$.

Proof:

$T_\varphi$ is a Banach algebra isomorphism on $L^\infty(X, \Sigma, \mu)$ and yields a Banach algebra isomorphism $\check{T}$ on $C(\check{X})$. As shown in (II.D.5) there exists a homeomorphism $\psi : \check{X} \to \check{X}$ such that $\check{T}f = f \circ \psi$ for $f \in C(\check{X})$. Obviously, $\psi$ leaves invariant the measure $\check{\mu}$. Then the desired isomorphism follows from (VI.2). ∎

Finally we return to a question already posed in (II.D.6):

Let $(X, \Sigma, \mu)$ be a probability space and consider an $\mathrm{FDS}(L^p(X, \Sigma, \mu); T)$. For which operators T is the FDS $(L^p(\mu); T)$ induced by an MDS ?

It is clear that T must be a Banach lattice isomorphism satisfying $T\mathbb{1} = \mathbb{1}$ and $T'\mathbb{1} = \mathbb{1}$, but if the measure space is pathological these conditions are not sufficient (see VI.D.1).

By the "Stone space" construction we may circumvent these difficulties.

Corollary:

Let T be a Banach lattice isomorphism on $L^p(X, \Sigma, \mu)$, $1 \leq p \leq \infty$, satisfying $T\mathbb{1} = \mathbb{1}$ and $T'\mathbb{1} = \mathbb{1}$. If X is compact and C(X) is canonically isomorphic to $L^\infty(X, \Sigma, \mu)$, then T is always induced by a bi-measure-preserving homeomorphism $\varphi : X \to X$.

References: Halmos [1974], Nagel [1973b], Semadeni [1971], Schaefer [1974].

## VII. Compact Operator Semigroups

Having investigated the asymptotic behavior of the Cesaro 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 occuring 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

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

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×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 App.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 <u>ideal</u> in a commutative semigroup S is a nonempty subset J such that $SJ := \{st : s \in S, t \in J\} \subset J$.

VII. 2 <u>Theorem</u>:
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

$$\emptyset \neq J_1 J_2 \cdots J_n \subset \bigcap_{i=1}^{n} J_i ,$$

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

$$K := \bigcap \{J : J \text{ closed ideal in } S\}$$

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 $s \in K$ 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

$$r\,q = r'sq = r's = r ,$$

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

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

Denote by $\mathcal{S} := \overline{\{T^n : n \in \mathbb{N}\}}$ the closure of $\{T^n : n \in \mathbb{N}\}$ in $\mathcal{L}(E)$ with respect to the weak operator topology. Then $\mathcal{S}$ and its closed convex hull $\overline{co}\,\mathcal{S}$ 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 $\mathcal{L}(E)$. 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 $\mathcal{S}$ commute with operators in $\{T^n : n \in \mathbb{N}\}$. Now take $0 \neq R_1, R_2 \in \mathcal{S}$, $f \in E$, $f' \in E'$ and $\varepsilon > 0$. Then there exists $R \in \{T^n : n \in \mathbb{N}\}$ such that

$$|\langle (R_2 - R)f, f'\rangle| \leq \frac{\varepsilon}{2\|R_1\|} \quad \text{and}$$

$$|\langle (R_2 - R)R_1 f, f'\rangle| \leq \frac{\varepsilon}{2} .$$

Therefore we have

$$|\langle (R_1R_2 - R_2R_1)f, f'\rangle| = |\langle (R_1R_2 - R_1R + RR_1 - R_2R_1)f, f'\rangle| \leq$$
$$|\langle R_1(R_2 - R)f, f'\rangle| + |\langle (R - R_2)R_1 f, f'\rangle| \leq \varepsilon ,$$

which implies $R_1R_2 = R_2R_1$.

Finally, the condition (*) implies that $\mathcal{S}$ is compact in $\mathcal{L}_w(E)$ (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 $co\,\mathcal{S}$ is a commutative semigroup, the same arguments as above apply to $\overline{co}\,\mathcal{S}$. ∎
Now we apply (VII.2) to the semigroups $\mathcal{S}$ and $\overline{co}\,\mathcal{S}$. Thereby the semigroup $\overline{co}\,\mathcal{S}$ 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, $$E = F \oplus F_0 ,$$

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

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

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
$$\mathcal{S} := \overline{\{T^n : n\in\mathbb{N}_0\}} \subset \mathcal{L}_w(E)$$
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 <u>Theorem:</u> Let (E;T) be an FDS satisfying (✱). Then there exists a projection
$$Q \in \mathcal{S} = \overline{\{T^n: n\in\mathbb{N}_0\}}$$
such that $$\mathcal{K} := Q\mathcal{S}$$
is the minimal ideal of $\mathcal{S}$ and a compact group with unit Q.

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

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

and $$G_0 := Q^{-1}(0) = \{f\in E: 0 \in \overline{\{T^n f: n\in\mathbb{N}_0\}}^{\sigma(E,E')}\}.$$

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 $f\in E$ the map $S\mapsto Sf$ is continuous from $\mathcal{L}_w(E)$ into $E_\sigma$ and since Q is contained in $\mathcal{S}$, we see that Qf = 0 implies

$0 \in \overline{\{T^n f : n \in \mathbb{N}_o\}}$. Conversely, if $0 \in \overline{\{T^n f : n \in \mathbb{N}_o\}}$, there exists an operator R in the compact semigroup $\mathcal{J}$ 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 $\mathcal{K} = Q\mathcal{J}$.

2. Next we prove that

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

Denote by $\hat{\mathcal{K}}$ the character group of $\mathcal{K}$ and define for every character $\gamma \in \hat{\mathcal{K}}$ the operator $P_\gamma$ by

$$P_\gamma(f) := \int_{\mathcal{K}} \overline{\gamma(S)}\, Sf \, dm(S), \quad f \in E.$$

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

$$\langle P_\gamma f, f' \rangle = \int_{\mathcal{K}} \overline{\gamma(S)} \langle Sf, f' \rangle \, dm(S) \quad \text{for every } f' \in E'.$$

$P_\gamma(f)$ is an element of the bi-dual $E''$ contained in $\overline{\text{co}}\,\{\overline{\gamma(S)} \cdot Sf : S \in \mathcal{K}\}$.

However by Krein's theorem (App. B.6) this set is $\sigma(E,E')$-compact and hence contained in E. Therefore $P_\gamma$ is a well-defined bounded linear operator on E.

Now take $R \in \mathcal{K}$ and observe that

$$\begin{aligned} RP_\gamma(f) &= R\left(\int_{\mathcal{K}} \overline{\gamma(S)}\, Sf \, dm(S)\right) = \int_{\mathcal{K}} \overline{\gamma(S)}\, RSf \, dm(S) \\ &= \gamma(R) \int_{\mathcal{K}} \overline{\gamma(RS)}\, RSf \, dm(RS) = \gamma(R)\, P_\gamma(f) \text{ for every } f \in E, \end{aligned}$$

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

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

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

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

$$\mathfrak{f} := (f_1, \ldots, f_n)$$

of normalized eigenvectors of T (and R) with

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

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

$$|1 - \lambda_i^m| \leqslant \varepsilon \quad \text{and consequently}$$

$$\|R^m f_i - f_i\| \leqslant \varepsilon \quad \text{simultaneously for } i = 1, \dots, n.$$

This proves that the set

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

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

The minimal ideal $\mathcal{K}$ of $\mathcal{S}$ 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 $Q = Id_H$. Moreover, the weak and strong topologies coincide on the one-dimensional orbits $\mathcal{S}f$ for every eigenvector f. Therefore the group $\mathcal{K}$ is even compact for the strong operator topology.
Operators for which H = E (and therefore $Q = Id_E$ and $\mathcal{S} = \mathcal{K}$) 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 MDS (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) $\mathcal{S} = \overline{\{T^n : n \in \mathbb{N}_0\}} \subset \mathcal{L}_w(E)$ is a compact group with unit $Id_E$.

(c) $\mathcal{S} = \overline{\{T^n : n \in \mathbb{N}_0\}} \subset \mathcal{L}_s(E)$ 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 space $l^2(\mathbb{Z})$ and the shift

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

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

VII. 8 Programmatic remark:

The semigroups

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

in $\mathcal{L}_w(L^p(X,\Sigma,\mu))$, $1 \leq p < \infty$, 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

$$\mathrm{Id} \geqslant Q \geqslant P \geqslant \mathbb{1} \otimes \mathbb{1} \quad ,$$

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 $P = \mathbb{1} \otimes \mathbb{1}$ we will study in the subsequent lectures the following "extreme" cases:

Lecture VIII: $\mathrm{Id} = Q > P = 1 \otimes 1$ ,

Lecture IX: $\mathrm{Id} > Q = P = 1 \otimes 1$ .

VII. D Discussion

VII. D.1 Semitopological semigroups:

One might expect that semigroups S - if topologized - should have jointly continuous multiplication, i.e.,

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

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 $\mathcal{L}(E)$:

In ergodic theory it is the semigroup $\{T^n : n \in \mathbb{N}_0\}$ - $T \in \mathcal{L}(E)$ 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 $\mathcal{L}(E)$. If we choose the norm topology or the strong operator topology, and if $\|T^n\| \leq c$, 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 of some subsequence will be obtained. The following example illustrates these remarks:

Take $E = l^2(\mathbb{Z})$ 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 is 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 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

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

generated by some FDS (E;T) contains an element whose powers are dense in $\mathcal{S}$. 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 $\Gamma$ 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 $\Gamma^n$, $n \in \mathbb{N}$, 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 $\Gamma$ (see App.D.8).

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

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

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

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

VII. D.4 <u>Compact semigroups generated by operators on $L^{p}(X,\Sigma,\mu)$, $1 \leq p < \infty$</u>:

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 $(X,\Sigma,\mu)$ and a positive operator

$$T : L^{1}(X,\Sigma,\mu) \longrightarrow L^{1}(X,\Sigma,\mu)$$

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

$$T_{p} : L^{p}(X,\Sigma,\mu) \longrightarrow L^{p}(X,\Sigma,\mu)$$

are contractive for $1 \leq p < \infty$. The semigroups

$$\mathcal{S}_{p} := \overline{\{T_{p}^{n} : n \in \mathbb{N}_{o}\}}$$

in $\mathcal{L}_{w}(L^{p}(\mu))$ are compact for $1 \leq p < \infty$ : 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 $1 \leq p < \infty$ will all be denoted by $\mathcal{S}$.
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

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

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

$$\varphi_a : z \mapsto az.$$

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 $T_{\varphi_a}$ induced on $L^p(\Gamma, \mathcal{B}, m)$, $1 \leq p < \infty$, 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 $1 \leq p < \infty$.

(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 $n \geq 2$ endow $E_n := \mathbb{C}^n$ with the norm

$$\|(x_1, \ldots, x_n)\| := \max \{(n+1-i)^{-1} |x_i| : 1 \leq i \leq n\}$$

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 $S_{(n)}$, $n \geq 2$, has discrete spectrum in $E_n$. An easy calculation shows that $\|S_{(n)}\| \leq 2$ and $\sup \{\|S^{n-1}_{(i)}\| : i \geq 2\} \leq \|S^{n-1}_{(n)}\| = n$ for all $n \geq 2$. Now, take the $l^1$-direct sum $E := \bigoplus_{n \geq 2} E_n$ and $T := \bigoplus_{n \geq 2} S_{(n)}$.

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 semi-topological 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, $\sigma$ a metrizable, compact Hausdorff topology on $G$ such that the mapping

$$(g,h) \mapsto gh : G \times G \to G$$

is separately continuous. Then $(G,\sigma)$ 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 $t$

$$\varepsilon \leq d(st, s_U t_V)$$

for some suitable $(s_U, t_V) \in U \times V$, and $d(\cdot,\cdot)$ a metric on $G$ generating $\sigma$. Since multiplication is separately continuous there exists a neighbourhood $U_o$ of $s$ and $V_o$ of $t$, such that

$$d(st, s't) \leq \varepsilon/4 \quad \text{for every } s' \in U_o ,$$

and

$$d(s_{U_o}t, s_{U_o}t') \leq \varepsilon/4 \text{ for every } t' \in V_o .$$

From this we obtain the contradiction

$$\varepsilon \leq d(st, s_{U_o}t_{V_o})$$
$$\leq d(st, s_{U_o}t) + d(s_{U_o}t, s_{U_o}t_{V_o}) \leq \varepsilon/2.$$

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 $(G,\sigma)$ 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 $T_\varphi$ 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 $T := T_\varphi \in \mathcal{L}(L^p(\mu))$, $1 \leq p < \infty$, and also the compact abelian semigroup

$$\mathcal{S} := \overline{\{T^n : n \in \mathbb{N}_0\}} \subset \mathcal{L}_w(L^p(\mu)) \qquad \text{(see VII.D.4).}$$

Since $\varphi$ is ergodic, the corresponding mean ergodic projection P is of the form

$$P = 1 \otimes 1 \in \overline{\text{co}}\,\mathcal{S} \qquad \text{(see IV.7).}$$

Since $\mathcal{S}$ is compact, there exists another projection

$$Q \in \mathcal{S}$$

such that $Q\mathcal{S}$ 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

$$Q = Id ,$$

i.e. $\mathcal{S}$ is a compact group in $\mathcal{L}_w(L^p(\mu))$ – or $\mathcal{L}_s(L^p(\mu))$, 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. $T_\varphi$ has discrete spectrum in $L^p(X,\Sigma,\mu)$, $1 \leq p < \infty$. 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 $T_\varphi$ 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 <u>Example</u>:
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 $\chi \in \hat{G}$ is an eigenfunction of $T_{\varphi_g}$ because

$$T_{\varphi_g} \chi(h) = \chi(gh) = \chi(h) \cdot \chi(h)$$

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 $T_{\varphi_g}$ has discrete spectrum in C(G).
Conversely, the following theorem shows that the example above is typical.

VIII. 2 <u>Theorem:</u>
Let $(X;\varphi)$ be a minimal TDS such that $T_\varphi$ 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

$$\mathcal{G} := \overline{\{T^n : n \in \mathbb{N}_0\}} \subset \mathcal{L}_s(C(X)).$$

We shall show that $(X;\varphi)$ is isomorphic to $(\mathcal{G}; \text{Rot } T)$.
The operator T is a Banach algebra isomorphism of C(X). Since $\mathcal{G}$ is a group, the same is true for every $S \in \mathcal{G}$. Therefore there exist homeomorphisms

$$\varphi_S : X \longrightarrow X$$

such that $Sf = f \circ \varphi_S$ for every $S \in \mathcal{G}$, $f \in C(X)$,
and $\varphi_{S_1 S_2} = \varphi_{S_1} \circ \varphi_{S_2}$ for $S_1, S_2 \in \mathcal{G}$ (see II.D.5).

Choose $x_0 \in X$ and define

$$\Theta : \mathcal{G} \longrightarrow X \text{ by } \Theta(S) := \varphi_S(x_0) \text{ for } S \in \mathcal{G}.$$

This map yields the isomorphism between $(\mathcal{G}; \text{Rot } T)$ and $(X;\varphi)$:

(1) $\Theta$ is continuous: If the net $(S_\alpha)_{\alpha\in A}$ converges to $S$ in the strong operator topology, then

$$f(\Theta(S_\alpha)) = S_\alpha f(x_o) \text{ converges to } Sf(x_o) = f(\Theta(S))$$

for every $f \in C(X)$. But this implies that $(\Theta(S_\alpha))_{\alpha\in A}$ converges to $\Theta(S)$ in $X$.

(2) $\Theta$ is surjective: $\Theta(\mathcal{G})$ is a closed subset of $X$ which is $\varphi$-invariant. From the minimality it follows that $\Theta(\mathcal{G}) = X$.

(3) $\Theta$ is injective: If $\Theta(S_1) = \Theta(S_2)$ for $S_1, S_2 \in \mathcal{G}$, we conclude that $\varphi_{S_1}(x_o) = \varphi_{S_2}(x_o)$ or $\varphi_{S_2^{-1}S_1}(x_o) = x_o$ and

$$\varphi_{S_2^{-1}S_1}(\varphi^n(x_o)) = \varphi^n(\varphi_{S_2^{-1}S_1}(x_o)) = \varphi^n(x_o) \quad \text{for all } n \in \mathbb{N}.$$

Again from the minimality of $\varphi$ it follows that $\{\varphi^n(x_o) : n \in \mathbb{N}\}$ is dense in $X$, and therefore that $\varphi_{S_2^{-1}S_1} = id_X$ or $S_2 = S_1$.

(4) The diagram

$$\begin{array}{ccc} X & \xrightarrow{\ \varphi\ } & X \\ \Theta\uparrow & & \uparrow\Theta \\ \mathcal{G} & \xrightarrow{\ \text{Rot } T\ } & \mathcal{G} \end{array}$$

commutes:

For $S \in \mathcal{G}$ we obtain $\varphi(\Theta(S)) = \varphi(\varphi_S(x_o)) = \Theta(TS)$. ∎

As an application of this representation theorem we can solve the isomorphism problem for minimal TDSs with discrete spectrum.

VIII. 3 <u>Corollary</u>:

(i) For minimal TDSs with discrete spectrum the point spectrum of the induced operator is a subgroup of the unit circle $\Gamma$, and as such a complete isomorphism invariant.

(ii) Let $\Gamma_o$ be an arbitrary subgroup of $\Gamma$ and endow $\Gamma_o$ with the discrete topology. The rotation on the compact group

$$G := \hat{\Gamma}_o$$

by the character $id : \lambda \longmapsto \lambda$ on $\Gamma_o$ is (up to isomorphism) the unique minimal TDS with discrete spectrum having $\Gamma_o$ 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 $T_\varphi$ is a subgroup of

$\Gamma$. 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 $(X_2;\varphi_2) \cong (G_2;\varphi_b)$. The next step is to show that the character group $\hat{G}_1$ is isomorphic to $P\sigma(T_{\varphi_a})$: Every $\gamma \in \hat{G}_1$ is a continuous eigenfunction of $T_{\varphi_a}$ with corresponding eigenvalue $\gamma(a)$. It is easy to see that

$$\Theta : \gamma \mapsto \gamma(a)$$

defines a group homomorphism from $\hat{G}_1$ into $P\sigma(T_{\varphi_a})$. Further more, $\Theta$ 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 $\gamma_1, \gamma_2$. The map $\Theta$ 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

$$f(a^{n+1}) = T_{\varphi_a} f(a^n) = \lambda f(a^n) = \lambda^{n+1}$$

for all $n \in \mathbb{N}$, and by continuity we conclude that f is a character on $G_1$ with $\Theta(f) = \lambda$.

Therefore, $\hat{G}_1$ is isomorphic to $P\sigma(T_{\varphi_a}) = P\sigma(T_{\varphi_1})$, and analogously $\hat{G}_2 \cong P\sigma(T_{\varphi_b}) = P\sigma(T_{\varphi_1})$. From $P\sigma(T_{\varphi_1}) = P\sigma(T_{\varphi_2})$ and Pontrjagin's duality theorem (App.D.6) we conclude $G_1 \cong G_2$.

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 eigenfunctions $f_\lambda$ for $T_{\varphi_a}$, resp. $g_\lambda$ for $T_{\varphi_b}$, normalized by $f_\lambda(a) = \lambda$, resp. $g_\lambda(b) = \lambda$. 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 $(G_1;\varphi_a) \cong (G_2;\varphi_b)$ by (VI.3).

(ii) By (i) it remains to show that $P\sigma(T_{\varphi_{id}}) = \Gamma_o$. But this follows from (App.D.6):

$$P\sigma(T_{\varphi_{id}}) \cong \hat{G} = \hat{\hat{\Gamma}}_o = \Gamma_o . \blacksquare$$

We have seen that the classification of minimal TDSs with discrete spectrum reduces to the classification of compact monothetic groups. The tori $\Gamma^n$, $n \in \mathbb{N}$, yield the standard examples (see VIII.D.2). 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-v.Neumann, 1942):
Let $(X,\Sigma,\mu;\varphi)$ be an ergodic MDS such that $T_\varphi$ has discrete spectrum in $L^p(X,\Sigma,\mu)$, $1 \leq p < \infty$. 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 $\lambda$, $|\lambda| = 1$, we conclude that

$$T\,|f| = |Tf| = |\lambda|\,|f| = |f| = c \cdot \mathbb{1}$$

since $\varphi$ 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 $\mathcal{A}$, is a commutative $C^*$-algebra with unit. By the Gelfand-Neumark theorem (App.C.9) there exists an isomorphism

$$j : \mathcal{A} \longrightarrow C(Y)$$

for some compact space Y.
The restriction of $T_\varphi$ to $\mathcal{A}$ is an algebra isomorphism on $\mathcal{A}$. Therefore, its isomorphic image $j \circ T_\varphi \circ j^{-1}$ 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: $T_\psi$ has discrete spectrum in C(Y) as $T_\varphi$ has in $\mathcal{A}$. Therefore, $T_\psi$ is mean ergodic by (VII.6) and (IV.4.c). Thus the fixed space of $T_\varphi$ in $\mathcal{A}$, and therefore of $T_\psi$ in C(Y) is one-dimensional. Since (the restriction of) $\mu$ is a strictly positive, $T_\varphi$-invariant linear form on $\mathcal{A}$, we obtain a strictly positive, $\psi$-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 $(Y;\psi)$ and obtain a homeomorphism

$$\Theta : G \longrightarrow Y ,$$

where G is a compact monothetic group with generating element a, making commutative the following diagram:

$$
\begin{array}{ccc}
\mathcal{A} & \xrightarrow{\;T_{\varphi}\;} & \mathcal{A} \\
{\scriptstyle j}\big\downarrow & & \big\downarrow{\scriptstyle j} \\
C(Y) & \xrightarrow{\;T_{\psi}\;} & C(Y) \\
{\scriptstyle T_{\theta}}\big\downarrow & & \big\downarrow{\scriptstyle T_{\theta}} \\
C(G) & \xrightarrow[\text{Rot } a]{} & C(G)
\end{array}
$$

where $(\text{Rot } a)f(g) := f(ag)$ for $f \in C(G)$.

But $\mathcal{A}$, $C(Y)$ 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 $T_{\theta}$ continuously to positive isometries (hence lattice isomorphisms, see App. C.4) on the corresponding $L^P$-spaces. Obviously, the same can be done for $T_{\varphi}$, $T_{\psi}$ and Rot a. Finally, we obtain an analogous diagram for the $L^P$-spaces, which proves the isomorphism of $(X,\Sigma,\mu;\varphi)$ and $(G,\mathcal{B}, m;$ 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 <u>Corollary</u>:

(i) For ergodic MDSs with discrete spectrum the point spectrum of the induced operator is a subgroup of $\Gamma$ and as such a complete isomorphism invariant.

(ii) Let $\Gamma_o$ be an arbitrary subgroup of $\Gamma$ and endow $\Gamma_o$ with the discrete topology.
The rotation on the compact group

$$G := \hat{\Gamma}_o$$

with normalized Haar measure m by the character id : $\lambda \longmapsto \lambda$ on $\Gamma_o$ is (up to isomorphism) the unique ergodic MDS with discrete spectrum having point spectrum $\Gamma_o$.

## VIII. D Discussion

### VIII. D.1 Rotations on the circle:

The simplest non-trivial examples of ergodic MDSs having discrete spectrum are the rotations

$$\varphi_a \; : \; z \longmapsto az$$

on the unit circle $\Gamma$ for $a \in \Gamma$, a not a root of unity. Their point spectrum is the group $\{a^n : a \in \mathbb{Z}\}$, hence always isomorphic to $\mathbb{Z}$. But two rotations $(\Gamma, \mathcal{B}, m; \varphi_a)$ and $(\Gamma, \mathcal{B}, m; \varphi_b)$ are isomorphic if and only if their point spectrum is equal, i.e. $a = b$ or $a = b^{-1}$.

### VIII. D.2 Monothetic tori:

We have seen in (VII.D.3) and (VIII.1) that the rotation by a generating element of the n-dimensional torus $\Gamma^n$, $n \in \mathbb{N}$, yields the standard example of a minimal TDS with discrete spectrum. More precisely: $a = (a_1, \dots, a_n) \in \Gamma^n$ is generating iff $\{a_1, \dots, a_n\}$ is linearly independent in the $\mathbb{Z}$ - module $\Gamma$, and the characters

$$(z_1, \dots, z_n) \longmapsto z_1^{m_1} z_2^{m_2} \dots z_n^{m_n}$$

for any n-tuple $(m_1, \dots, m_n) \in \mathbb{Z}^n$ are eigenfunctions for the rotation $\varphi_a$. Since $G := \Gamma^s$, s any cardinal number, is still a compact abelian group, the rotation $\varphi_a$, $a \in G$, defines a minimal TDS having discrete spectrum as soon as a is generating in G (see VII.1).

The following result shows that this is true only for "small" tori.

Proposition: $G = \Gamma^s$ is monothetic if and only if s is at most the cardinality c of the continuum.

Proof:

Assume that G is monothetic. Fr m (VIII.1) and (VIII.3) we conclude that the character group $\hat{G}$ is a subgroup of $\Gamma$, and in particular $|\hat{G}| \leq c$. If J is an index set of cardinality s, the projections

$$\pi_j \; : \; \Gamma^s \longrightarrow \Gamma \quad : \quad (x_i)_{i \in J} \longmapsto x_j \qquad (j \in J)$$

are different characters on $\Gamma^s$, and we obtain $s = |J| \leq |\hat{G}| \leq c$.

Conversely, if $s \leq c$ we can choose a set $\{a_k\} \subset \Gamma$ of cardinality s which is linearly independent in the $\mathbb{Z}$-module $\Gamma$. From (App.D.8) it follows that every open rectangle in $\Gamma^s$ (see App.A.3) contains at least one power of $a = (a_k)$, and therefore a is generating in $\Gamma^s$. ■

VIII. D.3 Disconnected monothetic groups:

In contrast to the tori $\Gamma^n$, the finite cyclic groups are monothetic and compact but not connected. The task of constructing monothetic compact groups which are totally disconnected but not finite is much more difficult but yields an interesting result:

Let $a = (a_0, a_1, \ldots)$ be a fixed sequence of integers $a_n \geq 2$.

Consider

$$\Delta_a := \prod_{n=0}^{\infty} \{0, 1, \ldots, a_n - 1\}$$

and define an addition for $x = (x_n)_{n \in \mathbb{N}_0}$, $y = (y_n)_{n \in \mathbb{N}_0} \in \Delta_a$ as follows:

We write $x_0 + y_0 = p_0 a_0 + z_0$ where $p_0 \in \mathbb{N}_0$ and $z_0 \in \{0, \ldots, a_0 - 1\}$.

If $p_0, p_1, \ldots, p_k$ and $z_0, z_1, \ldots, z_k$ have been defined, we write

$$x_{k+1} + y_{k+1} + p_k = p_{k+1} a_{k+1} + z_{k+1}$$

where $p_{k+1} \in \mathbb{N}_0$ and $z_{k+1} \in \{0, \ldots, a_{k+1} - 1\}$.

Finally, we use the values $z_n$ to define

$$x + y := z := (z_n)_{n \in \mathbb{N}_0} \in \Delta_a .$$

It is clear that the operation " + " is commutative and has $0 := (0, 0, \ldots)$ as neutral element. The inverse of $x = (x_n)_{n \in \mathbb{N}_0}$ is the element $y = (y_n)_{n \in \mathbb{N}_0}$ where

$$y_n := \begin{cases} 0 & \text{for } 0 \leq n < k \\ a_k - x_k & \text{for } n = k \\ a_n - x_n - 1 & \text{for } n > k \end{cases}$$

and where k denotes the index of the first non-zero component of x.

Finally, we refer to Hewitt-Ross [1979], 10.3 where it is shown that " + " is associative. Hence we obtained a group structure on $\Delta_a$.

The product topology on $\Delta_a$ makes it a compact (by Tychonov's theorem) and totally disconnected space (as a product of discrete spaces).

Theorem:

$\Delta_a$ endowed with the addition and topology described above is a metrizable totally disconnected compact monothetic group. Conversely, every infinite totally disconnected compact monothetic group is isomorphic to $\Delta_a$ for an appropriate sequence $a = (a_0, a_1, \dots)$.

Proof:

By Ellis' theorem (VII.D.6) it remains to show that the addition on $\Delta_a$ is separately continuous: Assume that $x_k \in \Delta_a$ converge to $x \in \Delta_a$. By the definition of the product topology, for every $n \in \mathbb{N}_0$ there exists $k_n \in \mathbb{N}$ such that the first n coordinates of $x_k$, $k \geq k_n$, coincide with those of x. This implies that for $k \geq k_n$ the n-th coordinate of $x_k + y$ agrees with that of $x + y$ for every $y \in \Delta_a$ and shows that $x_k + y$ converges to $x + y$.

Finally, it is not difficult to see that $u := (1, 0, 0, \dots)$ is a generating element in $\Delta_a$, hence $\Delta_a$ is monothetic.

For the proof of the second assertion we refer to Hewitt-Ross [1979], 25.16.

Remark:

If the sequence $a = (a_n)_{n \in \mathbb{N}_0}$ is constant, say $a_n = p \geq 2$ for all $n \in \mathbb{N}_0$, we write $\Delta_p$ instead of $\Delta_a$ and call it the <u>p-adic integers</u>.

VIII. D.4 <u>An MDS with discrete spectrum</u>:

Consider the MDS $([0,1), \mathcal{B}, m; \varphi)$ where $m$ is the Lebesgue measure on $[0,1)$ and

$$\varphi(x) := x - \frac{2^k - 3}{2^k} \quad \text{if } x \in \left[\frac{2^k - 2}{2^k}, \frac{2^k - 1}{2^k}\right] \text{ and } k \in \mathbb{N} .$$

We show that $T_\varphi$ has discrete spectrum in $L^2([0,1), \mathcal{B}, m)$ :

Define $\quad \chi_m := 1\!\!1_{[0,2^{-m})}$

and $\quad f_m := \sum_{k=0}^{2^m-1} \exp\left(-\frac{2\pi ik}{2^m}\right) T_\varphi^k \chi_m \qquad$ for $m \in \mathbb{N}_0$.

Then $$T_\varphi f_m = \sum_{k=0}^{2^m-1} \exp(-\frac{2\pi i k}{2^m})\ T_\varphi^{k+1}\chi_m$$
$$= \exp(\frac{2\pi i}{2^m}) \sum_{k=1}^{2^m} \exp(-\frac{2\pi i k}{2^m})\ T_\varphi^k \chi_m$$
$$= \exp(\frac{2\pi i}{2^m}) \cdot f_m \quad ,$$
which proves that $\left\{\exp(\frac{2\pi i}{2^m}) : m \in \mathbb{N}_0\right\} \subset P\sigma(T_\varphi)$.

Since $T_\varphi$ is an algebra homomorphism on $L^\infty(m)$, we conclude that $T_\varphi f_m^n = \exp(\frac{2\pi i n}{2^m})\ f_m^n$ for $n \in \{0, 1, \ldots, 2^m-1\}$.

A short calculation shows that
$$\chi_m = 2^{-m} \sum_{n=0}^{2^m-1} f_m^n \quad .$$

The "dyadic intervals" in $[0,1)$ generate $\mathcal{B}$, and hence the eigenfunctions $\left\{f_m^n : m \in \mathbb{N}_0;\ n = 0, 1, \ldots, 2^m - 1\right\}$ form a complete orthonormal system in $L^2(m)$.

Therefore $T_\varphi$ has discrete spectrum and
$$P\sigma(T_\varphi) = \left\{\exp(\frac{2\pi i n}{2^m}) : m \in \mathbb{N}_0;\ n = 0, 1, \ldots, 2^m-1\right\} .$$

Moreover, it is not difficult to show that $\varphi$ is ergodic. For a deeper understanding of the above example and an application of the Halmos-v.Neumann theorem the reader is advised to do the following exercise:

Prove that the MDS $([0,1), \mathcal{B}, m; \varphi)$ is isomorphic to a rotation on $\Delta_2$, see (VIII.D.3), and construct the isomorphism. (Hints: $\Delta_2$ is the dual group of $P\sigma(T_\varphi)$. Write $x \in [0,1)$ as dyadic number $0.x_0x_1x_2\ldots\ldots$, $x_i \in \{0, 1\}$, and observe that $(x_i)_{i\in\mathbb{N}_0} \in \Delta_2$).

VIII. D.5 Spectrum of Bernoulli shifts:

The opposite extreme to the dynamical systems with discrete spectrum are the systems having no other eigenvalues except 1, i.e. weakly mixing MDSs (see Lecture IX). The most import examples for this sort of dynamical system are the Bernoulli shifts. Hence, while the set of eigenvalues characterizes the ergodic MDSs with discrete spectrum, i.e. is a complete isomorphism invariant, this set if of no use for the investigation of Bernoulli shifts. In Lecture XII we shall introduse a new invariant, the entropy, which is complete for the Bernoulli shifts, but takes the value 0 for <u>all</u> ergodic MDSs with discrete spectrum (XIII.7) .

VIII. D.6 <u>Abstract dynamical systems with discrete spectrum</u>:
The representation theorems (VIII.2) and (VIII.4) can be extended in two directions:
The hypothesis that the induced operator T (= $T_\varphi$ ) has discrete spectrum is equivalent to the fact that

$$\overline{\{T^n : n \in \mathbb{N}_0\}} \subset \mathcal{L}_s(E)$$

is a compact group. Therefore, it is natural to investigate the analogous problem for arbitrary (even non-commutative) <u>compact operator groups</u>. With the appropriate definitions the operators in such groups can be represented as rotation operators (see Ellis [1969], ch.4).

The second way to extend and unify the above results consists in the investigation of <u>irreducible</u> positive operators T on a <u>Banach lattice</u> E. Clearly, the induced operators $T_\varphi$ on the spaces C(X) or $L^p(X,\Sigma,\mu)$ are the concrete examples behind such a purely functional-analytic approach. If the operator T has discrete spectrum then it can be proved that the Banach lattice E is "sandwiched" between a space C(X) and a space $L^1(X,\Sigma,\mu)$. More precisely:

<u>Theorem</u> (Nagel-Wolff, 1972):
Let E be any complex Banach lattice and let T be a positive irreducible operator on E.
Suppose that $\mathcal{G} := \overline{\{T^n : n \in \mathbb{N}_0\}} \subset \mathcal{L}_s(E)$ is a compact group with identity $Id_E$. If m denotes the normalized Haar measure on $\mathcal{G}$, the canonical injection $C(\mathcal{G}) \longrightarrow L^1(\mathcal{G}; m)$ can be factored through E, and the diagram

$$\begin{array}{ccccc} C(\mathcal{G}) & \longrightarrow & E & \longrightarrow & L^1(\mathcal{G};m) \\ \text{Rot } g \downarrow & & \downarrow T & & \downarrow \text{Rot } g \\ C(\mathcal{G}) & \longrightarrow & E & \longrightarrow & L^1(\mathcal{G};m) \end{array}$$

commutes, where Rot g is the rotation operator induced in $C(\mathcal{G})$, resp. $L^1(\mathcal{G}; m)$, by some generating element $g \in \mathcal{G}$.

Corollary:

Let S, T be positive irreducible operators on a Banach lattice E, both generating compact operator groups in $\mathcal{L}_s(E)$.
The FDSs (E; S) and (E; T) are isomorphic iff the operators S and T have the same point spectrum.

References: Ellis [1969], Lotz [1968], Nagel-Wolff [1972],

## 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 $L^p(X,\Sigma,\mu)$, $1 \le p < \infty$, generates a compact semigroup

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

in $\mathcal{L}(L^p(\mu))$ 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 a 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) $T_n \longrightarrow 1\!\!1 \otimes 1\!\!1$ in the weak operator topology.

(b) $\langle T_n f,g\rangle \longrightarrow (\int f\, d\mu)(\int g\, d\mu)$ for all $f,g \in L^\infty(X,\Sigma,\mu)$.

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

(d) 1 is a 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 $1\!\!1 \otimes 1\!\!1$ .
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) $T^n \longrightarrow 1\!\!1 \otimes 1$ in the weak operator topology.

(b) $\langle T^n f,g\rangle \longrightarrow (\int f\, d\mu)(\int g\, d\mu)$ for all $f,g \in L^\infty(X,\Sigma,\mu)$.

(c) $\mu(\varphi^{-n}A \cap B) \longrightarrow \mu(A)\cdot\mu(B)$ for all $A,B \in \Sigma$ .

IX. 3 **Definition:**
An MDS $(X,\Sigma,\mu;\varphi)$, resp. the transformation $\varphi$, 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 <u>Definition</u>:
An MDS $(X,\Sigma,\mu;\varphi)$, resp. the transformation $\varphi$, is called <u>weakly mixing</u> 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

$$\mathcal{S} := \overline{\{T^n : n \in \mathbb{N}\}}^{\sigma}$$

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 $\overline{\text{co}}\,\mathcal{S}$, and denote by $Q \in \mathcal{S}$ the projection generating the minimal ideal

$$\mathcal{K} = Q\mathcal{S}$$

of $\mathcal{S}$. The fact that 1 is a simple eigenvalue of T corresponds to the fact that $P = 1\!\!1 \otimes 1\!\!1$, see (IV.7), hence

$$1\!\!1 \otimes 1\!\!1 \in \overline{\text{co}}\,\mathcal{S} .$$

In (VII.5) we proved that Q is a projection onto the subspace spanned by all unimodular eigenvectors, hence

$$QE = PE = \langle 1\!\!1 \rangle .$$

From $Q \in \mathcal{S}$ it follows as in (IV.7) that

$$Q = P = 1\!\!1 \otimes 1\!\!1 ,$$

or equivalently

$$\{1\!\!1 \otimes 1\!\!1\} = \mathcal{K}$$

is the minimal ideal in $\mathcal{S}$. Briefly, weakly mixing systems are those for which the mean ergodic projection is already contained in $\mathcal{S}$ and is of the form $1\!\!1 \otimes 1\!\!1$.
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 $R := L^p(X, \Sigma, \mu)$, $1 \le p < \infty$, is separable, the following assertions are equivalent:

(a) $T^{n_i} \longrightarrow 1\!\!1 \otimes 1\!\!1$ for the weak operator topology and for some subsequence $\{n_i\} \subset \mathbb{N}$.

(a') $T^{n_i} \longrightarrow 1\!\!1 \otimes 1\!\!1$ for the weak operator topology and for some subsequence $\{n_i\} \subset \mathbb{N}$ having density 1.

(a") $\frac{1}{n} \sum_{i=0}^{n-1} |\langle T^i f, g\rangle - \langle f, 1\!\!1\rangle \langle 1\!\!1, g\rangle| \longrightarrow 0$ for all $f \in E$, $g \in E'$.

(b) $\langle T^{n_i} f, g\rangle \longrightarrow (\int f \, d\mu)(\int g \, d\mu)$ for all $f, g \in L^\infty(\mu)$ and for some subsequence $\{n_i\} \subset \mathbb{N}$.

(c) $\mu(\varphi^{-n_i} A \cap B) \longrightarrow \mu(A) \cdot \mu(B)$ for all $A, B \in \Sigma$ and for some subsequence $\{n_i\} \subset \mathbb{N}$.

(d) $\varphi$ is weakly mixing.

(e) $\varphi \otimes \varphi$ is ergodic.

(f) $\varphi \otimes \varphi$ is weakly mixing.

Remarks:

1. A subsequence $\{n_i\} \subset \mathbb{N}$ has density 1 if

$$\lim_{k \to \infty} \frac{1}{k} \left| \{n_i\} \cap \{1, 2, \ldots, k\} \right| = 1 \quad \text{(see App.E.1).}$$

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 Cesaro 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 $\Sigma$.

Proof:

The general considerations above imply that (d) is equivalent to $1\!\!1 \otimes 1\!\!1 \in \mathcal{S} = \overline{\{T^n : n \in \mathbb{N}\}}$. But by (VII.D.4), $\mathcal{S}$ is metrizable for the weak operator topology, hence there even exists a subsequence in

$\{T^n : n \in \mathbb{N}\}$ converging to $1\!\!1 \otimes 1\!\!1$, which shows the equivalence of (a) and (d).

(a) ⇒ (a"): We recall again that $\mathcal{S}$ is a commutative compact semigroup containing $1\!\!1 \otimes 1\!\!1$ as a zero, i.e. $R \cdot (1\!\!1 \otimes 1\!\!1) = 1\!\!1 \otimes 1\!\!1$ for all $R \in \mathcal{S}$.

Define the operator

$$\tilde{T} : C(\mathcal{S}) \longrightarrow C(\mathcal{S})$$

induced by the rotation by $T$ on $\mathcal{S}$, i.e.

$$\tilde{T}\tilde{f}(R) = \tilde{f}(TR) \quad \text{for } R \in \mathcal{S},\ \tilde{f} \in C(\mathcal{S}).$$

First, we show that this operator is mean ergodic with projection $\tilde{P}$ defined as

$$\tilde{P}\tilde{f}(R) = \tilde{f}(1\!\!1 \otimes 1\!\!1) \quad \text{for } R \in \mathcal{S},\ \tilde{f} \in C(\mathcal{S}):$$

Since multiplication by $T$ is (uniformly) continuous on $\mathcal{S}$, the mapping from $\mathcal{S}$ into $\mathcal{L}(C(\mathcal{S}))$ which associates to every $R \in \mathcal{S}$ its rotation operator $\tilde{R}$ is well defined. Consider a sequence $(S_k)_{k \in \mathbb{N}}$ in $\mathcal{S}$ converging to $S$.

Then $\tilde{S}_k\tilde{f}(R) = \tilde{f}(S_k R)$ converges to $\tilde{f}(SR) = \tilde{S}\tilde{f}(R)$ for all $R \in \mathcal{S}$, $\tilde{f} \in C(\mathcal{S})$. 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 $\mathcal{L}_w(C(\mathcal{S}))$, and the mapping $S \to \tilde{S}$ is continuous from $\mathcal{S}$ into $\mathcal{L}_w(C(\mathcal{S}))$. Therefore, from $T^{n_i} \longrightarrow 1\!\!1 \otimes 1\!\!1$ we obtain $\tilde{T}^{n_i} \to \widetilde{(1\!\!1 \otimes 1\!\!1)} = \tilde{P} \in \mathcal{L}_w(C(\mathcal{S}))$. Applying (IV.4.d) we conclude that the Cesaro means of $\tilde{T}^n$ converge strongly to $\tilde{P}$.

Take now $f \in E$, $g \in E'$ and define a continuous function $\tilde{f} \in C(\mathcal{S})$ by

$$\tilde{f}(R) := |\langle Rf,h\rangle - \langle f,1\!\!1\rangle \cdot \langle 1\!\!1,g\rangle| \quad .$$

Obviously, we have $\tilde{P}\tilde{f}(T) = \tilde{f}(1\!\!1 \otimes 1\!\!1) = 0$. Therefore

$$0 = \lim_{n\to\infty} \tilde{T}_n\tilde{f}(T) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |\langle T^i f,g\rangle - \langle f,1\!\!1\rangle \cdot \langle 1\!\!1,g\rangle| \quad .$$

(a") ⇒ (a ): Since $\mathcal{S}$ is metrizable and compact for the topology induced from $\mathcal{L}_w(E)$, 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 $\mathcal{S}$. By the assumption (a") and by (App.E.2) for every pair $(k,l)$ we obtain a subsequence

$$\{n_i\}^{k,l} \subset \mathbb{N}$$

with density 1, such that

$$\langle T^{n_i} f_k, g_l \rangle \longrightarrow \langle f_k, \mathbb{1} \rangle \cdot \langle \mathbb{1}, g_l \rangle \quad .$$

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.

<u>(a') ⇒ (a)</u> 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 $\mathcal{J}$ is compact. Therefore, these topologies coincide on $\mathcal{J}$.

<u>(c)⇒(f)</u>: Take $A, A', B, B' \in \Sigma$. For a suitable but fixed subsequence $(n_i) \subset \mathbb{N}$ $\mu(\varphi^{-n_i} A \cap B)$, 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 $\mu(A) \cdot \mu(B) \cdot \mu(A') \cdot \mu(B') = (\mu \otimes \mu)(A \times A') \cdot (\mu \otimes \mu)(B \times B')$. 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.

<u>(f)⇒(e)</u> is clear.

<u>(e)⇒(d)</u>: Assume that $T_\varphi f = \lambda f$, $|\lambda| = 1$, for $0 \neq f \in L^1(\mu)$. Then we have $T_\varphi \overline{f} = \overline{\lambda} \; \overline{f}$ and, for the function $f \otimes \overline{f}$: $(x,y) \mapsto f(x) \cdot \overline{f(y)}$, $(x,y) \in X \times X$, we obtain $T_{\varphi \otimes \varphi}(f \otimes \overline{f}) = \lambda f \otimes \overline{\lambda} \, \overline{f} = |\lambda|^2 (f \otimes \overline{f}) = f \otimes \overline{f}$. But 1 is a simple eigenvalue of $T_{\varphi \otimes \varphi}$ with eigenvector $\mathbb{1}_X \otimes \mathbb{1}_X$. Therefore we conclude $f = c \, \mathbb{1}_X$ and $\lambda = 1$, i.e. $\varphi$ is weakly mixing. ■

IX. 6 <u>Examples</u>:

While it is easy to find MDSs which are ergodic but not weakly mixing (e.g. the rotation $\varphi_a$, $a^n \neq 1$ for all $n \in \mathbb{N}$, on the circle $\Gamma$ has all powers of a as eigenvalues of $T_{\varphi_a}$), 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 $B(p_o,\dots,p_{k-1})$ 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 decribed 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 $X' := X \times \{0, 1\}$, $\Sigma'$ the obvious $\sigma$-algebra on $X'$ and $\mu'$ 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. $$\varphi'(x,j) := \begin{cases} (\varphi(x),j) & \text{for } x \in X \setminus S \\ (\varphi(x),1-j) & \text{for } x \in S, \end{cases}$$

for some fixed $S \in \Sigma$ with $0 < \mu(S) < 1$.

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:

(d) $(T^n f \mid f) \longrightarrow (f, \mathbb{1})^2$ for all $f \in L^\infty(X,\Sigma,\mu)$, where $(.\mid.)$ denotes the scalar product in $L^2(X,\Sigma,\mu)$.

(e) $\frac{1}{n}\sum_{i=0}^{n-1} T^{k_i} \longrightarrow \mathbb{1} \otimes \mathbb{1}$ in the weak operator topology for every subsequence $(k_i) \subset \mathbb{N}$.

Proof:

(d) $\Rightarrow$ (a): By (App.B.15) it suffices to show that $\langle T^n f, g\rangle$ converge to $\langle f, 1\!\!1\rangle \cdot \langle 1\!\!1, g\rangle$ 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

$$E_o := \overline{\text{lin}} \{1\!\!1, f, Tf, T^2f, \ldots\} \subset L^2(\mu).$$

The assertion is trivial for $g \in E_o^\perp$ 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 Cesaro sense. ∎

Certainly, the equivalence of (b) and (e) remains valid under much more general circumstances. But for operators induced by an MDS the <u>weak</u> operator convergence of $\frac{1}{n} \sum_{i=1}^{n} T^{k_i}$ as in (e) is equivalent to the <u>strong</u> operator convergence of these averages. This surprising result will be discussed in (IX.D.5).

IX. D.3 <u>Strong operator convergence of $T^n$</u>:

One of the striking features of the mean ergodic theorem (IV.4) is the equivalence of strong and weak operator convergence for the Cesaro means $T_n$. Since the weak operator convergence of the powers $T^n$ of the operator induced by an ergodic MDS $(X, \Sigma, \mu; \varphi)$ characterizes "strong mixing" one might think of introducing a possibly stronger mixing property by requiring $T^n$ to converge in the strong operator topology in $\mathcal{L}(L^1(\mu))$. But the limit operator is necessarily the one-dimensional projection $P = 1\!\!1 \otimes 1\!\!1$ and $T_\varphi$ is an isometry. For every characteristic function $1\!\!1_A$, $A \in \Sigma$, this implies

$$\| T_\varphi^n 1\!\!1_A - P\, 1\!\!1_A \| = \| 1\!\!1_A - \mu(A) \cdot 1\!\!1 \| = 2\mu(A) \cdot \mu(X \setminus A).$$

Therefore, as soon as $(X, \Sigma, \mu; \varphi)$ is ergodic and $0 < \mu(A) < 1$ for some $A \in \Sigma$, the powers of the induced operator $T_\varphi$ will not converge in the strong operator topology.

However, for positive operators $T \in \mathcal{L}(L^1(\mu))$ not induced by a transformation $\varphi: X \to X$ the strong operator and even the uniform convergence of T is an interesting and important property and will be investigated in App. X .

IX. D.4 Weak mixing implies "strong ergodicity":

Let $(X,\Sigma,\mu;\varphi)$ be an ergodic MDS. In general, its n-th power $(X,\Sigma,\mu;\varphi^n)$ is no longer ergodic. By the spectral mapping theorem and by (III.4) this is determined by the existence of an n-th root of unity different from 1 contained in the point spectrum of T. Therefore, a cyclic permutation of a set of n elements, n prime, yields an MDS such that $\varphi, \varphi^2, \dots, \varphi^{n-1}$ are ergodic but $\varphi^n$ is not.

On the other hand, the rotation $\varphi_a$, $a^n \neq 1$ for all $n \in \mathbb{N}$, on the unit circle has point spectrum $\{a^n : n \in \mathbb{Z}\}$ (see VIII.D.1), hence contains no non-trivial root of unity. Consequently, every power of $\varphi_a$ is ergodic. Such systems may be called "strongly ergodic". Clearly, every weakly mixing MDS is strongly ergodic, and the above rotation shows that the converse is not true.

But of course, the two notions coincide for finite-dimensional MDSs. This statement becomes somewhat less trivial, if we generalize it to operators not necessarily induced by an MDS. Such generalizations will be investigated in detail in (IX.D.6). Here we restrict ourselves to finite-dimensional spaces and consider the following class of matrices:

A positive matrix $T \in \mathcal{L}(\mathbb{C}^n)$, $n \in \mathbb{N}$, is called stochastic if $T\mathbb{1} = \mathbb{1}$ (i.e. each row sum is 1) and bi-stochastic if the same holds also for the transpose $T^t$ of T (i.e. each column sum of T is 1, too).

For example, every permutation matrix is bi-stochastic as is the

$$n \times n \text{ - matrix } \quad J_n := \begin{pmatrix} n^{-1} & \dots & n^{-1} \\ \vdots & & \vdots \\ n^{-1} & \dots & n^{-1} \end{pmatrix} .$$

Now, consider $\mathbb{C}^n$ as $L^1(X,\Sigma,\mu)$, where $X = \{1,\dots,n\}$, $\Sigma$ the power set of X and $\mu(\{x\}) = n^{-1}$ for $x \in X$. Then $J_n = \mathbb{1} \otimes \mathbb{1}$ as an operator on $\mathbb{C}^n$, and we present the following result showing the equivalence of the properties analogous to strong ergodicity, weak mixing and strong mixing. The "irreducibility" of a matrix generalizes "ergodicity" and has been defined in (IV.D.6) (see also Schaefer 1974, I.6 and (III.D.11)).

Proposition: For a bi-stochastic matrix $T \in \mathcal{L}(\mathbb{C}^n)$ the following are equivalent:

(i) $T^k \longrightarrow 1\!\!1 \otimes 1\!\!1$ for $k \to \infty$.
(ii) 1 is a simple and the unique unimodular eigenvalue of $T$.
(iii) $T$ is irreducible for all $k \in \mathbb{N}$.

Proof:
(i) $\Rightarrow$ (iii): If $T^k$ were reducible for some $k \in \mathbb{N}$ the same would be true for $T^{kr}$ for all $r \in \mathbb{N}$ and - if the limit exists - for $\lim_{r\to\infty} T^{kr}$. But $\lim_{k\to\infty} T^k = 1\!\!1 \otimes 1\!\!1$ is not reducible.
(iii) $\Rightarrow$ (ii): Since $T$ is irreducible, 1 is a simple eigenvalue. Assume that there exists a unimodular eigenvalue $\lambda \neq 1$.
By the theorem of Frobenius (Schaefer 1974, I.6.5) $\lambda$ is a m-th root of unity for some $m \in \mathbb{N}$, and therefore $T^m$ has at least two linearly independent fixed vectors, i.e. $T^m$ is not irreducible (use IV.D.7).
(ii) $\Rightarrow$ (i): By (IV.5) and (IV.3.ii/iv) $T$ is mean ergodic with corresponding projection $P = 1\!\!1 \otimes 1\!\!1$. Define $S := T - P$ and observe that $PS = 0 = SP$.
Therefore

$$T^k = (P + S)^k = P + S^k .$$

If $Sf = \alpha f$ for $\alpha \in \mathbb{C}$ with $|\alpha| \geq 1$, we obtain $\alpha Pf = PSf = 0$, hence $Pf = 0$ and $Tf = \alpha f$. For $\alpha = 1$ we conclude $Tf = f = Pf = 0$; for $\alpha \neq 1$ again we conclude $f = 0$ by (ii) and $r(T) = 1$. Therefore the spectrum $\sigma(S)$ is contained in $\{\lambda \in \mathbb{C} : |\lambda| < 1\}$ which implies $S^k \to 0$, i.e. $T^k \to P$ for $k \to \infty$. ■

Remark:
For a stochastic matrix $T \in \mathcal{L}(\mathbb{C}^n)$ we always have an invariant measure $p = (p_1,\dots,p_n)$ on $(X,\Sigma)$. If we consider $T$ as an operator on $L^1(X,\Sigma,p)$ we again have $T1\!\!1 = 1\!\!1$ and $T'1\!\!1 = 1\!\!1$ although the adjoint $T'$ of $T$ does no longer correspond to the transpose $T^t$ of the matrix $T$. Thus the equivalences of the proposition above hold equally for a stochastic matrix $T \in \mathcal{L}(L^1(p))$. The only specialty of bi-stochastic matrices is the fact that their invariant measure is the equidistribution $\mu$ on $X = \{1,\dots,n\}$.

IX. D.5 <u>Weak convergence implies strong convergence of averages</u>:
As we have seen in (IX.1), (IX.2) and (IX.5) mixing of an MDS $(X,\Sigma,\mu;\varphi)$ is a functional-analytic property of the induced $L^p$-operator $T_\varphi$. This observation has lead to a close investigation of

the asymptotic behavior of the powers of a bounded linear operator T on a Banach space E (see App. X). For example, it is clear that the powers of $T \in \mathcal{L}(E)$ converge - for the weak operator topology and necessarily to a projection P - if and only if the averages

$$\frac{1}{N} \sum_{i=1}^{N} T^{n_i}$$

converge to P for all infinite subsequences $(n_i)_{i \in \mathbb{N}}$ of $\mathbb{N}$ (see IX.D.2). But due to the special geometry of Hilbert spaces we have the following more surprising result.

Theorem 1 (Blum - Hanson, 1960):
Let T be a contraction on a Hilbert space H. Then T is mean ergodic with corresponding projection P, and the following are equivalent:

(a) $T^n$ converges to P in the weak operator topology.

(b) For all infinite subsequences $(n_i)_{i \in \mathbb{N}}$ of $\mathbb{N}$, the averages $\frac{1}{N} \sum_{i=1}^{N} T^{n_i}$ converge to P in the strong operator topology.

:. We refer to **Schaefer 1974 ,V.8.5 for a short proof.**

Corollary:
Let $(X, \Sigma, \mu)$ be a finite measure space and let $T \in \mathcal{L}(L^p(\mu))$, $1 \leq p < \infty$, be a positive operator satisfying $T \mathbb{1} \leq \mathbb{1}$ and $T' \mathbb{1} \leq \mathbb{1}$ . Then the assertions (a) and (b) above are equivalent.

Proof:
By the Riesz convexity theorem, T induces a contraction $T_2$ on $L^2(\mu)$. By (VII.D.4) T satisfies (a) if and only if $T_2$ satisfies (a). The same holds for property (b) since all $L^p$-topologies, $1 \leq p < \infty$ , coincide on $[-\mathbb{1}, \mathbb{1}]$ which is a total subset in $L^p(\mu)$.
Therefore, the $L^p$-result follows from Theorem 1 applied to $T_2$ on $L^2(\mu)$. ∎

The equivalence of the properties (a) and (b) has been investigated in great detail, and we quote one of the most beautiful results. As for the proof see **Schaefer 1974 ,V.8.7.**

Theorem 2 (Akcoglu - Sucheston, 1972):
For any contraction T on $L^1(\mu)$ the assertions (a) and (b) above are equivalent.

References: Akcoglu-Huneke-Rost [1974], Akcoglu-Sucheston [1972], [1975a], [1975b], Bellow [1975], Blum-Hanson [1960], Fong-Sucheston [1974], Krengel-Sucheston [1969], Lin [1981], Nagel [1974], Sato [1980].

IX. D.6 Weak mixing in Banach spaces:

Most of the properties characterizing weak mixing in Theorem (IX.5) are of purely functional-analytic nature and can be formulated for arbitrary operators on Banach spaces. In fact, the proof of the equivalence of (a), (a'), (a") and (d) in (IX.5) is essentially based on the relative compactness for the weak operator topology of the semigroup generated by the operator $T_{\varphi}$. The only ingredient from measure-theoretical ergodic theory is the special form of the limit operator $P = 1\!\!1 \otimes 1\!\!1$. This again shows the usefulness of the functional-analytic and in particular operator-semigroup approach to ergodic theory.

In the following we shall formulate the "weak mixing theorem" in its natural Banach space context.

Theorem:

Let $T \in \mathcal{L}(E)$, E a Banach space, and assume that

$$\mathcal{S} := \overline{\{T^n : n \in \mathbb{N}\}}$$

is compact and metrizable in $\mathcal{L}_w(E)$. Then T is mean ergodic with projection P and the following properties are equivalent:

(a) At most 1 is a unimodular eigenvalue of T.

(b) $P \in \mathcal{S}$.

(c) $T^{n_i} \to P$ for the weak operator topology and for some subsequence $(n_i) \subset \mathbb{N}$ having density 1.

(d) $\frac{1}{n} \sum_{i=1}^{n} |\langle T^i f, f' \rangle - \langle Pf, f' \rangle| \to 0$ for every $f \in E$, $f' \in E'$.

(e) $\frac{1}{m} \sum_{k=1}^{m} T^{n_k} \to P$ for the strong operator topology and for all subsequences $(n_k) \subset \mathbb{N}$ with lower density $> 0$.

Proof:

The properties (a), (b), (c) and (d) appeared already in Theorem (IX.5) (with $P = 1\!\!1 \otimes 1\!\!1$) and the proof of their equivalence given there extends verbatim to the situation treated here. Only condition (e) is new.

(e) $\Rightarrow$(d): Since the assertion is trivial for the elements of the fixed space of T, we may assume that P = 0. If (d) is not satisfied, we can find $f \in E$, $f' \in E'$, $\varepsilon > 0$ and a subsequence $(n_i) \subset \mathbb{N}$ with upper density $\alpha > 0$ such that the following is true:

for the sequence $\quad a_{n_i} := \langle T^{n_i} f, f' \rangle$

we have $\quad \operatorname{Re} a_{n_i} > \varepsilon \quad$ or $\quad -\operatorname{Re} a_{n_i} > \varepsilon$

or $\quad \operatorname{Im} a_{n_i} > \varepsilon \quad$ or $\quad -\operatorname{Im} a_{n_i} > \varepsilon$

for every $i \in \mathbb{N}$. Let us assume that $\operatorname{Re} a_{n_i} >$ for every $i \in \mathbb{N}$.

Now define another subsequence $(m_k)_{k \in \mathbb{N}} \subset \mathbb{N}$ as the union of $(n_i)_{i \in \mathbb{N}}$ and $(jn_o)_{j \in \mathbb{N}}$ for some fixed $n_o \in \mathbb{N}$. Then $(m_k)$ has lower density $\geq \frac{1}{n_o} > 0$, but for infinitely many $k_o \in \mathbb{N}$ we have

$$\begin{aligned} \operatorname{Re} \langle \frac{1}{k_o} \sum_{k=1}^{k_o} T^{m_k} f, f' \rangle &= \frac{1}{k_o} \sum_{k=1}^{k_o} \operatorname{Re} a_{m_k} \\ &\geq \frac{1}{k_o} \left( \frac{\alpha}{2} m_{k_o} \varepsilon - \frac{m_{k_o}}{n_o} \cdot M \right). \\ &= \frac{m_{k_o}}{k_o} \left( \frac{\alpha}{2} \varepsilon - \frac{M}{n_o} \right), \end{aligned}$$

where $M := \sup \{ |\langle T^n f, f' \rangle| : n \in \mathbb{N} \}$.
Since $n_o$ may be chosen arbitrarily large this contradicts (e).

(d) $\Rightarrow$ (e): Again, we assume P = 0 and renorm the Banach space E so as to obtain $\|T\| \leq 1$ (Hint: $\|g\| := \sup_{n \in \mathbb{N}} \|T^n g\|$ ). Take $f \in E$ fixed and consider the dual unit ball $U^o := \{ g' \in E' : \|g'\| \leq 1 \}$.

Then $$\tilde{T}\tilde{g}(g') := \tilde{g}(T'g'), \quad \tilde{g} \in C(U^o),$$
defines a contraction $\tilde{T}$ on $C(U^o)$. The assumption (d) implies that for the function
$$\tilde{f}(g') := |\langle f, g' \rangle|$$
the means $\frac{1}{n} \sum_{i=1}^{n} \tilde{T}^i \tilde{f}$ converge to 0 pointwise on $U^o$. As explained in (V.D.5) pointwise convergence of Cesaro means to a continuous function already implies norm convergence, hence
$$\left\| \frac{1}{n} \sum_{i=1}^{n} \tilde{T}^i \tilde{f} \right\| = \sup_{g' \in U^o} \frac{1}{n} \sum_{i=1}^{n} |\langle T^i f, g' \rangle| \longrightarrow 0.$$
Now let $(n_k)$ be a subsequence with positive lower density, i.e.

$\frac{n_k}{k} \leq c < \infty$.

Then

$$\left\|\frac{1}{m}\sum_{k=1}^{m} T^{n_k}f\right\| \leq \sup_{g'\in U^o} \frac{1}{m}\sum_{k=1}^{m} |\langle T^{n_k}f,g'\rangle| \leq \sup_{g'\in U^o} \frac{1}{m}\sum_{i=1}^{n_m} |\langle T^{i}f,g'\rangle|$$

$$\leq c\cdot \sup_{g'\in U^o} \frac{1}{n_m}\sum_{i=1}^{n_m} |\langle T^{i}f,g'\rangle| \longrightarrow 0 . \blacksquare$$

Remark:

Property (e) may be illustrated by saying that weak convergence of the powers $T^n$ along a thick sequence is equivalent to strong convergence of the means as long as the powers over which we average do not get too sparse. This is a beautiful supplement to the Akcoglu-Sucheston theorem in (IX.D.5).

In the above theorem we described a certain convergence property of $T^n$ for $T \in \mathcal{L}(E)$ and , $n \to \infty$, thereby answering question (1) stated at the beginning of Lecture IX. While this was possible within the general framework of Banach spaces the second question ("equidistribution") has no natural generalization, since it is not clear which elements of an abstract Banach space should take over the role of the "constant functions".

It is here again that Banach lattices and positive operators naturally enter ergodic theory: Let $T \in \mathcal{L}(E)$ be a positive operator on a Banach lattice E. If T is irreducible (see III.D.11) and if the fixed space F is non-trivial, then $F = \langle u\rangle$ for some quasi-interior point $u \in E_+$. If, in addition, T is mean ergodic then its corresponding projection P is of the form $\mu \otimes u$ for some strictly positive $T'$-invariant linear form $\mu \in E_+$. This is an analogue of the projection $\mathbb{1} \otimes \mathbb{1}$ appearing in all statements on mixing MDSs in this lecture.

For an abstract version featuring all aspects of the weak mixing theorem (IX.5) we have to assure only the compactness of the generated semigroup. This will be achieved by assuming that E has order continuous norm.

Corollary:

Let E be a separable Banach lattice with order continuous norm and let $T \in \mathcal{L}(E)$ denote a positive contraction satisfying $Tu = u$ and $T'\mu = \mu$ for some quasi-interior element $u \in E_+$ and some strictly positive linear form $\mu \in E'_+$. Then the following properties are equivalent:

(a) $T$ is irreducible and 1 is the unique unimodular eigenvalue of $T$.

(b) $\mu \otimes u$ is contained in the closure of $\{T^n : n \in \mathbb{N}\}$ in $\mathcal{L}_w(E)$.

(c) $\frac{1}{n} \sum_{i=1}^{n} |\langle T^i f, f' \rangle - \langle f, \mu \rangle \cdot \langle u, f' \rangle| \longrightarrow 0$ for every $f \in E$ and $f' \in E'$.

This theory of mixing operators on Banach lattices is worked out in Nagel [1974] and Schaefer 1974 ,V.8.

References: Jones [1971], Jones-Lin [1976], [1980], Krengel [1972], Nagel [1974].

IX. D.7 Mixing in C(X):

Mixing properties such as "ergodicity" and "weak mixing" of an MDS $(X;\Sigma,\mu;\varphi)$ can be characterized by spectral properties of the induced operator $T_\varphi$ on $L^p(X,\Sigma,\mu)$, $1 \leq p < \infty$. The underlying reason is the weak compactness of the semigroup generated by $T_\varphi$. Since for a TDS $(X;\varphi)$ the semigroup

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

in $\mathcal{L}(C(X))$ rarely is relatively weakly compact, there is no hope of obtaining similar results in the topological context. For "minimality" this has already been observed in (III.8.i) and (III.D.8). But by choosing appropriate concepts it might still be possible to prove a correspondence between topological mixing properties and spectral properties of an associated linear operator.

Definition:

A TDS $(X;\varphi)$ is called topologically ergodic if $\varphi(A) = A$ for some closed subset $A$ of $X$ implies $A = X$ or $A = \emptyset$.
It can be shown that for a TDS $(X;\varphi)$ on a compact metric space $X$, topological ergodicity is equivalent to the properties appearing in (III.D.8).
Now we will try to find a function space on X such that $\varphi$ is topologically ergodic if and only if the corresponding fixed space of the induced operator is one-dimensional. To that purpose the space C(X) is too small as is shown by the example $X = [0,1]$ and $\varphi(x) := x^2$.
An appropriate Banach space can be obtained by the following procedure:

For every bounded complex valued function f on X we denote by C(f) the set of continuity points of f. Then define

$$D_o(X) := \{f: X \to \mathbb{C} \ : \ C(f) \text{ contains a countable intersection of open dense sets}\}$$

and

$$N := \{f \in D_o(X): f(x) = 0 \text{ on a countable intersection of open dense sets}\}.$$

$D_o(X)$ is canonically a Banach algebra and a Banach lattice for the sup-norm, and N is a closed ideal. Therefore

$$D(X) := D_o(X)/N$$

is a Banach algebra and a Banach lattice which can be characterized in the following way.

Lemma: Let X be a compact space. The following Banach lattices are canonically isomorphic:

(i) D(X).

(ii) The Dedekind completion of C(X).

(iii) C(Y) where Y is the Stone space of the Boolean algebra of all regularly closed subsets of X.

The proof of this lemma belongs to the theory of vector lattices and may be found in Nakano-Shimogaki [1962] (see Peressini [1967],p.159 or Semadeni [1971],p.267). Its importance for us lies in the fact that D(X) is a (more or less) concrete function space which is order-theoretically associated to C(X). In particular, the lattice isomorphism $T_\varphi$ on C(X) can canonically be extended to $\tilde{T}_\varphi$ on the Dedekind completion D(X). This permits the desired result.

Theorem:

For a TDS $(X;\varphi)$ the following assertions are equivalent:

(i) $(X;\varphi)$ is topologically ergodic.

(ii) There is no band B in C(X) satisfying $T_\varphi B = B$ except $B = C(X)$ and $B = \{0\}$.

(iii) There is no projection band $\tilde{B}$ in D(X) satisfying $\tilde{T}_\varphi \tilde{B} = \tilde{B}$ except $\tilde{B} = D(X)$ and $\tilde{B} = \{0\}$.

(iv) The fixed space $\tilde{F}$ of $\tilde{T}_\varphi$ in D(X) is one-dimensional.

Proof:

(ii) is just a different way of expressing (i) since bands in C(X) are of the form $B_A = \{f \in C(X): f(A) \subset \{0\}\}$ for some regularly closed subset A of X. From results on the Dedekind completion of vector lattices it follows that there is a bijection between bands in C(X) and projection bands in D(X), hence (ii) is equivalent to (iii).
Finally, the fixed space $\tilde{F}$ of $\tilde{T}_\varphi$ in D(X) is a vector sublattice of D(X) and contains the order unit $\mathbb{1}_X$. Therefore $\tilde{F}$ is one-dimensional if and only if $\mathbb{1}_X$ cannot be decomposed into a non-trivial sum of two orthogonal positive invariants elements, i.e. (iii)$\Longleftrightarrow$(iv). ∎

Using the Banach lattice D(X) and arguments similar to those in the measure-theoretical case one can introduce the notions of <u>topological weak</u>- and <u>strong mixing</u>, and these properties can be characterized by spectral properties in D(X). We refer to the papers of Keynes-Robertson [1968], [1969], for more information and close with the following observation.

<u>Proposition</u>:
An MDS $(X,\Sigma,\mu;\varphi)$ is ergodic iff the TDS $(\check{X};\psi)$ is topologically ergodic, where $\check{X}$ is the Stone space of the measure algebra $\Sigma$ and $\psi$ the homeomorphism induced by $\varphi$ (see VI.D.6).

References: Keynes-Robertson [1968], [1968], Nakano-Shimogaki [1962].

## 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 $n_o \in \mathbb{N}$) if $\varphi^{n_o} x = x$ (and $\varphi^n x \neq x$ for $n = 1, \dots, n_o - 1$).

(ii) The transformation $\varphi$ is periodic (with period $n_o \in \mathbb{N}$) if $\varphi^{n_o} = \mathrm{id}$ (and $\varphi^n \neq \mathrm{id}$ for $n = 1, \dots, n_o - 1$).

(iii) The transformation $\varphi$ 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 X.D.1).
   Then we obtain
   $$\{x \in X : \varphi^n x = x\} = \bigcap_{k \in \mathbb{N}} (B_k \cap \varphi^n B_k) \cup ((X \setminus B_k) \cap \varphi^n (X \setminus B_k))$$
   for every $n \in \mathbb{N}$, and therefore we conclude that $A_n \in \mathcal{B}$.
3. An arbitrary transformation $\varphi$ 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 $A_{ap} := X \setminus \bigcup_{n \in \mathbb{N}} A_n$. Then X is the disjoint union of the $\varphi$-invariant sets $A_n$, $n \in \mathbb{N}$, and $A_{ap}$. The restriction of $\varphi$ to $A_n$ is periodic with period n and $\varphi$ 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 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 $A, \varphi A, \varphi^2 A, \dots, \varphi^{n-1}A$ are pairwise disjoint and $m(A) = \frac{1}{n}$.

(ii) If $\varphi$ is antiperiodic then for every $n \in \mathbb{N}$ and $\varepsilon > 0$ there exists $A \in \mathcal{B}$ such that $A, \varphi A, \varphi^2 A, \dots, \varphi^{n-1}A$ are pairwise disjoint and $m(\bigcup_{k=0}^{n-1} \varphi^k A) > (1-\varepsilon)$.

Proof:

(i) If $n > 1$ there exists a measurable set $C_1$ such that $m(C_1 \Delta \varphi C_1) > 0$ (use the existence of a separating base) and therefore
$m(C_1 \setminus \varphi C_1) = m(C_1) - m(C_1 \cap \varphi C_1) = m(\varphi C_1) - m(\varphi C_1 \cap C_1) =$
$= m(\varphi C_1 \setminus C_1) > 0$. Certainly, $B_1 := C_1 \setminus \varphi C_1$ is disjoint from $\varphi B_1$.
If $n > 2$ there exists $C_2 \subset B_1$ such that $m(C_2 \Delta \varphi^2 C_2) > 0$. For $B_2 := C_2 \setminus \varphi^2 C_2$ 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 $B_{n-1}, \varphi B_{n-1}, \dots, \varphi^{n-1} B_{n-1}$ are pairwise disjoint.
Consider the measure algebra $\check{\mathcal{B}}$ and the equivalence classes $\check{B} \in \check{\mathcal{B}}$ of sets $B \in \mathcal{B}$ such that $B, \varphi B, \dots, \varphi^{n-1} B$ are pairwise disjoint. Since $\check{\mathcal{B}}$ is a complete Boolean algebra (see A. 9) an application of Zorn's lemma yields $\check{A} \in \check{\mathcal{B}}$ which is maximal such that $A, \varphi A, \dots \varphi^{n-1} A$ are pairwise disjoint for some $A \in \check{A}$.
If we assume $m(A) < \frac{1}{n}$ we can apply the above construction to the $\varphi$-invariant set $X \setminus \bigcup_{i=0}^{n-1} \varphi^i A$ and obtain a contradiction to the maximality of $\check{A}$.
Therefore, $m(A) = \frac{1}{n}$, and the assertion is proved.

(ii) We may take $\varepsilon = \frac{1}{p}$ for some $p \in \mathbb{N}$. For $r := np$ and as in the proof of (i) we construct $B \in \mathcal{B}$ such that $B, \varphi B, \dots, \varphi^{r-1} B$ are pairwise disjoint and such that B is maximal relative to this property.
For $1 \leq k \leq r$ define

$$B_k := \{x \in \varphi^{r-1} B : \varphi^k x \in B \text{ and } \varphi^j x \notin B \text{ for } 1 \leq j < k\}.$$

These sets are pairwise disjoint, and the same holds for $B_k, \varphi B_k, \dots, \varphi^k B_k$ for any $k = 1, \dots r$.
Therefore, the maximality of B implies

$$(*) \quad m(\varphi^{r-1} B \setminus \bigcup_{k=1}^{r} B_k) = 0.$$

Moreover, the sets

$$\begin{array}{l} \varphi B_2 \\ \varphi B_3,\ \varphi^2 B_3 \\ \varphi B_4, \varphi^2 B_4,\ \varphi^3 B_4 \\ \cdots\cdots\cdots\cdots \\ \varphi B_r,\ \varphi^2 B_r, \ldots\ldots, \varphi^{r-1} B_r \end{array}$$

are disjoint from any $\varphi^k B$ for $0 \leq k \leq r-1$, since

$$\varphi^i B_j \cap \varphi^k B = \varphi^i (B_j \cap \varphi^{k-i} B) \subseteq \varphi^i (\varphi^{r-1} B \cap \varphi^{k-i} B) = \emptyset$$

if $0 < i < j \leq r$ and $i \leq k$

(resp. $\varphi^i B_j \cap \varphi^k B = \varphi^k (\varphi^{i-k} B_j \cap B) = \emptyset$ 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

$$m\left(\bigcup_{k=1}^{r} \varphi^k B_k\right) = m\left(\bigcup_{k=1}^{r} B_k\right) = m(\varphi^{r-1} B) = m(B).$$

Now, consider
$$B^* := \bigcup_{k=0}^{r-1} \varphi^k B \cup \bigcup_{1 \leq i \leq j \leq n} \varphi^i B_j ,$$

which is $\varphi$-invariant modulo $m$-null sets. Since B is maximal and $\varphi$ 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 $\bigcup_{l=0}^{n-1} \varphi^l A$ contains every $\varphi^k B$, $0 \leq k \leq r-1$. From $B^* = X$ it follows that $X \setminus \bigcup_{l=0}^{n-1} \varphi^l A$ is contained in $\bigcup_{k=0}^{p-1} \bigcup_{0<i\leq j\leq n} \varphi^{kn+i} B_{kn+j}$. Therefore, we conclude that

$$m\left(X \setminus \bigcup_{l=0}^{n-1} \varphi^l A\right) \leq n \cdot m(B) \leq \frac{n}{r} = \varepsilon . \quad \blacksquare$$

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 $\tilde{\mathcal{G}}$ the group of all bi-measure-preserving bijections on $(X, \mathcal{B}, m)$. Here we identify transformations which coincide $m$-almost everywhere.

The set $\mathcal{G} := \{ T_\varphi : \varphi \in \tilde{\mathcal{G}} \}$ of all induced operators

$$T_\varphi : L^1(X, \mathcal{B}, m) \longrightarrow L^1(X, \mathcal{B}, m)$$

is a group in $\mathcal{L}(L^1(X, \mathcal{B}, m))$.

The following lemma shows that the map $\varphi \mapsto T_\varphi$ from $\tilde{\mathcal{G}}$ onto is a group isomorphism.

X. 3 <u>Lemma:</u> If $\varphi \in \tilde{\mathcal{G}}$ 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 $A_n$, $n \geq 2$, or $A_{ap}$ defined in Remarks 2, 3 following (X.1) has non-zero measure. By (X.2) we obtain a measurable set A such that $m(A) > 0$ and $A \cap \varphi(A) = \emptyset$. This yields

$$A \cap \varphi^{-1}(A) = \emptyset \quad \text{and} \quad T_\varphi \mathbb{1}_A = \mathbb{1}_{\varphi^{-1}(A)} \neq \mathbb{1}_A. \qquad \blacksquare$$

On $\mathcal{G}$ we consider the topology which is induced by the strong operator topology on $\mathcal{L}(L^1(m))$. This topology coincides on $\mathcal{G}$ with the topology of pointwise convergence on all characteristic functions $\mathbb{1}_{B_k}$, $k \in \mathbb{N}$, where $\{B_k: k \in \mathbb{N}\}$ generates $\mathcal{B}$ (use B.11), and will be transfered to $\tilde{\mathcal{G}}$. In particular, $T_{\varphi_i}$ converges to $T_\varphi$ (resp. $\varphi_i$ converges to $\varphi$) if and only if $m(\varphi_i(A) \triangle \varphi(A)) \longrightarrow 0$ for every $A \in \mathcal{B}$. Since the multiplication on bounded subsets of $\mathcal{L}(L^1(m))$ is continuous for the strong operator topology, $\mathcal{G}$ (and $\tilde{\mathcal{G}}$ ) is a topological group which is metrizable. In (X.D.3) we shall see that $\mathcal{G}$ is complete, hence $\mathcal{G}$ and $\tilde{\mathcal{G}}$ are complete metric spaces, and Baire's category theorem is applicable (see A.6).

X. 4 <u>Proposition:</u> 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 $\tilde{\mathcal{G}}$.

Proof:

Consider $\varphi \in \tilde{\mathcal{G}}$, $\varepsilon > 0$ and characteristic functions $\chi_1, \dots, \chi_m \in L^1(m)$. We shall construct $\psi \in \tilde{\mathcal{G}}$ with period larger than n such that

$$\| T_\varphi \chi_i - T_\psi \chi_i \| \leq 3\varepsilon \quad \text{for } i = 1, \dots, m.$$

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 $m(\bigcup_{j > l} A_j) < \frac{\varepsilon}{2}$. Defining $B := A_1 \cup \dots \cup A_l$ we observe that $\psi|_B$ 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 $C \subset A_{ap}$ such that $C, \varphi C, \dots, \varphi^{k-1} C$ are pairwise disjoint and

$$\frac{1}{k}\left(1 - \frac{\varepsilon}{2}\right) \cdot m(A_{ap}) \leq m(C) \leq \frac{1}{k}\, m(A_{ap}) .$$

The transformation $\psi \in \tilde{\mathcal{G}}$ defined as

$$\psi(x) := \begin{cases} \varphi(x) & \text{for } x \in B \cup C \cup \varphi C \ldots \cup \varphi^{k-2}C \\ \varphi^{1-k}(x) & \text{for } x \in \varphi^{k-1}C \\ x & \text{for all other } x \in X, \end{cases}$$

is periodic with period $k > n$. But, $\psi$ coincides with $\varphi$ outside of a set R with measure

$$m(R) \leq \frac{1}{k} m(A_{ap}) + \frac{\varepsilon}{2} m(A_{ap}) + \frac{\varepsilon}{2} \leq 3 \cdot \frac{\varepsilon}{2}.$$

Therefore, we conclude $\| T_\varphi x_i - T_\psi x_i \| \leq 2 \cdot m(R) \leq 3\varepsilon$ for $i=1,\ldots,m$. ∎

X. 5 <u>Theorem</u> (Rohlin, 1948): The set $\tilde{\mathcal{S}}$ of all strongly mixing transformations on $(X, \mathcal{B}, m)$ is of first category in $\tilde{\mathcal{G}}$.

Proof: Let $A := [0, \frac{1}{2}] \subset X$. For every $k \in \mathbb{N}$,

$$\tilde{\mathcal{M}}_k := \{ \varphi \in \tilde{\mathcal{G}} : | m(A \cap \varphi^k A) - \tfrac{1}{4} | \leq \tfrac{1}{5} \}$$

is closed. If $\varphi \in \tilde{\mathcal{G}}$ is strongly mixing, we have

$$\lim_{k\to\infty} m(A \cap \varphi^k A) = m(A)^2 = \frac{1}{4} \qquad \text{(by IX.2)},$$

hence $\varphi \in \tilde{\mathcal{M}}_k$ for all $k \geq k_o$, or

$$\tilde{\mathcal{S}} \subset \bigcup_{n\in\mathbb{N}} \tilde{\mathcal{N}}_n \quad \text{for} \quad \tilde{\mathcal{N}}_n := \bigcap_{k\geq n} \tilde{\mathcal{M}}_k .$$

Since $\tilde{\mathcal{N}}_n$ is closed, it remains to show that $\tilde{\mathcal{G}} \setminus \tilde{\mathcal{N}}_n$ is dense in $\tilde{\mathcal{G}}$. If $\varphi$ is periodic, say $\varphi^k = id$, then

$$m(A \cap \varphi^k A) - \frac{1}{4} = \frac{1}{4}, \text{ hence } \varphi \in \tilde{\mathcal{G}} \setminus \tilde{\mathcal{M}}_k .$$

Therefore, $\bigcup_{k\geq n} \{ \varphi \in \tilde{\mathcal{G}} : \varphi^k = id \} \subset \tilde{\mathcal{G}} \setminus \bigcap_{k\geq n} \tilde{\mathcal{M}}_k = \tilde{\mathcal{G}} \setminus \tilde{\mathcal{N}}_k$,

and the assertion follows from (X.4). ∎

X. 6 <u>Proposition:</u> The set $\tilde{\mathcal{W}}$ of all weakly mixing transformations on $(X, \mathcal{B}, m)$ is dense in $\tilde{\mathcal{G}}$.

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 <u>Theorem</u> (Halmos, 1944): The set $\tilde{\mathcal{W}}$ of all weakly mixing transformations on $(X, \mathcal{B}, m)$ is of second category in $\tilde{\mathcal{G}}$.

Proof:
Since $\tilde{\mathcal{G}}$ is a complete metric space, Baire's category theorem (see A.6) asserts that $\tilde{\mathcal{G}}$ is of second category. Therefore and by (X.6) it is enough to show that $\tilde{\mathcal{W}}$ is the intersection of a sequence of open sets. We prove this assertion for the (induced) operator sets $\mathcal{W} := \{T_\varphi \in \mathcal{L}(L^1(m)) : \varphi \in \tilde{\mathcal{W}}\}$. Let $\{f_i\}_{i \in \mathbb{N}}$ be a subset of $L^\infty(m)$ which is dense in $L^1(m)$. Define

$$\mathcal{W}_{ijkn} := \left\{T_\varphi \in \mathcal{G} : |\langle T^n f_i, f_j\rangle - \langle f_i, 1\rangle \langle f_j, 1\rangle| < \tfrac{1}{k}\right\} \text{ for } i,j,k,n \in \mathbb{N}.$$

By (X.D.2) the sets $\mathcal{W}_{ijkn}$ and therefore $\mathcal{W}_{ijk} := \bigcup_{n \in \mathbb{N}} \mathcal{W}_{ijkn}$ are open. We shall show that $\mathcal{W} = \bigcap_{i,j,k} \mathcal{W}_{ijk}$.

The inclusion $\mathcal{W} \subset \bigcap_{i,j,k} \mathcal{W}_{ijk}$ is obvious by (IX.5.a).

On the other hand, if $\varphi$ is not weakly mixing, then there exists a non-constant eigenvector $h \in L^1(m)$ of $T_\varphi$ with unimodular eigenvalue $\lambda$. It is possible to choose $h$ with $\|h\| = 1$ and $\langle h, 1\rangle = 0$. Now, choose $k \in \mathbb{N}$ such that $\|h - f_k\| \leq \frac{1}{10}$.
We obtain

$$|\langle T_\varphi^n f_k, f_k\rangle - \langle f_k, 1\rangle \langle f_k, 1\rangle| =$$

$$|\langle T_\varphi^n (f_k - h), (f_k - h)\rangle - \langle (f_k - h), 1\rangle \langle (f_k - h), 1\rangle + \langle T^n h, h\rangle| \geq \frac{1}{2}$$

for every $n \in \mathbb{N}$.
This yields $T_\varphi \notin \mathcal{W}_{kk2}$, and the theorem is proved. ∎

Combining (X.5) and (X.7) 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 $C := (I_j)_{j=1,..,q}$ of height $q$ is a $q$-tuple of disjoint intervals $I_j = [a_j, b_j) \subset [0,1)$ of equal length.

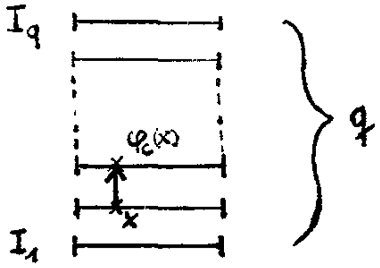
(ii) With a column $C$ there is associated a piecewise linear mapping

$$\varphi_C : \bigcup_{j=1}^{q-1} I_j \longrightarrow \bigcup_{j=2}^{q} I_j \quad \text{defined by}$$

$$\varphi_C(x) := (x - a_j) + a_{j+1} \quad \text{for } x \in I_j .$$

Remark:

A column is represented diagrammatically as follows:




Therefore the mapping $\varphi_C$ moves a point $x \in I_j$, $j \leq q-1$, 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 $C(n) = (I_j(n))_{j=1,..,q(n)}$ of columns. Then we use the associated mappings $\varphi_n := \varphi_{C(n)}$ to define $\varphi$ 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

$$C(1) := ([0, \tfrac{1}{4}), [\tfrac{1}{4}, \tfrac{1}{2}), [\tfrac{1}{2}, \tfrac{3}{4})) \quad \text{and } R(1) := [\tfrac{3}{4}, 1) .$$

In this way we proceed! More precisely, from $I_j(n) = [a_j(n), b_j(n)) \in C(n)$ we produce

$$I'_j(n) := \left[a_j(n), \frac{a_j(n) + b_j(n)}{2}\right)$$

and

$$I''_j(n) := \left[\frac{a_j(n) + b_j(n)}{2}, b_j(n)\right) ,$$

and from $R(n)$ we produce

$$R'(n) := \left[b_{q(n)}(n), \frac{b_{q(n)}(n) + 1}{2}\right)$$

and

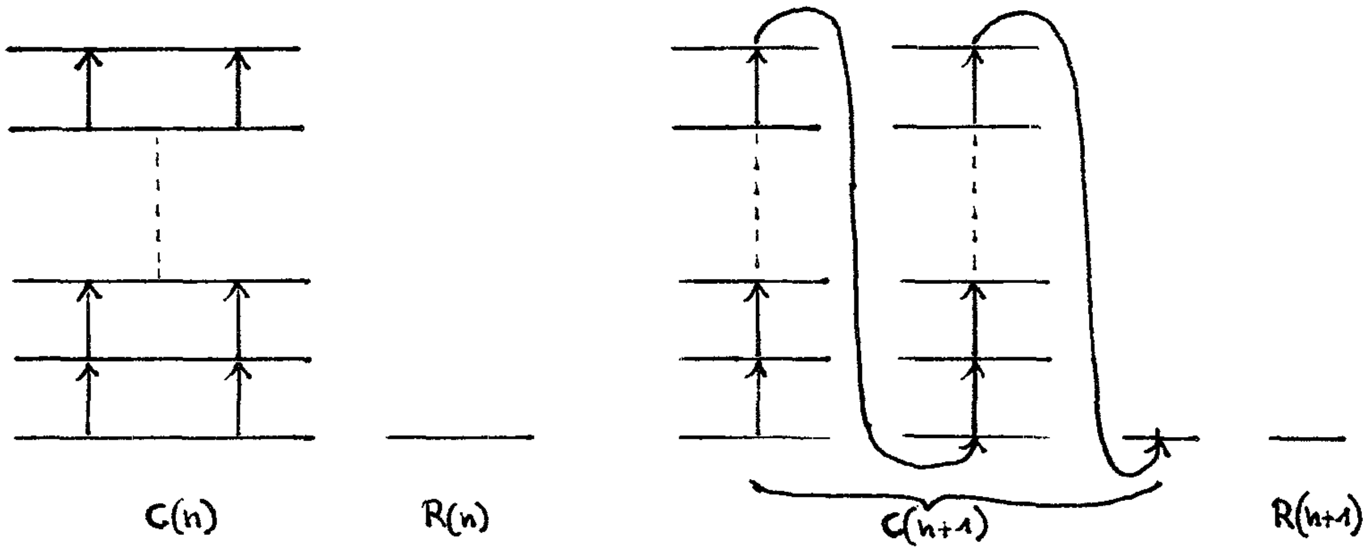
$$R''(n) := \left[\frac{b_{q(n)}(n) + 1}{2}, 1\right) .$$

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:

(0) $m(R(n)) = 2^{-(n+1)}$ converges to zero as n tends to infinity.

(1) Every interval $I_j(n) \in C(n)$ is a union of intervals in C(n+1).

(2) The $\sigma$-algebra $\sigma(\bigcup_{n=k}^{\infty} \bigcup_{j=1}^{q(n)} I_j(n))$, $k \in \mathbb{N}$, is equal to the Borel algebra $\mathcal{B}$ .

(3) The mapping $\varphi_{n+1}$ is an extension of $\varphi_n$.

(4) For every $x \in [0, 1)$ there exists $n_o \in \mathbb{N}$ such that

$$\varphi(x) := \varphi_n(x), \; n \geq n_o,$$

is defined.

Now: $(X, \mathcal{B}, m; \varphi)$ is an MDS if we take $\varphi$ as the mapping just defined on $X = [0, 1)$.

X. 9 <u>Theorem:</u> 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 $\varphi$ it follows $m(\varphi^{-q(n)}(I_j(n)) \cap I_j(n)) \geq \frac{1}{2} m(I_j(n))$.
Therefore

$$m(\varphi^{-q(n)}(A) \cap A) \geq \frac{1}{2} m(A) = \frac{1}{8} \text{ for every } n \in \mathbb{N}.$$

But if $\varphi$ 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 in 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) $(X, \mathcal{B}, m; \varphi)$ is ergodic: Assume $\varphi(A) = A \in \mathcal{B}$. This implies for any $j = 1, \dots, q(n)$ 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

$$\begin{aligned} 0 &= \lim_{n\to\infty} m(A(n) \triangle A) = \lim_{n\to\infty} (m(A(n)) + m(A) - 2\, m(A(n) \cap A)) \\ &= \lim_{n\to\infty} (m(A(n)) + m(A) - 2\, q(n)^{-1} |L_{n,A}| (m(A) - m(R(n) \cap A))) \\ &= m(A) + m(A) - 2\, m(A) \cdot m(A) \\ &= 2 m(A)(1 - m(A)) \end{aligned}$$

proves that $m(A) = 0$ or $m(A) = 1$, i.e. $\varphi$ is ergodic.

3) Finally, it remains to show that 1 is the only eigenvalue of the induced operator $T_\varphi$ (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 $L \subseteq \{1, \dots, q(n)\}$ such that $D' := \bigcup_{i \in L} I_i(n)$ satisfies $D \subseteq D'$ and $m(D' \setminus D) < \varepsilon \cdot m(D) \leq \varepsilon \cdot m(D')$ and $m(I_i(n)) < \delta$ for $i \in L$.
Now, define $I := I_j(n) \cap D$
where $m(I_j(n) \setminus D) < \varepsilon \cdot m(I_j(n))$ for a suitable $j \in L$.
From the definition of $\varphi$ it follows that

$$m(\varphi^{q(n)}(I_j(n)) \cap I_j(n)) \geq \frac{1}{2} m(I_j(n))$$

and

$$m(\varphi^{q(n)+1}(I_j(n)) \cap I_j(n)) \geq \frac{1}{4} m(I_j(n)).$$

Therefore, we conclude that

$$\begin{aligned} &m(\varphi^{q(n)}(I) \cap I) = m(\varphi^{q(n)}(I_j(n)) \cap I) - m(\varphi^{q(n)}(I_j(n) \setminus D) \cap I) \\ &\geq m(\varphi^{q(n)}(I_j(n)) \cap I_j(n)) - m(\varphi^{q(n)}(I_j(n)) \cap (I_j(n) \setminus D)) - \varepsilon \cdot m(I_j(n)) \\ &\geq \frac{1}{2} m(I_j(n)) - 2\varepsilon \cdot m(I_j(n)) > 0 \end{aligned}$$

and analogously

$$m(\varphi^{q(n)+1}(I) \cap I) \geq \frac{1}{4} m(I_j(n)) - 2\varepsilon \cdot m(I_j(n)) > 0.$$

If $x = \varphi^{q(n)}(y) \in \varphi^{q(n)}(I) \cap I$ we obtain

$$f(x) = f(\varphi^{q(n)}(y)) = \lambda^{q(n)} f(y) \quad \text{and} \quad |f(x) - f(y)| < \varepsilon .$$

If $x' = \varphi^{q(n)+1}(y') \in \varphi^{q(n)+1}(I) \cap I$ we obtain

$$f(x') = f(\varphi^{q(n)+1}(y')) = \lambda^{q(n)+1} f(y') \text{ and } |f(x') - f(y')| < \varepsilon .$$

Finally,

$$\lambda = \frac{\lambda^{q(n)+1}}{\lambda^{q(n)}} = \frac{f(x')}{f(y')} \cdot \frac{f(y)}{f(x)}$$

implies

$$|\lambda - 1| \leq \left| \frac{f(x')}{f(y')} \cdot \left(\frac{f(y)}{f(x)} - 1\right)\right| + \left|\frac{f(x')}{f(y')} - 1\right| \leq 2\varepsilon,$$

which proves that 1 is the only eigenvalue of $T_\varphi$. ■

## X. D <u>Discussion</u>

### X. D.1 <u>Separating bases:</u>

The results of this lecture have been formulated for the particular measure space $([0,1], \mathcal{B}, m)$. They remain true for more general spaces $(Y, \Sigma, \mu)$, but an essential hypothesis for defining (X.1) and proving (X.3) is the existence of a <u>separating base</u>, i.e. a sequence $\{B_n\}_{n \in \mathbb{N}}$ of measurable sets generating $\Sigma$ such that for different $x, y \in Y$, there exists $k \in \mathbb{N}$ with $x \in B_k$, $y \notin B_k$ or $x \notin B_k$, $y \in B_k$.

As a trivial counterexample to (X.3) we consider $(Y, \Sigma, \mu)$ where $Y = \{a,b\}$, $\Sigma = \{\emptyset, Y\}$, $\mu(Y) = 1$ and $\varphi : Y \to Y$ defined by $\varphi(a) = b$ and $\varphi(b) = a$. Then

$$\mu\{x \in Y : \varphi(x) \neq x\} = 1,$$

i.e. $\varphi \neq id$ in $\tilde{\mathcal{G}}$, but the induced operator $T_\varphi$ is the identity operator on $L^1(\mu)$.

Since it can be proved that every "Lebesgue space" is isomorphic to $([0, 1], \mathcal{B}, m)$ (see X.D.5), the category results (X.5) and (X.7) are valid for a large class of probability spaces.

### X. D.2 <u>Topologies on operator groups:</u>

Let $\tilde{\mathcal{G}}$ be the group of all bi-measure-preserving transformations on $([0, 1], \mathcal{B}, m)$, $\mathcal{B}$ the Borel algebra and $m$ the Lebesgue measure. On

$$\mathcal{G} = \{ T_\varphi \in \mathcal{L}(L^1(m)) : \varphi \in \tilde{\mathcal{G}} \}$$

we can consider three natural topologies: the norm topology, the strong and the weak operator topology, all induced from $\mathcal{L}(L^1(m))$.

While we used the strong operator topology in order to obtain the category theorems (X.5) and (X.7), the norm topology does not yield interesting results.

<u>Proposition 1:</u> $\mathcal{G}$ is discrete for the norm topology.

Proof:

Choose $\varphi \in \tilde{\mathcal{G}}$ with $\varphi \neq \mathrm{id}$. As in the proof of (X.3) we obtain a set $A \in \mathcal{B}$ such that $m(A) > 0$ and $T_\varphi 1_A \wedge 1_A = 0$. Defining $f := m(A)^{-1} \cdot 1_a$ we conclude that

$$\| T_\varphi - \mathrm{Id} \| \geq \| T_\varphi f - f \| = m(A)^{-1} \| T_\varphi 1_A - 1_A \| = 2.$$

Since $\mathcal{G}$ is a topological group, the proof is complete. ■

The weak operator topology is more interesting but coincides on $\mathcal{G}$ with the strong operator topology, a fact that has already been used in the proof of (X.7).

<u>Proposition 2</u> : The weak and strong operator topologies coincide on $\mathcal{G}$.

Proof:

Choose $T_n, T \in \mathcal{G} \subseteq \mathcal{L}(L^1(m))$ and assume that $T_n \longrightarrow T$ in the weak operator topology. The seperate continuity of multiplication yields $T^{-1}T_n \longrightarrow \mathrm{Id}$ in this topology. For $0 \leq f \in L^\infty(m)$ we observe that

$$(T^{-1}T_n f \mid f) \longrightarrow (f \mid f)$$

where $(\cdot|\cdot)$ denotes the scalar product on $L^2(m)$.

Since $T_n$ and $T$ induce isometries on $L^2(m)$ we conclude that

$$\begin{aligned} \|T_n f - Tf\|_2^2 &= \| T^{-1}T_n f - f \|_2^2 \\ &= ( T^{-1}T_n f - f \mid T^{-1}T_n f - f ) \\ &= 2 (f \mid f) - 2 (T^{-1}T_n f \mid f) \longrightarrow 0 \end{aligned}$$

as $n \to \infty$. By Hölder's inequality we obtain

$$\| T_n f - Tf \|_1 \longrightarrow 0.$$

An application of (B.11) yields that $T_n$ converges to $T$ in the strong operator topology on $\mathcal{G}$. ■

We remark that, in contrast to Proposition 2, $\mathcal{G}$ as a subset of $\mathcal{L}(L^1(m))$ is closed in the strong but not in the weak operator topology (see also X.D.3).

X. D.3 <u>The group of all bi-measure-preserving transformations is complete:</u>

The group $\tilde{\mathcal{G}}$ of all bi-measure-preserving transformations of a probability space $(X, \Sigma, \mu)$ can be topologized by the following subbase of open sets $U(\varphi, A, \varepsilon) := \{\psi \in \tilde{\mathcal{G}} : \mu(\psi^{-1}A \,\Delta\, \varphi^{-1}A) < \varepsilon\}$ for $\varphi \in \tilde{\mathcal{G}}$, $A \in \Sigma$, $\varepsilon > 0$. In order to prove the assertion stated in the title above, let us assume that $\tilde{\mathcal{G}}$ is isomorphic to the group of all algebra isomorphisms of the measure algebra $\check{\Sigma}$, i.e. that every algebra isomorphism of $\check{\Sigma}$ is induced by a bi-measure-preserving transformation of X, and that different elements in $\tilde{\mathcal{G}}$ induce different algebra isomorphisms (see X.3). In (VI.2) we have seen that every algebra isomorphism $\check{\varphi}$ of $\check{\Sigma}$ induces an isometric lattice isomorphism T of $L^1(X, \Sigma, \mu)$ satisfying $T\mathbb{1} = \mathbb{1}$, and conversely. Therefore $\tilde{\mathcal{G}}$ is isomorphic to $\mathcal{G} := \{T \in \mathcal{L}(L^1(\mu)) : T$ is an isometric lattice isomorphism with $T\mathbb{1} = \mathbb{1}\}$.

Moreover, the topology of $\tilde{\mathcal{G}}$ corresponds to the topology on $\mathcal{G}$ induced by the strong operator topology on $\mathcal{L}(L^1(\mu))$. Therefore, it suffices to show that $\mathcal{G}$ is complete: From the Banach-Steinhaus theorem (see Schaefer 1970 , III.4.6) it follows that the unit ball $\{S \in \mathcal{L}(L^1(\mu)) : \|S\| \leq 1\}$ is complete for the strong operator topology. In order to show that $\mathcal{G}$ is closed, let $T_i$ converge to S, i.e. $T_i f \xrightarrow{\|\ \|} Sf$ for every $f \in L^1(\mu)$. We conclude that $|T_i f| \to |Sf|$ and $T_i|f| \to S|f|$. Since $|T_i f| = T_i|f|$ we obtain $|Sf| = S|f|$, and S is a lattice homomorphism. Similarly, one shows $S\mathbb{1} = \mathbb{1}$ and $\|Sf\| = \|f\|$, hence $S \in \mathcal{G}$.

X. D.4 <u>The Rohlin lemma for ergodic MDSs:</u>

The proof of the second assertion in Rohlin's lemma (X.2) is much easier if we assume that the MDS is ergodic instead of antiperiodic.

<u>Lemma:</u> Let $([0,1], \mathcal{B}, m; \varphi)$ be an MDS. If $\varphi$ is ergodic then for every $n \in \mathbb{N}$ and $\varepsilon > 0$ there exists $A \in \mathcal{B}$ such that $A, \varphi A, \dots, \varphi^{n-1}A$ are pairwise disjoint and $m(\bigcup_{k=0}^{n-1} \varphi^k A) \geq 1 - \varepsilon$.

Proof: Choose $B \in \mathcal{B}$ such that $0 < m(B) < \frac{\varepsilon}{n}$ and consider the measurable disjoint sets

$$B_m := \{x \in X : \varphi^m x \in B;\ \varphi^i x \notin B \ \text{ if } i = 0, \dots, m-1\}$$

for $m \in \mathbb{N}_0$. Since $\varphi$ is bi-measure-preserving and ergodic we have

$$m\Big(\bigcup_{m=0}^{\infty} B_m\Big) = 1.$$

Define $A := \varphi^{-(n-1)}(\bigcup_{j=1}^{\infty} B_{jn})$ and observe that $A, \varphi A, \dots, \varphi^{n-1}A$ are pairwise disjoint and $\bigcup_{k=0}^{n-1} \varphi^k A \supseteq X \setminus \bigcup_{i=0}^{n-1} \varphi^{-i}B$. ∎

## X. D.5 Lebesgue spaces:

As announced in (X.D.1) we show that there is a large and important class of probability spaces which are isomorphic to $([0,1], \mathcal{B}, m)$, $m$ the lebesgue measure. Here the corresponding isomorphism will be induced by a bi-measure-preserving point to point map (compare VI.D.1).

Definition: A probability space $(X, \mathcal{B}_X, \mu)$ is called a Lebesgue space if X is a separable complete metric space, $\mathcal{B}_X$ the Borel algebra on X and $\mu$ a diffuse Borel measure.
The isomorphism of $(X, \mathcal{B}_X, \mu)$ and $([0,1], \mathcal{B}, m)$ will be obtained by showing that both probability spaces are isomorphic to $\{0, 1\}^{\mathbb{N}}$ with a suitable probability measure.

Lemma:
(i) Let $\nu$ be a diffuse probability measure on the product $-\sigma$-algebra $\hat{\Sigma}$ of $\hat{X} = \{0, 1\}^{\mathbb{N}}$. Then there is a diffuse probability measure $\mu$ on $([0,1], \mathcal{B})$ such that $(\hat{X}, \hat{\Sigma}, \nu)$ is isomorphic to $([0,1], \mathcal{B}, \mu)$.
(ii) For a diffuse probability measure $\mu$ on $[0, 1]$, $([0,1], \mathcal{B}, \mu)$ is isomorphic to $([0,1], \mathcal{B}, m)$.

Proof:
(i) This statement is proved by the mapping

$$\theta : \hat{X} \longrightarrow [0, 1] \qquad (x_k)_{k \in \mathbb{N}} \longmapsto \sum_{k=1}^{\infty} 2^{-k} x_k$$

of (VI.D.2), which is bijective except from two countable sets, and by taking $\mu = \nu \circ \theta^{-1}$.

(ii) Let $f(t) := \mu([0,t])$, then $f : [0,1] \to [0,1]$ is continuous, increasing and $f(0) = 0$ and $f(1) = 1$. Define $f^*(s) := \sup\{t: f(t) \leq s\}$ and $f_*(s) := \inf\{t: f(t) \geq s\}$ and $M := \{s \in [0,1]: f^*(s) \neq f_*(s)\}$.
Obviously, f is constant (equal to s) on the intervals $[f_*(s), f^*(s)]$ for $s \in M$, and therefore $\mu([f_*(s), f^*(s)]) = 0$. Since $\sum_{s \in M} (f^*(s) - f_*(s)) \leq 1$, M is at most countable and thus $m(M) = 0$ and $\mu(N) = \sum_{s \in M} \mu([f_*(s), f^*(s)]) = 0$ where $N := \bigcup_{s \in M} [f_*(s), f^*(s)]$. On the complement of N the function f is invertible and $m = \mu \circ f^{-1}$ since

$$\mu(f^{-1}([0,t])) = \mu([0, f^*(t)]) = f(f^*(t)) = t = m([0,t])$$

for all $t \in [0,1]$. ■

<u>Theorem:</u> Every Lebesgue space $(X, \mathcal{B}_X, \mu)$ is isomorphic to $([0,1], \mathcal{B}, m)$.

Proof:
By the lemma it suffices to prove that $(X, \mathcal{B}_X, \mu)$ is isomorphic to $(\hat{X}, \hat{\Sigma}, \nu)$ for some diffuse probability measure $\nu$ on $\hat{X} = \{0, 1\}^{\mathbb{N}}$. Since X is separable and metric, there exists a separating base $(A_n)_{n \in \mathbb{N}}$ of open $\varepsilon$-balls in X (see X.D.1). We define the embedding

$$i : X \longrightarrow \hat{X} \text{ by } (i(x))_n := \begin{cases} 1 & \text{if } x \in A_n \\ 0 & \text{if } x \notin A_n \end{cases},$$

which is injective since $(A_n)$ is separating.
Furthermore, the map i is measurable since $\hat{\Sigma}$ is generated by the sets $B_n := \{(x_m) \in \hat{X}: x_n = 1\}$ and $i^{-1}(B_n) = A_n$. A theorem of Kuratowsky (see Jacobs 1978 , XIII.2.18) yields that $i(X)$ is $\hat{\Sigma}$-measurable and $i : X \longrightarrow i(X)$ is bi-measurable. Then $\nu := \mu \circ i^{-1}$ is a diffuse probability measure on $\hat{X}$ and $\nu(\hat{X} \setminus i(X)) = 0$, hence $(X, \mathcal{B}_X, \mu)$ is isomorphic to $(\hat{X}, \hat{\Sigma}, \nu)$. ■

Remarks:
1. "Lebesgue spaces" can be defined without referring to topological concepts: it is the existence of a separating base and the measurability of the set i(X) appearing in the proof above for a suitable $\sigma$-algebra on X which is essential (see Haezendonck [1973], Proposition 6).

2. It is not difficult to deduce from the theorem above an analogous representation theorem for probability spaces $(X, \mathcal{B}_X, \mu)$, X separable complete metric, but $\mu$ not necessarily diffuse.

References: Haezendonck [1973], Jacobs [1978], Riecan [1978], Rohlin [1949].

# XI. Informatinn of Covers

In this lecture we continue our efforts to find further invariants for measure-theoretical or topological dynamical systems with the intention of thereby solving the isomorphism problem for these systems. But henceforth the methods and the language we use will change drastically: while up to now theroms and proofs had a strong functional-analytic touch, in the remaining lectures we shall develop the concept of "entropy" for dynamical systems mainly using ordinary set theory.
The fundamental concept will be that of a <u>cover</u> $\alpha$ of a set X, i.e. a <u>finite</u> <u>collection</u> $\alpha$ of subsets of X such that $X = \bigcup_{A \in \alpha} A$.
Intuitively, we may talk of the "information" that such covers $\alpha$ provide about the location of a point $x \in X$, meaning the knowledge obtained by specifying an element $A \in \alpha$ which contains the point x. Moreover, the "finer" is a cover the more informations it provides (see XI.1 below). These intuitive concepts will be made precise in this lecture in a purely "static" approach. In the next lecture the dynamics will be added, leading to the important Kolmogoroff-Sinai invariant for dynamical systems: the entropy.

XI. 1 <u>Definition:</u> Let $\alpha$ and $\beta$ be two covers of a set X. We call $\alpha$ <u>finer than</u> $\beta$, written $\alpha \succcurlyeq \beta$, if every $A \in \alpha$ is contained in some $B \in \beta$. The cover

$$\alpha \vee \beta := \{A \cap B : A \in \alpha, B \in \beta\}$$

is called the <u>common</u> <u>refinement</u> of $\alpha$ and $\beta$.

<u>Remark:</u>
$\gamma \succcurlyeq \alpha \vee \beta$ if and only if $\gamma \succcurlyeq \alpha$ and $\gamma \succcurlyeq \beta$. But since "$\succcurlyeq$" is not antisymmetric and therefore not an order relation we cannot call $\alpha \vee \beta$ <u>the</u> supremum of $\alpha$ and $\beta$ (for example we have $\alpha \succcurlyeq \alpha \cup (\alpha \vee \beta) \succcurlyeq \alpha$ although $\alpha \neq \alpha \cup (\alpha \vee \beta)$ in general).
As indicated above, we are looking for a numerical measure h of the "information" of a cover $\alpha$. It is reasonable to require h to satisfy the following two basic properties:

(*) Monotonicity, i.e. $\alpha \succcurlyeq \beta$ implies $h(\alpha) \geq h(\beta)$,

(**) Subadditivity, i.e. $h(\alpha \vee \beta) \leq h(\alpha) + h(\beta)$

for all covers $\alpha, \beta$ we are considering. With this in mind there are essentially two different ways of introducing such measures h, which lead to the notions of topological entropy and of measure-theoretical entropy, respectively.

The first way can be pursued in a purely set-theoretical context, but we restrict ourselves to the topological setting. Hence, we assume that X is a compact space and denote by $\widetilde{\mathcal{O}}$ the set of all <u>open</u> (finite) covers of X. The definition of the "information" of an open cover $\alpha \in \widetilde{\mathcal{O}}$ will be based on the simple idea of counting the elements of $\alpha$. Therefore, we denote by $|\alpha|$ the number of (different) elements in $\alpha$ and conclude from

$$|\alpha \vee \beta| \leq |\alpha| \cdot |\beta|$$

that

$$h_t^*(\alpha) := \log|\alpha|$$

satisfies (**). However, it does not have property (*) since for every subcover $\beta$ of $\alpha$ (i.e. $\beta \in \widetilde{\mathcal{O}}$ and $\beta \subset \alpha$) we have $\beta \succcurlyeq \alpha$ but $|\beta| \leq |\alpha|$. Therefore we introduce the following definition which forces the "information" to become monotone.

XI. 2 <u>Definition</u>: For $\alpha \in \widetilde{\mathcal{O}}$ we call $h_t(\alpha) := \inf\{h_t^*(\beta) : \alpha \preccurlyeq \beta \in \widetilde{\mathcal{O}}\}$ the <u>t - information</u> of the cover $\alpha$.

Formally, the above expression for $h_t(\alpha)$ is analogous to the expression appearing in Definition (XI.6), and we have the following proposition.

XI. 3 <u>Proposition</u>: The t-information $\alpha \longmapsto h_t(\alpha)$ is monotone and subadditive on $\widetilde{\mathcal{O}}$.

The Definition (XI.2) can be simplified considerably: As observed above, a subcover $\beta$ of $\alpha$ is always finer than $\alpha$. Moreover, for every open cover $\gamma$ finer than $\alpha$ there is a subcover $\beta$ of $\alpha$ such that $|\gamma| \geq |\beta|$ . There, $h_t(\alpha)$ may be computed as follows.

XI. 4 <u>Lemma</u>: The t-information $h_t(\alpha)$ of a cover $\alpha \in \widetilde{\mathcal{O}}$ is equal to $\log N(\alpha)$, where $N(\alpha)$ is the number of elements of a subcover of $\alpha$ which has smallest cardinality.

The second way of introducing a measure of the "information" of a cover requires a probability space $(X, \Sigma, \mu)$. We can then use Shannon's information (see Z.1) to define

$$h_\mu^*(\alpha) := -\sum_{A \in \alpha} \mu(A) \log \mu(A)$$

for every <u>disjoint</u> $\Sigma$-measurable cover $\alpha$ of X. If we denote by $\widetilde{\Sigma}$ the set of all <u>$\Sigma$-measurable</u> (finite) covers of X, and by $\widetilde{\Sigma}_d$ the subset of all disjoint (finite) covers (=partitions), then we observe that h behaves quite well on $\tilde{\Sigma}_d$.

XI. 5 Lemma: For $\alpha, \beta \in \tilde{\Sigma}_d$ we have $h^*_\mu(\alpha \vee \beta) \leq h^*_\mu(\alpha) + h^*_\mu(\beta)$. Equality holds if and only if $\mu(A \wedge B) = \mu(A) \cdot \mu(B)$ for every $A \in \alpha$, $B \in \beta$.

Proof:
Let $\alpha$ and $\beta$ be two disjoint covers of X. Then

$$h^*_\mu(\alpha \vee \beta) - h^*_\mu(\alpha) - h^*_\mu(\beta)$$
$$= -\sum_{A\in\alpha}\sum_{B\in\beta} \mu(A\cap B)\log\mu(A\cap B) + \sum_{A\in\alpha}\mu(A)\log\mu(A) + \sum_{B\in\beta}\mu(B)\log\mu(B)$$
$$= -\sum_{A\in\alpha}\sum_{B\in\beta} \mu(A\cap B)\log\mu(A\cap B) + \sum_{A\in\alpha}\Big(\sum_{B\in\beta}\mu(A\cap B)\Big)\log\mu(A) + \sum_{B\in\beta}\Big(\sum_{A\in\alpha}\mu(A\cap B)\Big)\log\mu(B)$$
$$= \sum_{A\in\alpha}\sum_{B\in\beta}\mu(A\cap B)\log\frac{\mu(A)\cdot\mu(B)}{\mu(A\wedge B)}$$
$$\leq \sum_{A\in\alpha}\sum_{B\in\beta}\mu(A\cap B)\left(\frac{\mu(A)\cdot\mu(B)}{\mu(A\wedge B)} - 1\right)\cdot\log e = 0,$$

where the inequality is true since $\log x \leq (x-1)\log e$, with equality only for $x = 1$. ■

To extend $h^*_\mu$ to all measurable covers $\alpha$ we use the same procedure as in (XI.2).

XI. 6 Definition: For $\alpha \in \tilde{\Sigma}$ we call

$$h_\mu(\alpha) := \inf\{h^*_\mu(\beta) : \alpha \leq \beta \in \tilde{\Sigma}_d\}$$ the $\mu$-information of the cover $\alpha$.

Again, $h_\mu$ is monotone by definition, and Lemma (XI.5) implies that

$$\inf\{h^*_\mu(\gamma) : \alpha\vee\beta \leq \gamma \in \tilde{\Sigma}_d\} \leq \inf\{h^*_\mu(\gamma_1 \vee \gamma_2) : \alpha \leq \gamma_1 \in \tilde{\Sigma}_d \text{ and } \beta \leq \gamma_2 \in \tilde{\Sigma}_d\}$$
$$\leq \inf\{h^*_\mu(\gamma_1) : \alpha \leq \gamma_1 \in \tilde{\Sigma}_d\} + \inf\{h^*_\mu(\gamma_2) : \beta \leq \gamma_2 \in \tilde{\Sigma}_d\}.$$

Therefore, this $\mu$-information also satisfies the desired properties (*) and (**).

XI. 7 Proposition: The $\mu$-information $\alpha \mapsto h_\mu(\alpha)$ is monotone and subadditive on $\tilde{\Sigma}$.

For computations as well as for theoretical reasons the following result is important.

XI. 8 Proposition: For disjoint covers $\alpha \in \tilde{\Sigma}$ we have

$$h_\mu(\alpha) = h^*_\mu(\alpha).$$

Proof:
It suffices to show that $\alpha \leq \beta$ implies $h^*_\mu(\alpha) \leq h^*_\mu(\beta)$ for $\alpha, \beta \in \tilde{\Sigma}_d$.
Let $\alpha = \{A_1, \ldots, A_n\}$ and $\beta = \beta_1 \vee \beta_2 \vee \ldots \vee \beta_n$ where $\beta_i = \{B \in \beta : B \subset A_i\}$.
Then

$$h^*_\mu(\beta) - h^*_\mu(\alpha) = -\sum_{B\in\beta} \mu(B) \log (B) + \sum_{i=1}^{n} \mu(A_i) \log \mu(A_i)$$
$$= \sum_{i=1}^{n} \left( - \sum_{B\in\beta_i} \mu(B) \log \mu(B) + \left(\sum_{B\in\beta_i} \mu(B)\right) \log \mu(A_i) \right)$$
$$= \sum_i \left( - \sum_{B\in\beta_i} \mu(B) \log \frac{\mu(B)}{\mu(A_i)} \right)$$
$$= \sum_i \mu(A_i) \left( - \sum_{B\in\beta_i} \frac{\mu(B)}{\mu(A_i)} \log \frac{\mu(B)}{\mu(A_i)} \right) \geq 0 . \blacksquare$$

In the final part of this lecture we deduce some relations between the different informations of covers which will be of great importance in the following. To that purpose let X be a compact space, $\mathcal{B}$ its Borel $\sigma$-algebra and $\mu$ a probability measure on X. We consider the set $\tilde{\mathcal{O}}$ of all open covers and $\tilde{\mathcal{B}}$ of all Borel measurable covers of X. In this setting it makes sense to compare t-information and $\mu$-information for $\alpha \in \tilde{\mathcal{O}}$.

XI. 9 <u>Theorem</u>:
(i) The t-information majorizes the $\mu$-information, i.e. $h_\mu(\alpha) \leq h_t(\alpha)$ for every $\alpha \in \tilde{\mathcal{O}}$.
(ii) The $\mu$-information is a concave function on the set of all probability measures on X, i.e. $h_{\lambda\mu+(1-\lambda)\nu}(\alpha) \geq \lambda \cdot h_\mu(\alpha) + (1-\lambda)\, h_\nu(\alpha)$ for $0 < \lambda < 1$, $\alpha \in \tilde{\mathcal{B}}$ and probability measures $\mu$ and $\nu$ on X.

Proof:
(i) If $\alpha$ is an open cover of X we may take a subcover $\bar{\alpha}$ of $\alpha$ having minimal cardinality. For $\bar{\alpha}$ we find a disjoint cover $\hat{\alpha} \in \tilde{\mathcal{B}}$ finer than $\bar{\alpha}$ and such that $|\hat{\alpha}| = |\bar{\alpha}|$ ("disjointification"). By the above definitions we have

$$h_\mu(\alpha) \leq h_\mu(\hat{\alpha})$$

and $$h_t(\alpha) = \log |\bar{\alpha}| = \log |\hat{\alpha}|.$$

Therefore,

$$h_\mu(\alpha) - h_t(\alpha) \leq h_\mu(\hat{\alpha}) - \log |\hat{\alpha}|$$
$$= - \sum_{\hat{A}\in\hat{\alpha}} \mu(\hat{A}) \log \mu(\hat{A}) - \sum_{\hat{A}\in\hat{\alpha}} \mu(\hat{A}) \log |\hat{\alpha}|$$
$$= \sum_{A\in\hat{\alpha}} \mu(A) \log \frac{1}{\mu(A)|\hat{\alpha}|}$$
$$\leq \sum_{A\in\hat{\alpha}} \mu(A) \left(\frac{1}{\mu(A)|\hat{\alpha}|} - 1\right) \cdot \log e = 0.$$

(ii) It suffices to prove the statement for disjoint covers $\alpha \in \widetilde{\mathcal{B}}$ :

$$\begin{aligned} h_{\lambda\mu+(1-\lambda)\nu}(\alpha) &= -\sum_{A\in\alpha} (\lambda\cdot\mu(A) + (1-\lambda)\cdot\nu(A)) \log(\lambda\cdot\mu(A) + (1-\lambda)\cdot\nu(A)) \\ &\geq -\sum_{A\in\alpha} (\lambda\cdot\mu(A) \log\mu(A) + (1-\lambda)\cdot\nu(A) \log\nu(A)) \\ &= \lambda\cdot h_\mu(\alpha) + (1-\lambda)\cdot h_\nu(\alpha), \end{aligned}$$

since $x \mapsto -x \log x$ is a concave function on $[0, 1]$. ∎

## XI. D Discussion

### XI. D.1 The t-information is like a $\mu$-information:

One might ask whether there is a closer connection between $h_t(\alpha)$ and $h_\mu(\alpha)$ besides the formal analogies in the structure of the definition. To this end we consider a compact space X and introduce a functional m by

$$m(A) := \begin{cases} 1 & \text{if } \emptyset \neq A \subset X \\ 0 & \text{if } \emptyset = A \end{cases} .$$

Let $\alpha$ be some open cover of X. Exactly as for the $\mu$-information we define

$$h_m^*(\alpha) := -\sum_{A\in\alpha} \frac{m(A)}{q} \log\frac{m(A)}{q}$$

for $q := \sum_{A\in\alpha} m(A)$ and

$$h_m(\alpha) := \inf\{h_m^*(\beta) : \alpha \preceq \beta \in \widetilde{\mathcal{O}}\} .$$

Fortunately, $h_m^*(\alpha) = \log|\alpha|$ and therefore $h_m(\alpha) = h_t(\alpha)$.
Hence, we obtain the t-information from the same expression as the $\mu$-information just by replacing the probability measure $\mu$ by the trivial functional m. The only difference is that disjoint covers don't appear in the topological case (unless the compact space X is disconnected).
For a common generalization of $h_\mu$ and $h_t$ see (XI.D.5).

### XI. D.2 The infimum in the definition of "information" is attained:

The t-information of $\alpha$, formally defined as $h_t(\alpha) = \inf\{h_t^*(\beta) : \alpha \preceq \beta \in \widetilde{\mathcal{O}}\}$ is in fact equal to $\min\{h_t^*(\beta) : \beta \subset \alpha\}$ by (XI.4).
In the measure-theoretical case we defined the $\mu$-information of $\alpha$ as

$$h_\mu(\alpha) = \inf\{h_\mu^*(\delta) : \alpha \preceq \delta \in \widetilde{\Sigma}_d\},$$

and again this infimum is attained at some "disjointification" of $\alpha$. The proof of this statement is not quite obvious and needs some preparation.

<u>Definition:</u> Let $(X, \Sigma, \mu)$ be a probability space.

(i) An element $\tilde{\alpha} = (A_1,\dots,A_n)$ of $\Sigma^n$ is called an <u>ordered cover</u> (of length n) if $\mu(\bigcup_{i=1}^n A_i) = 1$. The set of all ordered covers of length n is denoted by $\Sigma_c^n$. Moreover, we call $\tilde{\alpha}$ an <u>ordering</u> of a cover

$\alpha \in \tilde{\Sigma}$ if $\bar{\alpha}$ and $\alpha$ contain the same elements up to $\mu$-null sets, but not necessarily the same number of elements.

(ii) Consider the mapping $\Delta: \Sigma_c^n \longrightarrow \tilde{\Sigma}_d$ where

$$\Delta(\bar{\alpha}) = \{A_i \setminus \bigcup_{j=1}^{i-1} A_j : i = 1, \dots, n\}.$$

If $\bar{\alpha}$ is an ordering of $\alpha$, then $\Delta(\bar{\alpha})$ is called the <u>disjointification</u> of $\alpha$ corresponding to the ordering $\bar{\alpha}$.

Proposition: $h_\mu(\alpha) = \min\{h_\mu^*(\delta): \delta \text{ is a disjointification of } \alpha\}$.

Proof:

Let $\alpha \leq \gamma \in \tilde{\Sigma}_d$ and assume that $\gamma = \{C_1, \dots, C_n\}$ is such that $\mu(C_{i+1}) \leq \mu(C_i)$. Our aim is to construct a disjointification $\delta$ of $\alpha$ with $h_\mu^*(\delta) \leq h_\mu^*(\gamma)$:

There is an $A_1 \in \alpha$ such that $C_1 \subset A_1$. Consider

$$\gamma_2 := \{C_1 \cup (C_2 \cap A_1),\ C_2 \setminus A_1,\ C_3, \dots, C_n\}.$$

Since the $\mu$-information is concave (XI.9), we obtain

$$h_\mu^*(\gamma_2) \leq h_\mu^*(\gamma).$$

Then we construct $\gamma_3 := \{C_1 \cup (C_2 \cap A_1) \cup (C_3 \cap A_1),\ C_2 \setminus A_1, C_3 \setminus A_1, C_4, \dots, C_n\}$ and so on. We end up with $\gamma_n := \{A_1,\ C_2 \setminus A_1, \dots, C_n \setminus A_1\}$, since $A_1 = C_1 \cup \bigcup_{i=2}^{n}(C_i \cap A_1)$, and

$$h_\mu^*(\gamma_n) \leq h_\mu^*(\gamma).$$

For the next step we may assume that $C_1 = A_1$ and again $\mu(C_{i+1}) \leq \mu(C_i)$. There is $A_2 \in \alpha$ such that $C_2 \subset A_2$. Now $A_2 \setminus A_1 = C_2 \cup \bigcup_{i=3}^{n}(C_i \cap A_2)$ and as above, we replace successively $C_2$ by $A_2 \setminus A_1$ and $C_i$ by $C_i \setminus A_2$ for $i = 3, \dots, n$.

This time we end up with $\gamma_n = \{A_1, A_2 \setminus A_1,\ C_3 \setminus A_2, \dots, C_n \setminus A_2\}$ and

$$h_\mu^*(\gamma_n) \leq h_\mu^*(\gamma).$$

Continuing in this way we obtain the desired disjointification. ∎

XI. D.3 <u>Conditional information of a cover</u>:

In the proof of (XI.8) we obtained an interesting expression, namely a convex combination of sums that look like $h_\mu^*$. In fact, they represent the information (see App.Z.3) of the vector of conditional probabilities $\mu(B|A_i) := \frac{\mu(A_i \cap B)}{\mu(A_i)}$. In analogy, we consider $A \in \Sigma$, $\alpha, \beta \in \tilde{\Sigma}_d$ and define the <u>conditional $\mu$-information of $\beta$ given A</u>

$$h_\mu(\beta | A) := -\sum_{B \in \beta} \frac{\mu(B \cap A)}{\mu(A)} \log \frac{\mu(B \cap A)}{\mu(A)}$$

and the <u>conditional $\mu$-information of $\beta$ given $\alpha$</u>

$$h_\mu(\beta \mid \alpha) := \sum_{A\in\alpha} \mu(A)\, h_\mu(\beta \mid A) .$$

<u>Proposition</u>: Let $\alpha, \beta, \gamma \in \tilde{\tilde{\Sigma}}_d$. Then

(i) $h_\mu(\gamma \mid \alpha) = h_\mu(\alpha \vee \gamma) - h_\mu(\alpha)$,

(ii) $h_\mu(\alpha \vee \beta \mid \gamma) \leq h_\mu(\alpha \mid \gamma) + h_\mu(\beta \mid \gamma)$,

(iii) $h_\mu(\alpha \mid \beta \vee \gamma) = h_\mu(\alpha \vee \beta \mid \gamma) - h_\mu(\beta \mid \gamma)$ .

Proof:

(i) Apply the identity in the proof of (XI.8) to $\alpha \vee \gamma$.

(ii) As in the proof of (XI.5) one shows that $h_\mu(\alpha \vee \beta \mid C) \leq h_\mu(\alpha \mid C) + h_\mu(\beta \mid C)$ for every $C \in \gamma$ .

(iii) $h_\mu(\alpha \vee \beta \mid \gamma) - h_\mu(\beta \mid \gamma) = h_\mu(\alpha \vee \beta \vee \gamma) - h_\mu(\gamma) - h_\mu(\beta \vee \gamma) + h_\mu(\gamma) = h_\mu(\alpha \mid \beta \vee \gamma)$ ∎

<u>Corollary</u>: Let $\alpha, \beta, \gamma \in \tilde{\Sigma}_d$ . Then

(i) $\alpha \leq \beta$ implies $h_\mu(\alpha \mid \gamma) \leq h\,(\beta \mid \gamma)$,

(ii) $\beta \geq \gamma$ implies $h_\mu(\alpha \mid \beta) \leq h\,(\alpha \mid \gamma)$.

Proof:

For disjoint covers $\alpha, \beta$ and $\gamma$ we have

$$\begin{aligned} h_\mu(\beta \mid \gamma) - h_\mu(\alpha \mid \gamma) &= h_\mu(\alpha \vee \beta \mid \gamma) - h_\mu(\alpha \mid \gamma) \\ &= h_\mu(\beta \mid \alpha \vee \gamma) \geq 0 \text{ proving (i), and} \\ h_\mu(\alpha \mid \gamma) - h_\mu(\alpha \mid \beta) &= h_\mu(\alpha \mid \gamma) - h_\mu(\alpha \mid \beta \vee \gamma) \\ &= h_\mu(\alpha \mid \gamma) - h_\mu(\alpha \vee \beta \mid \gamma) + h_\mu(\beta \mid \gamma) \\ &\geq 0 \text{ proving (ii).} \end{aligned}$$ ∎

XI. D.4 <u>Two metrics for covers</u>:

If $(X, \Sigma, \mu)$ is a probability space, then there is a natural (quasi-) metric on the $\sigma$-algebra $\Sigma$ defined by

$$d(A,B) := \mu(A \triangle B) \quad \text{for } A,\, B \in \Sigma .$$

This metric induces in a canonical way a metric on $\tilde{\tilde{\Sigma}}$ (see below). On the other hand there is a second metric on $\tilde{\tilde{\Sigma}}$ which was introduced by Rohlin and whose definition is based on the concept of "conditional - information" (see XI.D.3). In the following we shall investigate and compare these metrics.

<u>Definition</u>:

(i) For two ordered covers $\tilde{\alpha} = (A_1, \ldots, A_n)$ and $\tilde{\beta} = (B_1, \ldots, B_n)$ of $\Sigma$ we define

$$d(\bar{\alpha}, \bar{\beta}) := \sum_{i=1}^{n} \mu(A_i \triangle B_i).$$

Then the distance between two covers $\alpha, \beta \in \tilde{\Sigma}$ is

$$d(\alpha, \beta) := \min \{ d(\bar{\alpha}, \bar{\beta}) : \bar{\alpha} \in \Sigma_c^n \text{ an ordering of } \alpha, \bar{\beta} \in \Sigma_c^n \text{ an ordering of } \beta \},$$

$$n = \max(|\alpha|, |\beta|).$$

(ii) The Rohlin - distance between two disjoint covers $\alpha, \beta \in \tilde{\Sigma}_d$ is given by

$$\varrho(\alpha, \beta) := h_\mu(\alpha | \beta) + h_\mu(\beta | \alpha).$$

It follows from the proposition in (XI.D.3) that $\varrho$ is in fact a metric on $\tilde{\Sigma}_d$. In the following we shall write simply h instead of $h_\mu$.

Proposition: The spaces $\tilde{\Sigma}^n := \{ \alpha \in \tilde{\Sigma} : |\alpha| \leq n \}$ and $\tilde{\Sigma}_d^n := \{ \alpha \in \tilde{\Sigma}_d : |\alpha| \leq n \}$ for fixed $n \in \mathbb{N}$ are complete for the uniform structure induced by the metric d.

Proof:

By (A.10) $\Sigma$ and therefore $\Sigma^n$ is complete. Let $(\alpha_i)$ be a Cauchy sequence in $\tilde{\Sigma}^n$. Then there is an increasing sequence $(k_m)$ such that $i, j \geq k_m$ implies $d(\alpha_i, \alpha_j) < 2^{-m}$. In particular, $d(\alpha_{k_m}, \alpha_{k_{m+1}}) < 2^{-m}$, and we can find orderings $\bar{\alpha}_m \in \Sigma^n$ of $\alpha_{k_m}$ and $\bar{\alpha}_{m+1} \in \Sigma^n$ of $\alpha_{k_{m+1}}$ such that $d(\bar{\alpha}_m, \bar{\alpha}_{m+1}) < 2^{-m}$. By the completeness of $\Sigma^n$ there exists an ordered cover $\bar{\alpha} \in \Sigma^n$ such that $\bar{\alpha}_m$ converges to $\bar{\alpha}$. For the cover $\alpha$ corresponding to $\bar{\alpha}$ we obtein $\lim \alpha_i = \alpha$.

The proof of the second part is left to the reader. ■

The following theorem shows the equivalence of d and $\varrho$ on the set of all partitions of a fixed cardinality. This equivalence gives us a greater flexibility and we can apply the metric most suited for the specific problem. Such applications will be made in (XII.D.6) and (XII.D.7).

Theorem: Let $(X, \Sigma, \mu)$ be a probability space.

(i) The metric $\varrho$ is finer than d on $\tilde{\Sigma}_d$.

(ii) The metrics $\varrho$ and d are equivalent on $\tilde{\Sigma}_d^n$ for every $n \in \mathbb{N}$.

(iii) For every $\varepsilon > 0$ there is $\delta > 0$ such that

$h(\beta | \alpha) \leq \delta$ implies $\beta \underset{\varepsilon}{\leq} \alpha$ (compare Definition XII.5)

and $\beta \underset{\delta}{\leq} \alpha$ implies $h(\beta | \alpha) \leq \varepsilon$

for any $\alpha, \beta \in \tilde{\Sigma}_d$ where $\beta$ has a fixed cardinality.

Remark:

The statement (iii) can be regarded as an "asymmetric" version of (ii).

Proof:

We proceed as follows: First we show the second part of (iii) and extend the proof to the corresponding part of (ii), i.e. d is finer than $\varrho$ on $\tilde{\Sigma}^n_d$. Then we prove the first part of (iii) and extend it to (i).

step 1:

In (XII.6) we shall construct a function $f : \mathbb{N} \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ such that $\lim_{\varepsilon \to 0} f(n, \varepsilon) = 0$ for every $n \in \mathbb{N}$, and $\beta \underset{\varepsilon}{\leq} \alpha$ implies the existence of $\gamma \in \tilde{\Sigma}$ with $\beta \leq \alpha \vee \gamma$ and $h(\gamma) \leq f(|\beta|, \varepsilon)$.

Thus $\beta \underset{\delta}{\leq} \alpha$ implies that

$h(\beta | \alpha) = h(\alpha \vee \beta) - h(\alpha) \leq h(\alpha \vee \gamma) - h(\alpha) \leq h(\gamma) \leq f(n, \delta) \leq \varepsilon$

for $\delta$ sufficiently small.

step 2:

If $d(\alpha, \beta) \leq \delta$ then $\alpha \underset{\delta}{\leq} \beta$ and $\beta \underset{\delta}{\leq} \alpha$ and therefore

$\varrho(\alpha, \beta) = h(\alpha | \beta) + h(\beta | \alpha) \leq \varepsilon$ for $\delta$ sufficiently small.

step 3:

Let $\delta := h(\beta | \alpha) > 0$ and $n := |\beta|$. For $c > 0$ we consider

$\alpha_0 := \{A \in \alpha : \mu(A) = 0 \text{ or } h(\beta | A) > c\}$.

Then the set $A_0 := \bigcup_{A \in \alpha_0} A$ has measure

$$(1) \quad \mu(A_0) \leq \frac{\delta}{c},$$

for otherwise $h(\beta | \alpha) \geq \sum_{A \in \alpha_0} \mu(A)\, h(\beta | A) > \frac{\delta}{c} c = \delta$.

For any $A \in \alpha \setminus \alpha_0$ there exists $B \in \beta$ such that

$$(2) \quad \frac{\mu(A \cap B)}{\mu(A)} \geq 1 - c,$$

because otherwise

$-\log \frac{\mu(A \cap B)}{\mu(A)} > -\log(1 - c) \geq c$ and therefore $h(\beta | A) > c$.

Now we define a mapping $b : \alpha \to \beta$ as follows:

For $A \in \alpha$ let $b(A)$ be an element $B \in \beta$ for which $\mu(B \cap A)$ is maximal.

From (2) we obtain

$$(3) \quad \mu(A \cap b(A)) \geq (1 - c) \cdot \mu(A) \quad \text{and}$$
$$\mu(A \setminus b(A)) \leq c \cdot \mu(A).$$

For $B \in \beta$ define $\alpha_B := \{A \in \alpha \setminus \alpha_0 : b(A) = B\}$ and $\bar{B} := \bigcup_{A \in \alpha_B} A$. Then

$$\gamma := \{A_0\} \cup \{\bar{B} : B \in \beta\}$$

is a partition which satisfies $\alpha \geq \gamma$. Moreover, we have

$$\mu(\bar{B} \setminus B) = \sum_{A \in \alpha_B} \mu(A \setminus b(A)) \leq c \cdot \sum_{A \in \alpha_B} \mu(A) = c\, \mu(\bar{B}) \quad \text{and}$$

$$\mu(B \setminus \bar{B}) = \mu\Big(\bigcup_{C \in \beta \setminus \{B\}} (B \cap \bar{C}) \cup (B \cap A_0)\Big)$$
$$\leq \sum_{C \in \beta \setminus \{B\}} \mu(\bar{C} \setminus C) + \mu(B \cap A_0)$$

$$\leq c \cdot \sum_{C\in\beta\setminus\{B\}} \mu(\bar{C}) + \mu(B\cap A_0)$$
$$\leq c + \mu(B\cap A_0) .$$

Now we take any $C \in \beta$ and define

$$\beta' := \{\bar{B} : B \in \beta\setminus\{C\}\} \cup \{\bar{C}\cup A_0\} \in \tilde{\Sigma}_d .$$

Then $\beta' \leq \alpha$, $|\beta'| \leq |\beta|$ and

$$\begin{aligned} d(\beta,\beta') &\leq \sum_{B\in\beta\setminus C} \mu(B\setminus\bar{B}) + \sum_{B\in\beta\setminus C} \mu(\bar{B}\setminus B) + \mu(\bar{C}\cup A_0\setminus C) + \mu(C\setminus\bar{C}\cup A_0) \\ &\leq \sum_{B\in\beta\setminus C} \mu(B\setminus\bar{B}) + \sum \mu(\bar{B}\setminus B) + \mu(\bar{C}\setminus C) + \mu(A_0\setminus C) + \mu(C\setminus\bar{C}) \\ &= \sum_{B\in\beta} \mu(B\setminus\bar{B}) + \sum_{B\in\beta} \mu(\bar{B}\setminus B) + \mu(A_0\setminus C) \\ &\leq cn + \frac{\delta}{c} + c + \frac{\delta}{c} \\ &\leq 3\sqrt{\delta n} + \sqrt{\frac{\delta}{n}} \qquad \text{for } c = \sqrt{\frac{\delta}{n}} . \end{aligned}$$

step 4:

Let $\delta := \varrho(\alpha,\beta) > 0$ and assume that $|\alpha| = |\beta| = n$. Then $h(\beta|\alpha)$ and $h(\alpha|\beta) \leq \delta$, and we can perform the construction of step 3 for both $\alpha$ and $\beta$. In particular, we have two mappings $b : \alpha \to \beta$ and $a : \beta \to \alpha$ and two sets $\alpha_0 \subset \alpha$ and $\beta_0 \subset \beta$ with $\mu(\bigcup_{A\in\alpha_0} A) \leq \frac{\delta}{c}$, $\mu(\bigcup_{B\in\beta_0} B) \leq \frac{\delta}{c}$ and the inequalities (3).

Take $\alpha_0$ as above and

$$\alpha_1 := \{A \in \alpha\setminus\alpha_0 : b(A) \in \beta_0\},$$
$$\alpha_2 := \{A \in \alpha\setminus\alpha_0 : b(A) \in \beta\setminus\beta_0,\ A \neq a(b(A))\} ,$$
$$\alpha_3 := \{A \in \alpha\setminus\alpha_0 : b(A) \in \beta\setminus\beta_0,\ A = a(b(A))\} ,$$

and $\beta_0, \beta_1, \beta_2, \beta_3$ analogously.

Then $\alpha = \alpha_0 \cup \alpha_1 \cup \alpha_2 \cup \alpha_3$ and $\beta = \beta_0 \cup \beta_1 \cup \beta_2 \cup \beta_3$ .

For $A \in \alpha_1$ we have from the inequality (3)

$$\mu(A) \leq \frac{1}{1-c}\,\mu(b(A)\cap A) \leq \frac{1}{1-c}\,\mu(b(A)) .$$

Therefore

$$(4) \quad \mu(\bigcup_{A\in\alpha_1} A) \leq \frac{1}{1-c}\,\mu(\bigcup_{B\in\beta_0} B) \leq \frac{\delta}{c(1-c)} .$$

Now we take any $B \in \beta\setminus\beta_0$ and consider

$$\alpha'_B := \{A \in \alpha_2 : b(A) = B\}.$$

Then $a(B) \notin \alpha'_B$ and therefore

$$\mu(B) \geq \sum_{A\in\alpha'_B} \mu(B\cap A) + \mu(B\cap a(B)) \geq (1-c)\Big(\sum_{A\in\alpha'_B} \mu(A) + \mu(B)\Big) \text{ by (3).}$$

This implies

$$(5) \quad \mu(\bigcup_{A\in\alpha_2} A) \leq \sum_{B\in\beta\setminus\beta_0} \Big(\sum_{A\in\alpha'_B} \mu(A)\Big) \leq \sum_{B\in\beta\setminus\beta_0} \frac{c}{1-c}\,\mu(B) \leq \frac{c}{1-c} .$$

Finally, again by (3), we get

$$(6) \quad \sum_{A\in\alpha_3} (\mu(A\setminus b(A)) + \mu(b(A)\setminus A)) \leq c \sum_{A\in\alpha_3} (\mu(A) + \mu(b(A))) \leq 2c.$$

In conclusion, we obtain from (1), (4), (5) and (6)

$$d(\alpha,\beta) \leq 2c + \sum_{A\in\alpha\setminus\alpha_3}\mu(A) + \sum_{B\in\beta\setminus\beta_3}\mu(B)$$
$$\leq 2c + 2\left(\frac{\delta}{c} + \frac{\delta}{c(1-c)} + \frac{c}{1-c}\right)$$
$$= 4\left(\sqrt{\delta} + \frac{\sqrt{\delta}}{1-\sqrt{\delta}}\right) \qquad \text{for } c := \sqrt{\delta}$$
$$\leq 16\sqrt{\delta} \quad \text{for } \delta \leq \tfrac{4}{9}. \qquad \blacksquare$$

References: Denken-Grillenberger-Sigmund [1976], Ch.11, Rohlin [1959], [1967]

XI. D.5 Weighted lattices and the information of covers:

The strong analogy in the definitions of the information of a cover in the topological and measure-theoretical cases leads us to ask for the common structure behind these two cases. It turns out that the structure of a "distributive lattice" of $\mathcal{O}$ and $\Sigma$ is the essential ingredient. We sketch this abstract approach to information.

Assume that V is a distributive lattice with 0 and 1 and assume we have a "weighting function"

$$m : V \longrightarrow \mathbb{R}_+ ,$$

which corresponds to the measure $\mu$ in the measure-theoretical case and to the functional m in (XI.D.1) in the topological case. The set

$$\widetilde{V} := \{\alpha \subset V : |\alpha| \text{ is finite and } \sup\{a : a\in\alpha\} = 1\}$$

is called the set of all "covers" of V. $\widetilde{V}$ will be (pre)ordered where $\alpha \leq \beta$ if for every $b\in\beta$ there is $a\in\alpha$ such that $b \leq a$. For $\alpha \in \widetilde{V}$ we define

$$h_m^*(\alpha) := -\sum_{a\in\alpha} \frac{m(a)}{m(\alpha)} \log \frac{m(a)}{m(\alpha)}$$

where $m(\alpha) := \sum_{a\in\alpha} m(a)$ .

Let us further define

$$\hat{h}_m(\beta) := \sup\{h_m^*(\alpha) : \beta \leq \alpha \text{ and } |\alpha| \leq |\beta|\} .$$

Then we obtain the "m-information" of a cover $\gamma \in \widetilde{V}$ as

$$h_m(\gamma) := \inf\{\hat{h}_m(\beta) : \gamma \leq \beta\} .$$

With some additional modifications that are irrelevant in the topological and in the measure-theoretical context (see Palm 1976 a,b ) one obtains monotonicity and subadditivity for $h_m$ . These are the basic properties for a further development of the theory as in lecture XII. Moreover, the above definition coincides with $h_\mu$, resp. $h_t$, if we take $V = \Sigma$ and $m = \mu$ for a probability space $(X,\Sigma,\mu)$, resp. $V = \mathcal{O}$ and m as in (XI.D.1) for a compact space X.

References: Palm [1976 a,b] .

# XII. Entropy of Dynamical Systems

Having introduced the necessary "static" concepts in the previous lecture we now consider the "dynamics" as given by the transformation

$$\varphi : X \longrightarrow X$$

of an MDS $(X, \Sigma, \mu ; \varphi)$ or of a TDS $(X; \varphi)$. In a natural way, $\varphi$ also induces dynamics on the set $\tilde{\Sigma}$ of all (finite) measurable covers or on the set $\tilde{\mathcal{O}}$ of all (finite) open covers of X. More precisely, to every (measurable or open) cover $\alpha$ of X we associate the cover

$$\tilde{\varphi}\alpha := \{ \varphi^{-1}(A) : A \in \alpha \} \quad .$$

Similarily, we define

$$\tilde{\varphi}^n \alpha := \{ \varphi^{-n}(A) : A \in \alpha \}$$

for $n \in \mathbb{Z}$ and use the following shorthand notation:

$$\alpha_k^n := \tilde{\varphi}^k \alpha \vee \tilde{\varphi}^{k+1} \alpha \vee \dots \vee \tilde{\varphi}^{n-1} \alpha = \bigvee_{i=k}^{n-1} \tilde{\varphi}^i \alpha \quad \text{for } k,n \in \mathbb{Z},\ k<n,$$

$$\alpha^n := \alpha_0^n = \bigvee_{i=0}^{n-1} \tilde{\varphi}^i \alpha \quad \text{for } n\ \mathbb{N}, \text{ and}$$

$$\alpha^0 := \alpha_0^0 := \{X\}.$$

As explained at the beginning of lecture XI, a cover $\alpha$ informs us of the location of a point $x \in X$ by specifying an element $A \in \alpha$ such that $x \in A$. Consequently, $\tilde{\varphi}\alpha$ provides the same information about the point $\varphi(x)$ and $\tilde{\varphi}^n\alpha$ about $\varphi^n(x)$ : $x \in \varphi^{-n}(A) \in \tilde{\varphi}^n\alpha$ iff $\varphi^n(x) \in A \in \alpha$ .
We have measured this information by $h_\mu(\alpha)$, resp. $h_t(\alpha)$.
Obviously $h_\mu(\tilde{\varphi}\alpha) = h_\mu(\alpha)$, resp. $h_t(\tilde{\varphi}\alpha) = h_t(\alpha)$ ,
and $h_\mu(\alpha^n) \leq h_\mu(\alpha^m)$, resp. $h_t(\alpha^n) \leq h_t(\alpha^m)$,
for $n \leq m$. The study of the asymptotic growth of $h_\mu(\alpha^n)$, resp. $h_t(\alpha^n)$, then leads to a new constant depending on $\alpha$ and the dynamics provided by $\varphi$. Variation over all suitable covers $\alpha$ yields the new isomorphism invariant, the "entropy".

XII. 1 <u>Definition</u>: Consider an MDS $(X, \Sigma, \mu ; \varphi)$.
The <u>$\mu$-entropy</u> of $\alpha \in \tilde{\Sigma}$ (<u>with respect to</u> $\varphi$) is defined by

$$h_\mu(\alpha ; \varphi) := \lim_{n \to \infty} \frac{1}{n} h_\mu(\alpha^n) \ .$$

The <u>$\mu$-entropy</u> of $(X, \Sigma, \mu ; \varphi)$ is defined by

$$h_\mu(X;\varphi) := \sup\{h_\mu(\alpha;\varphi) : \alpha\in\tilde{\Sigma}\}.$$

In the topological case we proceed analogously.

XII. 2 Definition: Consider a TDS $(X;\varphi)$.
The topological entropy of $\alpha\in\tilde{\mathcal{O}}$ (with respect to $\varphi$) is defined by

$$h_t(\alpha;\varphi) := \lim_{n\to\infty} \frac{1}{n} h_t(\alpha^n).$$

The topological entropy of $(X;\varphi)$ is defined by

$$h_t(X;\varphi) := \sup\{h_t(\alpha;\varphi) : \alpha\in\tilde{\mathcal{O}}\}.$$

Remarks:

1. We observe that the two definitions are identical, and the same will be true for some of the following proofs. Therefore, from now on we write h in all statements valid for both $h_\mu$ and $h_t$.
2. The limits in the above definitions exist since the subadditivity of h (see XI.4 and XI.7) implies that
$$h(\alpha^{n+m}) \leq h(\alpha^n) + h(\alpha^m).$$
In addition, the elementary Lemma (E.5) shows that
$$h(\alpha;\varphi) = \inf_n \frac{1}{n} h(\alpha^n).$$
3. The $\mu$-entropy was first introduced by Kolmogorov in 1958, the t-entropy by Adler, Konheim and McAndrew in 1965.

XII. 3 Proposition: The entropy $h(X;\varphi)$ is an isomorphism invariant.

Proof:
An isomorphism $\Theta$ between two MDSs (resp. TDSs) induces a bijective mapping

$$\tilde{\Theta}: \alpha \mapsto \{\Theta^{-1}A : A\in\alpha\}$$

that preserves the pre-order relation "$\leq$" on $\tilde{\Sigma}$ (resp. $\tilde{\mathcal{O}}$) and satisfies $h(\alpha) = h(\tilde{\Theta}\alpha)$.
Then, step by step, one shows the equality of the entropy of two isomorphic systems.
Still, there remains the troublesome fact that the above definition of entropy involves several steps, and it seems to be difficult to do all the required computations in concrete examples. Therefore, it is our first aim to become familiar with the definition and to find rules to faciliate these computations. A first list of such rules is contained in the following proposition.

XII. 4 Proposition: For the entropy of a dynamical System $(X, \Sigma, \mu; \varphi)$, resp. $(X;\varphi)$, and a measurable, resp. open, cover $\alpha$ the following statements are valid:

(i) $h(X;\varphi) = \sup\{h_\mu(\alpha;\varphi) : \alpha \in \tilde{\Sigma}_d\}$,

(ii) $h(\alpha;\varphi) = h(\alpha_k^m;\varphi)$ for $k < m \in \mathbb{Z}$,

(iii) $h(\alpha;\varphi) = \lim_{n\to\infty} \frac{1}{2n} h(\alpha_{-n}^n)$,

(iv) $h(\alpha;\varphi)$ is the Cesaro limit of the sequence $(h(\alpha^{n+1}) - h(\alpha^n))_{n\in\mathbb{N}}$,

(v) $h(X;\varphi^m) = |m|\cdot h(X;\varphi)$ for $m \in \mathbb{Z}$.

Proof:

(i) The set $\tilde{\Sigma}_d$ of disjoint measurable covers is cofinal in $\tilde{\Sigma}$, i.e. for every $\alpha \in \tilde{\Sigma}$ there exists $\sigma \in \tilde{\Sigma}_d$ which is finer than $\alpha$. Therefore, the assertion follows from the monotonicity of $h_\mu$.

(ii) For any cover $\beta$ we have $h(\beta) = h(\tilde{\varphi}\beta)$. Therefore
$$\frac{1}{n} h((\alpha_k^m)^n) = \frac{1}{n} h(\alpha_k^{n+m-1}) = \frac{1}{n} h(\alpha^{n+m-k-1}) =$$
$$= \frac{n+m-k-1}{n} \frac{1}{n+m-k-1} h(\alpha^{n+m-k-1})$$ converges to $h(\alpha;\varphi)$.

(iii) The sequence $(\frac{1}{2n} h(\alpha_{-n}^n))_{n\in\mathbb{N}} = (\frac{1}{2n} h(\alpha^{2n}))_{n\in\mathbb{N}}$ converges to $h(\alpha;\varphi)$.

(iv) It suffices to observe that
$$\frac{1}{n} \sum_{i=0}^{n-1} (h(\alpha^{i+1}) - h(\alpha^i)) = \frac{1}{n}(h(\alpha^n) - h(\alpha^0)) = \frac{1}{n} h(\alpha^n).$$

(v) Assume first $0 < m \in \mathbb{N}$. By (ii) we obtain
$$h(\alpha;\varphi^m) = h(\alpha^m;\varphi^m) = \lim \frac{1}{n} h(\alpha^m \vee \tilde{\varphi}^m \alpha^m \vee \dots \vee \tilde{\varphi}^{m(n-1)} \alpha^m)$$
$$= \lim \frac{1}{n} h(\alpha^{mn}) = m \lim \frac{1}{mn} h(\alpha^{mn}) = m\, h(\alpha;\varphi).$$
For $m = 0$ we have $h(\alpha;\mathrm{id}) = \lim \frac{1}{n} h(\alpha) = 0$. Finally, it suffices to show that
$$h(\alpha;\varphi^{-1}) = h(\alpha;\varphi) : \lim \frac{1}{n} h(\alpha_{1-n}^1) = \lim \frac{1}{n} h(\tilde{\varphi}^{n-1}\alpha_{1-n}^1) = \frac{1}{n} h(\alpha_0^n).$$

Remark: The term
$$h(\alpha^{n+1}) - h(\alpha^n) = h(\alpha_{-n}^1) - h(\alpha_{-n}^0)$$ occuring in (iv),

can be interpreted as follows:

Given all the information provided by $\alpha$ about the points $\varphi^i(x)$ "for a long time in the past" ($i = -1, \dots, -n$), this may still be insuffient to predict the position of x, because the additional information provided by $\alpha$ about x is $h(\alpha_{-n}^1) - h(\alpha_{-n}^0) \geq h(\alpha;\varphi)$. Therefore, if $h(\alpha;\varphi) > 0$ for every $\alpha$ this implies that the dynamics $\varphi$ is not fully predictable, although the mapping $\varphi$ is well defined. By (XI.D.3), $h(\alpha_{-n}^1) - h(\alpha_{-n}^0) = h(\alpha|\alpha_{-n}^0)$ decreases monotonically to $h(\alpha;\varphi)$ as n .

But even with these "computation rules" in mind the calculation of $h(X;\varphi)$ for non-trivial exaples remains dificult, if not impossible.

Exercise: Calculate $h(X;\varphi)$ for periodic $\varphi$.

The main difficulty is caused by the supremum in the Definitions (XII.1) and (XII.2), since it requires to calculate the entropy for <u>all</u> covers - a seemingly hopeless task. However, if we are able to find a cover whose entropy dominates the entropy of all covers, we can overcome this problem. For instance, assume that our dynamical system admits a cover having the following property:

(✱) For every cover $\beta$ there exists $m \in \mathbb{N}$ such that $\beta \leq \alpha^m$.

Using (XII.4.ii) and the monotonicity of h we conclude that

$$h(\beta;\varphi) \leq h(\alpha^m;\varphi) = h(\alpha;\varphi),$$

hence $H(X;\varphi) = h(\alpha;\varphi)$. Therefore, it suffices to calculate the entropy of a single cover $\alpha$, a task solvable in many cases.

Unfortunately, dynamical systems possessing a cover $\alpha$ with property (✱) are rather rare.

But - and this was the essential observation of Kolmogorov and Sinai in 1956 - for MDSs it will be possible to weaken the condition (✱) such that, firstly, it will be satisfied in many cases and, secondly, it still permits the same conclusion. The reason for this is the fact, that in the measure-theoretical case (in contrast to the topological case, see Keynes-Robertson [1969]) the metric induced by the measure $\mu$ on $\Sigma$ is non-discrete and yields a non-trivial metric on $\tilde{\Sigma}$ thus permitting us to say when covers are "almost equal" or "almost finer".

XII. 5 <u>Definition:</u> Let $(X, \Sigma, \mu;\varphi)$ be an MDS and $\alpha, \beta \in \tilde{\Sigma}$.
Then we introduce the following notation.

(i) $d(\alpha,\beta) := \min\left\{\sum_{A\in\alpha}\mu(A \,\Delta\, \pi(A)) : \pi : \alpha \to \beta \text{ is bijective}\right\}$; here we assume that $|\alpha| = |\beta|$ which may be obtained by adding a suitable number of $\mu$-null sets to $\alpha$ or $\beta$.

(ii) $\beta \underset{\varepsilon}{\leq} \alpha$ ("$\alpha$ <u>is finer than</u> $\beta$ <u>up to</u> $\varepsilon$") if there is a cover $\beta' \in \tilde{\Sigma}$ such that $\beta' \leq \alpha$, $|\beta| = |\beta'|$ and $d(\beta,\beta') < \varepsilon$ for $\varepsilon > 0$.

(iii) A subset W of $\tilde{\Sigma}$ is called <u>generating</u> if for every $\varepsilon > 0$ and every $\beta \in \tilde{\Sigma}$ there is $\alpha \in W$ and $n \in \mathbb{N}$ such that $\beta \underset{\varepsilon}{\leq} \alpha^{n}_{-n}$.

(iv) If a generating subset W only contains one element $\alpha$, then $\alpha$ is called a <u>generator</u>.

Certainly, the condition in (XII.5iii) is a weakening of the property (✱) above. In order to prove that the entropy of $(X,\Sigma,\mu;\varphi)$, i.e. the "supremum" in (XII.1), is already attained on a single cover $\alpha$ satisfying (XII.5.iii), it is essential to show that $\beta \underset{\varepsilon}{\leq} \alpha$ implies some estimate of $h_\mu(\beta)$ by $h_\mu(\alpha)$. This is achieved by the following lemma.

XII. 6 <u>Lemma:</u> The function $f : \mathbb{N} \times \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$f(n,\varepsilon) := -(1-\varepsilon)\log(1-\varepsilon) - \varepsilon\log\varepsilon + \varepsilon\log n$$

has the following two properties:

(i) $\lim_{\varepsilon\to 0} f(n,\varepsilon) = 0$ for every $n \in \mathbb{N}$ .

(ii) Let $\varepsilon > 0$ and $\alpha,\beta \in \tilde{\Sigma}$ . Then $\beta \underset{\varepsilon}{\leq} \alpha$ implies the existence of $\gamma \in \tilde{\Sigma}$ such that $\beta \leq \alpha \vee \gamma$ and $h_\mu(\gamma) \leq f(|\beta|, \varepsilon)$.

Proof:

The first assertion follows from analyzing the function

$$x \mapsto x \log x \quad \text{for } x \in [0, 1] .$$

For $\beta = \{B_1,\dots, B_n\}$ there exists $\beta' = \{B_1',\dots, B_n'\}$ such that $\beta' \leq \alpha$ and $d(\beta;\beta') = \sum_{i=1}^{n}\mu(B_i \triangle B_i') < \varepsilon$ . Consider $D := \bigcup_{i=1}^{n} B_i \triangle B_i'$, $C := X \setminus D$ and $\gamma := \{C, D\cap B_i : i = 1,\dots, n\} \in \tilde{\Sigma}$ . We claim that $\beta \leq \alpha\vee\gamma$: Take an element of $\alpha\vee\gamma$ . If it is of the form $A\cap D\cap B_i$, $A \in \alpha$ , then it is contained in $B_i \in \beta$ . If it is of the form $A\cap C$, $A \in \alpha$ , then there is a $B_i' \in \beta'$ such that $A \subset B_i'$. This implies that

$$A\cap C \subset B_i'\cap C = (B_i' \cup (B_i \triangle B_i'))\cap C = (B_i \cup (B_i \triangle B_i'))\cap C = B_i\cap C \subset B_i \in \beta .$$

Now we estimate $h_\mu(\gamma)$ : Let $\delta := \{C, D_1,\dots, D_n\}$ be a "disjointification" (see XI.D.2) of $\gamma$ such that $D_i \subset D\cap B_i$. Then $\gamma \leq \delta$ and

$$h_\mu(\gamma) \leq h^*_\mu(\delta) = -\mu(C)\log\mu(C) - \sum_{i=1}^{n}\mu(D_i)\log\mu(D_i)$$

$$= -\mu(C)\log\mu(C) - \mu(D)\log\mu(D) - \mu(D)\sum_{i=1}^{n}\mu(D_i)\log\mu(D_i) \leq f(|\beta|, \varepsilon)$$

(use XI.9).

After this preparation the main result, on which most calculations of $h_\mu$ are based, is a rather simple consequence.

XII. 7 <u>Theorem:</u> If $(X,\Sigma,\mu;\varphi)$ is an MDS and W a generating subset of $\tilde{\Sigma}$, then $h_\mu(X;\varphi) = \sup\{h_\mu(\alpha;\varphi) : \alpha \in \tilde{W}\}$.

Proof:

For every $\varepsilon > 0$ and $\beta \in \tilde{\Sigma}$ there exists $\alpha \in W$ and $n \in \mathbb{N}$ such that $\beta \underset{\varepsilon}{\leq} \alpha^{n}_{-n}$. By (XII.6) we can find $\gamma \in \tilde{\Sigma}$ such that $\beta \leq \alpha^{n}_{-n} \vee \gamma$ and $h_\mu(\gamma) \leq f(|\beta|, \varepsilon)$. By the standard computation we obtain

$$h_\mu(\beta^k) \leq h_\mu(\alpha^{n}_{-n} \vee \tilde{\varphi}\alpha^{n}_{-n} \vee \dots \vee \tilde{\varphi}^{k-1}\alpha^{n}_{-n}) + k\, h_\mu(\gamma)$$

and

$h_\mu(\beta;\varphi) \leq h_\mu(\alpha^n_{-n};\varphi) + h_\mu(\gamma) \leq h_\mu(\alpha;\varphi) + f(|\beta|;\varepsilon)$,
hence $h_\mu(\beta;\varphi) \leq h_\mu(\alpha;\varphi)$.

XII. 8 Corollary (Kolmogorov - Sinai, 1958): Let $(X,\Sigma,\mu;\varphi)$ be an MDS and $\alpha \in \tilde{\Sigma}_d$. If $\Sigma$ is generated by $\bigcup_{n=0}^{\infty}\alpha^n_{-n}$, then $h_\mu(X;\varphi) = h_\mu(\alpha;\varphi)$.

Proof:
In (A.11) we mentioned that for any subset W of $\Sigma$ the closure with respect to the metric d of the algebra $\mathfrak{A}(W)$ generated by W is the $\sigma$-algebra $\sigma(W)$ generated by W. Therefore

$$\Sigma = \sigma(\bigcup_{n=0}^{\infty}\alpha^n_{-n}) = \overline{\mathfrak{A}(\bigcup_{n=0}^{\infty}\alpha^n_{-n})} = \overline{\bigcup_{n=0}^{\infty}\mathfrak{A}(\alpha^n_{-n})} .$$

Thus for $B \in \Sigma$ and $\varepsilon > 0$ there exists $n \in \mathbb{N}_0$ and $B' \in \mathfrak{A}(\alpha^n_{-n})$ such that $\mu(B \vartriangle B') < \varepsilon$. As a consequence, for $\varepsilon > 0$ and $\beta \in \tilde{\Sigma}$ we find $n \in \mathbb{N}_0$ and $\beta' \in \tilde{\Sigma}$, $\beta' \subset \mathfrak{A}(\alpha^n_{-n})$ such that $d(\beta,\beta') < \varepsilon$. Since $\alpha$ is disjoint, $\beta' \subset \mathfrak{A}(\alpha^n_{-n})$ implies that $\beta' \leq \alpha^n_{-n}$, and we have shown that $\alpha$ is a $\mu$-generator. ■

The most important concrete examples for which we are able to compute the entropy are Markov- and, in particular, Bernoulli-shifts.

XII. 9 Proposition: Let $(\hat{X},\hat{\Sigma},\hat{\mu};\varphi)$ be the Markov shift with invariant distribution $\mu = \begin{pmatrix} p_0 \\ p_{k-1} \end{pmatrix}$ and transition matrix $T = (a_{ij})$ (see II.6).

Then $h_{\hat{\mu}}(\hat{X};\varphi) = -\sum_{i,j} p_i\, a_{ij} \log a_{ij}$ .

Proof:
On $\hat{X} = \{0,\dots,k-1\}^{\mathbb{Z}}$ choose the disjoint cover $\alpha = \{A_0,\dots,A_{k-1}\}$ where $A_i := \{(x_j)_j \in \hat{X} : x_0 = i\}$. From the definition of the product $\sigma$-algebra on X it follows that we can apply (XII.8) to obtain $h_{\hat{\mu}}(\alpha;\varphi) = h_{\hat{\mu}}(\hat{X};\varphi)$. To calculate this number, we use (XII.4.iv):

$$\begin{aligned}
& h_{\hat{\mu}}(\alpha^{n+1}) - h_{\hat{\mu}}(\alpha^n) \\
&= -\sum_{A\in\alpha^n}\sum_{B\in\varphi(\alpha)}\mu(A\cap B)\log\mu(A\cap B) + \sum_{A\in\alpha^n}\mu(A)\log\mu(A) \\
&= -\sum_{A\in\alpha^n}\sum_{B\in\varphi(\alpha)}\mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(A)} \\
&= -\sum_{i_0,\dots,i_{n-1}}\sum_{i_n} p_{i_0}\, a_{i_0 i_1}\, a_{i_1 i_2}\cdots a_{i_{n-1} i_n}\log a_{i_{n-1} i_n} \\
&= -\sum_{i_{n-1}}\sum_{i_n} p_{i_{n-1}}\, a_{i_{n-1} i_n}\log a_{i_{n-1} i_n}, \text{ since } \sum_i p_i\, a_{ij} = p_j .
\end{aligned}$$
■

XII.10 <u>Corollary:</u> Let $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \varphi)$ denote the Bernoulli shift $B(p_0, \ldots, p_{k-1})$. Then $h_\mu(\hat{X}; \varphi) = -\sum_{i=0}^{k-1} p_i \log p_i$ .

Proof:

This is a special case of (XII.9) for $a_{ij} = p_j$. Therefore,

$-\sum_{i,j} p_i\, a_{ij} \log a_{ij} = -\sum_j p_j \log p_j$ . ∎

<u>Remark:</u>

In the next lecture we shall show that MDSs arising from rotations on the unit circle $\Gamma$ have entropy zero.

Theorem (XII.7) may be applied immediately to obtain a relation between $\mu$-entropy and topological entropy.

XII.11 <u>Corollary</u> (Goodwyn, 1969): Let $(X;\varphi)$ be a TDS and $(X, \mathcal{B}, \mu; \varphi)$ an MDS for a $\varphi$-invariant probability measure $\mu$ on X. Then

$$h_\mu(X;\varphi) \leq h_t(X;\varphi) .$$

Proof:

Let $\mathcal{O}$ denote the family of all open subsets of X. Then $\tilde{\mathcal{O}}$ satisfies the assumptions of (XII.7). Since the $\mu$-information of open covers is dominated by the t-information (see XI.9), we conclude that

$h_\mu(X;\varphi) = \sup\{h_\mu(\alpha;\varphi) : \alpha \in \tilde{\mathcal{O}}\} \leq \sup\{h_t(\alpha;\varphi) : \alpha \in \tilde{\mathcal{O}}\} = h_t(X;\varphi)$. ∎

Finally we show that the dynamical entropy is affine.

XII.12 <u>Proposition:</u> Let $(X;\varphi)$ be a TDS. If $\mu$ and $\nu$ are $\varphi$-invariant probability measures on X, we have

$h_{\lambda\mu + (1-\lambda)\nu}(\alpha;\varphi) = \lambda h_\mu(\alpha;\varphi) + (1-\lambda) h_\nu(\alpha;\varphi)$

for $0 < \lambda < 1$ and every cover $\alpha \in \tilde{\mathcal{B}}_d$.

Proof:

For $A \in \mathcal{B}$ and $m := \lambda\cdot\mu(A) + (1-\lambda)\nu(A)$ we have

$m \log m \quad - \lambda\mu(A) \log\mu(A) - (1-\lambda)\nu(A) \log\nu(A)$

$\quad = \lambda\mu(A)(\log m - \log\mu(A)) + (1-\lambda)\nu(A)(\log m - \log\nu(A))$

$\quad = \lambda\mu(A)(\log m - \log(\lambda\mu(A))) + (1-\lambda)\nu(A)(\log m - \log((1-\lambda)\nu(A)))$

$\qquad + \mu(A)\lambda\log\lambda + \nu(A)(1-\lambda)\log(1-\lambda)$

$\quad \geq \mu(A)\lambda\log\lambda + \nu(A)(1-\lambda)\log(1-\lambda)$.

For every cover $\alpha \in \tilde{\mathcal{B}}_d$ this yields

$-h_{\lambda\mu+(1-\lambda)\nu}(\alpha^n) + \lambda h_\mu(\alpha^n) + (1-\lambda) h_\nu(\alpha^n)$

$\geq \lambda\log\lambda + (1-\lambda)\log(1-\lambda) \geq -\log 2$ by Appendix Z, Property 3.

Now the first inequality follows by dividing by $n \in \mathbb{N}$.
In view of (XI.9.ii) we have the converse inequality, and the assertion is proved. ■

## XII. D Discussion

### XII. D.1 Entropy is a complete invariant for Bernoulli shifts.

The fact that we are able to compute the entropy of a Bernoulli shift quite easily in (XII.10) shows that there are infinitely many non-isomorphic shifts: For any number $r \in [0,1]$ we can produce a probability vector $(p_1, p_2)$ such that $B(p_1,p_2)$ has the entropy $r = - p_1 \log p_1 - p_2 \log p_2$. Thus we can distinguish different Bernoulli shifts (for example: $B(\frac{1}{2},\frac{1}{2})$ and $B(\frac{1}{3},\frac{2}{3})$) although they are all spectrally isomorphic (see VI.D.5).

But we are faced with a remaining problem: The Bernoulli shifts $B(\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{2})$ and $B(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$ have the same entropy 2 log 2 and therefore might be isomorphic.

Do we need another isomorphism invariant to distinguish between them ? The answer is "no". Meshalkin [1959] and Blum-Hanson [1963] showed that in fact $B(\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{2})$ is isomorphic to $B(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$ .

Finally, in 1970 Ornstein gave the most general answer to this problem.

Theorem: (Ornstein, 1970): Two Bernoulli shifts are isomorphic if and only if the have the same entropy, i.e. the entropy is a complete isomorphism invariant for Bernoulli shifts.

In general the entropy is far from being a complete invariant as can be seen from (XIII.7).

The subsequent table shows that we have solved the isomorphism problem for two extremely opposite cases:

| | point spectrum $P_\sigma(T_\varphi)$ | entropy $h_\mu$ |
|---|---|---|
| ergodic and discrete spectrum | any group $G \subset \Gamma$ (complete isomorphism invariant) | 0 |
| $B(p_0,\dots,p_{k-1})$ | $\{1\}$ | any $r > 0$ (complete isomorphism invariant) |

XII. D.2 The usual definition of a generator. Usually a generator $\alpha$ of an MDS $(X, \Sigma, \mu; \varphi)$ is defined by the property appearing in (XII.8):

$$\sigma\Big(\bigcup_{n\in\mathbb{N}} \alpha^{n}_{-n}\Big) = \Sigma .$$

But for disjoint covers this coincides with our more general definition in (XII.5).

Proposition: Let $(X, \Sigma, \mu; \varphi)$ be an MDS. For $\alpha \in \widetilde{\Sigma}_d$ the following are equivalent:

(a) $\alpha$ is a generator.

(b) $\sigma\Big(\bigcup_{n=1}^{\infty} \alpha^{n}_{-n}\Big) = \Sigma .$

Proof:

For the implication (b) $\Rightarrow$ (a) see (XII.8).

(a) $\Rightarrow$ (b): Let $\alpha$ be a generator and $B \in \Sigma$. For $\beta := \{B, X\setminus B\}$ and $\varepsilon > 0$ we find $n \in \mathbb{N}$ and $\beta' \in \hat{\widetilde{\Sigma}}$ such that $\beta' \leq \alpha^{n}_{-n}$ and $d(\beta, \beta') < \varepsilon$. We may assume that $\beta' = \{B_1, B_2\}$ and $\mu(B_1 \triangle B) + \mu(B_2 \triangle (X\setminus B)) < \varepsilon$.

Then we have

$$\begin{aligned}
\mu(B_1 \cap B_2) &\leq \mu(B \cap B_2) + \quad ((B_1 \triangle B) \cap B_2) \\
&\leq \mu(B \cap (X\setminus B)) + \mu(B \cap (B_2 \triangle (X\setminus B))) + \mu((B_1 \triangle B) \cap B_2) \\
&\leq \mu(B_2 \triangle (X\setminus B)) + \mu(B_1 \triangle B) < \varepsilon .
\end{aligned}$$

If $B'$ denotes the union of all $A \in \alpha^{n}_{-n}$ such that $A \subset B_1$, then

$$\mu(B_1 \triangle B') = \mu(B_1 \setminus B') \leq \mu(B_1 \cap B_2) < \varepsilon ,$$

and therefore

$$\mu(B \triangle B') \leq \mu(B \triangle B_1) + \mu(B_1 \triangle B') < 2\varepsilon .$$

Thus $B \quad \overline{\bigcup_{n\in\mathbb{N}} \mathfrak{A}(\alpha^{n}_{-n})} = \sigma\Big(\bigcup_{n\in\mathbb{N}} \alpha^{n}_{-n}\Big)$ (see XII.3).

XII. D.3 Which MDSs have a generator ? We have seen that it is much easier to calculate the entropy of an MDS once we have found a generator. But the existence of a generator seems to be rather restrictive, hence the following theorem of Krieger [1970] is a relief:

Theorem (Krieger, 1970): Let $(X, \Sigma, \mu; \varphi)$ be an ergodic MDS with separable measure algebra and finite entropy $h_\mu(X; \varphi)$. Then there is a disjoint generator $\alpha$ with $|\alpha| \leq 2^{h_\mu} + 1$.

Once we have found such a disjoint generator $\alpha$, we may represent $(X, \Sigma, \mu; \varphi)$ as a shift: For $\alpha = \{A_1, \ldots, A_n\}$ we define

$$\psi : X \longrightarrow \{1, \ldots, n\}^{\mathbb{Z}} := S$$

by $\psi(x) := (k_i)_{i \in \mathbb{Z}}$ where $\varphi^i(x) \in A_{k_i}$. Then

$$\psi \circ \varphi = \tau \circ \psi$$

where $\tau$ is the shift on S. The transferred measure

$$\nu := \mu \circ \psi^{-1} \quad \text{on} \quad \mathcal{T} := \psi(\Sigma)$$

is $\tau$-invariant. The MDS $(S, \mathcal{T}, \nu; \tau)$ is called a "general shift".

In view of Krieger's theorm this shows that every ergodic separable MDS with finite entropy is isomorphic to a general shift. Therefore we can see that it is neither the sequence space $\{0, \ldots, k-1\}$ nor the shift transformation that is so special about a Bernoulli (or Markov) shift. In fact, the essence lies in certain "independence" properties of the measure, which will be characterized in more detail in (App.T ).

XII. D 4 The entropy of $\varphi \otimes \psi$:

Given two MDSs $(X, \Sigma, \mu; \varphi)$ and $(Y, \mathcal{T}, \nu; \psi)$ we may consider $X \times Y$ endowed with the product $\sigma$-algebra and the product measure $\mu \otimes \nu$. The product transformation $\varphi \otimes \psi$ defined by

$$\varphi \otimes \psi(x,y) := (\varphi(x), \psi(y))$$

preserves $\mu \otimes \nu$, and we are able to compute its entropy.

Proposition: $h_{\mu\otimes\nu}(X \times Y; \varphi \otimes \psi) = h_\mu(X; \varphi) + h_\nu(Y; \psi)$.

Proof: For any $\alpha \in \tilde{\Sigma}_d$ we define

$$\alpha \times Y := \{A \times Y : A \in \alpha\}$$

and analogously $X \times \beta$ for $\beta \in \tilde{\mathcal{T}}_d$. Then we have

$$h_{\mu\otimes\nu}(\alpha \times Y \vee X \times \beta) = h_{\mu\otimes\nu}(\alpha \times Y) + h_{\mu\otimes\nu}(X \times \beta) \quad \text{by (XI.5)}$$
$$= h_\mu(\alpha) + h_\nu(\beta),$$

and analogously,

$$h_{\mu\otimes\gamma}((\alpha \times Y \vee X \times \beta)^n) = h_{\mu\otimes\nu}((\alpha \times Y)^n \vee (X \times \beta)^n)$$
$$= h_{\mu\otimes\nu}(\alpha^n \times Y \vee X \times \beta^n)$$
$$= h_\mu(\alpha^n) + h_\nu(\beta^n).$$

This shows that

$$h_\mu(X; \varphi) + h_\nu(Y; \psi) \leq h_{\mu\otimes\nu}(X \times Y; \varphi \otimes \psi).$$

For the converse inequality consider

$$M := \{A \times B : A \in \Sigma,\ B \in \mathcal{T}\} \quad \text{and} \quad \tilde{M} := \{\alpha \in \tilde{\Sigma} : \alpha \subset M\}$$

which is a generating subset of $\hat{\Sigma}$.

Thus

$$h_{\mu\otimes\nu}(X \times Y; \varphi \otimes \psi) = \sup\{h_{\mu\otimes\nu}(\gamma; \varphi \otimes \psi) : \gamma \in M,\ \gamma \text{ cover of } X \times Y\}.$$

Let $\gamma = \{A_1 \times B_1, \ldots, A_n \times B_n\} \subset M$, $\alpha := \{A_1, \ldots, A_n\}$ and $\beta := \{B_1, \ldots, B_n\}$.

The Boolean algebras generated by $\alpha$ and by $\beta$ are both finite, and we denote the covers made up of their atoms by $\alpha_1 \in \tilde{\Sigma}_d$ and $\beta_1 \in \tilde{\mathcal{T}}_d$, respectively.
Then we obtain

$$\gamma \leq \alpha_1 \times Y \vee X \times \beta_1$$

which implies

$$h_{\mu\otimes\nu}(\gamma) \leq h_{\mu\otimes\nu}(\alpha_1 \times Y \vee X \times \beta_1) = h_\mu(\alpha_1) + h_\nu(\beta_1)$$

and the desired inequality. ■

XII. D.5 Topological entropy of the shift:
Consider $\hat{X} := \{0,1,\dots,k-1\}^{\mathbb{Z}}$
as a compact metric space with the metric

$$d(x,y) := \sum_{i=0}^{\infty} \delta(x_i,y_i)\cdot 4^{-i} + \sum_{i=-1}^{-\infty} \delta(x_i,y_i)\cdot 4^{i}$$

where $\delta(r,s) := \begin{cases} 0 & \text{if } r = s \\ 1 & \text{if } r \neq s \end{cases}$ for $r,s \in \{0,1,\dots,k-1\}$.

The (left) shift $\tau : (x_i) \rightarrow (x_{i+1})$
is a homeomorphism on $\hat{X}$, hence $(\hat{X};\tau)$ is a TDS called the topological k - shift.
For any closed $\tau$-invariant subset $\hat{Y}$ of $\hat{X}$ we may consider the TDS $(\hat{Y};\tau)$ called a topological subshift.

Proposition: Let $(\hat{Y};\tau)$ be a topological subshift and consider the open cover $\alpha := \{A_0, A_1,\dots, A_{k-1}\}$

where $A_i := \{x \in \hat{Y} : x_0 = i\}$ for $i = 0, 1,\dots, k-1$. Then $h_t(\hat{Y};\tau) = h_t(\alpha;\tau)$.

Proof:
Take $\beta \in \tilde{\mathcal{O}}$. By Lebesgue's lemma (A.5) there exists $\varepsilon = 4^{-m}$, $m \in \mathbb{N}$, such that for every $x \in X$ the open $\varepsilon$-ball $U_\varepsilon(x)$ is contained in some $B \in \beta$. By definition of the metric we have $y \in U_\varepsilon(x)$ if and only if the coordinates $x_i = y_i$ for $i = -m,\dots, m$.
This implies that $U_\varepsilon(x) \in \alpha_{-m}^{m}$, hence $\beta \leq \alpha_{-m}^{m}$. Now we immediately obtain $h_t(\beta;\tau) \leq h_t(\alpha_{-m}^{m};\tau) = h_t(\alpha;\tau)$ by (XII.4.ii). ■

Corollary: The topological k-shift has the topological entropy log k.

Proof:

For the cover $\alpha$ appearing in the above proposition we have

$$h_t(\alpha ; \tau) = \lim_{m\to\infty} \frac{1}{m} \log |\alpha^m| = \log |\alpha| = \log k. \quad ■$$

We remark that the cover $\alpha$ may be called a topological generator, and refer to Keynes-Robertson [1969] for more information.

XII. D.6 Another approach to the Kolmogorov-Sinai theorem:

The proof of the Kolmogorov-Sinai theorem (XII.7 and XII.8) requires essentially two steps:

(i) the hypothesis " $\sigma(\bigcup_{n=1}^{\infty} \alpha_{-n}^{n}) = \Sigma$ " implies that any cover $\beta$ can be approximated by $\alpha_{-n}^{n}$ for sufficiently large n.

(ii) a continuity argument for the entropy ensures that $h_\mu(\alpha ; \varphi)$ is close to $h_\mu(\beta ; \varphi)$ is $\alpha$ is close to $\beta$.

We have used the metric d on $\tilde{\Sigma}$. Therefore the first step was rather simple (see XII.8), whereas the second step required some trickery (see XII.6).

Proposition 1: The entropy mapping

$$\alpha \longmapsto h_\mu(\alpha ; \varphi)$$

is continuous on $\tilde{\Sigma}^n$ for the metric d.

We refer to (XI.D.4) for the notation; the proof follows immediately from (XII.6).

Another approach to the Kolmogorov-Sinai theorem uses the Rohlin metric (see XI.D.4). In this case the first step becomes difficult (see,for exsmple, Walters [1975], IV.4.8) whereas the second step is quite simple.

Proposition 2: The entropy mapping

$$\alpha \longmapsto h_\mu(\alpha ; \varphi)$$

is continuous on $\tilde{\Sigma}_d$ with the metric $\varrho$.

Proof:

Assume $\varrho(\alpha, \beta) = h_\mu(\alpha | \beta) + h_\mu(\beta | \alpha) < \varepsilon$ . Then

$$\begin{aligned} h_\mu((\alpha \vee \beta)^n) - h_\mu(\alpha^n) &= h_\mu((\alpha \vee \beta)^n | \alpha^n) \\ &\leq \sum_{i=1}^{n} h_\mu(\tilde{\varphi}^i(\alpha \vee \beta) | \alpha^n) \\ &\leq \sum_{i=1}^{n} h_\mu(\tilde{\varphi}^i(\alpha \vee \beta) \tilde{\varphi}^i \alpha) \text{ (by XI.D.3, Corollary)} \end{aligned}$$

$$= n \cdot h_\mu(\alpha \vee \beta \mid \alpha)$$
$$= n \cdot h_\mu(\beta \mid \alpha).$$

Therefore $0 \le h_\mu(\alpha \vee \beta ; \varphi) - h_\mu(\alpha ; \varphi) \le h_\mu(\beta \mid \alpha)$
and $0 \le h_\mu(\alpha \vee \beta ; \varphi) - h_\mu(\beta ; \varphi) \le h_\mu(\alpha \mid \beta)$, which implies

$$|h_\mu(\alpha ; \varphi) - h_\mu(\beta ; \varphi)| \le h_\mu(\alpha \mid \beta) + h_\mu(\beta \mid \alpha) < \varepsilon . \blacksquare$$

References: Rohlin [1967], Walters [1975].

XII. D.7 Entropy and discrete spectrum:
Intuitively, the ergodic MDSs with discrete spectrum are at the opposite extreme to the ergodic MDSs having large entropy (compare XII.D.1). But it is very surprising that the purely functional-analytic property " discrete spectrum" can be characterized by the concept of entropy.

Let $(X, \Sigma, \mu ; \varphi)$ be an MDS. As before we denote by $\tilde{\Sigma}^n_d$ the set of all disjoint measurable covers of X having cardinality at most n. On $\tilde{\Sigma}^n_d$ we consider the topology induced by the metric $\varrho$ (or d) (see XI.D.4).

Lemma: For a subset K of $\tilde{\Sigma}^2_d$ the following assertions are equivalent:
(a) $\overline{K}$ is compact.
(b) $\lim_{n\to\infty} \frac{1}{n} h_\mu(\bigvee_{i=1}^{n} \alpha_i) = 0$ for every sequence $(\alpha_i)_{i\in\mathbb{N}} \subset K$.

Proof:
(a) $\Rightarrow$ (b): Assume that $\overline{K}$ is compact and take $(\alpha_i)_{i\in\mathbb{N}} \subset K$ and $\varepsilon > 0$. The oben $\varepsilon$-neighborhoods $U_\varepsilon(\alpha_i)$ of $\alpha_i$, $i \in \mathbb{N}$, for the metric $\varrho$ form an open cover of the closure of $\{\alpha_i : i \in \mathbb{N}\}$. Therefore there exist $i_1 < i_2 < \dots < i_n$ such that

$$\overline{\{\alpha_i : i \in \mathbb{N}\}} \subset \bigcup_{j=1}^{n} U_\varepsilon(\alpha_{i_j}) .$$

Then for any $k > i_n$ there is an $i \le i_n$ such that $\varrho(\alpha_k, \alpha_i) < \varepsilon$. Therefore
$h_\mu(\bigvee_{j=1}^{k} \alpha_j) - h_\mu(\bigvee_{j=1}^{k-1} \alpha_j) = h_\mu(\alpha_k \mid \bigvee_{j=1}^{k-1}\alpha_j) \le h_\mu(\alpha_k \mid \alpha_i) \le \varrho(\alpha_k, \alpha_i) < \varepsilon$,
which implies the assertion.
(b) $\Rightarrow$ (a): If $\overline{K}$ is not compact, then there is an $\varepsilon > 0$ and a sequence $(\alpha_i) \subset K$ such that $\varrho(\alpha_i, \alpha_j) \ge \varepsilon$ for $i \ne j$.
Next, we construct $\delta > 0$ and a subsequence $\alpha_{i_n}$ such that

$$h_\mu(\alpha_{i_n} \mid \bigvee_{j=1}^{n-1} \alpha_{i_j}) > \delta \qquad \text{for all } n \in \mathbb{N} :$$

First, we recall that by (iii) in Theorem (XI.D.4) there is $\delta > 0$ such

that $h_\mu(\beta \mid \alpha) \leq \delta$ implies $\beta \underset{\varepsilon/2}{\leq} \alpha$ for any $\alpha \in \tilde{\Sigma}_d$ and $\beta \in \tilde{\Sigma}_d^2$.
Now we set $i_1 := 1$ and assume that $i_1, \ldots, i_{n-1}$ have been chosen suitably. Consider the set

$$M_n := \{ \beta \in \tilde{\Sigma}_d^2 : h_\mu(\beta \mid \bigvee_{j=1}^{n-1} \alpha_{i_j}) \leq \delta \}.$$

This set is contained in a finite union of $\frac{\varepsilon}{2}$ - balls, hence $M_n \cap \{\alpha_i : i \in \mathbb{N}\}$ is finite, and we can find $i_n > i_{n-1}$ such that $\alpha_{i_n} \notin M_n$, i.e. $h_\mu(\alpha_{i_n} \mid \bigvee_{j=1}^{n-1} \alpha_{i_j}) > \delta$.

By (i) in Proposition (XI.D.3) and elementary analysis, this implies that

$$\lim_{n \to \infty} \frac{1}{n} h_\mu(\bigvee_{j=1}^{n} \alpha_{i_j}) \geq \delta . \blacksquare$$

<u>Theorem</u> (Kushnirenko, 1967) : For an MDS $(X, \Sigma, \mu; \varphi)$ and a sequence $s := (i_j)_{j \in \mathbb{N}}$ of positive integers we define

$$h_s(\alpha; \varphi) := \lim \frac{1}{n} h_\mu(\bigvee_{j=1}^{n} \tilde{\varphi}^{i_j} \alpha) \quad \text{and}$$

$$h_s(X; \varphi) := \sup_\alpha h_s(\alpha; \varphi) .$$

Then, $(X, \Sigma, \mu; \varphi)$ has discrete spectrum if and only if $h_s(X; \varphi) = 0$ for every sequence s.

Proof:
The MDS $(X, \Sigma, \mu; \varphi)$, resp. $T_\varphi$, has discrete spectrum if and only if $\{T_\varphi^n 1_A : n \in \mathbb{N}\}$ is relatively compact in $L^1(\mu)$ for every $A \in \Sigma$ (see VII.6).

Consider the mapping

$$\Theta : \{1_A : A \in \Sigma\} \longrightarrow \tilde{\Sigma}_d^2$$

which associates to every characteristic function $1_A$ the disjoint cover

$$\Theta(1_A) := \alpha_A := \{A, X \setminus A\}.$$

$\Theta$ is continuous since $\| 1_A - 1_B \| = \frac{1}{2} d(\alpha_A, \alpha_B)$, see (XI.D.4). Moreover, a subset $K \subseteq \{1_A : A \in \Sigma\}$ is compact iff $\Theta(K)$ is compact. Therefore, we observe that $T_\varphi$ has discrete spectrum if and only if $\overline{\{\tilde{\varphi}^n \alpha_A : n \in \mathbb{N}\}}$ is compact for every $A \in \Sigma$.

By the above lemma this holds iff $h_s(\alpha_A; \varphi) = 0$ for every $A \in \Sigma$ and every sequence s in $\mathbb{N}$. But this is equivalent to $h_s(X; \varphi) = 0$ for every sequence s, because for an arbitrary $\alpha \in \tilde{\Sigma}_d$ we have $\alpha \leq \bigvee_{A \in \alpha} \alpha_A$ and $h_s(\alpha; \varphi) \leq \sum_{A \in \alpha} h_s(\alpha_A; \varphi) = 0$. $\blacksquare$

Reference: Kushnirenko 1967 .

# XIII. Uniform Entropy and Comparison of Entropies

In our final lecture we consider a TDS $(X;\varphi)$, the Borel algebra $\mathcal{B}$ and the various $\varphi$-invariant probability measures on X. While we have seen in (XII.11) that the topological entropy $h_t(X;\varphi)$ is an upper bound of all measure-theoretical entropies $h_\mu(X;\varphi)$, we now prove that it is in fact the least upper bound.
To that purpose we need new methods and, in particular, a new entropy (due to Dinaburg and Bowen) based on the uniform structure of the compact space X. For greater simplicity we assume in this lecture that X is a metric space with metric d, and for every $n \in \mathbb{N}$ we introduce a new, but equivalent metric

$$d_n(x,y) := \max\{d(\varphi^i x, \varphi^i y) : 0 \leq i \leq n-1\}$$

and denote its open $\varepsilon$-balls with center $x \in X$ by

$$U_{n,\varepsilon}(x) := \{y \in X : d_n(x,y) < \varepsilon\} .$$

Finally, it is convenient to call a subset R of X an <u>$(n,\varepsilon)$-net</u> if $X = \bigcup \{U_{n,\varepsilon}(x) : x \in R\}$. With this notation the new entropy of $(X;\varphi)$ will be defined in three steps.

XIII. 1 <u>Definition:</u> Let $(X;\varphi)$ be a TDS. For $n \in \mathbb{N}$ and $\varepsilon > 0$ we consider

$$r_n(\varepsilon) := \min \{|R| : R \text{ is an } (n,\varepsilon)\text{-net in } X\} \text{ and}$$

$$r(\varepsilon) := \overline{\lim}\, \frac{1}{n} \log r_n(\varepsilon).$$

Then the <u>uniform</u> <u>entropy</u> of $(X;\varphi)$ is defined by

$$h_u(X;\varphi) := \lim_{\varepsilon \searrow 0} r(\varepsilon) .$$

To justify the above definition we have to observe that $r(\varepsilon)$ is increasing as $\varepsilon \searrow 0$, and therefore the limit exists, though it may be $+\infty$. Even though there is much we would say about this new entropy (see Bowen [1971], [1973], Dinaburg [1970], Denker-Grillenburger-Sigmund [1976], Walters [1975], Ch.6) we shall use it only to prove the following inequality between $h_t$ and $h_\mu$.
As in (IV.S.4) we denote by $\mathcal{P}_\varphi$ the set of all $\varphi$-invariant probability measures in M(X).

XIII. 2 <u>Theorem:</u> $h_t(X;\varphi) \leq h_u(X;\varphi) \leq \sup\{h_\mu(X;\varphi) : \mu \in P_\varphi\}$.

Proof:

a) For the proof of the first inequality we choose an open cover $\alpha \in \tilde{\mathcal{O}}$ and $n \in \mathbb{N}$. By the compactness of X we can find $\varepsilon > 0$ such that for every $x \in X$ there exists $A \in \alpha$ with $U_{1,\varepsilon}(x) \subset A$. Now choose an $(n,\varepsilon)$-net R for n and $\varepsilon$ as above. By definition, we have $y \in U_{n,\varepsilon}(x)$ if and only if $y \in \bigcap_{i=0}^{n-1} \varphi^{-i} U_{1,\varepsilon}(\varphi^i x)$. But by the choice of $\varepsilon$ there exists $A_{k_i} \in \alpha$ such that $\bigcap_{i=0}^{n-1} \varphi^{-i} U_{1,\varepsilon}(\varphi^i x) \subset \bigcap_{i=0}^{n-1} \varphi^{-i} A_{k_i}$,
which is an element of $\alpha^n$. Therefore the open cover $\{U_{n,\varepsilon}(x) : x \in R\}$ is finer than $\alpha^n$.

Now we choose the $(n,\varepsilon)$-net R in such a way that $|R| = r_n(\varepsilon)$.
Denoting the minimal cardinality of a subcover of $\beta \in \tilde{\mathcal{O}}$ by $N(\beta)$, we obtain

$$N(\alpha^n) \leq N(\{U_{n,\varepsilon}(x) : x \in R\}) \leq |R| = r_n(\varepsilon)$$

and

$$h_t(\alpha;\varphi) = \lim \frac{1}{n} \log N(\alpha^n) \leq \overline{\lim} \frac{1}{n} \log r_n(\varepsilon) \leq h_u(X;\varphi)$$

for every $\alpha \in \tilde{\mathcal{O}}$.

b) The proof of the second inequality is much more difficult. We present a proof due to Misiurewicz 1976 and show the following:
For every $\varepsilon > 0$ there exists a $\varphi$-invariant probability measure $\mu$ on X and a disjoint cover $\alpha \in \tilde{\mathcal{B}}_d$ such that $h_\mu(\alpha;\varphi) \geq r(\varepsilon)$.
We fix $\varepsilon > 0$ and proceed in several steps.

step 1. <u>The construction of $\mu$:</u>
For every $n \in \mathbb{N}$ choose a subset $S_n \subset X$ maximal with respect to the following property:

$$x, y \in S_n \text{ and } x \neq y \text{ emplies } d_n(x,y) \geq \varepsilon .$$

By the compactness of X, $S_n$ is finite and an $(n,\varepsilon)$-net, hence $|S_n| \geq r_n(\varepsilon)$.
Define the following probability measures on X:

$$p_n := \frac{1}{|S_n|} \sum_{x \in S_n} \delta_x \quad \text{and} \quad \mu_n := \frac{1}{n} \sum_{i=0}^{n-1} T'^{\,i}_\varphi p_n$$

where $\delta_x$ denotes the Dirac measure at $x \in X$, $T_\varphi$ the operator on $C(X)$ induced by $\varphi$ and $T'_\varphi$ its adjoint on $C(X)'$. Finally, we choose a subsequence $(n_k)$ of $\mathbb{N}$ for which $\lim_{k\to\infty} \frac{1}{n_k} \log r_{n_k}(\varepsilon) = r(\varepsilon)$ and a $\sigma(C(X)', C(X))$ - cluster point $\mu$ of $\{\mu_{n_k} : k \in \mathbb{N}\}$.

step 2. <u>The $\varphi$-invariance of $\mu$:</u>
This follows from standard arguments (use IV.3.0) and the weak*-continuity of $T'_\varphi$) and may be left to the reader.

step 3. The construction of $\alpha$:

Take an open cover of X by $\varepsilon/4$-balls. For every $x \in X$ there exists an open neighborhood $U_x$, contained in such an $\varepsilon/4$-ball and such that the $\mu$-measure of the boundary of $U_x$ is zero.

This is possible since every ball $U_{\varepsilon'}(x)$ is the disjoint uncountable union of the boundaries of $U_r(x)$, $0 \le r < \varepsilon'$, where only countable many of these can have non-zero measure. From the open cover $\{U_x : x \in X\}$ choose a finite subcover, form a disjointification and obtain $\alpha \in \tilde{\mathcal{B}}_d$ such that for every $A \in \alpha$ the dismeter $d(A)$ is less than $\varepsilon/2$ and $\mu(\text{boundary } A) = 0$.

step 4. The continuity of $h_{\mu_{n_k}}$ in $\alpha$:

For every $A \in \alpha$ there exist a compact set K and an open set U such that

$$K \subset \mathring{A} \subset A \subset \bar{A} \subset U,$$

where $\mathring{A}$ is the interior and $\bar{A}$ the closure of A. By Urysohn's lemma (A.4) we find $0 \le f,\ g \in C(X)$ such that

$$1\!\!1_K \le f \le 1\!\!1_{\mathring{A}} \le 1\!\!1_A \le 1\!\!1_{\bar{A}} \le g \le 1\!\!1_U$$

Since $\mu(\mathring{A}) = \mu(A) = \mu(\bar{A})$ by step 3, and since $\mu$ is a regular Borel measure, we may choose f and g in such a way that $\langle f - g, \mu \rangle$ becomes arbitrarily small. Therefore, the fact that $\mu$ is a $\sigma(C(X)', C(X))$-cluster point of $\{\mu_{n_k}\}$ implies that $\mu(A)$ is a cluster point of $\{\mu_{n_k}(A) : k \in \mathbb{N}\}$ for every $A \in \alpha$. Recalling the definition of the entropy for a disjoint cover (see XI.8), we see that it depends continuously on the measure. Therefore, we obtain that $h_\mu(\alpha)$ is a cluster point of $\{h_{\mu_{n_k}}(\alpha) : k \in \mathbb{N}\}$.

step 5. Auxiliary estimates:

Using the subadditivity of $h_{p_n}$ and (XI.9.i) we obtain for $n, k \in \mathbb{N}$ the estimates

$$h_{p_n}(\alpha^k) \le \sum_{i=0}^{k-1} h_{p_n}(\tilde{\varphi}^i \alpha) \le k \log |\alpha| .$$

On the other hand, it follows from the construction of $\alpha$ that $|B \cap S_n| \le 1$ for every $B \in \alpha^n$. Therefore

$$\begin{aligned} h_{p_n}(\alpha^n) &= -\sum_{B \in \alpha^n} p_n(B) \log p_n(B) \\ &= -\sum_{B \cap S_n \neq \emptyset} \frac{1}{|S_n|} \log \frac{1}{|S_n|} \\ &= \log |S_n| \\ &\ge \log r_n(\varepsilon), \text{ by step 1.} \end{aligned}$$

step 6. The estimate for $h_\mu(\alpha;\varphi)$:
We want to show that $h_\mu(\alpha^m) \geq m\, r(\varepsilon)$. By the continuity proved in step 4 it suffices to show an analogous statement for $h_{\mu_n}$ and large $n \in \mathbb{N}$.
To that end we take $0 < k \leq m < n$ and consider

$$l(k) := \left[\frac{n-k}{m}\right],$$

i.e. $l(k)$ is the maximal number of consecutive blocks of length m starting at k and ending before n.
Then, we calculate:

$$\begin{aligned}
h_{\mu_n}(\alpha^m) &\geq \frac{1}{n}\sum_{i=0}^{n-1} h_{T_\varphi'^i p_n}(\alpha^m), \quad \text{by (XI.9.ii)},\\
&= \frac{1}{n}\sum_{i=0}^{n-1} h_{p_n}(\tilde{\varphi}^i \alpha^m)\\
&= \frac{1}{n}\sum_{k=0}^{m-1}\sum_{j=0}^{l(k)} h_{p_n}(\tilde{\varphi}^{jm+k}\alpha^m)\\
&\geq \frac{1}{n}\sum_{k=0}^{m-1} h_{p_n}(\alpha_k^n), \quad \text{by the subadditivity of h,}\\
&\geq \frac{1}{n}\sum_{k=0}^{m-1} (h_{p_n}(\alpha^n) - h_{p_n}(\alpha^k)), \quad \text{again by the subadditivity of h,}\\
&\geq \frac{1}{n}\sum_{k=0}^{m-1} (\log |S_n| - k \log|\alpha|), \quad \text{by step 5,}\\
&\geq \frac{m}{n}\log r_n(\varepsilon) - \frac{m(m-1)}{2n}\log|\alpha| .
\end{aligned}$$

Letting $n \to \infty$, we obtain the desired inequality. ∎

After these efforts, now we are able to draw some beautiful and interesting conclusions. Since a careful inspection of the above proof shows that it also works for compact not necessarily metric spaces (see XIII.D.1), we formulate these conclusions for arbitrary TDSs.

XIII.3 Theorem (Dinaburg-Goodman-Goodwyn, 1969/71):
For a TDS $(X;\varphi)$ we have

$$h_t(X;\varphi) = h_u(X;\varphi) = \sup\{h_\mu(X;\varphi) : \mu \in P_\varphi\}.$$

Proof:
The above equality is a consequence of the inequalities in (XII.11) and (XIII.2). ∎

XIII. 4 Corollary: In the TDS $(X;\varphi)$ is minimal and $T_\varphi$ is mean ergodic on $C(X)$, then

$$h_t(X;\varphi) = h_\mu(X;\varphi).$$

for the only $\varphi$-invariant probability measure $\mu$ on X.

Proof:
Here, we apply (IV.8) and (XIII.3). ■

XIII. 5 Corollary: If $\varphi$ is an isometry on the compact metric space X, then $h_t(X;\varphi) = h_u(X;\varphi) = h_\mu(X;\varphi) = 0$ for every $\varphi$-invariant probability measure $\mu$ on X.

Proof:
If $\varphi$ is an isometry it follows that $r_n(\varepsilon) = r_m(\varepsilon)$ for $n, m \in \mathbb{N}$, hence $h_u(X;\varphi) = 0$. Then apply (XIII.3). ■

XIII. 6 Corollary: Let G be a compact metrizable group, $\varphi: G \to G$ an automorphism and $\varrho := L_g$, $g \in G$, a left rotation on G. Then $h_t(G;\varphi\circ\varrho) = h_t(G;\varphi)$.

Proof:
Take a left-invariant metric d inducing the topology on G (see Hewitt-Ross 1979 , II.8.3). Then the metrics

$$d_n(x,y) = \max\{d(\varphi^i x, \varphi^i y) : 0 \le i \le n-1\} \text{ and}$$
$$d_n'(x,y) = \max\{d((\varphi\circ\varrho)^i x, (\varphi\circ\varrho)^i y) : 0 \le i \le n-1 \text{ coincide,}$$

since

$$\begin{aligned} d(\varphi\circ\varrho(x), \varphi\circ\varrho(y)) &= d(\varphi(gx), \varphi(gy)) \\ &= d(\varphi(g)\varphi(x), \varphi(g)\varphi(y)) \\ &= d(\varphi(x), \varphi(y)) . \end{aligned}$$

Hence the assertion follows from the definition of $h_u$ and from (XIII.3). ■

XIII. 7 Corollary: Irreducible dynamical systems, i.e. ergodic MDSs or minimal TDSs, with discrete spectrum have zero entropy.

Proof:
By (VIII.2) or (VIII.4) the dynamical system is isomorphic to a rotation on a compact group G. This group is metrizable if C(G), resp. $L^1(G,\mathcal{B},m)$, is separable (see Schaefer [1974], II.7.5). In this case the assertion follows from (XIII.6). The general result may be obtained from (XII.D.7) and (XIII.3). ■

XIII. D Discussion

XIII. D.1 Uniform entropy for general uniform spaces:

In (XIII.1) we defined the uniform entropy only for metric spaces. Here we briefly indicate how this definition carries over to the general case (the same is true for the proof of $h_t(X;\varphi) \leq h_u(X;\varphi)$). A uniform structure on a set X is defined as a filter $\mathcal{U}$ on $X \times X$ having special properties (see Schaefer [1966], prerequisites B.6). Now consider a TDS $(X;\varphi)$ and the (unique) uniform structure $\mathcal{U}$ that generates the topology of X. For a vicinity $U \in \mathcal{U}$ we can define

$$U_n := \bigcap_{i=0}^{n-1} (\varphi \otimes \varphi)^{-i}(U) = \{(x,y) \in X \times X : (\varphi^i(x), \varphi^i(y)) \in U \text{ for } i = 0, \dots, n-1\}$$

and $$U_n(x) := \{y \in X : (x,y) \in U_n\} \quad \text{for } x \in X.$$

A subset R of X is called a $U_n$-net, if

$$X = \bigcup \{U_n(x) : x \in R\} .$$

Then the uniform entropy is defined as follows:

For $U \in \mathcal{U}$ and $n \in \mathbb{N}$ let

$$r_n(U) := \min \{|R| : R \text{ is a } U_n\text{-net}\} ,$$

$$r(U) := \overline{\lim} \frac{1}{n} \log r_n(U) ,$$

and finally

$$h_u(X;\varphi) := \lim_{U \in \mathcal{U}} r(U) = \sup_{U \in \mathcal{U}} r(U) .$$

XIII. D.2 Entropy and dimension of a metric space:

At first sight and for a beginner, the definition of the uniform entropy $h_u(X;\varphi)$ of a TDS $(X;\varphi)$ is complicated. But there is a forerunner of this concept initiated by Pontrjagin-Schnirelman [1932] in their endeavour to measure the dimension of a metric space. In the terminology of (XIII:1), for $\varepsilon > 0$ the number

$$\log r_n(\varepsilon)$$

is the $\varepsilon$-entropy of the metric space $(X, d_n)$, and the size of X with respect to the metric $d_n$ is measured by the increase of this $\varepsilon$-entropy as $\varepsilon$ tends to 0.

As an example we mention that a subset of a Banach space is k-dimensional if and only if its $\varepsilon$-entropy increases as $(\frac{1}{\varepsilon})^k$ (see Pietsch [1972], ch.9). In our attempt to describe the dynamics of a homeomorphism $\varphi$ on a metric space X, we apply the above idea but we reverse the order of the limiting procedures. First we measure the size of X for the metric $d_n$ and $\varepsilon > 0$. As n tends to infinity these new metrics reflect the dynamics of $\varphi$. Only thereafter do we take the limit as $\varepsilon \to 0$. Briefly, the uniform entropy of $(X;\varphi)$ is

$$h_\mu(X;\varphi) = \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} [\varepsilon\text{-entropy of}(X,d_n)].$$

For more information we refer to Kolmogorov-Tichomirov [1961] and Pietsch [1971].

XIII. D.3 Entropy of homeomorphisms of the unit circle:

It is obvious from (XIII.3) that as soon as we have calculated one type of entropy this information can be carried over to the other types of entropy. We present the following example.

Proposition: Let $\varphi$ be a homeomorphism of the unit circle $\Gamma$ and $\mu$ a $\varphi$-invariant probability measure. Then $h_\mu(\Gamma;\varphi) = 0$.

Proof:

It suffices to prove that $h_t(\Gamma;\varphi) = 0$. To do this we use the fact that $\varphi$ maps "segments" (a,b) of $\Gamma$ onto segments. For any open cover $\alpha$ consisting of segments, define $B(\alpha) := \{a\in\Gamma : (a,b)\in\alpha \text{ for some } b\in\Gamma\}$. Clearly $N(\alpha) \le |B(\alpha)|$. If $\beta\in\tilde{\mathcal{O}}$ is arbitrary, then we can find a cover $\alpha \succ \beta$ consisting of segments (use Lebesgue's lemma (A.5)).

Now $N(\alpha^n) \le |B(\alpha^n)| \le \sum_{i=0}^{n-1} |B(\varphi^{-i}\alpha)| = n\cdot|B(\alpha)|$.

This implies that $0 \le h_t(\alpha;\varphi) \le \lim \frac{1}{n}\log(n\cdot|B(\alpha)|) = 0$. ∎

XIII. D.4 Subshifts of finite type are intrinsically ergodic:

Let $C = (c_{ij})$ be a $k\times k$-matrix with $c_{ij}\in\{0, 1\}$ and assume that no row and no column of C consists only of zeros. Consider

$$Y := \{x = (x_n)_{n\in\mathbb{Z}} : x_n\in\{0,1\ldots,k-1\} \text{ and } x_n = i, x_{n+1} = j \text{ implies } c_{ij} = 1\}$$

which is a closed shift-invariant subst of $X := \{0,1,\ldots,k-1\}^{\mathbb{Z}}$.

Therefore, $(Y;\tau)$ for the left shift $\tau$ is a TDS, called a subshift of finite type (compare XII.D.5).

Proposition 1: The topological entropy of a subshift of finite type defined by the matrix C is $\log\lambda$, where $\lambda$ is the largest (real) eigenvalue of C.

Proof:

We apply (XII.D.5), Proposition, and consider the open cover $\alpha = \{A_0,\ldots, A_{k-1}\}$ where $A_i := \{(x_n)\in Y : x_0 = i\}$.

Since $Y \cap \bigcap_{\ell=0}^{m-1} \tau^{-1}A_{i_\ell} \neq \emptyset$ iff $\prod_{\ell=0}^{m-2} c_{i_\ell i_{\ell+1}} = 1$ we obtain

$$|\alpha^m| = \sum_{i_0=0}^{k-1} \cdots \sum_{i_{m-1}=0}^{k-1} \prod_{\ell=0}^{m-2} c_{i_\ell i_{\ell+1}} = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} (C^{m-1})_{ij} = \| C^{m-1} \|_1 ,$$

i.e. $|\alpha^m|$ is equal to the $L^1$-norm of the (m-1)-st power of $C \in \mathbb{R}^{k^2}$. Therefore

$$h_t(\alpha ; \tau) = \lim_m \frac{1}{m} \log |\alpha^m| = \lim \frac{1}{m} \log \| C^{m-1} \|_1$$
$$= \lim_m \log \| C^m \|_1^{m^{-1}} = \log r(C).$$

But by the Perron-Frobenius theorem (see Schaefer [1974], I.2.3), the spectral radius $r(C)$ of $C$ is an eigenvalue of $C$. ∎

Subshifts of finite type are a second example of how to associate a dynamical system to a positive matrix. The first has been the Markov shift corresponding to a transition matrix (see II.6), whose measure-theoretical entropy has been calculated in (XII.9).
An irreducible (0,1)-matrix $C = (c_{ij})$ (see IV.D.6) defines a subshift of finite type $(Y; \tau)$. We may also construct a Markov shift from $C$ if we observe that $C$ and its adcoint $C'$ both have strictly positive eigenvectors $x$ and $y$ corresponding to the spectral radius $\lambda$ of $C$ (see Schafer [1974], I.6.2). The new matrix

$$P = (p_{ij}) \text{ defined by } p_{ij} := \frac{x_j}{\lambda x_i} c_{ij}$$

is stochastic, again irreducible, and has a unique strictly positive probability vector $\pi$ with $\pi_i := x_i y_i$ such that $P'\pi = \pi$. (Let $X, \Sigma, \mu; \tau$) be the Markov shift defined by $P$ and $\pi$.
Next we restrict $\mu$ to $Y$ and to the $\sigma$-algebra $\mathcal{T}$ induced by $\Sigma$. Since $\mu(Y) = 1$, we obtain an MDS $(Y, \mathcal{T}, \mu; \tau)$ isomorphic to $(X, \Sigma, \mu; \tau)$. By (XII.9), its measure-theoretical entropy is $\log \lambda$ and, by Proposition 1, equal to the topological entropy of the TDS $(Y; \tau)$.
In view of Theorem (XIII.3) we have thus <u>constructed</u> a $\tau$-invariant measure $\mu$ on $Y$ on which the maximum in

$$h_t(Y; \tau) = \sup \{ h_\nu(Y; \tau) : \nu \in \mathcal{P}_\tau \}$$

is actually attained ($\mathcal{P}_\tau$ denotes the set of all $\tau$-invariant probability measures on $Y$). Although $\nu \longmapsto h_\nu(Y; \tau)$ is an affine function (see XII.12) on the compact convex set $\mathcal{P}_\tau$ (see also IV.S.4), the existence of such a measure is not at all clear in general, since $\nu \longmapsto h_\nu(Y; \tau)$ does not necessarily have the "right" continuity properties (compare Denker-Grillenberger-Sigmund [1976]).
But for subshifts of finite type we shall show even more.

<u>Definition:</u> A TDS $(X; \varphi)$ is called <u>intrinsically</u> <u>ergodic</u> if the function $\mu \longmapsto h_\mu(X; \varphi)$ on $\mathcal{P}_\varphi$ attains its maximum at exactly one measure.

**Proposition 2:** Any subshift of finite type is intrinsically ergodic, and the measure on which its entropy attains its maximum is the measure constructed above.

Proof:

We will use the notation introduced above. In particular, $\mu$ will be the "maximal" measure on Y constructed above.

Assume that $h_\nu(Y;\tau) > h_\mu(Y;\tau)$ for some $\tau$-invariant probability measure $\nu \neq \mu$ and write $\nu = c\nu_1 + (1-c)\nu_2$ for $0 \le c \le 1$, $\nu_2$ absolutely continuous with respect to $\mu$ and $\nu_1$ singular with respect to $\mu$. Since $\nu_2$ is $\tau$-invariant and C is irreducible we may assume that $\nu_2 = \mu$. Therefore, since $\nu \neq \mu$ i.e. $c \neq 0$, we obtain $h_{\nu_1}(Y;\tau) > h_\mu(Y;\tau)$.

Next we take a set $M \in \Sigma$ with $\mu(M) = 0$ and $\nu_1(M) = 1$. For the open cover $\alpha = \{A_0, \ldots, A_{k-1}\}$ considered above we find $M_n \in \sigma(\alpha^n_{-n})$ such that $\lim_{n\to\infty} (\mu + \nu_1)(M_n \Delta M) = 0$ and therefore $\mu(M_n) \to 0$ and $\nu_1(M_n) \to 1$.

For $A \in \alpha^n_{-n}$ we have

$$\lambda^{-(2n+1)}r \le \mu(A) \le \lambda^{-(2n+1)}s$$

for $r := \min\{\lambda x_i y_i,\ 0 \le i \le k-1\}$ and $s := \max\{\lambda x_i y_i : 0 \le i \le k-1\}$.

Then

$$\begin{aligned}
\log \lambda^{2n+1} &= (2n+1)h_\mu(Y;\tau) \le (2n+1)h\nu_1(Y;\tau) \le h\nu_1(\alpha^n_{-n}) \\
&\le -\sum_{A\subset M_n} \nu_1(A)\log\nu_1(A) - \sum_{A\subset M_n^c}\nu_1(A)\log\nu_1(A) \\
&= -\nu_1(M_n)\sum_{A\subset M_n}\frac{\nu_1(A)}{\nu_1(M_n)}\log\frac{\nu_1(A)}{\nu_1(M_n)} - \nu_1(M_n^c)\sum_{A\subset M_n^c}\frac{\nu_1(A)}{\nu_1(M_n^c)}\log\frac{\nu_1(A)}{\nu_1(M_n^c)} \\
&\quad - \nu_1(M_n)\log\nu_1(M_n) - \nu_1(M_n^c)\log\nu_1(M_n^c) \\
&\le \nu_1(M_n)\log|\{A\in\alpha^n_{-n} : A\subset M_n\}| \\
&\quad + \nu_1(M_n^c)\log|\{A\in\alpha^n_{-n} : A\subset M_n^c\}| \\
&\quad - \nu_1(M_n)\log\nu_1(M_n) - \nu_1(M_n^c)\log\nu_1(M_n^c) \\
&\le \nu_1(M_n)\log\left(\frac{\lambda^{2n+1}}{r}\mu(M_n)\right) + \nu_1(M_n^c)\log\left(\frac{\lambda^{2n+1}}{r}\mu(M_n^c)\right) \\
&\quad - \nu_1(M_n)\log\nu_1(M_n) - \nu_1(M_n^c)\log\nu_1(M_n^c).
\end{aligned}$$

and therefore

$$0 \le \nu_1(M_n)\log\frac{\mu(M_n)}{r} + \nu_1(M_n^c)\log\frac{\mu(M_n^c)}{r} - \nu_1(M_n)\log\nu_1(M_n) - \nu_1(M_n^c)\log\nu_1(M_n^c).$$

On the other hand, this term converges to - as n , and we have a contradiction. ∎

This proposition incidentally proves the theorem of Dinaburg-Goodman-Goodwyn (XIII.3) for subshifts of finite type. In fact, this theorem was first known for subshifts of finite type (see Goodwyn [1969]) and its range of validity was subsequently enlarged, until it was finally shown to hold for arbitrary TDSs.

References: Adler-Weiss [1970], Denker-Grillenberger-Sigmund [1976].

XIII. D.6 Concluding example:

Spectral invariants and entropy have been investigated in detail in the previous lectures. They have been also used to solve the isomorphism problem for some subclasses of MDSs (see Lecture VIII and XII.D.1). But we have already remarked that both invariants together are far from being complete (e.g. VI.D.4). Here we present a simple but illusion-shattering example proving just that.

Take $X = \Gamma \times \Gamma$ and $\varphi_1 : (u,v) \mapsto (au, u^p v)$, $\varphi_2 : (u,v) \mapsto (au, u^q v)$ where $a \in \Gamma$ is not a root of unity, and $p, q \in \mathbb{Z}$. The normalized Haar measure m on X is invariant under $\varphi_1$ and $\varphi_2$, and by (XIII.D.5) we have

$$h_t(\varphi_1) = h_m(\varphi_1) = 0 = h_t(\varphi_2) = h_m(\varphi_2) .$$

Next we show that $(X, \mathcal{B}, m; \varphi_1)$ and $(X, \mathcal{B}, m; \varphi_2)$ are spectrally isomorphic:

Consider the basis

$$\{g_{m,n} : n, m \in \mathbb{Z}\}$$

of $L^2(X, \mathcal{B}, m)$, where $g_{m,n}(u,v) := u^m v^n$ .

Then $T_{\varphi_1} g_{m,n}(u,v) = a^m u^m u^{pn} v^n$, i.e. $T_{\varphi_1} g_{m,n} = a^m g_{m,n}$ for $n = 0$ and orthogonal to $g_{m,n}$ otherwise. Thus $L^2(X, \mathcal{B}, m)$ has a basis consisting of eigenvectors of $T_{\varphi_1}$ corresponding to the simple eigenvalues $a^m$ and of a countable union of "orbits" $\{T^j_{\varphi_1} h_i : j \in \mathbb{Z}\}$, $h_i \in L^2(m)$ (compare VI.D.5). The same reasoning applies to $T_{\varphi_2}$, and the correspondence between the two bases yields the desired isomorphism.

Finally we show that $(X, \mathcal{B}, m; \varphi_1)$ and $(X, \mathcal{B}, m; \varphi_2)$, while being spectrally isomorphic and having the same entropy, are not isomorphic in general.

Assume that $\Theta : (u,v) \mapsto (f(u,v), g(u,v))$ is a point isomorphism on $\Gamma \times \Gamma$ satisfying

$$\Theta \circ \varphi_1 = \varphi_2 \circ \Theta \quad ,$$

or explicitly

(i) $\quad f \circ \varphi_1 = a f$ and

(ii) $\quad g \circ \varphi_1 = f^q g$ .

As a consequence of (i) we find that

$$f(u,v) = c \cdot g_{1,0}(u,v) = c \cdot u \quad \text{for some } c \in \Gamma \text{ (use the above basis)},$$

and therefore (ii) becomes

$$g \circ \varphi_1(u,v) = c^q u^q g(u,v) .$$

Using the basis $\{g_{m,n} : m, n \in \mathbb{Z}\}$ again, we see that g has to be of the form

$$g = d \cdot g_{n,qp^{-1}} \quad ,$$
$$c = a^{nq^{-1}} \quad \text{and } d \in \Gamma.$$

Therefore q is a multiple of p and

$$\Theta(u,v) = (a^{nq^{-1}} u, \; du^n c^{qp^{-1}})$$

for some $n \in \mathbb{Z}$, $d \in \Gamma$ , which is invertible only if $qp^{-1} = \pm 1$ .
Otherwise the two MDSs are not isomorphic.

References: Anzai [1951] , Walters [1975].

## 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 $(X,\mathcal{O})$, $\mathcal{O}$ 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 $\mathcal{O}$ and if $\mathcal{O}'$ is another topology on X, coarser than $\mathcal{O}$ but still Hausdorff, then $\mathcal{O} = \mathcal{O}'$.

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 $\{x = (x_\alpha)_{\alpha \in A} : x_{\alpha_i} \in O_{\alpha_i} \text{ for } i = 1,..,n\}$ for $\alpha_1,\ldots,\alpha_n \in A$, $n \in \mathbb{N}$ and $O_{\alpha_i}$ open in $X_{\alpha_i}$. Then Tychonov's theorem asserts that for this topology, X is compact if and only if each $X_\alpha$, $\alpha \in A$, 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 $f(A) \subset \{0\}$ and

$f(B) \subset \{1\}$.

A.5 Lebesgue's covering lemma:
If $(X,d)$ is a compact metric space and $\alpha$ is a finite open cover of $X$, then there exists a $\delta > 0$ such that every set $A \subseteq X$ with diameter $d(A) < \delta$ is contained in some element of $\alpha$.

A.6 Category:
A subset $A$ of a topological space $X$ is called nowhere dense if the closure of $A$, denoted by $\bar{A}$, has empty interior: $\mathring{\bar{A}} = \emptyset$.
$A$ 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 [1978]. 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 $(X, \Sigma, \mu)$ is a measure space if $X$ is a set, $\Sigma$ a $\sigma$-algebra of subsets of $X$ and $\mu$ a measure on $\Sigma$, i.e.

$$\mu : \Sigma \longrightarrow \mathbb{R}_+ \cup \{\infty\}$$

is $\sigma$-additive and satisfies $\mu(\emptyset) = 0$.
If $\mu(X) < \infty$ (resp. $\mu(X) = 1$), $(X, \Sigma, \mu)$ is called a finite measure space (resp. a probability space); it is called $\sigma$-finite, if $X = \bigcup_{n \in \mathbb{N}} A_n$ with $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$.
A set $N \in \Sigma$ is a $\mu$-null set if $\mu(N) = 0$.
Properties, implications, conclusions etc. are valid "$\mu$-almost-everywhere" or for "almost all $x \in X$" if they are valid for all $x \in X \setminus N$

where N is some $\mu$-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 $(X, \Sigma, \mu)$ be a $\sigma$-finite measure space and $\nu$ another measure on $\Sigma$. $\nu$ is called absolutely continuous with respect to $\mu$ if every $\mu$-null set is a $\nu$-null set. $\nu$ is equivalent to $\mu$ iff $\nu$ is absolutely continuous with respect to $\mu$ and conversely. The measures which are absolutely continuous with respect to $\mu$ can be characterized by the Radon - Nikodym theorem (see Halmos [1950], §31).

A.9 The measure algebra:
In a measure space $(X, \Sigma, \mu)$ the $\mu$-null sets form a $\sigma$-ideal $\mathfrak{N}$. The Boolean algebra

$$\check{\Sigma} := \Sigma / \mathfrak{N}$$

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,

$$A \Delta B := (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$$

denotes the symmetric difference of A and B, and

$$d(A,B) := \mu(A \Delta B)$$

defines a quasi-metric on $\Sigma$ vanishing on the elements of $\mathfrak{N}$ (if $\mu(X) < \infty$). Therefore we obtain a metric on $\check{\Sigma}$ still denoted by d.

A.10 Proposition:
The measure algebra $(\check{\Sigma}, d)$ of a finite measure space $(X, \Sigma, \mu)$ is a complete metric space.

Proof:
It suffices to show that $(\Sigma, d)$ 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 all $k, l \geq n_i$. 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

$$d(B_m, A_{n_m}) \leq \sum_{j=m}^{\infty} \mu(A_{n_{j+1}} \setminus A_{n_j}) \leq \sum_{j=m}^{\infty} 2^{-j} \leq 2 \cdot 2^{-m},$$

and

$$\begin{aligned} d(A,B_m) &\leq \sum_{j=m}^{\infty} \mu(B_j \setminus B_{j+1}) \\ &\leq \sum_{j=m}^{\infty} \left( d(B_j,A_{n_j}) + d(A_{n_j},A_{n_{j+1}}) + d(A_{n_{j+1}},B_{j+1}) \right) \\ &\leq \sum_{j=m}^{\infty} \left( 2 \cdot 2^{-j} + 2^{-j} + 2 \cdot 2^{-(j+1)} \right) \\ &\leq 8 \cdot 2^{-m} \quad . \end{aligned}$$

Therefore

$$d(A,A_k) \leq d(A,B_m) + d(B_m,A_{n_m}) + d(A_{n_m},A_k) \leq 11 \cdot 2^{-m}$$

for $k \geq n_m$ .

A.11 For a subset $\check{W}$ of $\check{\Sigma}$ we denote by $a(\check{W})$ the <u>Boolean algebra generated by</u> $\check{W}$, by $\sigma(\check{W})$ the <u>Boolean $\sigma$-algebra generated by</u> $\check{W}$.
$\check{\Sigma}$ is called countably generated, if there exists a countable subset $\check{W} \subseteq \check{\Sigma}$ such that $\sigma(\check{W}) = \check{\Sigma}$ .

The metric d relates $a(\check{W})$ and $\sigma(\check{W})$. More precisely, using an argument as in (A.10) one can prove that in a finite measure space

$$\sigma(\check{W}) = \overline{a(\check{W})}^{d} \qquad \text{for every } \check{W} \subseteq \check{\Sigma} .$$

A.12 <u>The Borel algebra</u>:
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 <u>always</u> take the $\sigma$-algebra $\mathcal{B}$ generated by the open sets, called the <u>Borel algebra</u> 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 <u>regular Borel measures</u>: here, $\mu$ 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 $\mu(A \setminus K) < \varepsilon$ and $\mu(U \setminus A) < \varepsilon$ .

A.13 <u>Example</u>:
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

$$\mathcal{D} := \{ [k \cdot 2^{-i}, (k+1) \cdot 2^{-i}] \ : \ i \in \mathbb{N}; \ k = 0,\dots, 2^i-1 \} .$$

$\mathcal{D}$ is called a <u>separating base</u> because it generates $\mathcal{B}$ and for any $x, y \in X$, $x \neq y$, there is $D \in \mathcal{D}$ 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 $(X, \Sigma, \mu)$ and $(Y, T, \nu)$. 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 $A \in T$ (abbreviated: $\mu \circ \varphi^{-1} = \nu$).

For real-valued measurable functions f and g on $(X, \Sigma, \mu)$, where $\mathbb{R}$ is endowed with the Borel algebra $\mathcal{B}$, we use the following notation:

$$[f \in B] := f^{-1}(B) \qquad \text{for } B \in \mathcal{B},$$
$$[f = g] := \{x \in X: f(x) = g(x)\},$$
$$[f \leq g] := \{x \in X: f(x) \leq g(x)\}.$$

Finally,

$$\mathbb{1}_A : x \mapsto \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$ denotes the characteristic

function of $A \subseteq X$. If $A = X$, we often write $\mathbb{1}$ instead of $\mathbb{1}_X$.

A.15 Continuous vs. measurable functions:

Let X be compact, $\mathcal{B}$ the Borel algebra on X and $\mu$ 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 $\mu(A^c) < \varepsilon$ 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 $\mu(V_j \setminus f^{-1}(U_j)) < \frac{\varepsilon}{2} 2^{-j}$.

If we take $B := \bigcup_{j=1}^{\infty} (V_j \setminus f^{-1}(U_j))$, we obtain $\mu(B) < \frac{\varepsilon}{2}$, and we show that $g := f|_{B^c}$ is continuous. To this end observe that

$$V_j \cap B^c = V_j \cap (V_j \setminus f^{-1}(U_j))^c \cap B^c = V_j \cap (V_j^c \cup f^{-1}(U_j)) \cap B^c$$
$$= V_j \cap f^{-1}(U_j) \cap B^c = f^{-1}(U_j) \cap B^c = g^{-1}(U_j).$$

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 $\mu(B^c \setminus A) < \frac{\varepsilon}{2}$, and conclude that f is continuous on A and that $\mu(X \setminus A) = \mu(B) + \mu(B^c \setminus A) < \varepsilon$ . ∎

A.16 <u>Convergence of integrable functions</u>:
Let $(X, \Sigma, \mu)$ be a finite measure space and $1 \leq p < \infty$ . A measurable (real) function f on X is called <u>p-integrable</u>, if $\int |f|^p \, d\mu < \infty$ (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 <u>$\mu$ - almost everywhere</u> 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 <u>in the p - norm</u> if
$$\lim_{n \to \infty} \int |f_n - f|^p \, d\mu = 0 \qquad \text{(see B.20)}$$
3. $(f_n)_{n \in \mathbb{N}}$ converges to f <u>$\mu$ - stochastically</u> if
$$\lim_{n \to \infty} \mu[|f_n - f| \geq \varepsilon] = 0 \qquad \text{for every } \varepsilon > 0.$$

<u>Proposition</u>:
Let $(f_n)_{n \in \mathbb{N}}$ be p-integrable functions and f be measurable.

(i) If $f_n \longrightarrow f$ $\mu$-almost everywhere or in the p-norm, then $f_n \longrightarrow f$ $\mu$-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 $(f_{n_k})_{k \in \mathbb{N}}$ converging to f $\mu$-a.e. (see Bauer 1972 , 2.7.5).

(iii) If $(f_n)_{n \in \mathbb{N}}$ converges to f $\mu$-a.e. and if there is a p-integrable function g such that $|f_n(x)| \leq g(x)$ $\mu$-a.e., then $f_n \longrightarrow f$ 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 <u>Product spaces</u>:
Given a countable family $(X_\alpha, \Sigma_\alpha, \mu_\alpha)_{\alpha \in A}$ of probability spaces, we can consider the cartesian product $X = \prod_{\alpha \in A} X_\alpha$ and the so-called product $\sigma$-algebra $\Sigma = \bigotimes_{\alpha \in A} \Sigma_\alpha$ on X which is generated by the set of all <u>measurable rectangles</u>, i.e. sets of the form

$R_{\alpha_1,\ldots,\alpha_n}(A_{\alpha_1},\ldots,A_{\alpha_n}) := \{ x = (x_\alpha)_{\alpha\in A} : x_{\alpha_i} \in A_{\alpha_i}$ for $i=1,\ldots,n$

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 $\Sigma$ 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], § 38 Theorem B).

Then $(X,\Sigma,\mu)$ is called the <u>product (measure) space</u> 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).

<u>Theorem</u>:

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 $\Sigma$ the product $\sigma$-algebra on $X = \prod_{n\in\mathbb{Z}} X_n$ , by $\mathcal{F}_m$ the set of all measurable sets in X whose elements depend only on the coordinates $-m, \ldots, 0, \ldots, m$. Finally we put $\mathcal{F} := \bigcup_{m\in\mathbb{N}} \mathcal{F}_m$. If $\mu$ is a function on $\mathcal{F}$ such that it is a regular probability measure on $\mathcal{F}_m$ for each $m\in\mathbb{N}$. then $\mu$ has a unique extension to a probability measure on $\Sigma$ .

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 $\mu$ has to be extended from $\mathcal{F}$ to $\sigma(\mathcal{F}) = \Sigma$ . By the classical Caratheodory 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 $\mathcal{F}$ satisfying $\bigcap_{i\in\mathbb{N}} C_i = \emptyset$ . Assume that $\mu(C_i) \geq \varepsilon$ for all $i\in\mathbb{N}$ and some $\varepsilon > 0$. For each $C_i$ there is an $n\in\mathbb{N}$ such that $C_i \in \mathcal{F}_n$ 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 $\nu_n(A_i \setminus B_i) \leq \frac{\varepsilon}{2} \cdot 2^{-i}$.

Then $D_i := \varphi_n^{-1}(B_i)$ is compact in X and $\mu(C_i \setminus D_i) \leq \frac{\varepsilon}{2}\, 2^{-i}$.

Now the sets $G_k := \bigcap_{i=1}^{k} D_i$ form a decreasing sequence of compact subsets of $X$, and we have

$$G_k \subseteq C_k \text{ and } \mu(G_k) = \mu(C_k) - \mu(C_k \setminus G_k) = \mu(C_k) - \mu\Big(\bigcup_{i=1}^{k} (C_i \setminus D_i)\Big)$$

$$\geq \mu(C_k) - \sum_{i=1}^{k} \mu(C_i \setminus D_i) \geq \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}.$$

Hence $G_k \neq \emptyset$ and therefore $\bigcap_{i \in \mathbb{N}} C_i$, which contains $\bigcap_{i \in \mathbb{N}} D_i$, 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 $U := \{f \in E: \|f\| \leq 1\}$. We associate to E its dual E' consisting of all continuous linear functionals on E. Usually, E' will be endowed with the dual norm

$$\|f'\| := \sup\{|\langle f,f'\rangle| : \|f\| \leq 1\},$$

where $\langle\cdot,\cdot\rangle$ denotes the canonical bilinear form

$$(f,f') \longmapsto \langle f,f'\rangle := f'(f) \quad \text{on } E \times E'.$$

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 of Alaoglu-Bourbaki:
The dual unit ball $U^{\circ} := \{f' \in E' : \|f'\| \leq 1\}$ 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 $\|Tf\| \leq \|f\|$, 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 $E = PE \oplus P^{-1}(0)$.

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 $\Phi : E' \rightarrow (PE)'$ 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 $(PE)' \cong E'/P'^{-1}(0) \cong P'E'$. ∎

B.8 The space $\mathcal{L}(E)$ of all bounded linear operators on E becomes a Banach space if endowed with the operator norm

$$\|T\| := \sup \{\|Tf\| : \|f\| \leq 1\} .$$

But other topologies on $\mathcal{L}(E)$ will be used as well. We write $\mathcal{L}_s(E)$ if we endow $\mathcal{L}(E)$ with the <u>strong</u> <u>operator</u> <u>topology</u> 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 $T_\alpha f \xrightarrow{\|\cdot\|} Tf$ for all $f \in E$. Observe that the strong operator topology is the topology on $\mathcal{L}(E)$ induced from the product topology on $(E, \|\cdot\|)^E$.
The <u>weak</u> <u>operator</u> <u>topology</u> on $\mathcal{L}(E)$ - write $\mathcal{L}_w(E)$ - is the topology of simple convergence on E with respect to $\sigma(E,E')$. Therefore,

$$T_\alpha \text{ converges to } T \text{ in the weak operator topology}$$
$$\text{iff} \quad \langle T_\alpha f, f'\rangle \rightarrow \langle Tf, f'\rangle \quad \text{for all } f \in E,\ f' \in E' .$$

Again, this topology is the topology on $\mathcal{L}(E)$ inherited from the product topology on $(E, \sigma(E,E'))^E$.

B.9 <u>Bounded subsets of $\mathcal{L}(E)$</u>:
For $M \subset \mathcal{L}(E)$ 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. $\sup\{\|T\| : T \in M\}$ .
(d) M is equicontinuous for $\|\cdot\|$ .

Proof:
See Schaefer [1971],III.4.1, Corollary, and III.4.2 for (b)$\Leftrightarrow$(c)$\Leftrightarrow$(d); for (a)$\Leftrightarrow$(b) observe that the duals of $\mathcal{L}_s(E)$ and $\mathcal{L}_w(E)$ 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 $\mathcal{L}(E)$, then the closure of M as subset of the product $(E, \|\cdot\|)^E$ is still contained in $\mathcal{L}(E)$ (Schaefer [1971],III.4.3).

B.11 On bounded subsets M of $\mathcal{L}(E)$, 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 $\mathcal{L}(E)$ is that more subsets of $\mathcal{L}(E)$ become compact. Therefore, the following assertions (B.12) - (B.15) are of great importance.

B.12 Proposition:

For $M \subset \mathcal{L}(E)$, $g \in E$, we define the orbit $Mg := \{Tg: T \in M\} \subset E$, and the subspaces $G_s := \{f \in E: Mf \text{ is relatively } \|\cdot\|\text{-compact}\}$

and $G_\sigma := \{f \in E: Mf \text{ is relatively } \sigma(E,E')\text{-compact}\}$.

If $M$ is bounded, then $G_s$ and $G_\sigma$ are $\|\cdot\|$-closed in $E$.

Proof:

The assertion for $G_s$ follows by a standard diagonal procedure. The argument for $G_\sigma$ is more complicated: Let $(f_n)_{n \in \mathbb{N}}$ be a sequence in $G_\sigma$ 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 $(T_{k_{i_2}})$ of $(T_{k_{i_1}})$ such that $(T_{k_{i_2}} f_2)$ weakly converges to $g_2 \in E$, 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 $T_{k_i} f_n \xrightarrow{i \to \infty} g_n \in E$ weakly for every $n \in \mathbb{N}$. From

$$\begin{aligned}\|g_n - g_m\| &= \sup \{ \langle g_n - g_m, f' \rangle \quad : \|f'\| \leq 1\} \\ &= \sup \{ \lim_{i \to \infty} |\langle T_{k_i} f_n - T_{k_i} f_m, f' \rangle| \; : \|f'\| \leq 1\} \\ &\leq \|T_{k_i}\| \cdot \|f_n - f_m\|\end{aligned}$$

it follows that $(g_n)_{n \in \mathbb{N}}$ is a Cauchy sequence, and therefore converges to some $g \in E$. A standard $3\varepsilon$-argument shows $T_{k_i} f \xrightarrow{i \to \infty} g$ for $\sigma(E,E')$. ∎

B.13 Proposition:

For a bounded subset $M \subset \mathcal{L}(E)$ 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 $T \to Tf$ from $\mathcal{L}_s(E)$

into E.
(b)$\Leftrightarrow$(c) follows from (B.12), and (b)$\Rightarrow$(a) is a consequence of (A.3) and (B.10). ■

B.14 Proposition:
For a bounded subset $M \subset \mathcal{L}(E)$ 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 $M \subset \mathcal{L}_w(E)$ be compact and choose a total subset $A \subset E$ and a $\sigma(E', E)$-total subset $A' \subset W'$. 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 $\mathcal{L}(E)$:
In Lecture VII the multiplication

$$(S, T) \longmapsto S \circ T$$

in $\mathcal{L}(E)$ plays an important role. Therefore, we state its continuity properties: The multiplication is jointly continuous on $\mathcal{L}(E)$ for the nórm 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 $\mathcal{L}_s(E)$ (see Schaefer [1971], p.183).

B.17 Spectral theory:
Let E be a complex Banach space and $T \in \mathcal{L}(E)$. The resolvent set $\varrho(T)$ consists of all complex numbers $\lambda$ for which the resolvent

$R(\lambda, T) := (\lambda - T)^{-1}$ exists. The mapping $\lambda \mapsto R(\lambda, T)$ is holomorphic on $\varrho(T)$. The <u>spectrum</u> $\sigma(T) := \mathbb{C} \setminus \varrho(T)$ is a non-empty compact subset of $\mathbb{C}$, and two subsets of $\sigma(T)$ are of special interest: the <u>point spectrum</u>

$$P\sigma(T) := \{\lambda \in \sigma(T) : (\lambda - T) \text{ is not injective}\}$$

and the <u>approximate point spectrum</u>

$$A\sigma(T) := \{\lambda \in \sigma(T) : (\lambda - T)f_n \to 0 \text{ for some normalized sequence } (f_n)\}.$$

A complex number $\lambda$ is called an (approximate) eigenvalue if $\lambda \in P\sigma(T)$ (resp. $\lambda \in A\sigma(T)$), and $F_\lambda := \{f \in E : (\lambda - T)f = 0\}$ is the eigenspace corresponding to the eigenvalue $\lambda$ ; $\lambda$ is a simple eigenvalue if $\dim F_\lambda = 1$ .

The real number $r(T) := \sup\{|\lambda| : \lambda \in \sigma(T)\}$ is called the <u>spectral radius</u> of T, and may be computed from the formula $r(T) = \lim_{n\to\infty} (\|T^n\|)^{\frac{1}{n}}$. 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 <u>The spaces C(X) and its duals M(X)</u>:

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 $\delta_x$, $x \in X$, defined by $\langle f, \delta_x \rangle := f(x)$ 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 $\|f_n\| \le c$ 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 <u>Sequence spaces</u>:

Let D be a set and take $1 \le p < \infty$ . The sequence space $l^p(D)$ is defined by

$$l^p(D) := \{(x_d)_{d\in D} : \sum_{d\in D} |x_d|^p < \infty\}$$

where $x_d$ are real (or complex) numbers.

Analogously, we define

$$l^\infty(D) := \{(x_d)_{d\in D} : \sup_{d\in D} |x_d| < \infty\}.$$

The vector space $l^p(D)$, resp. $l^\infty(D)$, becomes a Banach space if endowed with the norm

$$\|(x_d)_{d\in D}\| := (\sum_{d\in D} |x_d|^p)^{1/p},$$

resp. $$\|(x_d)_{d\in D}\| := \sup_{d\in D} |x_d|.$$

In our lectures, D equals $\mathbb{N}$, $\mathbb{N}_0$ or Z. Instead of $l^p(D)$ we write $l^p$ if no confusion is possible.

B.20 <u>The spaces $L^p(X,\Sigma,\mu)$</u>:

Let $(X,\Sigma,\mu)$ be a measure space and take $1 \leq p < \infty$.
By $\mathcal{L}^p(X,\Sigma,\mu)$ we denote the vector space of all real- or complex-valued measurable functions on X with $\int_X |f|^p \, d\mu < \infty$.

Then
$$\|f\|_p := (\int_X |f|^p d\mu)^{1/p}$$
is a semi-norm on $\mathcal{L}^p(X,\Sigma,\mu)$, and
$$N_\mu := \{f \in \mathcal{L}^p(X,\Sigma,\mu) : \|f\|_p = 0\}$$
is a closed subspace.
The quotient space
$$L^p(X,\Sigma,\mu) = L^p(\mu) := \mathcal{L}^p(X,\Sigma,\mu)/N_\mu$$
endowed with the quotient norm is a Banach space. Analogously, one denotes by $\mathcal{L}^\infty(X,\Sigma,\mu)$ the vector space of $\mu$-essentially bounded measurable functions on X. Again,
$$\|f\|_\infty := \inf\{c\in\mathbb{R}^+ : \mu[|f| > c] = 0\}$$
yields a semi-norm on $\mathcal{L}^\infty(X,\Sigma,\mu)$, and the subspace
$$N_\mu := \{f\in\mathcal{L}^\infty(X,\Sigma,\mu) : \|f\|_\infty = 0\}$$
is closed.
The quotient space
$$L^\infty(X,\Sigma,\mu) = L^\infty(\mu) := \mathcal{L}^\infty(X,\Sigma,\mu)/N_\mu$$
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 $f \in \mathcal{L}^p(X,\Sigma,\mu)$ 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 $\mathcal{L}^p(X,\Sigma,\mu)$. However, if they leave invariant $N_\mu$, we can and shall consider the induced operators on $L^p(X,\Sigma,\mu)$.

B.21 For $1 \leq p < \infty$, 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 $(X,\Sigma,\mu)$ is finite, then

$$L^\infty(\mu) \subset L^{p_2}(\mu) \subset L^{p_1}(\mu) \subset L^1(\mu)$$

for $1 \leq p_1 \leq p_2 \leq \infty$ .

B.23 Let $(X,\Sigma,\mu)$ be $\sigma$-finite.
Then the dual of $L^p(X,\Sigma,\mu)$, $1 \leq p < \infty$ , is isomorphic to $L^q(X,\Sigma,\mu)$, where $\frac{1}{p} + \frac{1}{q} = 1$, and the canonical bilinear form is given by

$$\langle f,g \rangle = \int f \cdot g \, d\mu \qquad \text{for } f \in L^p(\mu),\ g \in L^q(\mu).$$

Analogously, the dual of $L^1(\mu)$ is isomorphic to $L^\infty(\mu)$.

B.24 <u>Conditional expectation</u>:
Given a measure space $(X,\Sigma,\mu)$ and a sub-$\sigma$-algebra $\Sigma_0 \subset \Sigma$ .
We denote by J the canonical injection from $L^p(X,\Sigma_0,\mu)$ into $L^p(X,\Sigma,\mu)$ for $1 \leq p \leq \infty$ . J is contractive and positive (see C.4).
Its (pre-)adjoint

$$P \ : \ L^q(X,\Sigma,\mu) \rightarrow L^q(X,\Sigma_0,\mu)$$

is a positive contractive projection satisfying

$$P(f \cdot g) = g \cdot P(f) \qquad \text{for } f \in L^q(X,\Sigma,\mu),\ g \in L^\infty(X,\Sigma_0,\mu).$$

Proof:
P is positive and contractive since J enjoys the same properties.
The above identity follows from

$$\langle P(fg),h \rangle = \langle fg, Jh \rangle = \int fgh \, d\mu = \langle f, J(gh) \rangle = \langle (Pf)\, g,\ h \rangle$$

for all (real) $h \in L^p(X,\Sigma_0,\mu)$. ∎

We call P the <u>conditional</u> <u>expectation</u> <u>operator</u> corresponding to $\Sigma_0$ .

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 $1 \leq p < \infty$. The $l^p$ - direct sum of $(E_i)_{i \in \mathbb{N}}$ is defined by

$$E := \bigoplus_p E_i := \left\{ (x_i)_{i \in \mathbb{N}} : x_i \in E_i \text{ for all } i \in \mathbb{N} \text{ and } \sum_{i \in \mathbb{N}} \|x_i\|^p < \infty \right\}.$$

E is a Banach space under the norm

$$\|(x_i)_{i \in \mathbb{N}}\| := \left( \sum_{i \in \mathbb{N}} \|x_i\|^p \right)^{\frac{1}{p}}.$$

Given $S_i \in \mathcal{L}(E_i)$ with $\sup_{i \in \mathbb{N}} \|S_i\| < \infty$,

then

$$\bigoplus S_i : (x_i)_{i \in \mathbb{N}} \longmapsto (S_i x_i)_{i \in \mathbb{N}}$$

is a bounded linear operator on E with $\|\bigoplus S_i\|_E = \sup\{\|S_i\| : i \in \mathbb{N}\}$.
Analogously one defines the $l^\infty$ - 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 $L^p(X,\Sigma,\mu)$, $1 \leq p \leq \infty$. 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 $f \geq 0$, if $f(x) \geq 0$ 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 $\mu$-almost all $x \in X$.

Using the positive cone $E_+ := \{f \in E: f \geq 0\}$ we define an order relation on $E$ by $f \geq g$ 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 $0 \leq f \leq g$ implies $\|f\| \leq \|g\|$, and $\| |f| \| = \|f\|$ for every $f \in E$.
If we consider a complex function space E then the order relation "$\leq$" 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 "$\leq$" making it into a vector lattice (i.e. $|f| = f \vee (-f)$ exists for every $f \in E$) and satisfying the compatibility condition:

$$|f| \leq g \text{ implies } \|f\| \leq \|g\| \text{ for all } f, g \in E.$$

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 $[g, h] := \{f \in E : g \leq f \leq h\}$ 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 $L^p(\mu)$, $1 \leq p \leq \infty$, while $C([0, 1])$ is not order complete.

C.4 Positive operators:
Let E, F be (real or complex) Banach lattices and

$$T : E \rightarrow F$$

a continuous linear operator.
T is positive if $TE_+ \subset F_+$, or equivalently, if $T|f| \geq |Tf|$ 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. $0 \leq f < g$ implies $\|f\| < \|g\|$; e.g. $E = L^p(\mu)$ for $1 \leq p < \infty$) then every positive isometry T on E is a lattice homomorphism. In fact, in that case
$|Tf| \leq T|f|$ and $\| |Tf| \| = \|Tf\| = \|f\| = \| |f| \| = \| T|f| \|$ imply $|Tf| = T|f|$.
Finally, T is called order continuous (countably order continuous) if $\inf_{\alpha \in A} Tx_\alpha = 0$ for every downward directed net (sequence) $(x_\alpha)_{\alpha \in A}$

with $\inf_{\alpha \in A} x_\alpha = 0$.

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

$$M_g : C(X) \longrightarrow C(X) \text{ (resp. } L^p(X,\Sigma,\mu) \to L^p(X,\Sigma,\mu))$$
$$f(x) \longmapsto f(x)\cdot g(x), \quad x \in X,$$

is a lattice homomorphism for every $0 \le g \in C(X)$ (resp. $0 \le g \in L^\infty(X,\Sigma,\mu)$).
The operators

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

induced in $C(X)$ or $L^p(X,\Sigma,\mu)$, $1 \le p \le \infty$, by suitable transformations

$$\varphi : X \longrightarrow X$$

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 $\mathbb{C}$ can be transfered to $E$: for $f, g \in E$ we define
$f\cdot g$, the product of $f$ and $g$, by $(f\cdot g)(x) := f(x)\cdot g(x)$ 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 $\mathbb{1}$, defined by $\mathbb{1}(x) := 1$ for all $x \in X$, is the neutral element of the above commutative multiplication. The operation "$*$" is an involution.

C.7 A $C^*$-algebra $\mathfrak{A}$ is a complex Banach space and an algebra with involution $*$ satisfying

$$\|f \cdot f^*\| = \|f\|^2$$

for all $f \in \mathfrak{A}$.
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 $l^\infty$.

C.8 Multiplicative operators:

Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two $C^*$-algebras.

The morphisms

$$T : \mathcal{A}_1 \to \mathcal{A}_2$$

corresponding to the $C^*$-algebra structure of $\mathcal{A}_1$ and $\mathcal{A}_2$ are continuous linear operators satisfying

$$T(f\cdot g) = (Tf)\cdot(Tg)$$

and

$$T(f^*) = (Tf)^*$$

for all $f, g \in \mathcal{A}_1$.

Let $\mathcal{A} = C(X)$, 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 $\mathcal{A}$ satisfying $T_\varphi 1\!\!1 = 1\!\!1$ and $T_\varphi f^* = (T_\varphi f)^*$ (see II.4).

C.9 Representation theorem of Gelfand-Neumark:

Every commutative $C^*$-algebra $\mathcal{A}$ 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 $\mathcal{A}$, endowed with the weak* topology (see Sakai [1971], 1.2.1).

We remark that for $\mathcal{A} = l^\infty(\mathbb{N})$ the space $X$ is homeomorphic to the Stone-Cech compactification $\beta\mathbb{N}$ of $\mathbb{N}$ (see Schaefer [1974], p.106), and for $\mathcal{A} = L^\infty(Y,\Sigma,\mu)$, $X$ 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 $\Gamma^n$. 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

$$(g,h) \mapsto g\cdot h \quad \text{on } G\times G$$

and

$$g \mapsto g^{-1} \quad \text{on } G$$

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. $m = R'_g m = L'_g m$ 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. (Y.10.1) or Schaefer [1974], 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 $\chi$ from G into the unit circle $\Gamma$ is called a <u>character</u> of G. The set of all characters of G is called the <u>character group</u> or <u>dual group</u> of G, denoted by $\hat{G}$. Endowed with the pointwise multiplication and the compact-open topology $\hat{G}$ 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 G is a discrete abelian group, $\hat{G}$ 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 $\Gamma$ is a compact group. Moreover, each character of $\Gamma$ 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 $\Gamma$.

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 $\qquad g \mapsto \hat{\hat{g}}$ with $\hat{\hat{g}}(\chi) := \chi(g)$
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 $\int \chi(g)\ dm(g) = 0$ for $\chi \neq 1$. Choose $h \in G$ with $\chi(h) \neq 1$. Then we have

$$\int \chi(g)\ dm(g) = \int \chi(hg)\ dm(g) = \chi(h) \int \chi(g)\ dm(g)$$

and hence $$\int \chi(g)\ dm(g) = 0.$$

Clearly, every character is a normalized function in $L^2(G,\mathcal{B},m)$. Let $g, h \in G$, $g \neq h$, 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 generated by $\hat{G}$, 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 $\Gamma^n$. 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 $a := (a_1, \dots, a_n) \in \Gamma^n$ such that $\{a_1, \dots, a_n\}$ is linearly independent in the $\mathbb{Z}$ - module $\Gamma$, i.e. $1 = a_1^{z_1} \dots a_n^{z_n}$, $z_i \in \mathbb{Z}$, implies $z_i = 0$ for $i = 1, \dots, n$. Then the subgroup $\{a^z : z \in \mathbb{Z}\}$ is dense in $\Gamma^n$.

Proof:
Endow $\hat{\mathbb{Z}} = \Gamma$ with the discrete topology and form the dual group $\widehat{(\hat{\mathbb{Z}})_d} = \hat{\Gamma}_d$.
$\hat{\Gamma}_d$ is a compact subgroup of the product $\Gamma^\Gamma$ - note that here the compact-open topology on $\hat{\Gamma}_d$ is the topology induced from the product $\Gamma^\Gamma$.
We consider the continuous monomorphism

$$\Phi : \mathbb{Z} \longrightarrow \widehat{(\hat{\mathbb{Z}})_d}$$
$$z \longmapsto \Phi(z) \quad \text{defined by } \Phi(z)(\gamma) := \gamma^z \text{ for all } \gamma \in \Gamma = \hat{\mathbb{Z}}.$$

Then the duality theorem yields that $\Phi(\mathbb{Z})$ is dense in $\widehat{(\hat{\mathbb{Z}})_d}$.

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 $\Gamma$ there exists a $\mathbb{Z}$-linear mapping

$$\chi \in \hat{\Gamma}_d \qquad \text{with } \chi(a_i) = b_i \quad \text{for } i = 1, \ldots, n.$$

By definition of the product topology on $\Gamma^\Gamma$ and by denseness of $\Phi(\mathbb{Z})$ in $\hat{\Gamma}_d$ we obtain $z \in \mathbb{Z}$ such that

$$|a_i^z - b_i| = |\Phi(z)(a_i) - \chi(a_i)|$$

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:**
(i) A sequence $(x_n)_{n\in\mathbb{N}_0}$ of real (or complex) numbers is called Cesaro - summable if $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}x_i$ exists.
(ii) Let $(n_i)_{i\in\mathbb{N}}$ be a subsequence of $\mathbb{N}_0$. Then $(n_i)_{i\in\mathbb{N}}$ has density $a\in[0, 1]$, denoted by $d((n_i)_{i\in\mathbb{N}}) = a$, if

$$\lim_{k\to\infty}\frac{1}{k}\left|\{n_i: i\in N\}\cap\{0, 1,\dots, k-1\}\right| = a$$

where $|\cdot|$ denotes the cardinality.

**E.2 Lemma:**
For $(x_n)_{n\in\mathbb{N}_0}\in l^\infty(\mathbb{N}_0)$ the following conditions are equivalent:
(i) $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}|x_i| = 0$ .
(ii) There exists a subsequence $N$ of $\mathbb{N}_0$ with $d(N) = 1$ such that $\lim_{n\in N} x_n = 0$ .

Proof:
We define $N_k := \{0, 1,\dots, k-1\}$.
(i) $\Rightarrow$ (ii): Let $J_k := \{n\in\mathbb{N}_0 : |x_n| \geq \frac{1}{k}\}$ , $k>0$, and observe that $J_1\subseteq J_2\subseteq\dots$ . Since $\frac{1}{n}\sum_{i=0}^{n-1}|x_i| \geq \frac{1}{n}\frac{1}{k}|J_k\cap N_n|$, each $J_k$ has density 0.
Therefore, we can choose integers $0 = n_0 < n_1 < n_2 < \dots$ such that

$$\frac{1}{n}\,|J_{k+1} \cap N_n| < \frac{1}{k+1} \qquad \text{for } n \geq n_k .$$

Define $J := \bigcup_{k\in\mathbb{N}} (J_{k+1} \cap (N_{n_{k+1}} \setminus N_{n_k}))$ and show $d(J) = 0$.

Let $n_k \leq n < n_{k+1}$. Then, we obtain

$$J \cap N_n = (J \cap N_{n_k}) \cup (J \cap (N_n \setminus N_{n_k})) \subseteq (J_k \cap N_{n_k}) \cup (J_{k+1} \cap N_n),$$

and conclude that

$$\frac{1}{n}\,|J \cap N_n| \leq \frac{1}{k} + \frac{1}{k+1} .$$

If $n$ tends to infinity, the same is true for k, and hence, J has density 0. Obviously, the sequence $N := \mathbb{N} \setminus J$ has the desired properties.

(ii) $\Rightarrow$ (i): Let $\varepsilon > 0$ and $c := \sup\{|x_n| : n \in \mathbb{N}_0\}$.

Because of (ii) and $d(\mathbb{N} \setminus N) = 0$ there exists $n_\varepsilon \in \mathbb{N}$ such that $n \geq n_\varepsilon$ implies $|x_n| < \varepsilon$ for $n \in N$ and $\frac{1}{n}\,|(\mathbb{N} \setminus N) \cap N_n| < \varepsilon$. If $n \geq n_\varepsilon$ we conclude that

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{n-1} |x_i| &= \frac{1}{n}\sum_{i\in(\mathbb{N}\setminus N)\cap N_n} |x_i| + \frac{1}{n}\sum_{i\in N\cap N_n} |x_i| \\ &\leq \frac{c}{n}\,|(\mathbb{N} \setminus N)\cap N_n| + \varepsilon \\ &\leq (c+1)\cdot\varepsilon . \quad \blacksquare\end{aligned}$$

E.3 <u>Lemma:</u>

Take a sequence $(z_n)_{N\in\mathbb{N}}$ of complex numbers such that

$$\sum_{n=1}^{\infty} n\,|z_{n+1} - z_n|^2 < \infty . \text{ If } \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} z_i = 0, \text{ then } \lim_{n\to\infty} z_n = 0.$$

Proof:

Define $c_n := \sum_{k=n} k\,|z_{k+1} - z_k|^2$. Then

$$\max\{|z_{n+k} - z_n| : 1 \leq k \leq n-2\} \leq \sum_{k=n}^{2n-3} |z_{k+1} - z_k| \leq \Big(\sum_{k=n}^{2n-3} |z_{k+1} - z_k|^2 (n-2)\Big)^{1/2}$$
$$\leq c_n$$

and

$$|z_n| = \Big|b_{n-1} - 2b_{2n-2} + \frac{1}{n-1}\sum_{k=1}^{n-2}(z_{n+k} - z_k)\Big| \quad \text{for } b_n := \frac{1}{n}\sum_{i=1}^{n} z_i . \blacksquare$$

E.4 <u>Lemma:</u>

Let $N_i$, $i = 1, 2, \ldots$ , be subsequences 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 $N \setminus N_i$ 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

$$1 - 2^{-i} \leq \frac{1}{k}|N_i \cap \{0, \ldots, k-1\}| \quad \text{for all } k \geq k_i .$$

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 <u>Lemma:</u>

If $(x_n)_{n \in \mathbb{N}}$ is a sequence of positive reals satisfying $x_{n+m} \leq x_n + x_m$ 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 $0 \leq m < n$. Then

$$\frac{x_j}{j} = \frac{x_{kn+m}}{kn+m} \leq \frac{x_{kn}}{kn} + \frac{x_m}{kn} \leq \frac{k\, x_n}{kn} + \frac{x_m}{kn} = \frac{x_n}{n} + \frac{x_m}{kn} .$$

If $j \to \infty$ then $k \to \infty$ , too, and we obtain

$$\limsup_{j \to \infty} \frac{x_j}{j} \leq \frac{x_n}{n} , \text{ and even } \limsup_{j \to \infty} \frac{x_j}{j} \leq \inf_{n \in \mathbb{N}} \frac{x_n}{n} .$$

On the other hand, $\inf_{n \in \mathbb{N}} \frac{x_n}{n} \leq \liminf_{n \to \infty} \frac{x_n}{n}$ , 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 $\nu$ on X which is invariant under $\varphi$. Such an invariant measure allows the application of the measure-theoretical results in the topological context.
It is even more important to obtain a $\varphi$-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 <u>Theorem</u> (Krylov-Bogoliubov, 1937):
Let X be compact and $\varphi: X \longrightarrow X$ continuous. There exists a probability measure $\nu \in C(X)'$ which is $\varphi$-invariant.

Proof:
Consider the induced operator $T := T_\varphi$ on $C(X)$. Its adjoint $T'$ leaves invariant the weak$^*$ compact set $\mathcal{P}$ of all probability measures in $M(X)$. If $\nu_o \in \mathcal{P}$, then the sequence

$$\{T_n' \nu_o : n \in \mathbb{N}\}$$

has a weak$^*$ accumulation point $\nu$. It is easy to see (use IV.3.o) that $T'\nu = \nu$, i.e. $\nu$ is $\varphi$-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 $\mu$ some $\varphi$-invariant probability measure. Moreover, the set $\mathcal{P}_\varphi$ of all $\varphi$-invariant measures in $\mathcal{P}$ is a convex $\sigma(C(X)',C(X))$-compact subset of $C(X)'$. Therefore, the Krein - Milman theorem yields many extreme points of $\mathcal{P}_\varphi$ called "ergodic measures". The reason for that nomenclature lies in the following characterization.

S.2 <u>Corollary:</u>
Let $(X;\varphi)$ be a TDS. $\mu$ is an extreme point of $\mathcal{P}_\varphi$ 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 $\varphi(X \setminus A) = X \setminus A$. Define two different measures $\mu_1, \mu_2 \in \mathcal{P}_\varphi$ by

$$\mu_1(B) := \frac{\mu(B \cap A)}{\mu(A)}$$

$$\mu_2(B) := \frac{\mu(B \cap (X \setminus A))}{\mu(X \setminus A)} \quad \text{for } B \in \mathcal{B}.$$

Clearly, $\mu = \mu(A) \cdot \mu_1 + (1 - \mu(A)) \cdot \mu_2$, and $\mu$ is not an extreme point of $\mathcal{P}_\varphi$.
On the other hand, assume $(X, \mathcal{B}, \mu ; \varphi)$ to be ergodic. If $\mu = \frac{1}{2}(\mu_1 + \mu_2)$ for $\mu_1, \mu_2 \in \mathcal{P}_\varphi$, then $\mu_1 \leq 2\mu$ and hence $\mu_1 \in L^1(\mu)' = L^\infty(\mu)$. But the fixed space of $T'_\varphi$ in $L^\infty(\mu)$ contains $\mu$ and $\mu_1$ and is one-dimensional by (IV.6), (IV.4.e) and (III.4).
Therefore we conclude $\mu = \mu_1$, i.e. $\mu$ must be an extreme point of $\mathcal{P}_\varphi$.
∎

The question, whether there exist $\varphi$-invariant probability measures equivalent to some distinguished measure, is more difficult and will be converted into a "mean ergodic" problem.

S.3 Theorem:
Let $\mu$ 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 $\mu$ (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 $\varphi$-invariant probability measure $\nu$ on X which is equivalent to $\mu$.
(b) For the induced operator $T := T_\varphi$ on $L^\infty(X, \mathcal{B}, \mu)$ the Cesaro means $T_n$ converge in the $\sigma(L^\infty, L^1)$-operator topology to some strictly positive projection $P \in \mathcal{L}(L^\infty(\mu))$, 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 $\varphi$ imply the $T = T_\varphi$ is a well-defined positive

contraction on $L^{\infty}(\mu)$ having a pre-adjoint $T'$ on $L^1(\mu)$ (see Schaefer [1974], III.9, Example 1).

(a) $\Rightarrow$ (c): By the Radon-Nikodym theorem the $\varphi$-invariant probability measure $\nu$ equivalent to $\mu$ corresponds to a normalized strictly positive $T$-invariant function $u \in L^1(\mu)$. But for such functions the order interval

$$[-u, u] := \{ f \in L^1(\mu) : -u \leq f \leq u \}$$

is weakly compact and total in $L^1(\mu)$. Therefore $T u = u$ 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

$$\nu_o := \mu \circ P$$

which is a strictly positive $\varphi$-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 $\nu$ as the band component of $\nu_o$ in $L^1(\mu)$.

By Ando [1968], Lemma 1, $\nu$ is still strictly positive and hence defines a measure equivalent to $\mu$. Moreover, $T'\nu$ is contained in $L^1(\mu)$ and dominated by $\nu_o$, hence $T'\nu \leq \nu$. From $T \mathbb{1} = \mathbb{1}$ we conclude that $T'\nu = \nu$ and that $\nu$ is $\varphi$-invariant. Normalization of $\nu$ 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 <u>piecewise $C^2$</u>, i.e. there is a finite partition of $[0,1]$ in intervals $A_i$ such that $\varphi$ can be extended continuously from the interior $\mathring{A}_i$ to the closure $\overline{A}_i$ and the resulting function $\varphi_i$ is twice continuously differentiable on $\overline{A}_i$. moreover we assume that the derivates $\dot{\varphi}_i$ do not vanish on $\mathring{A}_i$, hence $\varphi_i$ is increasing or decreasing.

In this case, $\varphi$ 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 $T\mathbb{1} = \mathbb{1}$

and having a pre-adjoint $T'$ on $L^1(m)$.
As a consequence of this theorem, one concludes that $\varphi$ possesses an invariant probability measure which is absolutely continuous with respect to $m$ iff $\dim F(T') \geq 1$. In particular, this follows if $T'$ is mean ergodic.
To find out under which conditions on $\varphi$ this holds, we observe that the pre-adjoint $T'$ can be written as

$$T'f(x) = \sum_i f \circ \varphi_i^{-1}(x) \quad \sigma_i(x) \; \mathbb{1}_{B_i}(x) \ ,$$

where $B_i = \varphi_i(\overline{A_i})$ and $\sigma_i$ is the absolute value of the derivate of $\varphi_i^{-1}$.
In fact: For every $x \in (0,1)$,

$$\int_0^x T'f \, dm = \int_0^1 f \cdot \mathbb{1}_{(0,x)} \circ \varphi \, dm = \int_{\varphi^{-1}(0,x)} f \, dm .$$

Thus $T'f$ is the ~~derivative~~ $\dot{g}$ of the function $g(x) = \int_{\varphi^{-1}(0,x)} f \, dm$.

If $\varphi$ 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

$$v(f) := \sup_{n \in \mathbb{N}} \left\{ \sum_{j=1}^{n} |f(t_j) - f(t_{j-1})| \; : \; a=t_0 < t_1 < \dots < t_n = b \right\}.$$

With this concept and using some elementary analysis, one proves that

$$(*) \quad v(f \cdot g) \leq v(f) \, \|g\|_\infty + \int_a^b |f \cdot \dot{g}| \, dm$$

if f is piecewise continuous and g continuously differentiable.

After these preparations we present the main result.

S.4 <u>Proposition:</u>
Let $\varphi : [0,1] \longrightarrow [0,1]$ be piecewise $C^2$ such that

$$s := \inf \{ |\dot{\varphi}(t)| \; : \; t \in (0,1) \text{ and } \varphi \text{ differentiable at } t \} > 1.$$

Then there exists a $\varphi$-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 $T'_\varphi$ of $T_\varphi$ 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, $\inf \{ |\dot\phi(t)| : t \in (0,1) \text{ and } \phi \text{ differentiable at } t \} \geq s^n > 2$.

Now we estimate the variation $v(T'_\phi f)$ for any piecewise continuous function $f : [0,1] \longrightarrow \mathbb{R}$.

To this purpose we need some constants determined by the function $\phi$.

Take the partition of $[0,1]$ into intervals $A_i$ corresponding to $\phi$ and write

$$T'_\phi f(x) = \sum_{i=1}^{m} f \circ \phi_i^{-1}(x) \; \sigma_i(x) \; 1_{B_i}(x)$$

where $B_i = \phi_i(\overline{A_i})$ and $\sigma_i(x) = |(\dot\phi_i^{-1})(x)|$.

1. For $\sigma_i$ we have $\sigma_i(x) \leq s^{-n} < \frac{1}{2}$ for every $x \in B_i$.
2. Put $k := \max \{ |\dot\sigma_i(x)| : x \in \overline{B_i};\ i=1,..,m \} \cdot \max \{ |\dot\phi_i(x)| : x \in \overline{A_i}; i=1,...,m \}$.
3. For the interval $\overline{A_i} = [a_{i-1}, a_i]$ we estimate

$$\begin{aligned} |f(a_{i-1})| + |f(a_i)| &\leq 2 \inf \{ |f(x)| : x \in A_i \} + v(f|_{A_i}) \\ &\leq \frac{2}{m(A_i)} \int_{A_i} |f| \, dm + v(f|_{A_i}) \\ &\leq 2h \int_{A_i} |f| \, dm + v(f|_{A_i}) \end{aligned}$$

for $h := \max \{ \frac{1}{m(A_i)} : 1 \leq i \leq m \}$.

Now, we can calculate:

$$\begin{aligned} v(T'_\phi f) &\leq \sum_{i=1}^{m} v(f \circ \phi_i^{-1}(x) \cdot |\sigma_i(x)| \cdot 1_{B_i}(x)) \\ &\leq \sum_{i=1}^{m} ( \|\sigma_i\|_\infty \cdot v(f \circ \phi_i^{-1}(x) \cdot 1_{B_i}(x)) + \int_{B_i} |f \circ \phi_i^{-1} \cdot \dot\sigma_i| \, dm \\ &\qquad\qquad \text{(by the inequality } (*) \text{ above)} \\ &\leq \sum_{i=1}^{m} (s^{-n} ( |f(a_{i-1})| + |f(a_i)| + v(f|_{A_i})) + k \int_{B_i} |f \circ \phi_i^{-1}| \cdot \sigma_i \, dm) \\ &\qquad (\text{since } \max\{ |\dot\phi_i(x)| : x \in \overline{A_i}; i=1,..m \} = \min\{ \sigma_i(x) : x \in B_i; =1,..,m \}) \\ &\leq \sum_{i=1}^{m} (s^{-n}(2h \int_{A_i} |f| \, dm + 2 \, v(f|_{A_i})) + k \int_{A_i} |f| \, dm) \\ &\leq (h + k) \|f\|_1 + 2 \, s^{-n} \, v(f) . \end{aligned}$$

Observing that $v(\mathbb{1}) = 0$ and $T'^{\,r}_\phi \mathbb{1}$ is again piecewise continuous, we obtain by induction

$$v(T_\phi'^{\,r}\mathbb{1}) \le (h+k)\sum_{i=0}^{r-1}(2\,s^{-n})^i \le \frac{h+k}{1-2\,s^{-n}} \quad \text{for every } r\in\mathbb{N},$$

and therefore

$$\| T_\phi'^{\,r}\mathbb{1}\|_\infty \le \| T_\phi'^{\,r}\mathbb{1}\|_1 + v(T_\phi'^{\,r}\mathbb{1}) \le 1 + \frac{h+k}{1-2\,s^{-n}} \quad ,$$

i.e. $T_\phi'^{\,r}\mathbb{1} \le M\cdot\mathbb{1}$ for $r\in\mathbb{N}$ and some $M>0$.

For the final conclusion the abstract mean ergodic theorem (IV.6) implies that $T_\phi'$ is mean ergodic. Since $T_\phi' = T_\varphi'^{\,n}$, the same is true for $T_\varphi'$ by (IV.D.2). ∎

In conclusion, we present some examples showing the range of the above proposition.

S.5 Examples:

1. The transformation

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

satisfies the assumptions of our proposition and has a $\varphi$-invariant measure. In fact, $m$ itself is invariant.

2. For

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

The assumption $|\dot\varphi(t)| > 1$ is violated at $t=0$. In fact, there is no $\varphi$-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 $\varphi(t) := 4t\cdot(1-t)$ the assumption $|\dot\varphi(t)| > 1$ is strongly violated, nevertheless there is a $\varphi$-invariant measure:

Indeed, the equation $\int_{[0,x]} f\,dm = \int_{\varphi^{-1}[0,x]} f\,dm$ together with the plausible assumption that $f(t) = f(1-t)$ leads to

$$F(x) := \int_0^x f(t)\,dt = 2\cdot \int_0^{\frac{1}{2}-\frac{1}{2}\sqrt{1-x}} f(t)\,dt = 2\,F(\tfrac{1}{2}-\tfrac{1}{2}\sqrt{1-x}) .$$

By substituting $x = \sin^2\xi$ we obtain

$$F(\sin^2\xi) = 2\,F(\tfrac{1}{2}-\tfrac{1}{2}\cos\xi) = 2\,F(\sin^2\tfrac{\xi}{2})$$

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 $\varphi$-invariant measure $f \cdot m$ on $[0,1]$.

4. Finally, $\varphi(t) := 2(t - 2^{-i})$ for $2^{-i} < t \leq 2^{1-i}$, $i \in \mathbb{N}$, has $\dot{\varphi}_i(t) = 2$, but infinitely many discontinuities. Again there exists no $\varphi$-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 T: Asymptotic Independence

In (IX.2) we have seen that an MDS $(X,\Sigma,\mu;\varphi)$ is strongly mixing iff $\mu(\varphi^{-n}A\cap B)$ converges to $\mu(A)\cdot\mu(B)$ for every $A, B\in\Sigma$. Since in probability theory the "events" A and B are called independent if $\mu(A\cap B)=\mu(A)\cdot\mu(B)$, mixing can be interpreted as asymptotic independence with respect to $\varphi$ of any two measurable sets. In the following we transfer this notion of asymptotic independence to covers and define stronger mixing properties.

T.1 Definition:

Let $(X,\Sigma,\mu;\varphi)$ be an MDS and $\alpha,\beta\in\tilde{\Sigma}_d$.

(i) The dependence between $\alpha$ and $\beta$ is defined by

$$\mathrm{dep}(\alpha,\beta) := \sum_{A\in\alpha}\sum_{B\in\beta}|\mu(A\cap B)-\mu(A)\cdot\mu(B)| .$$

(ii) The disjoint cover $\alpha$ is called mixing (or asymptotic independent) if

$$\lim_{k\to\infty}\mathrm{dep}(\alpha^n,\tilde{\varphi}^k\alpha^m)=0 \quad \text{for every } n, m\in\mathbb{N}.$$

(iii) $\alpha$ is called a K(olmogorov)-partition if

$$\lim_{k\to\infty}\sup_{m\in\mathbb{N}}\mathrm{dep}(\alpha^n,\tilde{\varphi}^k\alpha^m)=0 \quad \text{for every } n\in\mathbb{N}.$$

(iv) is called a w(eak) B(ernoulli)-partition if

$$\lim_{k\to\infty}\sup_{m,n\in\mathbb{N}}\mathrm{dep}(\alpha^n,\tilde{\varphi}^{k+n}\alpha^m)=0.$$

Clearly, for $\alpha\in\tilde{\Sigma}_d$ wB implies K, and K implies mixing. Moreover, in Definition (ii) and (iv) one might as well choose n = m. This is due to the fact that $\alpha\leq\alpha'\in\tilde{\Sigma}_d$ implies that $\mathrm{dep}(\alpha,\beta)\leq\mathrm{dep}(\alpha',\beta)$.

For the canonical generator $\alpha$ of a Bernoulli shift we have $\mathrm{dep}(\alpha,\tilde{\varphi}^k\alpha)=0$ for every $k\in\mathbb{N}$. Therefore $\alpha$ has all the above properties.

**T.2 Proposition:**

An MDS $(X, \Sigma, \mu; \varphi)$ is strongly mixing if and only if every partition $\alpha \in \tilde{\Sigma}_d$ is mixing.

Proof:

"$\Leftarrow$": For $A, B \in \Sigma$ consider $\alpha := \{A, X \setminus A\} \vee \{B, X \setminus B\}$. Then

$$|\mu(\varphi^{-k}A \cap B) - \mu(A)\cdot\mu(B)| = \Big|\sum_{A' \subseteq A, A' \in \alpha} \sum_{B' \subseteq B, B' \in \alpha} (\mu(\varphi^{-k}A' \cap B') - \mu(A')\cdot\mu(B'))\Big|$$

$$\leq \sum_{\alpha \ni A' \subseteq A} \sum_{\alpha \ni B' \subseteq B} |\mu(\varphi^{-k}A' \cap B') - \mu(\varphi^{-k}A')\ (B')|$$

$$\leq \text{dep}\,(\alpha, \tilde{\varphi}^k \alpha), \text{ which tends to zero.}$$

"$\Rightarrow$": Fix $\alpha \in \tilde{\Sigma}_d$ and $n, m \in \mathbb{N}$. For $\varepsilon > 0$ we can find $k \in \mathbb{N}$ such that

$$|\mu(\varphi^{-k}A \cap B) - \mu(\varphi^{-k}A)\cdot\mu(B)| \leq \frac{\varepsilon}{|\alpha^m|\cdot|\alpha^n|}$$

for every $A \in \alpha^m$ and $B \in \alpha^n$. Therefore $\text{dep}(\alpha^n, \tilde{\varphi}^k \alpha^m) < \varepsilon$. ∎

**T.3 Proposition:**

An MDS $(X, \Sigma, \mu; \varphi)$ with a mixing generator $\alpha \in \tilde{\Sigma}_d$ is strongly mixing.

Proof:

The set $\{1_A : A \in \alpha_{-n}^n \text{ for some } n \in \mathbb{N}\}$ is dense in $\{1_B : B \in \Sigma\}$. For $A \in \alpha_{-n}^n$ and $B \in \alpha_{-m}^m$ we have $\varphi^{-n}A \in \alpha^{2n+1}$ and $\varphi^{-m}B \in \alpha^{2m+1}$, and therefore

$$|\mu(\varphi^{-k}A \cap B) - \mu(A)\cdot\mu(B)|$$

$$= |\mu(\varphi^{-k-m+n}(\varphi^{-n}A) \cap \varphi^{-m}B) - \mu(\varphi^{-n}A)\cdot\mu(\varphi^{-m}B)|$$

$$\leq \text{dep}(\tilde{\varphi}^{k+m-n}\alpha^{2n+1}, \alpha^{2m+1}) \to 0 \text{ as } k \to \infty.$$

By (IX.2) this shows that $(X, \Sigma, \mu; \varphi)$ is strongly mixing. ∎

Before introducing mixing properties based on (T.1(iii) and (iv), we show some interesting equivalences.

**T.4 Proposition:**

For a partition $\alpha \in \tilde{\Sigma}_d$ the following are equivalent:

(i) $\lim_{k\to\infty} \sup_{n\in\mathbb{N}} \operatorname{dep}(\beta, \tilde{\varphi}^k \alpha^n) = 0$ holds for every $\beta \in \tilde{\Sigma}_d$.

(ii) $\alpha$ is a K - partition.

(iii) For every $n \in \mathbb{N}$ and $A \in \alpha^n$

$$\lim_{k\to\infty} \sup \{ |\mu(A \cap B) - \mu(A)\cdot\mu(B)| : B \in \tilde{\varphi}^k \alpha^m, m \in \mathbb{N} \} = 0.$$

(iv) $\alpha$ satisfies Kolmogorov's "0 - 1 - law", i.e. for every

$$A \in \operatorname{Tail}(\alpha) := \bigcap_{k=0}^{\infty} \sigma(\bigcup_{m=1}^{\infty} \tilde{\varphi}^k \alpha^m) \text{ we have } \mu(A) \in \{0, 1\}.$$

Proof:

The implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) are obvious.

(iii) $\Rightarrow$ (iv): Let $0 < \varepsilon < 1$ and define $\Sigma_k := \sigma(\bigcup_{m\in\mathbb{N}} \tilde{\varphi}^k \alpha^m)$.
If $A \in \operatorname{Tail}(\alpha)$ then $A \in \Sigma_k$ for every $k \in \mathbb{N}_0$. Therefore, there is $B_k \in \tilde{\varphi}^k \alpha^{m_k}$ such that $d(A, B_k) < \varepsilon$. In particular, $B_0 \in \alpha^{m_0}$ and by (iii) it follows that

$$|\mu(B_0 \cap B_k) - \mu(B_0)\cdot\mu(B_k)| < \varepsilon$$

for sufficiently large k.
Thus

$$\begin{aligned} |\mu(A) - \mu(A)^2| &\leq |\mu(A) - \mu(B_0 \cap B_k)| + |\mu(B_0 \cap B_k) - \mu(B_0)\mu(B_k)| \\ &\quad + |\mu(B_0)\mu(B_k) - \mu(A)^2| \\ &\leq 2\varepsilon + \varepsilon + 2\varepsilon + \varepsilon^2 \\ &\leq 6\varepsilon, \end{aligned}$$

and we obtain $\mu(A) = \mu(A)^2$.

(iv) $\Rightarrow$ (i): Fix $\beta \in \tilde{\Sigma}_d$ and define $\Sigma_k$ as above. The semigroup consisting of the corresponding conditional expectation operators

$$P_k : L^1(X, \Sigma, \mu) \longrightarrow L^1(X, \Sigma_k, \mu) \quad (k \in \mathbb{N})$$

has trivial fixed space (by hypothesis) and is mean ergodic by (Y.10.4).

As in (Y.14) it follows that

$$P_k f \longrightarrow \langle f, 1 \rangle 1$$

for every $f \in L^1(X, \Sigma, \mu)$.
Therefore, for $\varepsilon > 0$ we can find k sufficiently large such that

$$\| P_k 1_B - \mu(B) 1 \| < \frac{\varepsilon}{|\beta|}$$

for every $B \in \beta$.
Since $A \in \tilde{\varphi}^k \alpha^m \subseteq \Sigma_k$ implies that $\mu(A \cap B) = \langle 1_A, 1_B \rangle = \langle 1_A, P_k 1_B \rangle$ we obtain

$$
\begin{aligned}
\mathrm{dep}(\beta, \tilde{\varphi}^k \alpha^m) &= \sum_{B\in\beta} \sum_{A\in\tilde{\varphi}^k\alpha^m} |\mu(A \cap B) - \mu(A)\mu(B)| \\
&= \sum_{B\in\beta} \sum_{A\in\tilde{\varphi}^k\alpha^m} |\langle 1_A, P_k 1_B\rangle - \langle 1_A, 1\rangle \mu(B)| \\
&\leq \sum_B \sum_A \int 1_A \cdot |P_k 1_B - \mu(B)\, 1|\, d\mu \\
&= \sum_{B\in\beta} \| P_k 1_B - \mu(B)\, 1 \| \\
&< \varepsilon . \quad \blacksquare
\end{aligned}
$$

T.5 Definition:
An MDS $(X, \Sigma, \mu; \varphi)$ in which every partition $\alpha \in \tilde{\Sigma}_d$ is a K-partition is called a K(olmogorov) - system .

The next result is the analogue of Proposition T.3 for K-systems. We present it without proof (see Smorodinsky [1971], Remark to Theorem 7.13 and Lemma 7.14).

T.6 Proposition:
An MDS $(X, \Sigma, \mu; \varphi)$ possessing a generator $\alpha \in \tilde{\Sigma}_d$ which is a K-partition, is a K-system.

From this proposition we can easily infer that every Bernoulli shift is a K-system. It is clear that every K-system is strongly mixing, but the converse is not true. In fact, the "Kolmogorov-property" contains much more complete information about the spectrum.

T.7 Proposition:
Every K-system $(X, \Sigma, \mu; \varphi)$ with separable infinite-dimensional $L^2(\mu)$ has countable Lebesgue spectrum.

Proof:
To simplify the proof we assume that the K-system $(X, \Sigma, \mu; \varphi)$ possesses a generator $\alpha$ (compare XII.D.2). From Proposition T.4 we infer that $\mathrm{Tail}(\alpha) := \bigcap_{k=0}^{\infty} \Sigma_k$ contains only sets of measure 0 or 1.
Here $\Sigma_k = \sigma(\bigcup_{m=1}^{\infty} \tilde{\varphi}^k \alpha^m)$, $k \in \mathbb{Z}$, and the corresponding projections $P_k$ satisfy $\lim_{k\to\infty} P_k f = \langle f, 1\rangle 1$ for every $f \in E := L^2(\mu)$ and $\bigcap_{k=1}^{\infty} P_k E = \langle 1 \rangle$.

Since $\alpha$ is a generator, $\bigcup_{k=0}^{-\infty} P_k E$ is dense in E.

We write $P_o E = P_1 E \oplus F$, where $F := (Id - P_1)P_o E$.
Then

$$TF = (TP_o - TP_1)E = (P_1 - P_2)E = (Id - P_2)P_1 E$$

and $\quad P_1 E = P_2 E \oplus TF$. Continuing this process we see that

$$E = \langle 1 \rangle \oplus \bigoplus_{i=-\infty}^{\infty} T^i F .$$

Now we show that $\Sigma_o$ has no atoms. Indeed, if A were an atom of $\Sigma_o$ and $\mu(A) > 0$, then $\varphi^k A \in \Sigma_k \subsetneq \Sigma_o$, and we have either

(i) $\varphi^k(A) = A$ for some $k \in \mathbb{N}$ or

(ii) $\varphi^k(A) \cap A = \emptyset$ for every $k \in \mathbb{N}$.

In case (i) then $A \in Tail(\alpha)$ and therefore $\mu(A) = 1$.
This implies that $P_o E$ is one-dimensional. But then also $P_k E = T^k P_o E$ is one-dimensional for every $k \in \mathbb{Z}$, and hence E is one-dimensional.

In case (ii) the sets $(\varphi^k(A))_{k \in \mathbb{N}}$ are pairwise disjoint and hence $\mu(\bigcup_{k \in \mathbb{N}} \varphi^k(A)) = \infty$ whih is also impossible.

Next we show that $\dim F = \infty$. Let $0 \neq g \in F$ and consider $B := [g \neq 0] \in \Sigma_o$. Since $\mu(B) > 0$ and $\Sigma_o$ has no atoms,

$$1_B \cdot P_o E := \{ 1_B \cdot h : h \in P_o E \}$$

is infinite-dimensional, and

$$1_B \cdot P_o E = 1_B \cdot F + 1_B \cdot P_1 E .$$

If F were finite-dimensional, then $1_B \cdot F$ would be too.

By $\langle g \cdot f_1, f_2 \rangle = \langle g, f_1 \cdot f_2 \rangle = 0$ for $f_1, f_2 \in P_1 E$ we obtain that $g \cdot P_1 E$ is contained in F. Therefore $g \cdot P_1 E$ and also $1_B \cdot P_1 E$ would be finite-dimensional which gives a contradiction.

Finally, if $(f_i)_{i \in \mathbb{Z}}$ is an orthonormal basis for F then E has an orthonormal basis of the form $\{1, T^j f_i : i, j \in \mathbb{Z}\}$. ∎

On the other hand, the "Kolmogorov-property" is no longer a "spectral" mixing property, but it can be characterized in terms of the entropy. This is the content of the following theorem which is stated without proof.

T.8 Theorem (Rohlin-Sinai, 1961):
An MDS $(X, \Sigma, \mu; \varphi)$ is a K-system if and only if $h_\mu(\alpha; \varphi) > 0$ for every $\alpha \in \tilde{\Sigma}_d$.

The Property (iv) in Definition T.1 appeared in the investigations of Ornstein and Friedman on the classification of Bernoulli shifts and leads to a still more restricted class of "mixing" MDSs.

T.9 Definition:
An MDS $(X, \Sigma, \mu; \varphi)$ is called a weak Bernoulli system if it possesses a generator that is a weak Bernoulli partition.

Remark:
Here the requirement that every $\alpha \in \tilde{\Sigma}_d$ be weak Bernoulli would be too restrictive, since even for Bernoulli shifts there are partitions which are not weak Bernoulli (see Smorodinsky [1971]).

T.10 Theorem:

For the Markov shift $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ obtained from a matrix T and invariant distribution $p \gg 0$ the following are equivalent:

(a) $T^n \longrightarrow 1 \otimes 1$ as $n \longrightarrow \infty$.

(b) $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau^n)$ is ergodic for every $n \in \mathbb{N}$.

(c) $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ is weakly mixing.

(d) $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ is weak Bernoulli.

Proof:
(d) $\Rightarrow$ (c) is clear, since weak Bernoulli implies strong mixing by Proposition T.3 and strong mixing implies weak mixing by (IX.5).
(c) $\Rightarrow$ (b) follows from (IX.D.4).
(b) $\Rightarrow$ (a): As in (IV.D.7) we can infer that $T^n$ is irreducible for every $n \in \mathbb{N}$. Then we use (IX.D.4), Remark.
(a) $\Rightarrow$ (d): Define $A_m := \{ (x_i) \in \hat{X} : x_0 = m \}$ for $m \in X = \{0, \ldots, k-1\}$.

Then $\alpha := \{A_o, \ldots, A_{k-1}\}$ is the canonical generator. If $1_m$ is the characteristic function of $\{m\} \subseteq X$, then $a_{ij} = \langle 1_i, T1_j \rangle$ and (a) implies that

$$\begin{aligned} |\hat{\mu}(\varphi^{-n} A_i \cap A_j) - \hat{\mu}(A_i) \cdot \hat{\mu}(A_j)| \\ = \Big| \sum_{i_1 \ldots i_{n-1}} p_i a_{i i_1} a_{i_1 i_2} \ldots a_{i_{n-2} i_{n-1}} a_{i_{n-1} j} - p_i p_j \Big| \\ = p_i \cdot |\langle 1_i, T^n 1_j \rangle - p_j| \end{aligned}$$

converges to zero as $n \to \infty$.

Now we show that $\alpha$ is a wB-partition:

For $\varepsilon > 0$ we find $n$ sufficiently large such that

$$|\langle 1_i, T^n 1_j \rangle - p_j| < \frac{\varepsilon}{c}$$

for every $i, j \in X$, where $c := \max \{ p_j^{-1} : j \in X \}$.

For $A, B \in \alpha^m$, say $A = \bigcap_{i=o}^{k-1} \varphi^{-i} A_{m_i}$, and analogously $B$, we define

$p_A := a_{m_o m_1} \ldots\ldots a_{m_{k-2} m_{k-1}}$, and analogously $p_B$. Then

$\hat{\mu}(A) = p_x p_A$ for some $x \in X$,

$\hat{\mu}(B) = p_y p_B$ for some $y \in X$ and

$\hat{\mu}(A \cap \varphi^{-(n+m)} B) = p_x p_A \cdot \langle 1_z, T^{n+1} 1_y \rangle \cdot p_B$ for some $z \in X$.

Thus

$$\begin{aligned} |\hat{\mu}(A \cap \varphi^{-(n+m)} B) - \hat{\mu}(A)\hat{\mu}(B)| &= p_x p_A p_B \cdot |\langle 1_z, T^{n+1} 1_y \rangle - p_y| \\ &\leq \hat{\mu}(A)\hat{\mu}(B)\, c \cdot |\langle 1_z, T^{n+1} 1_y \rangle - p_y| \\ &< \hat{\mu}(A)\hat{\mu}(B)\, \varepsilon \quad \text{, and therefore} \end{aligned}$$

$dep(\alpha^m, \tilde{\varphi}^{n+m} \alpha^m) < \varepsilon$ for every $m \in \mathbb{N}$. ■

Theorem T.10 shows that for a Markov shift all mixing properties coincide, namely weak mixing, strong mixing, the K-property and the weak Bernoulli property. This can be compared with the coincidence of different mixing properties of the matrix T (see IX.D.4 and also U.13).

The main reason why the weak Bernoulli property became famous, lies in the following result, which we cannot prove here.

T.11 <u>Theorem</u>:

Every weak Bernoulli system is isomorphic to a Bernoulli shift.

This theorem has been used by several mathematicians to show that various MDSs are isomorphic to Bernoulli shifts (e.g. Katznelson [1971], Ornstein-Weiss [1973]).
One important result of this kind is the following immediate corollary of Theorems T.10 and T.11.

T.12 Corollary:
Every weakly mixing Markov shift is isomorphic to a Bernoulli shift.

References: Friedman-Ornstein [1971], Katznelson [1971], Rohlin-Sinai [1961], Ornstein-Weiss [1973], Shields [1974], Smorodinsky [1971] .

## Appendix U: Dilation of Positive Operators

In (II.6) we constructed an MDS $(\hat{X},\hat{\Sigma},\hat{\mu};\hat{\varphi})$ starting from a positive operator

$$T : L^1(X,\Sigma,\mu) \to L^1(X,\Sigma,\mu) \text{ with } T\mathbb{1} = \mathbb{1} \text{ and } T'\mathbb{1} = \mathbb{1}.$$

Actually, in (II.6) the space $L^1(X,\Sigma,\mu)$ was just $\mathbb{R}^k$ and $\mu$ was given by a probability vector $\begin{pmatrix} p_o \\ \vdots \\ p_{k-1} \end{pmatrix}$. Then we pointed out in (II.D.9) a particularly important property of this construction: There exists a canonical imbedding of $L^1(X,\Sigma,\mu)$ into $L^1(\hat{X},\hat{\Sigma},\hat{\mu})$ relating T (and its powers) to the induced lattice isomorphism

$$\hat{T}\hat{f} := \hat{f}\circ\tau^{-1} .$$

Such "extensions" of a given operator on a given space to an isomorphism on a larger space are called "dilations". For our purposes the interesting spaces are Banach lattices of type $L^p$, and therefore we make the following definition.

U.1 <u>Definition</u>:
Let T be a positive operator on $L^p(X,\Sigma,\mu)$ for $1 \leq p < \infty$ and $\mu(X) = 1$. Assume that there exists a probability space $(\hat{X},\hat{\Sigma},\hat{\mu})$, a positive isometry J, a positive operator T and a positive contraction Q such that the diagram

$$\begin{array}{ccc} L^p(X,\Sigma,\mu) & \xrightarrow{\;T^n\;} & L^p(X,\Sigma,\mu) \\ J\downarrow & & \uparrow Q \\ L^p(\hat{X},\hat{\Sigma},\hat{\mu}) & \xrightarrow[\;\hat{T}^n\;]{} & L^p(\hat{X},\hat{\Sigma},\hat{\mu}) \end{array}$$

commutes for n = 0, 1, ... . The FDS $(\hat{T}; L^p(\hat{X},\hat{\Sigma},\hat{\mu}))$ is called a <u>positive dilation</u> of $(T; L^p(X,\Sigma,\mu))$. It is called a <u>lattice dilation</u> if T also is a Banach lattice isomorphism.

From several points of view the construction of lattice dilations seems to be important:

(i) A lattice dilation is the Banach lattice analogue of the "unitary dilations" which are a fundamental tool in the investigation of Hilbert space operators.
We refer to Sz.-Nagy-Foias 1970 .

(ii) The dilated operator $\hat{T}$ has nicer properties but is still closely related to the original operator T. In particular, $\hat{T}$ reflects the asymptotic behavior of the powers of T (see U.12 below). For a representation theorem of lattice isomorphisms on $L^p$-spaces we refer to Schaefer 1974 , III.Exercise 27.

In this appendix we extend the construction of (II.6) from $L^1$-spaces to $L^p$-spaces for $1 \leq p < \infty$ and from positive operators satisfying $T\mathbb{1} = \mathbb{1}$ and $T'\mathbb{1} = \mathbb{1}$ to arbitrary positive contractions. To do this we proceed in several steps and construct several different (positive and lattice) dilations which will be composed to yield the final lattice dilation.
In the following we always consider $E := L^p(X, \Sigma, \mu)$ for some finite measure space and for $1 \leq p < \infty$. The notation $0 \ll u \in E$ indicates that the function u is strictly positive, i.e. $u(x) > 0$ a.e. on X. The dual $E'$ will be canonically identified with $L^q(X, \Sigma, \mu)$, where $\frac{1}{p} + \frac{1}{q} = 1$. Finally, the operator $T \in \mathcal{L}(E)$ is assumed to be a positive contraction.

U.2 <u>Proposition</u> (first dilation):
Assume there exist $0 \ll u \in E$ and $0 \ll v \in E$ such that $Tu \leq v$ and $T'v^{p-1} \leq u^{p-1}$. Then $(E;T)$ has a <u>positive dilation</u> $(\hat{E};\hat{T})$ with $0 \ll \hat{v} \in \hat{E}$ such that $\hat{T}'\hat{v}^{p-1} = \mathbb{1}$ and $\hat{T}\mathbb{1} = \hat{v}$ .

Proof:
Define the new measure space $(\hat{X}, \hat{\Sigma}, \hat{\mu})$ by adding two different points, i.e. $\hat{X} := X \cup \{y, z\}$ , and $\hat{\mu} := u^p \cdot \mu + \alpha \cdot \delta_y + \beta \cdot \delta_z$, where $\delta_y$, $\delta_z$ are the Dirac measures in y, z and

$$\alpha := (2^p - 1)^{-1} \cdot \langle u, u^{p-1} - T'v^{p-1} \rangle$$

$$\beta := (1 - 2^{-p})^{-1} \cdot \langle v - Tu, v^{p-1} \rangle \quad .$$

In the following we identify a function f on X with the function on $\hat{X}$ which is zero on y and z and coincides with f on X. Then we define a positive isometry

$$J : E \longrightarrow \hat{E} := L^p(\hat{X}, \hat{\Sigma}, \hat{\mu})$$

by $$Jf := u^{-1} \cdot f$$

and a positive contraction

$$Q : \hat{E} \longrightarrow E \qquad \text{by} \qquad Q\hat{f} := \hat{f} \cdot u .$$

For the element $\hat{f} := f + \eta\, 1_y + \varphi\, 1_z \in \hat{E}$, $f \in L^p(X, \Sigma, u^p\mu)$ and $\eta, \varphi \in \mathbb{R}$, we define the operator $\hat{T}$ by

$$\begin{aligned}
\hat{T}\hat{f} &:= u^{-1}T(uf) + \beta^{-1}\langle \hat{f}, 1_z\rangle \,(u^{-1}(v - Tu) + 2^{-1}1_z) \\
&\quad + 2^{1-p}\,\alpha^{-1}\langle \hat{f}, 1_y + 1_X - u^{1-p}T'v^{p-1}\rangle\, 1_y \\
&= u^{-1}T(uf) + \varphi\,(u^{-1}(v - Tu) + 2^{-1}1_z) \\
&\quad + 2^{1-p}(\eta + \alpha^{-1}\int_X f(1_X - u^{1-p}T'v^{p-1})\,d\hat{\mu})\,1_y
\end{aligned}$$

where we may assume that $0 \neq \alpha, \beta$ .

Then we put $\hat{v} := u^{-1}v + 2\cdot 1_y + \frac{1}{2}\, 1_z$ and obtain

$$\begin{aligned}
\hat{T}\,1 &= u^{-1}Tu + u^{-1}(v - Tu) + \tfrac{1}{2}\,1_z + 2^{1-p}\,\alpha^{-1}(\int_X u(u^{p-1} - T'v^{p-1})\,d\mu + \alpha)1_y \\
&= u^{-1}v + 2^{1-p}\,\alpha^{-1}(\alpha\,(2^p - 1) + \alpha)\,1_y + \tfrac{1}{2}\,1_z \\
&= u^{-1}v + 2\cdot 1_y + \tfrac{1}{2}\,1_z = \hat{v}
\end{aligned}$$

and

$$\begin{aligned}
\langle \hat{T}\hat{f}, \hat{v}^{\,p-1}\rangle &= \int_{\hat{X}} u^{-1}T(uf)\cdot u^{1-p}v^{p-1}\,d\hat{\mu} + \varphi \int_{\hat{X}} u^{-1}(v - Tu)u^{1-p}v^{p-1}\,d\hat{\mu} \\
&\quad + \varphi\,\beta\cdot 2^{-p} + \eta\cdot\alpha + \int_{\hat{X}} f(1_X - u^{1-p}T'v^{p-1})\,d\hat{\mu} \\
&= \int_X T(uf)v^{p-1}\,d\mu + \varphi\int_X (v - Tu)v^{p-1}\,d\mu \\
&\quad + \varphi\cdot\beta\cdot 2^{-p} + \eta\cdot\alpha + \int_X uf(u^{p-1} - T'v^{p-1})\,d\mu \\
&= \langle uf, T'v^{p-1}\rangle + \varphi\cdot\beta + \eta\cdot\alpha + \langle uf, u^{p-1} - T'v^{p-1}\rangle \\
&= \int_{\hat{X}} \hat{f}\,d\hat{\mu} \\
&= \langle \hat{f}, 1_{\hat{X}}\rangle .
\end{aligned}$$

Finally, it is easy to verify that $T^n f = Q\,\hat{T}^n J f$ for $n = 0, 1, \ldots$ . ∎

The above dilation is of a preparatory nature, and allows us to obtain the desired lattice dilation by a construction similar to that in (II.6).

U.3 Proposition (second dilation):
Assume that there exists $0 \ll v \in E$ such that $T1 = v$ and $T'v^{p-1} = 1$. Then $(E;T)$ has a lattice dilation $(\hat{E};\hat{T})$.

Proof:
Define $\hat{X} := X^{\mathbb{Z}}$ with product $\sigma$-algebra $\hat{\Sigma}$. We denote by $\tau$ the (left) shift on $\hat{X}$. With the projection onto the i-th coordinate

$$\pi_i : \begin{cases} \hat{X} \longrightarrow X \\ (x_j)_{j\in\mathbb{Z}} \longmapsto x_i \end{cases}$$

we obtain $\pi_i = \pi_0 \circ \tau^i$ for $i \in \mathbb{Z}$.
Next we define positive operators

$$S_+, S_- : L^\infty(X,\Sigma,\mu) \longrightarrow L^\infty(X,\Sigma,\mu)$$

by $S_+f := v^{-1}Tf$ and $S_-f := T'(v^{p-1}f)$. Then $S_+1 = S_-1 = 1$ and in view of (A.17), Theorem, we can define a measure $\hat{\mu}$ on $(\hat{X},\hat{\Sigma})$ by

$$\hat{\mu}\left(\prod_{i=-n}^{n} f_i\circ\pi_i\right) := \mu\left(f_0\left(\prod_{i=-1}^{-n} S_-M_{f_i}\right)1\left(\prod_{i=1}^{n} S_+M_{f_i}\right)1\right)$$

where $M_{f_i}g := f_i\cdot g$ for $f_i \in L^\infty(X,\Sigma,\mu)$.
Now take

$$J : \begin{cases} E \longrightarrow \hat{E} := L^p(\hat{X},\hat{\Sigma},\hat{\mu}) \\ f \longmapsto f\circ\pi_0 \end{cases}$$

which is a lattice homomorphism with $J\,1_X = 1_{\hat{X}}$. From $\int_{\hat{X}} Jf\,d\hat{\mu} = \int_X f\,d\mu$ follows that J is even an isometry from $L^p(X,\Sigma,\mu)$ into $L^p(\hat{X},\hat{\Sigma},\hat{\mu})$ for every $1 \le p < \infty$. Consequently, we can consider J on $L^q(X,\Sigma,\mu)$ and its adjoint

$$Q := J' : L^p(\hat{X},\hat{\Sigma},\hat{\mu}) \longrightarrow L^p(X,\Sigma,\mu)$$

is a positive contraction.
Finally we define

$$\hat{T} : \hat{E} \longrightarrow \hat{E}$$

by

$$\hat{T}\hat{f} := Jv\cdot\hat{f}\circ\tau \quad .$$

Then $\hat{T}$ is a lattice homomorphism with dense image. It is even a Banach lattice isomorphism since it acts isometrically on the elements $\hat{f} := \prod_{i=-n}^{n} f_i \circ \pi_i$, $f_i \in L^\infty(X,\Sigma,\mu)$ : Denote $g_i := |f_i|^p$ . Then

$$\int_{\hat{X}} |\hat{f}|^p \, d\hat{\mu} = \int_X \prod_{i=-1}^{-n} S_- M_{g_i} 1\!\!1 \cdot g_o \cdot \prod_{i=1}^{n} S_+ M_{g_i} 1 \, d\mu$$

$$= \int_X T'(v^{p-1} \cdot g_{-1} \cdot \prod_{i=-2}^{-n} S_- M_{g_i} 1\!\!1) \cdot g_o \cdot \prod_{i=1}^{n} S_+ M_{g_i} 1\!\!1 \, d\mu$$

$$= \int_X v^{p-1} g_{-1} \prod_{i=-2}^{-n} S_- M_{g_i} 1\!\!1 \cdot T(g_o \cdot \prod_{i=1}^{n} S_+ M_{g_i} 1\!\!1) \, d\mu$$

$$= \int_X v^p g_{-1} \cdot \prod_{i=-2}^{-n} S_- M_{g_i} 1\!\!1 \cdot \prod_{i=o}^{n} S_+ M_{g_i} 1\!\!1 \, d\mu$$

$$= \int_{\hat{X}} J v^p \cdot |\hat{f} \circ \tau|^p \, d\hat{\mu}$$

$$= \int_{\hat{X}} |\hat{T} \hat{f}|^p \, d\hat{\mu} .$$

With all these definitions in mind we can show that $\hat{T}$ is in fact a dilation of T:

Choose $f, g \in L^\infty(X,\Sigma,\mu)$ and $n \in \mathbb{N}_o$ . Then

$$\langle Q\hat{T}^n Jf, g \rangle = \langle \hat{T}^n Jf, Jg \rangle$$

$$= \langle \prod_{i=o}^{n-1} v \circ \pi_i \cdot f \circ \pi_n, g \circ \pi_o \rangle$$

$$= \int_X v \cdot (S_+ M_v)^{n-1} S_+ M_f 1 \cdot g \, d\mu$$

$$= \int_X g (M_v S_+)^n f \, d\mu$$

$$= \langle T^n f, g \rangle . \blacksquare$$

Remark:

For $v = 1\!\!1$ the above construction coincides with the usual construction of the Markov processes from "transition" operators.

As soon as the assumptions of (U.2) are satisfied we obtain a lattice dilation simply as a composition of the "first" and "second" dilations. For $p = 1$ and a positive contraction T it is easy to choose $0 \ll u \in E$ and $0 \ll v \in E$ satisfying $Tu \le v$ and $T' v^{p-1} = T' 1 \le 1 = u^{p-1}$ .

Therefore we have proved the first main result.

U.4 <u>Theorem</u> ($L^1$ - dilation):
Every positive contraction on $L^1(X, \Sigma, \mu)$ has a lattice dilation.

The question of whether every positive contraction on $L^p(\mu)$, $1 < p < \infty$, possesses a lattice dilation is more delicate. But at least for positive contractions on finite dimensional $L^p$-spaces we are able (see U.8 below) to verify the assumptions of (U.2). Again, by composition of (U.2) and (U.3) we obtain the desired result.

U.5 <u>Theorem</u> (finite - dimensional $L^p$ - dilation):

Every positive contraction on a finite-dimensional $L^p$-space has a lattice dilation.

Remarks:

1. In general, the assumptions of (U.2) are not satisfied (compare Akcoglu - Kopp [1977]). But by approximating $L^p(X, \Sigma, \mu)$ by finite-dimensional sublattices and using (U.5) and ultra-power techniques we are able to construct a lattice dilation for every positive $L^p$ - contraction (see Akcoglu - Sucheston [1977]). In addition, for many applications, the finite-dimensional case is sufficient (see App.V.6 and App.V.7).
2. The lattice dilation constructed above is not unique, but at least the "second" dilation has some remarkable properties (see Kern-Nagel-Palm [1977]).

It remains to prove that in fact, every positive contraction on a finite-dimensional $L^p$-space satisfies the assumptions of (U.2). This will be achieved in (U.8) and in the course of the proof we shall freely use lattice-theoretical notation and arguments. In particular, $f \perp g$ means $|f| \wedge |g| = 0$ and $\{f\}^{\perp} := \{g \in L^p : f \perp g\}$ is the band orthogonal to f.

U.6 <u>Lemma</u>:

Let T be a positive contraction on $L^p(X, \Sigma, \mu)$, $1 < p < \infty$.
If there exists $0 < f \in L^p$ satisfying $\|Tf\| = \|f\|$, then

(i) $T'h^{p-1} = f^{p-1}$ for $h := Tf$ .
(ii) $g \perp f$ implies $Tf \perp Tg$ .
(iii) $g \perp Tf$ implies $T'g \perp f$ .

Proof:
(i) For $\|f\| = 1$ we have
$\langle f, f^{p-1}\rangle = \|f\|^p = \|Tf\|^p = \langle Tf, (Tf)^{p-1}\rangle = \langle f, T'h^{p-1}\rangle$ .
Since $\|T'h^{p-1}\|_q \leq 1$ and $\|f^{p-1}\|_q = 1$ this implies $T'h^{p-1} = f^{p-1}$ .
(ii) By the positivity of T we obtain
$0 \leq \langle |Tg|, (Tf)^{p-1}\rangle \leq \langle T|g|, h^{p-1}\rangle = \langle |g|, f^{p-1}\rangle = 0$ .
(iii) follows similarily. ■

Now we proceed as follows:
If $L^p(\mu)$ is finite-dimensional and T has norm one, then we can find $0 \leq f, h$ with $\|Tf\| = \|f\|$ and $T'h^{p-1} = f^{p-1}$ (apply U.6).
We shall try to add successively more positive components to the
klein vectors f and h until we obtain strictly positive vectors $u \geq f$ and $v \geq h$ satisfying $Tu \leq v$ and $T'v^{p-1} \leq u^{p-1}$. To do this we need another lemma.

U.7 <u>Lemma</u>:
Let T be a positive contraction with norm one on a finite-dimensional space $L^p(X, \Sigma, \mu)$, $1 < p < \infty$ . Choose $0 < f \in L^p(\mu)$ satisfying $\|Tf\| = \|f\|$ . If

$$\{f\}^{\perp} \neq \{0\} \neq \{Tf\}^{\perp} ,$$

then there exists a contraction $S \in \mathcal{L}(L^p(\mu))$ with $T \leq S$, and $0 < f_o \in L^p(\mu)$ satisfying $\|Sf_o\| = \|f_o\|$ and $\{f_o\}^{\perp} \subsetneqq \{f\}^{\perp}$ .

Proof:
Define $c := \sup\{\|Tg\| : \|g\| \leq 1, g \in \{f\}^{\perp}\}$ . If $c \neq 0$, we can find $0 < g_o \in \{f\}^{\perp}$ with $\|g_o\| = 1$ and $\|Tg_o\| = c$. Then we define

$$S(f_1 + f_2) := Tf_1 + c^{-1} \cdot Tf_2$$

where $f_1 \in \{f\}^{\perp\perp}$ , $f_2 \in \{f\}^{\perp}$ and we take $f_o := f + g_o$. Then $0 < T \leq S$ and $\|Sf_o\|^p = \|Tf\|^p + \|c^{-1}Tg_o\|^p = \|f\|^p + 1 = \|f_o\|^p$.

If $c = 0$, we take $0 < g_1, g_2 \in \{f\}^{\perp}$ and $0 < g_2'$ in the dual of $\{f\}^{\perp}$ satisfying $\|g_2\| = \|g_2'\| = 1$ and $\langle g_1, g_2' \rangle = \|g_1\|$. Now we define $f_o := f + g_1$ and $S := T + g_2 \otimes g_2'$ which have the desired properties. ∎

U.8 Proposition:

If T is a positive contraction on a finite-dimensional space $L^p(X, \Sigma, \mu)$, $1 < p < \infty$, then there exist strictly positive functions $0 \ll u, v \in L^p(\mu)$ such that $Tu \leq v$ and $T' v^{p-1} \leq u^{p-1}$.

Proof:

Applying (U.7) a finite number of times we eventually obtain an operator $S \geq T$ and $0 < g$ with $\|Sg\| = \|g\|$ such that either $\{g\}^{\perp} = \{0\}$ or $\{Sg\}^{\perp} = \{0\}$. If $0 \ll g$, i.e. $\{g\}^{\perp} = \{0\}$, then we take $u := g$ and $v := Sg + h \gg 0$, where $h \perp Sg$.

If $0 \ll Sg$, i.e. $\{Sg\}^{\perp} = \{0\}$, then we take $v := Sg$ and $u := g+h \gg 0$, where $h \perp g$. In both cases it is easy to verify the desired properties (use U.6). ∎

In the final part of this appendix we shall verify the statement that our lattice dilation reflects the asymptotic behavior of the original operator. More precisely, we show that T and $\hat{T}$ possess the same ergodic properties such as ergodicity or mixing. To do this we have to restrict our attention to positive operators $T : L^1(X,\Sigma,\mu) \longrightarrow L^1(X,\Sigma,\mu)$ satisfying $T\mathbb{1} = \mathbb{1}$ and $T'\mathbb{1} = \mathbb{1}$, so-called bi-Markov or doubly stochastic operators. Then we can apply the construction of (U.3) with $v = \mathbb{1}$ and obtain a lattice dilation $(L^1(\hat{X}, \hat{\Sigma}, \hat{\mu}); T)$. Moreover, the operators $S_+$, $S_-$ appearing in the construction are simply T and $T'$, respectively, and $\hat{T}$ is the operator induced by the left shift $\tau$ on $\hat{X}$. In fact, the construction in (U.3) coincides with the construction of the Markov shift in (II.6).

We start with some observations: By the Riesz convexity theorem (see Schaefer, [1974], V.8.2), T induces a contraction on $L^2(X, \Sigma, \mu)$ as does

$$T_\lambda := \bar{\lambda}\, T \qquad \text{for } |\lambda| = 1 .$$

By (IV.5), $T_\lambda$ is mean ergodic, hence the Cesaro means $(T_\lambda)_n$ converge strongly to the orthogonal projection $P_\lambda$ onto the fixed space $F(T_\lambda)$ of $T_\lambda$ which is the eigenspace of T corresponding to $\lambda$. Since $T_\lambda$ is contractive, it follows from Hilbert space arguments that $F(T_\lambda) = F(T_\lambda')$. Therefore the Cesaro means $(T_\lambda')_n$ also converge to $P_\lambda$.

Analogous considerations are valid for $\hat{T}$, which gives rise to a unitary operator on $L^2(\hat{X}, \hat{\Sigma}, \hat{\mu})$. Hence the Cesaro means $(\hat{T}_\lambda)_n$ and $(\hat{T}_\lambda^{-1})_n$ converge strongly to the projection $\hat{P}_\lambda$ onto the eigenspace of $\hat{T}$ corresponding to $\lambda$. It is our first aim to establish a relation between $P_\lambda$ and $\hat{P}_\lambda$. From $Q\hat{T}^n J = T^n$ and the strong convergence of the Cesaro means we obtain $Q\hat{P}_\lambda J = P_\lambda$. But it is more interesting that $\hat{P}_\lambda$ can be calculated from $P_\lambda$.

U.9 <u>Lemma</u>:

Define $\hat{E}^+$ (resp. $\hat{E}^-$) as the closed sublattice of $L^2(\hat{X}, \hat{\Sigma}, \hat{\mu})$ generated by all functions $\prod_{i=0}^{n} f_i \circ \pi_i$ (resp. $\prod_{i=0}^{-n} f_i \circ \pi_i$) where $n \in \mathbb{N}$ and $f_i \in L^\infty(X, \Sigma, \mu)$. Then we have

$$\hat{E}^+ \cap \hat{E}^- = J(L^2(X, \Sigma, \mu)) .$$

Proof:

One inclusion is trivial. For the other we observe that $Q(\hat{f}\hat{g}) = Q\hat{f} \cdot Q\hat{g}$ if $\hat{f} \in \hat{E}^+ \cap L^\infty(\hat{X}, \hat{\Sigma}, \hat{\mu})$ and $\hat{g} \in \hat{E}^- \cap L^\infty(\hat{X}, \hat{\Sigma}, \hat{\mu})$. Now take $\hat{f} \in \hat{E}^+ \cap \hat{E}^- \cap L^\infty(\hat{X}, \hat{\Sigma}, \hat{\mu})$. Then $JQ(|\hat{f}|^2) = |JQ\hat{f}|^2$ and therefore

$$\int |\hat{f}|^2 \, d\hat{\mu} = \int Q|\hat{f}|^2 \, d\hat{\mu} = \int JQ|\hat{f}|^2 \, d\hat{\mu} = \int |JQ\hat{f}|^2 \, d\hat{\mu}.$$

Since JQ is the orthogonal projection from $L^2(\hat{X}, \hat{\Sigma}, \hat{\mu})$ onto $J(L^2(X, \Sigma, \mu))$ we conclude that $\hat{f} = JQ\hat{f}$. ∎

U.10 <u>Proposition</u>: $\hat{P}_\lambda = J P_\lambda Q$ for any $\lambda \in \Gamma$.

Proof:

For $\hat{f} := \prod_{i=-k}^{k} f_i \circ \pi_i$ and $n > k$ we have $\hat{T}^n \hat{f} \in \hat{E}^+$ and $\hat{T}^{-n}\hat{f} \in \hat{E}^-$, and therefore

$$\hat{P}_\lambda \hat{f} = \lim_{n\to\infty} (\hat{T}_\lambda)_n \hat{f} = \lim_{n\to\infty} (\hat{T}_\lambda^{-1})_n \hat{f} \in \hat{E}^+ \wedge \hat{E}^- = J(L^2(X,\Sigma,\mu)) .$$

Thus $\hat{P}_\lambda = JQ\hat{P}_\lambda JQ = JP_\lambda Q$ . ∎

Using this relation between $P_\lambda$ and $\hat{P}_\lambda$ one shows easily that $P_\lambda f = f$ (resp. $\hat{P}_\lambda \hat{f} = \hat{f}$) implies $\hat{P}_\lambda Jf = Jf$ (resp. $P_\lambda Q\hat{f} = Q\hat{f}$), i.e. J is an isomorphism from the $\lambda$-eigenspace of T onto the $\lambda$-eigenspace of $\hat{T}$ (with inverse induced by Q). We collect this information in the following theorem and refer to Lecture VII for the spectral-theoretical background.

U.11 <u>Theorem</u>:

Let T be a bi-Markov operator on $L^1(X,\Sigma,\mu)$ and let $(L^1(\hat{X},\hat{\Sigma},\hat{\mu});\hat{T})$ be the corresponding lattice dilation as in (U.3). Denote by G (resp. $\hat{G}$) the closed subspace in $L^2(X,\Sigma,\mu)$ (resp. $L^2(\hat{X},\hat{\Sigma},\hat{\mu})$) generated by all eigenvectors of T (resp. $\hat{T}$) with unimodular eigenvalues. Then $J(G) = \hat{G}$ and the FDS $(G; T|_G)$ is isomorphic to $(\hat{G}; \hat{T}|_{\hat{G}})$ .

Another way of expressing the result is by saying that T and its dilation $\hat{T}$ possess the same unimodular eigenvalues with the same multiplicities. Since ergodicity and weak mixing of T and $\hat{T}$ are determined by the eigenvalues (see IX.1 and IX.5) we have the desired result.

U.12 <u>Corollary</u>:

The FDS $(L^1(X,\Sigma,\mu);T)$, T bi-Markov, is irreducible (resp. weakly mixing, resp. strongly mixing) if and only if its lattice dilation $(L^1(\hat{X},\hat{\Sigma},\hat{\mu});\hat{T})$ is irreducible (resp. weakly mixing, resp. strongly mixing).

Proof:

After the previous considerations only the strong mixing case needs an additional argument: Take

$$\hat{f} := \prod_{i=-n}^{n} f_i \circ \pi_i$$

and $$\hat{g} := \prod_{i=-m}^{m} g_i \circ \pi_i \qquad \text{for } f_i, g_i \in L^\infty(X,\Sigma,\mu).$$

For $k > n + m$ we observe that $\hat{T}^{-m}\hat{g} \in \hat{E}^-$, $\hat{T}^n\hat{f} \in \hat{E}^+$ and $\hat{T}^{k-m}\hat{f} \in \hat{E}^+$. Therefore

$$\langle \hat{T}^k\hat{f},\hat{g}\rangle = \langle \hat{T}^{k-m}\hat{f},\hat{T}^{-m}\hat{g}\rangle = \int_X Q(\hat{T}^{k-m}\hat{f}\ \hat{T}^{-m}\hat{g})\, d\mu = \int_X Q\hat{T}^{k-m}\hat{f}\cdot Q\hat{T}^{-m}\hat{g}\, d\mu$$
$$= \int_X T^{k-m-n}Q\hat{T}^n\hat{f}\cdot Q\hat{T}^n\hat{g}\, d\mu$$
$$= \langle T^{k-m-n}\tilde{f}, \tilde{g}\rangle \quad \text{where } \tilde{f} := Q\hat{T}^n\hat{f} \text{ and } \tilde{g} := Q\hat{T}^n\hat{g}\ .$$

Hence the weak convergence of $T^n$ is equivalent to the weak convergence of $\hat{T}^n$ . ∎

As an application, we show that for a Markov shift with finite state space (see II.6), the different mixing properties coincide.

U.13 <u>Corollary</u>:

Let $(\hat{X}, \hat{\Sigma}, \hat{\mu}\ ; \tau)$ be the Markov shift generated by the transition matrix T. The following conditions are equivalent:

(a) $\tau^k$ is ergodic for all $k \in \mathbb{N}$.
(b) $\tau$ is weakly mixing.
(c) $\tau$ is strongly mixing.

Proof:

The Markov shift, resp. $(L^1(\hat{X}, \hat{\Sigma}, \hat{\mu}); \hat{T})$, is the lattice dilation of the positive contraction T on some finite-dimensional $L^1(\mu)$. Therefore, the assertion follows from (U.12) and the Proposition in (IX.D.4).

References: Akcoglu [1975a], Akcoglu - Kopp [1977], Akcoglu - Sucheston [1977], Kern - Nagel - Palm [1977], Nagel - Palm [1982], Peller [1978].

## Appendix V: Akcoglu's Individual Ergodic Theorem

It is not difficult to show that the linear operator $T_\varphi$ induced by an MDS $(X,\Sigma,\mu;\varphi)$ is mean ergodic on $L^2(X,\Sigma,\mu)$. In fact the original v.Neumann mean ergodic theorem has been extended to much more general situations (see App.Y). In contrast, the proof of the individual ergodic theorem, even in its classical form, requires considerable effort and fewer generalizations have been obtained. For instance, no complete and satisfactory answer to the following question is available:

Let T be a contraction on $L^p(X,\Sigma,\mu)$, $1<p<\infty$. Then T is always mean ergodic (see IV.5). Characterize classes of operators T which are also individually ergodic !

It was soon conjectured that the positivity of T might somehow be essential for individual ergodicity. A.Jonescu-Tulcea [1964] succeeded in showing that positive isometries on $L^p(\mu)$, $1<p<\infty$, are individually ergodic. Thereafter, many partial results were obtained, but the final (positive) answer was given only after Akcoglu [1975b] constructed lattice dilations for positive constructions on $L^p(\mu)$.

In this appendix we present a complete proof of the individual ergodic theorem for positive contractions on reflexive $L^p$-spaces. As in the proof of (V.3), it is easy to obtain a.e.-convergence of the Cesaro means on a dense subset of $L^p(\mu)$. The difficulty consists in establishing some "equicontinuity" of the Cesaro means for a.e.-convergence. The "dominated estimate" of (App.V.3), however, does just that.

We shall be considering throughout the following situation:
Let $(X,\Sigma,\mu)$ be a probability space. Choose $1<p<\infty$ and consider $E = L^p(X,\Sigma,\mu)$. Let $T\in\mathcal{L}(E)$ be a positive operator satisfying $\|T\|\leq 1$. As before we denote by

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

its Cesaro means, and we call T individually ergodic if $T_nf$ converges a.e. for every $f\in E$.

V.1 Lemma:

There exists a dense subspace $E_o$ in E such that $T_n f$ converges a.e. for every $f \in E_o$.

Proof:

The contraction T is mean ergodic by (IV.5). Therefore

$$E_o := F + (Id - T)E$$

is dense in E, where F denotes the fixed space of T. Since the assertion is trivial on F, we choose $f := g - Tg$ for some $g \in E$. By (IV.3.0) it suffices to show that $\frac{1}{n} T^n g$ converges a.e. to 0. For $0 \leq g \in E$ this is implied by the following estimate:

$$\int \sum_{n=1}^{\infty} (\tfrac{1}{n}T^n g)^p \, d\mu = \sum_{n=1}^{\infty} \int (\tfrac{1}{n}T^n g)^p \, d\mu = \sum_{n=1}^{\infty} \|\tfrac{1}{n}T^n g\|^p \leq (\sum_{n=1}^{\infty} \frac{1}{n^p}) \|g\|^p < \infty .$$

Now, the assertion holds for every $g \in E$, since

$$|T^n g| \leq T^n |g|,$$

T being positive. ■

V.2 Definition:

$T \in \mathcal{L}(E)$ satisfies a dominated estimate if there exists $0 < M$ such that

$$\|f^*\| \leq M \cdot \|f\|$$

for $f \in E$ and $f^* := \sup \{T_k f : k \in \mathbb{N}\}$.

Remark:

It suffices to require that for every $n \in \mathbb{N}$

$$\|f_n^*\| \leq M \cdot \|f\|$$

for $f_n^* := \sup \{T_k f : 1 \leq k \leq n\}$.

V.3 Lemma:

Every positive contraction $T \in \mathcal{L}(E)$ satisfying a dominated estimate is individually ergodic.

Proof:

We take $f \in E$ and show that

$$h_f(x) := \lim_{n,m \in \mathbb{N}} \sup |T_n f(x) - T_m f(x)| = 0 \text{ for almost all } x \in X.$$

First, we observe that $h_f = h_{f-f_o}$ and $h_{f-f_o} \leq 2 \cdot |f - f_o|^*$ for every $f_o$ contained in the dense subspace $E_o$ of Lemma (App.V.1). For $\varepsilon > 0$ we choose $f_o \in E_o$ such that

$$\frac{2M}{\varepsilon} \int |f - f_o| \, d\mu \quad < \quad \varepsilon$$

and, by the dominated estimate,

$$\frac{2}{\varepsilon} \int |f - f_o|^* \, d\mu \quad < \quad \varepsilon.$$

This implies that

$$\mu[h_f > \varepsilon] \;=\; \mu[h_{f-f_o} > \varepsilon] \;\leq\; \mu\left[|f - f_o|^* > \tfrac{\varepsilon}{2}\right] < \varepsilon,$$

and $h_f = 0$ almost everywhere. ∎

We are now going step by step to extend the class of operators on E satisfying a dominated estimate.

V.4 Lemma:

The shift operator

$$T : (x_n)_{n \in \mathbb{Z}} \longmapsto (x_{n+1})_{n \in \mathbb{Z}}$$

on $l^p(\mathbb{Z})$ satisfies a dominated estimate with $M = \frac{p}{p-1}$ .

Proof:

For $0 \leq f := (x_k)_{k \in \mathbb{Z}} \in l^p$ we write

$$y_m := \sup \left\{ x_m, \tfrac{1}{2}(x_m + x_{m+1}), \ldots, \tfrac{1}{n}(x_m + x_{m+1} + \ldots + x_{m+n-1}) \right\}$$

for the m-th coordinate of $f_n^*$, $n \in \mathbb{N}$ fixed. Choose $\gamma > 0$ and consider $A_{n,\gamma}(f) := [f_n^* > \gamma]$ where the sequence $(a_\gamma(k))_{k \in \mathbb{Z}}$ denotes the characteristic function of $A_{n,\gamma}$. By the maximal ergodic lemma (V.5) we have

$$\sum_{k \in \mathbb{Z}} a_\gamma(k) \leq \frac{1}{\gamma} \sum_{k \in \mathbb{Z}} x_k \, a_\gamma(k)$$

or

$$\gamma^{p-1} \sum_{k \in \mathbb{Z}} a_\gamma(k) \leq \gamma^{p-2} \sum_{k \in \mathbb{Z}} x_k \, a_\gamma(k) .$$

Integrating the left hand expression yields

$$\int_0^\infty \gamma^{p-1} \sum_{k \in \mathbb{Z}} a_\gamma(k) \, d\gamma = \sum_{k \in \mathbb{Z}} \int_0^\infty \gamma^{p-1} a_\gamma(k) \, d\gamma = \sum_{k \in \mathbb{Z}} \int_0^{y_k} \gamma^{p-1} d\gamma = \frac{1}{p} \sum_{k \in \mathbb{Z}} y_k^p = \frac{1}{p} \|(y_k)\|_p^p,$$

while from the right hand term we obtain

$$\int_0^\infty \gamma^{p-2} \sum_{k\in\mathbb{Z}} x_k\, a_\gamma(k)\, d\gamma = \sum_{k\in\mathbb{Z}} \int_0^\infty \gamma^{p-2}\, a_\gamma(k)\, d\gamma = \sum_{k\in\mathbb{Z}} x_k \int_0^{1/k} \gamma^{p-2}\, d\gamma$$

$$= \frac{1}{p-1} \sum_{k\in\mathbb{Z}} x_k\, y_k^{p-1}$$

$$\leq \frac{1}{p-1} \|(x_k)_{k\in\mathbb{Z}}\|_p \cdot \|(y_k^{p-1})_{k\in\mathbb{Z}}\|_q, \text{ by Hölder's inequality,}$$

$$= \frac{1}{p-1} \|(x_k)_{k\in\mathbb{Z}}\|_p \cdot \|(y_k)_{k\in\mathbb{Z}}\|_p^{p-1} .$$

Therefore a dominated estimate holds for positive $f \in l^p$ and $M = \frac{p}{p-1}$ . For arbitrary $f \in l^p$ the assertion follows since $\|f\| = \| |f| \|$ and $\|f_n^*\| \leq \| |f_n|^* \|$ . ■

V.5 Lemma:

Every Banach lattice isomorphism T on $L^p(X,\Sigma,\mu)$ satisfies a dominated estimate with $M = \frac{p}{p-1}$ .

Proof:

With the notation introduced above and for $0 \leq f \in L^p(\mu)$ and $n \in \mathbb{N}$ fixed we obtain

$$(T^k f)_n^* = T^k(f_n^*),$$

and

$$\int (T^k f)_n^{*\,p}\, d\mu = \int (T^k f_n^*)^p\, d\mu = \int f_n^{*p}\, d\mu ,$$

since T is a lattice homomorphism and an $L^p$-isometry.

Therefore $$\int f_n^{*p}\, d\mu = \frac{1}{2N+1} \sum_{k=-N}^{N} \int (T^k f)_n^{*\,p}\, d\mu .$$

For $\varepsilon > 0$ and $x \in X$ define

$$x_k := \begin{cases} T^k f(x) & \text{for } |k| \leq N \\ 0 & \text{otherwise} \end{cases} ,$$

where N is such that $\frac{2N+2n+1}{2N+1} < 1 + \varepsilon$ . Applying (App.V.4) to $(x_k)_{k\in\mathbb{Z}} \in l^p(\mathbb{Z})$ we obtain

$$\int f_n^{*p}\, d\mu \leq \frac{1}{2N+1} \left(\frac{p}{p-1}\right)^p \int \sum_{k=-N-n}^{N+n} (T^k f(x))^p\, d\mu \leq (1+\varepsilon) \left(\frac{p}{p-1}\right)^p \int f^p\, d\mu .$$

■

Remark:

It suffices to assume that T is a positive isometry on $L^p(\mu)$ (use C.4).

The possibility of "dilating" a positive contraction on $L^p(\mu)$ to a Banach lattice isomorphism enlarges considerably the class of operators admitting a dominated estimate.

V.6 Lemma:
Every positive operator on $L^p$ having a lattice dilation as in App. U admits a dominated estimate with $M = \frac{p}{p-1}$.

Proof:
With the notation of (U.1) we have $T^k = Q \cdot \hat{T}^k \cdot J$. Moreover, $\hat{T}$ admits a dominated estimate with $\frac{p}{p-1}$ by Lemma (App.V.5). Therefore

$$f_n^* = \sup\{Q\hat{T}_k Jf : 1 \leq k \leq n\} \leq Q(\sup\{\hat{T}_k Jf : 1 \leq k \leq n\}) = Q((Jf)_n^*),$$

and

$$\|f_n^*\| \leq \|Q((Jf)_n^*)\| \leq \|(Jf)_n^*\| \leq \frac{p}{p-1}\|Jf\| = \frac{p}{p-1}\cdot\|f\| . \quad \blacksquare$$

The above simple observation and the ingenious construction of lattice dilations were the essential achievements by which M.A.Akcoglu in 1975 was able to prove his individual ergodic theorem for positive $L^p$-contractions. But on the basis of the dilation results presented in Appendix U there remains to surmount one further difficulty. In fact, not every positive $L^p$-contraction satisfies the assumptions necessary for our construction of a lattice dilation (see Remark 1, following App.U.5). Fortunately, the finite-dimensional dilation theorem as presented in (U.5) is sufficient.

V.7 Lemma:
Assume that every positive contraction on a finite-dimensional $L^p$-space satisfies a dominated estimate with $\frac{p}{p-1}$. Then the same holds for every positive contraction $T \in \mathcal{L}(L^p(X,\Sigma,\mu))$.

Proof:
We may assume that $L^p(X,\Sigma,\mu)$ and therefore the measure algebra is separable (If not, consider a separable, T-invariant closed sublattice). Then there exist finite subalgebras $\Sigma_n \subset \Sigma_{n+1} \subset \ldots \subset \Sigma$ and positive and contractive projections

$$P_{(n)} : L^p(X,\Sigma,\mu) \longrightarrow L^p(X,\Sigma_n,\mu)$$

such that $f(n) := P_{(n)}f \xrightarrow{\|\cdot\|} f$ for every $f \in L^p(X,\Sigma,\mu)$ (use Schaefer [1974], p.211 and apply a martingale convergence theorem as in (Y.14)). Define

$$T_{(n)}f := P_{(n)}Tf$$

for $f \in L^p(X,\Sigma_n,\mu)$. Then

$$\begin{aligned}\|P_{(n)}TP_{(n)}f - Tf\| &\le \|P_{(n)}TP_{(n)}f - P_{(n)}Tf\| + \|P_{(n)}Tf - Tf\| \\ &\le \|P_{(n)}T\| \cdot \|P_{(n)}f - f\| + \|P_{(n)}Tf - Tf\|\end{aligned}$$

implies that $T_{(n)}P_{(n)}f \xrightarrow{\|\cdot\|} Tf$ for every $f \in L^p(X,\Sigma,\mu)$.

Similarily one shows that $T^i_{(n)}P_{(n)}f \xrightarrow{\|\cdot\|} T^if$ for every $i \in \mathbb{N}_o$.

Therefore

$$f^*_{(n),m} := \sup\left\{\frac{1}{k}\sum_{i=o}^{k-1} T^i_{(n)}f(n) : 1 \le k \le m\right\}$$

converges to $f^*_m := \sup\left\{\frac{1}{k}\sum_{i=o}^{k-1} T^if : 1 \le k \le m\right\}$ for every $m \in \mathbb{N}$.

On the other hand, $T_{(n)}$ satisfies a dominated estimate, i.e.

$$\|f^*_{(n),m}\| \le \frac{p}{p-1} \|f(n)\|$$

for every $n \in \mathbb{N}$.

The convergence proved above shows that

$$\|f^*_m\| \le \frac{p}{p-1} \|f\| . \quad \blacksquare$$

Now the desired individual ergodic theorem for a positive $L^p$-contraction T will be obtained by putting together the various pieces: First, we have a.e. convergence on a dense subspace (App.V.1) which can be extended to all of $L^p(\mu)$ as soon as we have a dominated estimate (App.V.3). Such a dominated estimate can be achieved for finite-dimensional $L^p$-spaces since positive contractions thereon possess a lattice dilation (U.5), and the dilated operator admits a dominated estimate (App.V.5) which is transferred to the original operator (App.V.6). Finally, the dominated estimate for T on an infinite-dimensional $L^p(\mu)$ is inherited by its finite-dimensional approximants $T_{(n)}$ (App.V.7), and the proof of Akcoglu's ergodic theorem is complete.

V.8 Theorem (Akcoglu, 1975):
Every positive contraction on a reflexive $L^p$-space is individually ergodic.

After having proved individual ergodicity for positive contractive operators on $L^p(\mu)$ one might ask for which larger classes of operators the same conclusion remains true. Since positivity was very essential in the proof it is not surprising that Akcoglu's ergodic theorem extends to all "regular" contractions $T \in \mathcal{L}(L^p(\mu))$, i.e., to operators $T$ having a modulus $|T| \in \mathcal{L}(L^p)$ such that $\| |T| \| \leq 1$.
On the other hand it was soon realized that the positivity assumption may not be dropped without any replacement. In fact, this followed from earlier results of Burkholder [1962], and in the last part of this appendix we shall present a concrete example of an $L^2$-contraction which is not individually ergodic. The basic tool for the construction of such an example is the following classical theorem (Kolmogorov - Menchov [1927]) from the theory of orthogonal series (see Olevskii [1975] for details):

V.9 Lemma:
In the Hilbert space $L^2([0,1],\mathcal{B},m)$ there exists a complete orthonormal system $(e_n)_{n\in\mathbb{N}}$ and a function $f_o$ such that the series $\sum_{n=1}^{\infty}(f_o|e_n)\, e_n(x)$ diverges for almost every $x\in[0,1]$.
Next we shall denote by Q the set of all positive sequences $(a_n)_{n\in\mathbb{N}}$, strictly increasing to 1, such that there exists a strictly increasing subsequence $(r_n)_{n\in\mathbb{N}}$ of $\mathbb{N}$ with
$(a_{n-1}^{r_n})_{n>1} \in l^2$ and $(1 - a_n^{r_n})_{n\in\mathbb{N}} \in l^2$. An example of an element in Q is furnished by $a_n := t^{((2n)!)^{-1}}$ for $0<t<1$ with $r_n := (2n-1)!$ .
With this notation we obtain the desired result.

V.10 Proposition:
Take $(a_n)_{n\in\mathbb{N}} \in Q$ with corresponding $(r_n)_{n\in\mathbb{N}}$, and choose an orthonormal system $(e_n)_{n\in\mathbb{N}}$ in $L^2([0,1],\mathcal{B},m)$ such that for some $f_o \in L^2(m)$ the series $\sum_{n=1}^{\infty}(f_o|e_n)\, e_n(x)$ diverges a.e.. Then the operator $T$ on $L^2([0,1],\mathcal{B},m)$ defined by

$$Tf := \sum_{n=1}^{\infty} a_n (f|e_n)\, e_n \ , \ f \in L^2(m),$$

is contractive, hence mean ergodic, but not individually ergodic.

Proof:

Clearly T is contractive and therefore mean ergodic by (IV.5), and we observe that the only fixed vector is 0. Since $T^k f = \sum_{n=1}^{\infty} a_n^k (f|e_n)e_n$ we have

$$\int_0^1 \sum_{k=1}^{\infty} k|T^k f - T^{k+1}f|^2 \, dm = \sum_{k=1}^{\infty} k \|T^k f - T^{k+1} f\|^2$$
$$= \sum_{k=1}^{\infty} k \sum_{n=1}^{\infty} (a_n^k - a_n^{k+1})^2 |(f|e_n)|^2$$
$$= \sum_{k=1} \sum_{n=1}^{\infty} k(a_n^k - a_n^{k+1})^2 |(f|e_n)|^2$$
$$\leq \|f\|^2 \ .$$

In particular this permits us to conclude that $\sum_{k=1}^{\infty} k|T^k f - T^{k+1} f|^2(x)$ is finite for almost every $x \in [0,1]$. Now assume that $T_n f_o$ converges a.e.. By the above remark it must converge to 0 and by the elementary Lemma (E.3) we obtain that even $\lim_{k\to\infty} T^k f_o = 0$ a.e.. Writing

$$y_k := \sum_{n=1}^{\infty} a_n^k (f_o|e_n)e_n \quad \text{and} \quad z_k := \sum_{n=k}^{\infty} (f_o|e_n)e_n \quad \text{we get}$$

$$\|y_{r_k} - z_k\|^2 = \sum_{n=1}^{k-1} a_n^{2r_k} |(f_o|e_n)|^2 + \sum_{n=k}^{\infty} (1 - a_n^{r_k})^2 |(f_o|e_n)|^2$$
$$\leq \|f_o\|^2 \, (a_{k-1}^{2r_k} + (1 - a_k^{r_k})^2) \ .$$

As before we conclude that $\sum_{k=2}^{\infty} |y_{r_k} - z_k|^2$ is finite a.e., and therefore $(y_{r_k} - z_k)_{k\in\mathbb{N}}$ converges to 0 a.e., as does $y_{r_k}$ by the above considerations. From this follows that $\sum_{n=1}^{\infty} (f_o|e_n)e_n(x)$ converges a.e., in contradiction to our hypothesis. ∎

In conclusion, we remark that it seems to be unknown whether every $L^p$-contraction for $1 < p < \infty$, $p \neq 2$, is individually ergodic.

References: Akcoglu [1975], [1979], Ionescu-Tulcea [1964], Krengel [1983].

## Appendix W: Uniformly Ergodic Operators

For a bounded linear operator T with bounded powers on a Banach space E the <u>weak</u> and <u>strong</u> operator convergence of its Cesaro means

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

are equivalent, while the <u>norm</u> convergence is a much stronger property. Operators having this convergence property will be discussed in detail in this appendix.

In the following let E always denote a (complex) Banach space. We recall the basic concept from (IV.D.1).

W.1 <u>Definition</u>:

An operator $T \in \mathcal{L}(E)$ is called <u>uniformly ergodic</u> if the sequence of its Cesaro means $(T_n)_{n \in \mathbb{N}}$ converges in the operator norm.

Remarks:

1. If T is uniformly ergodic, then T is mean ergodic, and there exists a corresponding projection P and a decomposition

$$E = PE \oplus P^{-1}(0) ,$$

where $PE = F(T) := \{f \in E : Tf=f\}$ and $P^{-1}(0) = \overline{(Id - T)E}$ (see IV.3).

2. If T is uniformly ergodic, then $\frac{1}{n} \|T^n\| \to 0$ for $n \to \infty$, and $r(T) \leq 1$ for the spectral radius $r(T)$ (use IV.3.0).
3. If $r(T) < 1$ then T is uniformly ergodic with $P = 0$.

We recall that mean ergodicity of T is determined by a spectral property (i.e. by the dimension of the fixed spaces of T and T', see IV.4.e). In a stronger sense the same will be true for uniformly ergodic operators. As a preparation to the main characterization theorem the following lemma describes first the spectrum of a uniformly ergodic operator.

W.2 Lemma:

Let $T \in \mathcal{L}(E)$ be uniformly ergodic.

(i) The following are equivalent:

(a) $1 \in \sigma(T)$,

(b) $1 \in P\sigma(T)$,

(c) 1 is a pole of order one of $R(\lambda, T)$.

(ii) $(Id - T)E$ is closed.

Proof:

Consider $E_0 := \overline{(Id - T)E} = (Id - P)E$ and $S := T|_{E_0}$. If we assume $1 \in \sigma(S)$, then $1 \in \sigma(S_n)$ and $\|S_n\| \geq 1$ for every $n \in \mathbb{N}$. On the other hand $S$ is uniformly ergodic with corresponding projection 0. Therefore $1 \notin \sigma(S)$.

(i) The implications (c) $\Rightarrow$ (b) $\Rightarrow$ (a) are evident. For (a) $\Rightarrow$ (c) we observe that

$$R(\lambda, T) = (\lambda - 1)^{-1}P + R(\lambda, S)(Id - P)$$

for $\lambda \neq 1$ in some neighborhood of 1. Hence 1 is a pole of order one of $R(\lambda, T)$.

(ii) The operator $(Id - S)$ is invertible and therefore

$$E_0 = (Id - S)E_0 = (Id - T)E_0 \subset (Id - T)E \quad . \quad \blacksquare$$

The following theorem shows that under a weak boundedness condition the uniform ergodicity of T is characterized by the behavior of $R(\lambda, T)$ in a neighborhood of 1.

W.3 Theorem:

Assume that $T \in \mathcal{L}(E)$ satisfies $\lim_{n \to \infty} \frac{1}{n} \|T^n\| = 0$.

The following are equivalent:

(a) T is uniformly ergodic.

(b) There exists a zero-element $P \in \overline{co}^{\|\cdot\|} \{T^n : n \in \mathbb{N}_0\}$, i.e. $PT = TP = P = P^2$ .

(c) $1 \notin \sigma(T)$ or 1 is a pole of first order of $R(\lambda, T)$.

(d) $(Id - T)E$ is closed.

(e) The Abel averages $(\lambda - 1)R(\lambda, T)$ norm converge for $\lambda \downarrow 1$ .

Remarks:

1. The boundedness condition in the above theorem is necessary to deduce uniform ergodicity from (b),(c) or (d): take $E=\mathbb{C}$ and $Tx := -2x$.
2. If one of the conditions in (W.3) is satisfied, then the mean ergodic projection P is the zero element in (b) and the residuum of $R(\lambda, T)$ at $\lambda = 1$.

Proof:

The implication (a) $\Rightarrow$ (b) is clear.

(b) $\Rightarrow$ (c): The zero element P is a projection onto the fixed space of T. Consider $E_o := (Id - P)E$ and $S := T|_{E_o}$.

Since the resolvent of T has the representation

$$R(\lambda, T) = (\lambda - 1)^{-1}P + R(\lambda, S)(Id - P) \quad \text{for } 1 \neq \lambda \in \varrho(S)$$

it suffices to show that $1 \notin \sigma(S)$: But if $1 \in \sigma(S)$, then $1 \in \sigma(V)$ and $1 \leq \|V\|$ for any $V \in \mathrm{co}\{S^n : n \in \mathbb{N}_o\}$. This contradicts the assumption $0 = P|_{E_o} \in \overline{\mathrm{co}}^{\|\cdot\|}\{S^n : n \in \mathbb{N}_o\}$.

(c) $\Rightarrow$ (d): If $1 \notin \sigma(T)$ then (d) holds. Let 1 be a pole of first order of $R(\lambda, T)$ and consider $E_o := \overline{(Id - T)E}$ and $S := T|_{E_o}$. Then we have

$$R(\lambda, T) = (\lambda - 1)^{-1}P + H(\lambda)$$

for $1 \neq \lambda$ in some neighborhood of 1, where P is a projection onto the fixed space of T and $\lambda \mapsto H(\lambda)$ is holomorphic. From the Neumann series it follows that $R(\lambda, S) = R(\lambda, T)|_{E_o} = H(\lambda)|_{E_o}$, hence $1 \notin \sigma(S)$. This yields

$$E_o = (Id - S)E_o \subset (Id - T)E .$$

(d) $\Rightarrow$ (a): By the open mapping theorem there exists $c > 0$ satisfying the following: For every $g \in (Id - T)E$ there exists $f \in E$ such that $\|f\| \leq c \cdot \|g\|$ and $g = (Id - T)f$. Thus

$$\left\| \frac{1}{n} \sum_{i=o}^{n-1} T^i g \right\| = \frac{1}{n} \left\| \sum_{i=o}^{n-1} T^i (Id - T) f \right\| \leq c \cdot \frac{1 + \|T^n\|}{n} \cdot \|g\|$$

and $S := T|_{E_o}$, $E_o := \overline{(Id - T)E}$, is uniformly ergodic with projection 0. As above we conclude by the spectral mapping theorem that $(Id - S)$ is invertible on $E_o$. For $f \in E$ there exists $g \in E_o$ such that $(Id - T)f = (Id - T)g$. Therefore $f = g+(f-g)$ and $T(f-g) = f-g$, i.e.

$E = F(T) \oplus (Id - T)E$, and $T$ is uniformly ergodic.
Finally, the implications (c) $\Rightarrow$ (e) $\Rightarrow$ (b) are easy to verify. ∎

W.4 <u>Examples</u>:

1. $T$ is uniformly ergodic as soon as some power $T^n$ is uniformly ergodic. This follows analogously to (IV.D.2), and implies that periodic operators (i.e. $T^{n_\circ} = Id$) are uniformly ergodic.

2. If $K \in \mathcal{L}(E)$ is compact and satisfies $\frac{1}{n}\|K^n\| \to 0$, then any pole of $R(\lambda, K)$ on the unit circle $\Gamma$ has order $\leq 1$ (Dunford-Schwartz [1958], VIII.8.1). Therefore (W.3.c) implies that such operators are uniformly ergodic.

3. If $T$ is uniformly ergodic, then $T^n$ is not necessarily uniformly ergodic for any $n \geq 2$. Take $E = l^2$ and $T$ the multiplication operator

$$T((x_n)) := (\alpha_n x_n)$$

where $\overline{\{\alpha_n : n \in \mathbb{N}\}} = \{\gamma \in \Gamma : \frac{\pi}{2} \leq \arg \gamma \leq \pi\}$.
Then $1 \notin \sigma(T)$, but $1 \in \sigma(T^n)$ is not a pole of $R(\lambda, T^n)$ for every $n \geq 2$. Another example is given in (IV.D.2).

4. Convex combinations of commuting, uniformly ergodic operators are uniformly ergodic. The proof follows from (IV.D.4) and the imbedding procedure developed in Schaefer [1974], V.1 if we observe that the original operator $T$ is uniformly ergodic if and only if the extended operator $\hat{T}$ on the $\mathcal{F}$-product $\hat{E}$, $\mathcal{F}$ the Frechet filter, is mean ergodic.

As a consequence we remark that compact operators with bounded powers yield an important class of uniformly ergodic operators. The main reason for this is the particular structure of the spectrum of compact operators. On the other hand it is clear by (W.3.c) that the convergence of $T_n$ - if $r(T) \leq 1$ - is determined by the behavior of $R(\lambda, T)$ on the unit circle $\Gamma$ only. Therefore one might expect that a useful class of operators might consist of operators having spectrum like a compact operator on and outside the unit circle and arbitrary spectrum in the interior.
In the following definition we make precise this idea.

W.5 Definition:

An operator $T \in \mathcal{L}(E)$ is called quasi-compact if there exists a compact operator $K \in \mathcal{L}(E)$ and $k \in \mathbb{N}$ such that

$$\| T^k - K \| < 1.$$

W.6 Examples:

1. $T \in \mathcal{L}(E)$ is quasi-compact if some power of T is compact.

2. T is quasi-compact if some power of T has norm less than one.

3. Every power of a quasi-compact operator is quasi-compact (see Neveu [1964], Lemma V.3.1).

The first step toward the main characterization theorem (W.10) is the following lemma.

W.7 Lemma:
The eigenspaces corresponding to unimodular eigenvalues of a quasi-compact operator T are finite-dimensional.

Proof:
We show that the fixed space F(S) is finite-dimensional where $S := T^k$ and $\| S - K \| < 1$ for some compact operator $K \in \mathcal{L}(E)$: Take $x_n \in F(S)$, $\| x_n \| = 1$ and $U := S - K$. Then $Kx_n = (Id - U)x_n$ and $(Id - U)$ is invertible. Therefore $x_n = (Id - U)^{-1}Kx_n$. Since K is compact this proves that $(x_n)_{n \in \mathbb{N}}$ has a convergent subsequence, hence $\dim F(T) \leq \dim F(S) < \infty$. For arbitrary eigenvalues $\lambda \in \mathbb{C}$, $|\lambda| = 1$, we consider $\lambda^{-1}T$ which is quasi-compact iff T is quasi-compact. ∎

Our next aim is to establish a relation between quasi-compactness and uniform ergodicity. This will be done by showing that quasi-compact operators satisfy (W.3.d).

W.8 Lemma:

If $T \in \mathcal{L}(E)$ is quasi-compact, then $(Id - T)E$ is closed.

Proof:

Take $K \in \mathcal{L}(E)$, K compact, and $k \in \mathbb{N}$ such that

$$\| T^k - K \| < 1 .$$

The operator $V := T^k - K$ has norm less than 1 and $Id - V$ is invertible. Now recall the following identities:

$$(*) \quad Id = (Id - V)^{-1} K + (Id - V)^{-1} \sum_{i=0}^{k-1} T^i (Id - T) ,$$

$$(**) \quad Id - T = Id - (Id - V)^{-1} KT - (Id - V)^{-1} \sum_{i=1}^{k} T^i (Id - T) .$$

Take $f_n \in E$ such that $(Id - T)f_n$ converges to some $g \in E$.

Suppose first that $(f_n)_{n \in \mathbb{N}}$ contains a bounded subsequence, or without loss of generality suppose that the sequence $(f_n)_{n \in \mathbb{N}}$ itself is bounded. From equation $(*)$ we obtain

$$f_n = (Id - V)^{-1} K f_n + (Id - V)^{-1} \sum_{i=0}^{k-1} T^i (Id - T) f_n$$

$$=: \quad a_n \quad + \quad b_n \quad .$$

$(a_n)_{n \in \mathbb{N}}$ contains a convergent subsequence $(a_{n_i})_{i \in \mathbb{N}}$ since K is compact, and $(b_{n_i})_{i \in \mathbb{N}}$ converges by assumption. Therefore $(f_{n_i})_{i \in \mathbb{N}}$ converges to $f \in E$ and $(Id - T)f = g$.

If $(f_n)_{n \in \mathbb{N}}$ contains no bounded subsequence we consider

$$d_n := \inf \{ \| f_n - h \| : h \in F(T) \}, \ F(T) \text{ the fixed space of } T,$$

and choose $g_n \in F(T)$ such that

$$d_n \leq \| f_n - g_n \| \leq 2 d_n .$$

If the sequence $(d_n)_{n \in \mathbb{N}}$ were unbounded, we can assume that

$\lim \frac{(Id - T)f_n}{\| f_n - g_n \|} = 0$, while still $\lim_{n \to \infty} (Id - T) f_n = g$ .

By identity $(**)$ and by writing $h_n := \frac{f_n - g_n}{\| f_n - g_n \|}$ we obtain

$$(Id - T)h_n = h_n - (Id - V)^{-1} K T h_n - (Id - V)^{-1} \sum_{i=1}^{k} T^i (Id - T) h_n$$

$$=: h_n - \quad \tilde{a}_n \quad - \quad \tilde{b}_n \quad .$$

Since $(Id - T)h_n = \frac{(Id - T)f_n}{\|f_n - g_n\|}$, the sequence $(\tilde{b}_n)_{n \in \mathbb{N}}$ converges to zero. Since KT is compact and $(h_n)_{n \in \mathbb{N}}$ is bounded, we find a subsequence satisfying $\lim_{i \to \infty} (\tilde{a}_{n_i}) =: a$. Consequently the subsequence $(h_{n_i})$ converges to a, too.

On the other hand we have

$$0 = \lim_{i \to \infty} (Id - T)h_{n_i} = a - Ta \qquad \text{and}$$

$$\|h_n - a\| = \frac{1}{\|f_n - g_n\|} \|f_n - (g_n + \|f_n - g_n\| a)\| \geq \frac{d_n}{2d_n} = \frac{1}{2}.$$

This yields a contradiction and shows that $(f_n - g_n)_{n \in \mathbb{N}}$ has to be bounded. Since $(Id - T)(f_n - g_n) = (Id - T)f_n$ we prove the assertion as above. ∎

As an immediate application of Lemma (W.8) and Condition (W.3.d) we obtain the following theorem, which remains true under the formally weaker assumption that $\frac{1}{n} \cdot T^n$ converges to zero in the weak operator topology (see Dunford-Schwartz [1958], VII.8.4).

W.9 Theorem:

A quasi-compact operator $T \in \mathcal{L}(E)$ satisfying $\lim_{n \to \infty} \frac{1}{n} \|T^n\| = 0$ is uniformly ergodic and has finite-dimensional fixed space.

Actually, if $T \in \mathcal{L}(E)$ is quasi-compact and satisfies the above boundedness condition, then $\lambda T$ is uniformly ergodic for every $\lambda \in \Gamma$. Conversely, conditions on the spectral behavior of T in each point $\lambda \in \Gamma$ - instead of just at $\lambda = 1$ - characterize quasi-compactness.

W.10 Theorem:

For $T \in \mathcal{L}(E)$ satisfying $\frac{1}{n} \|T^n\| \to 0$ the following assertions are equivalent:

(a) T is quasi-compact.

(b) $\lambda T$ is uniformly ergodic for every $\lambda \in \Gamma$ and the corresponding projections $P_\lambda$ have finite-dimensional range.

(c) $\sigma(T) \cap \Gamma$ only contains poles of $R(\lambda, T)$ of first order with finite-dimensional eigenspaces.

(d) $T = S + R$ , where $S \in \mathcal{L}(E)$ is of finite rank and $R \in \mathcal{L}(E)$ has spectral radius $r(R) < 1$.

Proof:
The implications (a) $\Rightarrow$ (b) $\Rightarrow$ (c) have been proved in (W.3) and (W.9).
(c) $\Rightarrow$ (d): The spectral projection Q corresponding to the spectral set $\sigma(T) \cap \Gamma$ is of finite rank. Therefore $S := TQ$ and $R := T(Id - Q)$ satisfy (d).
(d) $\Rightarrow$ (a): There exists $n \in \mathbb{N}$ such that $\|R^n\| < 1$. Therefore

$$T^n = (S + R)^n = U + R^n$$

where U is of finite rank, hence compact. ∎

Remark:
The equivalence of (a) and (d) holds without any boundedness condition on $\{T^n : n \in \mathbb{N}\}$. For the most far-reaching result see Brunel-Revuz [1974].

To summarize the results obtained so far we state that uniform ergodicity of an operator $T \in \mathcal{L}(E)$ with $r(T) \leq 1$ is determined by its spectral behavior at (resp. in a neighborhood of) $\lambda = 1$, while quasi-compactness depends on a specific structure of all of the peripheral spectrum $\sigma(T) \cap \Gamma$.
Therefore it is quite surprising and satisfactory that there exists an important class of operators for which the peripheral spectrum is largely determined by the behavior of the resolvent near the spectral radius. In fact, this idea is the "Leitmotiv" of the so-called "Perron-Frobenius theory" of <u>positive</u> operators on <u>Banach lattices</u> and has been confirmed by many beautiful results. We refer to Schaefer [1974], chap. V for a presentation of the general theory.
In the following second part of this appendix we will prove some uniform ergodic theorems for positive operators on abstract and concrete Banach lattices. These are based on the above mentioned spectral theory of positive operators and lead us closer to the situations occurring in ergodic theory.
The first result in this direction is nothing else but a reformulation of Schaefer [1974], V.5.5 and shows that for positive operators on Banach lattices the converse of Theorem (W.9) holds.

W.11 Theorem (Lotz-Schaefer, 1968):
Let T be a positive operator on a Banach lattice E. If T is uniformly ergodic with finite-dimensional fixed space, then T is quasi-compact.

Another example showing which strong properties can be deduced from the uniform ergodicity of positive operators is obtained if we consider lattice isomorphisms. Obviously, this result applies to operators $T_\varphi$ induced by an MDS $(X, \Sigma, \mu; \varphi)$.

W.12 Proposition:
Let T be an isometric lattice isomorphism on a Banach lattice E. T is uniformly ergodic if and only if T is periodic, i.e. $T^n = Id$ for some $n \in \mathbb{N}$.

Proof:
By assumption $\sigma(T) \subset \Gamma$. If T is uniformly ergodic, then 1 is isolated in $\sigma(T)$. The results on the cyclic structure of the peripheral spectrum of positive operators (see Schaefer [1974], V.4.9) imply that $\sigma(T)$ is finite and $\sigma(T^n) = \{1\}$ for some $n \in \mathbb{N}$. From Schaefer-Wolff-Arendt [1978] we conclude that $T^n = Id$.
The converse implication has been discussed in (W.4.1). ∎

As another feature of positive operators we mention that the equivalences of (W.3) hold under the weaker assumption that $r(T) \leq 1$ (see Karlin [1959]). We are now turning our attention to positive operators on concrete function spaces and first reformulate a classical result from the theory of Markov processes due to Doeblin [1937] (see Jacobs [1960]). It will permit us to obtain rather surprising ergodic theorems on $L^1$- and $L^\infty$-spaces. Our presentation follows Lotz [1981].

W.13 Proposition:
Let T denote a Markov operator (i.e. a positive operator with $T\mathbb{1} = \mathbb{1}$) on a Banach lattice C(X), X compact. If T satisfies (D) there exists $m \in \mathbb{N}$, $0 < \mu \in M(X)$ and $0 < \gamma < 1$, such that

$$T^m f - \mu(f)\mathbb{1} \leq \gamma \cdot \mathbb{1} \qquad \text{for all } 0 \leq f \leq \mathbb{1},$$

then T is quasi-compact.

Proof:

If we assume that C(X) is order complete we know from Schaefer [1974], IV.1.5 that $\mathcal{L}(C(X))$ is a vector lattice. In particular, it follows from the concrete representation of the lattice operations for operators (see Schaefer [1974],p.229), that condition (D) implies

$$(T^m - \mu\otimes 1\!\!1)^+\cdot 1\!\!1 = \sup\{T^m f - \mu(f)\cdot 1\!\!1 : 0 \leq f \leq 1\!\!1\}$$
$$\leq \gamma\cdot 1\!\!1 .$$

Since $(T^m - \mu\otimes 1\!\!1)^+$ is a positive operator we obtain $\|(T^m - \mu\otimes 1\!\!1)^+\| \leq \gamma$ . In the vector lattice $\mathcal{L}(C(X))$ we can decompose

$$T^m = U + V$$

where $U := (T^m - \mu\otimes 1\!\!1)^+$ and $V := T^m \wedge \mu\otimes 1\!\!1$ . The operator V is dominated by $T^m$, hence is contractive, and dominated by $\mu\otimes 1\!\!1$, hence is weakly compact, what can be proved as follows: from $V \leq \mu\otimes 1\!\!1$ one obtains that V can be extended to $\tilde{V}$ on $L^1(X, \mathcal{B}, \mu)$, hence $V = \tilde{V}\circ J$ where J denotes the canonical imbedding from C(X) into $L^1(\mu)$. Therefore V is weakly compact since the same is true for J (Schaefer [1974],p.129 Example 5).

In order to show that T is quasi-compact we consider

$$T^{nm} = (U + V)^n$$
$$= U^n + (U^{n-1} V + U^{n-2} VU + \dots + VU^{n-1}) + K_n,$$

where $K_n$ is compact by the Dunford-Pettis property of C(X) (see Schaefer [1974],II.9.9). On the other hand

$$\|T^{nm} - K_n\| \leq \gamma^n + n\cdot\gamma^{n-1} < 1$$

for n large enough, so that T is quasi-compact.

If C(X) is not order complete we consider the biadjoint $T''$ on $C(X)''$ which still satisfies (D) if T does. From the proof above we conclude that $T''$ is quasi-compact. The characterization (W.10.c) implies the quasi-compactness of T. ∎

It is not difficult to show that the condition (D) is in fact an abstract version of Doeblin's condition for a transition probability (see Lotz [1981]), but we shall pursue the investigation of abstract Markov operators.

In the following we assume T to be an irreducible Markov operator on $C(X)$, i.e. $0 \leq T$, $T\mathbb{1} = \mathbb{1}$ and T leaves invariant no closed non-trivial (lattice) ideal. In that case the dual fixed space $F' := F(T')$ is a non-trivial sublattice of $\mathcal{M}(X)$. In fact, the existence of $T'$-invariant probability measures is proved as Theorem 1 in ( App. S ), and for every $\mu \in F'$ we have AA

$$|\mu| = | T'\mu | \leq T'|\mu| \text{ and}$$

$$< \mathbb{1}, T'|\mu| - |\mu| > = 0, \text{ hence } |\mu| \in F' .$$

Moreover, every $0 < \mu \in F'$ is strictly positive, since otherwise $J_\mu := \{ f \in C(X) : < |f|, \mu > = 0\}$ would be a non-trivial T-invariant ideal. Analogously to the argument for $F'$ this implies that $F := F(T)$ is a sublattice of $C(X)$. If $\dim F \geq 2$ then there exist two positive orthogonal functions $f, g \in F$ generating two non-trivial T-invariant ideals. Therefore F must be one-dimensional and is spanned by the constant functions (see Schaefer [1974], V.5.2).

Finally, if T is mean ergodic it has been shown in (III.D.11) that the corresponding projection P is of the form $P = \mu \otimes \mathbb{1}$ for the strictly positive $T'$-invariant probability measure $\mu$ spanning the fixed space $F'$.

The following main result confirms once again the claim that for positive operators quasi-compactness is implied by much weaker ergodic properties. We present a graded list of equivalent properties showing that an irreducible Markov operator on $C(X)$ is already quasi-compact if its adjoint is mean ergodic having not too large fixed space.

W.14 <u>Theorem</u> (Lotz, 1981):
Let T be an irreducible Markov operator on $C(X)$.
The following are equivalent:

(a) T is quasi-compact.

(b) T is uniformly ergodic.

(c) T and $T'$ are mean ergodic.

(d) $\dim F' = 1 = \dim F''$, where $F'' := F(T'')$.

(e) $T'$ is mean ergodic and the vector lattice $F'$ has a weak order unit.

Proof:

The implications (a) $\Rightarrow$(b) $\Rightarrow$(c) $\Rightarrow$(d) $\Rightarrow$(e) are obvious and show that (e) is (formally) much weaker than (a).

However we are able to show that T satisfies a Doeblin condition (D) if (e) holds. Consequently (e) implies (a):

Take a weak order unit $\mu$ in $F'$ and define functions

$$g_n(x) := \inf \{ T^n(\mathbb{1} - f)(x) + \mu(f) : 0 \leq f \leq \mathbb{1} \}$$

for $x \in X$, $n \in \mathbb{N}$. As a pointwise infimum of continuous functions every $g_n$ is upper semi-continuous. On the other hand,

$$\begin{aligned} g_n(x) &= \inf \{ < g, T'^n \delta_x > + \mu(f) : 0 \leq f, g \in C(X), f + g = \mathbb{1} \} \\ &= (T'^n \delta_x \wedge \mu)\mathbb{1} \\ &= \| T'^n \delta_x \wedge \mu \| \end{aligned}$$

where $\delta_x$ denotes the Dirac measure at $x \in X$ (again Schaefer [1974], p.229(1)).

Finally, $0 \leq Tf \leq \mathbb{1}$ for $0 \leq f \leq \mathbb{1}$ implies that $(g_n)$ is an increasing sequence, i.e.

$$\begin{aligned} g_n(x) &\leq \inf \{ T^n(\mathbb{1} - Tf)(x) + \mu(Tf) : 0 \leq f \leq \mathbb{1} \} \\ &= \inf \{ T^{n+1}(\mathbb{1} - f)(x) + \mu(f) : 0 \leq f \leq \mathbb{1} \} \\ &= g_{n+1}(x). \end{aligned}$$

Now define $H_n := \{ f \in C(X) : 0 \leq f \leq g_n \}$. It follows immediately that $H_n \subset H_{n+1}$ and $TH_n \subset H_{n+1}$. Therefore

$$H := \bigcup_{n \in \mathbb{N}} H_n$$

is a T-invariant subset of C(X) and

$$A := \{ x \in X : h(x) = 0 \text{ for all } h \in H \}$$

is a closed subset of X. Since T is irreducible the closed ideal generated by H is either equal to $\{0\}$ or equal to C(X), hence

$$A = X \quad \text{or} \quad A = \emptyset .$$

Assume that A = X and consider the closed sets

$$B_{n,k} := \{ x : g_n(x) \geq k^{-1} \}$$

for $n,k \in \mathbb{N}$. The interior of $B_{n,k}$ is empty since $H_n = \{0\}$ for every $n \in \mathbb{N}$. Since X is a Baire space there exists $y \in X \setminus \bigcup_{n,k=1}^{\infty} B_{n,k}$ .

For this point and every $n \in \mathbb{N}$ we have

$$0 = g_n(y) = \| T'^n \delta_y \wedge \mu \| \quad \text{and therefore}$$

$$T_n' \delta_y \wedge \mu = 0 .$$

By assumption the sequence $(T_n')_{n\in\mathbb{N}}$ converges strongly to a projection P on $\mathcal{M}(X)$. Since the lattice operations are norm-continuous in $\mathcal{M}(X)$ we conclude that $P\delta_y \wedge \mu = 0$. But $\mu$ is a weak order unit in $F'$, hence $P\delta_y = 0$. This is a contradiction to $0 < \langle 1\!\!1, \delta_y\rangle = \langle 1\!\!1, T'\delta_y\rangle = \langle 1\!\!1, P\delta_y\rangle$.

Therefore A must be empty. By the compactness of X we can find $h_1,\dots,h_j \in H$ such that

$$\sum_{i=1}^{j} h_i(x) > 0 \quad \text{for every } x\in X.$$

Moreover, we can find $m\in\mathbb{N}$ such that $h_1,\dots h_j \in H_m$ and $h := \frac{1}{j}\sum_{i=1}^{j} h_i \in H_m$. h is a strictly positive continuous function on X, and there exists $0 < \delta < 1$ such that

$$\delta 1\!\!1 \leq h \leq g_m \leq T^m(1\!\!1 - f) + (\mu \otimes 1\!\!1)\, f \quad \text{for every } 0 \leq f \leq 1\!\!1 .$$

Taking $\gamma \in (1-\delta, 1)$ we obtain the Doeblin condition

$$T^m f - (\mu \otimes 1\!\!1) f < \gamma \cdot 1\!\!1 . \qquad \blacksquare$$

Even if condition (e) above is formally much weaker than quasi-compactness it might still be difficult to check it in concrete cases. In particular it is not simple to control the adjoint $T'$ on the rather large dual Banach space $\mathcal{M}(X)$. Therefore we continue our search for additional hypotheses faciliating this task.
First we observe that the dual fixed space $F'$ has a weak order unit if it is separable. Moreover we recall that a Banach space E is called a <u>Grothendieck space</u> if every weak* convergent sequence in $E'$ is weakly convergent. A space C(X) is a Grothendieck space if two open disjoint $F_\sigma$-sets in X have disjoint closures, in particular, if X is $\sigma$-Stonian (see Schaefer [1974], II.10 or Seever [1968]).

W.15 <u>Lemma</u>:
Let T be a Markov operator on a Grothendieck space C(X). If the dual fixed space $F'$ is separable then it is finite-dimensional, and $T'$ is mean ergodic.

Proof:
The dual fixed space $F'$ is weak*-closed, hence it is the dual of $C(X)/_G$ where $G := \{f \in C(X) : \langle f, \nu\rangle = 0 \text{ for every } \nu \in F'\}$ .

The separability of $F'$ implies that $C(X)/_G$ is separable, hence it follows as in Schaefer [1974], II.10.4, Cor. 2, that $C(X)/_G$ is reflexive.
The dual of a reflexive Banach space is reflexive, too. On the other hand, $F'$ is a closed sublattice of the AL-space $M(X)$, therefore is an AL-space and hence has the Dunford-Pettis property (Schaefer [1974], II.9.9). This and the reflexivity imply that $F'$ is finite-dimensional.

Now we show that $T'$ is mean ergodic, i.e. $T_n'\nu$ norm-converges for every $\nu \in M(X)$. By the mean ergodic theorem (IV.4) it suffices to show that $(T_n'\nu)_{n\in\mathbb{N}}$ has a weakly converging or, since $C(X)$ is a Grothendieck space, weak*-converging subsequence.
As we have seen, the subspace $G$ of $C(X)$ has finite co-dimension, hence there exists a finite-dimensional subspace $H$ such that $C(X) = G + H$. Obviously we can find a subsequence $(T_{n_k}'\nu)_{k\in\mathbb{N}}$ such that $(\langle h, T_{n_k}'\nu\rangle)_{k\in\mathbb{N}}$ converges for every $h \in H$. On the other hand, every accumulation point of $(\langle g, T_{n_k}'\nu\rangle)_{k\in\mathbb{N}}$ for $g \in G$ is equal to $\langle g, \bar{\nu}\rangle$ for some weak accumulation point $\bar{\nu}$ of $(T_{n_k}'\nu)_{k\in\mathbb{N}}$. But such accumulation points $\bar{\nu}$ are contained in $F'$ and hence $\langle g, \bar{\nu}\rangle = 0$ and $(\langle g, T_{n_k}'\nu\rangle)_{k\in\mathbb{N}}$ converges to zero. Since $C(X) = G + H$ we have proved the weak*-convergence of $(T_{n_k}'\nu)_{k\in\mathbb{N}}$ and therefore the norm convergence of $(T_n'\nu)_{n\in\mathbb{N}}$. ∎

From (W.14) and the above lemma it follows immediately that an irreducible Markov operator $T$ on a Grothendieck space $C(X)$ is quasi-compact as soon as its dual fixed space $F'$ is separable. A rather concrete situation is described in the following corollary for which we recall that $L^\infty(X,\Sigma,\mu)$ is isomorphic to some $C(Y)$, $Y$ hyperstonian (see VI.D.10).

W.16 <u>Corollary</u>:
Every irreducible contraction $T$ on $L^\infty(X,\Sigma,\mu)$, $\mu(X) = 1$, is quasi-compact.

Proof:

The assertion follows if $r(T) < 1$. For $r(T) = 1$, the dual fixed space $F'$ contains a positive linear form $f'$ (Krein-Rutman theorem, see Schaefer [1966], App.2.6, Cor.) which is strictly positive since T is irreducible (compare the reasoning preceeding (W.14)). Since T is contractive we have $T\mathbb{1} \leq \mathbb{1}$ and

$$\langle \mathbb{1} - T\mathbb{1}, f' \rangle = 0,$$

which implies that T is in fact a Markov operator.

Next we observe that the dual fixed space $F'$ of a Markov operator is a sublattice in the dual of $L^\infty(X, \Sigma, \mu)$. If $\dim D' > 1$, there exist at least two positive, orthogonal, invariant linear forms $f', g' \in F'$, both being strictly positive. Decompose $f'$ into its order continuous and singular part, i.e. $f' = \nu_n + \nu_s$ where $\nu_n = g \cdot \mu$ for some $g \in L^1(\mu)$ and for every $\varepsilon > 0$ there exists $A \in \Sigma$ such that $\mu(A) < \varepsilon$ and the support of $\nu_s$ is contained in A.

Therefore the order continuous component of $f'$ is strictly positive on $L^\infty(\mu)$ and corresponds to a weak order unit in $L^1(\mu)$. The orthogonality of $f'$ and $g'$ implies that $g'$ has trivial order continuous component, which contradicts the strict positivity.

Therefore we obtain that $F'$ is one-dimensional. Since $L^\infty(\mu)$ is a Grothendieck space we can apply (W.15) and the assertion follows from (W.14). ∎

Remark:

The operator $T_\varphi$ induced by an ergodic MDS $(X, \Sigma, \mu; \varphi)$ is in general not irreducible on $L^\infty(\mu)$ - while it is on $L^1(\mu)$. In fact, $T_\varphi$ leaves invariant no non-trivial projection band in $L^\infty(\mu)$ but there are many closed ideals invariant under $T_\varphi$ (compare IV.D.10).

On the other hand, kernel operators

$$Tf(x) := \int k(x,y) f(y) \, d\mu(y)$$

are irreducible on $L^\infty(\mu)$ as soon as $k(x,y) \geq \varepsilon > 0$ for $x, y \in X$.

In the next corollary we transfer the previous results to operators on $L^1$-spaces.

W.17 <u>Corollary</u>:

Let T denote an irreducible positive contraction on $L^1(X,\Sigma,\mu)$, $(X,\Sigma,\mu)$ a probability space. The following are equivalent:

(a) T is quasi-compact.

(b) T is uniformly ergodic.

(c) $T'$ is uniformly ergodic.

(d) $T'$ is mean ergodic.

Proof:

It suffices to show that (d) implies (a), and we may assume that $r(T) = r(T') = 1$. Then the $T''$-fixed space in the bi-dual of $L^1(X,\Sigma,\mu)$ is non-trivial by the Krein-Rutman theorem. If $T'$ is mean ergodic on $L^\infty(\mu)$, then $(T_n''f'')_{n\in\mathbb{N}}$ is weak* convergent for every $f''$ in the dual of $L^\infty(\mu)$. But $L^\infty(\mu)$ is a Grothendieck space and therefore $(T_n''f'')_{n\in\mathbb{N}}$ is weakly convergent, and $T''$ is mean ergodic with corresponding projection $P'' \neq 0$. In particular, the original operator T is mean ergodic on $L^1(\mu)$. Since T was supposed to be irreducible, the corresponding projection P is of the form $P = g' \otimes f$ for strictly positive functions $f \in L^1(\mu)$, $g' \in L^\infty(\mu)$. Since f is a strictly positive linear form on $L^\infty(\mu)$ and since $T'\mathbb{1} \leq \mathbb{1}$ we conclude $g' = \mathbb{1}$ and $P' = f \otimes \mathbb{1}$ for $P'$ the projection corresponding to $T'$. The proposition in (III.D.11) implies that $T'$ is irreducible on $L^\infty(X,\Sigma,\mu)$ and hence quasi-compact by (W.16). ∎

In conclusion we remark that most of the previous results can be extended from irreducible operators to operators having finite-dimensional fixed spaces. We refer to Lotz [1981] and quote the following final theorem.

W.18 <u>Theorem</u> (Lotz, 1981):

Let T be a Markov operator on $L^\infty(X,\Sigma,\mu)$. If the $T'$-fixed space $F'$ is contained in $L^1(X,\Sigma,\mu)$, then T is quasi-compact and uniformly ergodic.

References: Ando [1968], Axmann [1980], Brunel-Revuz [1974], Fortet [1978], Horowitz [1972], Lin [1974],[1975],[1978], Lotz [1981], Revuz [1975], Yoshida-Kakutani [1941].

## Appendix X: Asymptotic Behavior of Markov Operators

One of the central themes of these lectures has been the asymptotic behavior of operators $T_\varphi$ induced by an MDS $(X,\Sigma,\mu;\varphi)$. Actually, we could extend many of the results to more general classes of operators such as contractions on Hilbert spaces (e.g.IV.5 or IX.D.5) or positive operators on Banach lattices (e.g. IX.D.6 ). A very important class of operators generalizing the $T_\varphi$'s are the positive operators $T$ on $L^1(X,\Sigma,\mu)$ satisfying $T\mathbb{1} = \mathbb{1}$ and $T'\mathbb{1} = \mathbb{1}$. In applications such operators are induces by transition probabilities $p(.,.)$ having invariant probability measure $\mu$. They have been called <u>bi-Markov operators</u> in (App.U). They are the functional-analytic model of a Markov process with finite invariant measure, and admit a very intuitive interpretation:

If there is a canonical choice of a representative in each equivalence class $T^n\mathbb{1}_A$, then the value

$$T^n\mathbb{1}_A(x) \qquad \text{(or } T'^n\mathbb{1}_A(x)\text{ )}$$

may be viewed as the "probability of the corresponding Markov process being in $A\in\Sigma$ at time $n\in\mathbb{N}$ when starting at $x\in X$ at time 0" (Yoshida-Kakutani [1941] or Lamperti [1977]). This interpretation yields one of the reasons, why we are interested in the limit of $T^n$ for $n\to\infty$ .

Concrete examples are provided by doubly stochastic matrices on $\mathbb{R}^n$ , by the induced operators $T_\varphi$ on $L^1(X,\Sigma,\mu)$ and by operators defined by measurable "kernels":

X.1 <u>Example</u>:
Let $(X,\Sigma,\mu)$ be a probability space and consider a positive measurable function $k(.,.)$ on $X\times X$ such that

$$\int k(x,y)\,d\mu(y) = 1 \qquad \text{for almost all } x\in X$$

and

$$\int k(x,y)\,d\mu(x) = 1 \qquad \text{for almost all } y\in X.$$

Then $$Tf(x) := \int k(x,y)f(y)d\mu(y), \quad f \in L^1(X,\Sigma,\mu),$$

defines a bi-Markov operator on $L^1(X,\Sigma,\mu)$.
The operators of the above form are called <u>kernel operators</u> and we refer to Schaefer [1974], IV.9 for a detailed investigation.

In this appendix we shall present a systematic account on the asymptotic behavior of bi-Markov operators. More precisely, we investigate the following.

X.2 <u>Problem and conjecture</u>:
Let $(X,\Sigma,\mu)$ be a probability space and T a bi-Markov operator on $L^1(X,\Sigma,\mu)$. Moreover, we assume in most occasions that T is <u>irreducible</u>, i.e. the fixed space

$$F(T) := \{f \in L^1 : Tf = f\}$$

is one-dimensional (compare III.D.11). We ask the following question: Under which conditions and with respect to which of the standard operator topologies does $T^n$ converge as $n \to \infty$ ? Motivated by many results in the previous lectures (e.g. Lecture IX) we conjecture that this convergence is determined by that part of the spectrum of T which is situated on the unit circle $\Gamma$.

Before dealing with the convergence of the powers $T^n$ it might be useful to recall the results from Lecture IV and App.W on the convergence of the Cesaro means $T_n := \frac{1}{n}\sum_{i=0}^{n-1} T^i$.

X.3 <u>Theorem</u> (convergence of means):
Let T be an irreducible bi-Markov operator on $L^1(X,\Sigma,\mu)$. Then
(i) T is always mean ergodic and $T_n$ converges strongly (= weakly) to the procetion $P = \mathbb{1} \otimes \mathbb{1}$.
(ii) T is uniformly ergodic, i.e. $T_n$ converges uniformly if and only if 1 is a pole of $R(\lambda,T)$.

The assertion (i) has been proved in (IV.6) and is the only convergence property holding in general; (ii) is a rather simple equivalence treated in (App.W.3). More important is the following consequence for the structure of the spectrum of T (see W.11 or Schaefer [1974], V.5.2).

X.4 Corollary:
If $T$ is an irreducible bi-Markov operator on $L^1(X,\Sigma,\mu)$ which is uniformly ergodic, then $\sigma(T)\cap\Gamma$ is the group $\Gamma_k$ of all k-th roots of unity for some $k\in\mathbb{N}$ and every $\lambda\in\Gamma_k$ is an eigenvalue of $T$.

For our considerations the above corollary shows that - as soon as $k>1$ - we may have uniform convergence of the Cesaro means $T_n$ while the powers $T^n$ do not converge even in the weak operator topology. In fact, if $\lambda$ is a primitive k-th root of unity and

$$Tf = \lambda f$$

for $0\neq f\in L^1(\mu)$, then $T$ defines a cyclic permutation on $\{f, \lambda f,\dots, \lambda^{k-1}f\}$ and therefore $T^nf$ does not converge. On the other hand one might still hope that the powers of $T$ behave asymptotically as such a cyclic permutation on the eigenfunctions of $T$. The subsequent concept will help us to pursue this idea.

X.5 Definition:
An operator $S\in\mathcal{L}(E)$, $E$ Banach space, is called partially periodic if there exists $n_o\in\mathbb{N}$ such that $S(Id - S^{n_o}) = 0$.

The above statement implies that $S^{n_o}(Id - S^{n_o}) = 0$, hence $S^{n_o}$ is a projection and therefore an equivalent property is the following:
$E$ is the direct sum of two closed $S$-invariant subspaces $E_o$ and $E_p$ such that $S|_{E_o} = 0$ and $(S|_{E_p})^{n_o} = Id_{E_p}$. In particular, a periodic operator is partially periodic with $E_o = \{0\}$.

Using this notion and Theorem (App.X.3.ii) we are able to describe the asymptotic behavior of an irreducible bi-Markov operator with respect to the uniform operator topology.

X.6 Theorem (uniform convergence):
Let $T$ be an irreducible bi-Markov operator on $L^1(X,\Sigma,\mu)$. The following are equivalent:
(a) 1 is a pole of $R(\lambda,T)$.
(b) There exists a partially periodic operator $S\in\mathcal{L}(L^1(\mu))$ such that $T^n - S^n \xrightarrow{\|\cdot\|} 0$ for $n\to\infty$.

Remark:
We repeat again that the assumption (a) referring to the spectral radius of $T$ implies that all other spectral values of $T$ on $\Gamma$ are simple poles of $R(\lambda, T)$ (Schaefer [1974], V.5.4).

Proof:
(a) $\Rightarrow$ (b): T is uniformly ergodic and even quasi-compact by (W.3) and (W.11). Moreover, we remarked in (App.X.4) that $\sigma(T) \cap \Gamma = \Gamma_{n_o}$ for some $n_o \in \mathbb{N}$. Therefore $R := T^{n_o}$ is uniformly ergodic with $\sigma(R) \cap \Gamma = \{1\}$. The corresponding projection $Q$ is a positive projection onto the fixed space of $R$ which coincides with the linear span of the eigenspaces of $T$ corresponding to the eigenvalues in $\Gamma_{n_o}$. Define the operator

$$S := T Q = Q T$$

which is positive and partially periodic, since $S^{n_o} = T^{n_o} Q = Q$. On the other hand $T^n - S^n = T^n(Id - Q) \quad (T(Id - Q))^n$ norm converges to zero as $n \to \infty$ since the spectral radius $r(T(Id - Q))$ is less than 1. This proves (b).
(b) $\Rightarrow$ (a): If S is partially periodic then $Q := S^{n_o}$ is a projection for some $n_o \in \mathbb{N}$. By assumption the powers of $T^{n_o}$ norm converge to Q, hence $T^{n_o}$ is uniformly ergodic. As in (IV.D.2) one deduces that $T$ is uniformly ergodic, too. ∎

The analogous result for weak convergence is slightly more complicated since we obtain - as in (IX.5) - convergence for a "dense" subsequence only. But still, its proof is based on the same arguments:
Consider $R := T^{n_o}$ for some appropriate $n_o \in \mathbb{N}$, apply the spectral mapping theorem and then use the weak mixing theorem (IX.D.6).
We leave the details to the reader.

X.7 <u>Theorem</u> (weak convergence):
Let $T$ be an irreducible bi-Markov operator on $L^1(X, \Sigma, \mu)$.
The following assertions are equivalent:

(a) 1 is isolated in $P\sigma(T) \cap \Gamma$.

(b) There exists a partially periodic positive operator $S \in \mathcal{L}(L^1(\mu))$ and a subsequence $\{n_i\} \subset \mathbb{N}$ with density 1 such that $T^{n_i} - S^{n_i} \to 0$ in the weak operator topology as $i \to \infty$.

Remark:

By the theorem in (III.D.11), (a) is equivalent to the fact that the peripheral point spectrum $P\sigma(T) \cap \Gamma$ is equal to some finite subgroup $\Gamma_{n_0}$ of $\Gamma$.

As can be seen from its proof the uniform convergence theorem (App.X.6) is essentially a consequence of the established and well known theory of uniformly ergodic and quasi-compact operators (see App.W). On the other hand the weak convergence theorem (App.X.7) is a corollary of the weak mixing theorem (see IX.5 and IX.D.6). The intermediate case of strong convergence is much less known, and for its investigation we have to develop new and very interesting tools. But first we state the theorem which expresses exactly what could be guessed by "interpolating" (App.X.6) and (App.X.7) with the difference that we can prove only the implication (a) $\Rightarrow$ (b).

X.8 Theorem (strong convergence):

Let $T$ be an irreducible bi-Markov operator on $L^1(X, \Sigma, \mu)$.

Then the condition (a) implies (b):

(a) 1 is isolated in $\sigma(T) \cap \Gamma$.

(b) There exists a partially periodic positive operator $S \in \mathcal{L}(L^1(\mu))$ such that $T^n - S^n \to 0$ in the strong operator topology as $n \to \infty$.

Remark:

By Schaefer [1974], V.4.6 the condition (a) is equivalent to the fact that the peripheral spectrum $\sigma(T) \cap \Gamma$ is equal to a finite union of finite subgroups of $\Gamma$.

As announced the proof of this theorem is long and complicated. Essentially it is based on a generalization of the "0-2 law" of Ornstein-Sucheston [1970] to non-irreducible Markov operators. We first present this theorem and only thereafter return to the proof of (App.X.8). The reader may already be advised that a good knowledge of the order-theoretical structure of the vector space of all operators on $L^1(X,\Sigma,\mu)$ will be necessary. All the needed information can be found in Schaefer [1974], IV.1: First we recall that for the Banach lattice $L^1(X,\Sigma,\mu)$ the space $\mathcal{L}(L^1(\mu))$ is a vector lattice with lattice operation $T \to |T|$ and with the infimum $S \wedge R$ of two positive operators S, R defined as in Schaefer [1974],p.229. In particular one has $|Sf| \leq |S|\,|f|$ and $|RS| \leq |R|\,|S|$ for $R, S \in \mathcal{L}(L^1(\mu))$ and $f \in L^1(\mu)$, and two operators are called <u>orthogonal</u> if $|R| \wedge |S| = 0$.

Now let $T$ be a bi-Markov operator on $L^1(X,\Sigma,\mu)$. Then the fixed space $F(T)$ is a vector sublattice of $L^1(\mu)$:

$$Tf = f \text{ implies } T|f| \geq |f|.$$

Since $\langle T|f|, 1\!\!1\rangle = \langle |f|, T'1\!\!1\rangle = \langle |f|, 1\!\!1\rangle$ we have $T|f| = |f|$.
For $k, n \in \mathbb{N}$ we obtain

$$|T^n(Id - T^k)|$$

as a positive operator on $L^1(\mu)$. Choose $0 < v \in F(T)$ and consider the sequence $(|T^n(Id - T^k)|\, v)_{n\in\mathbb{N}}$. It is decreasing since

$$0 \leq |T^{n+1}(Id-T^k)|\, v \leq |T^n(Id-T^k)|\, Tv = |T^n(Id-T^k)|\, v \leq 2v,$$

and its infimum $w := \inf\limits_{n\in\mathbb{N}} |T^n(Id-T^k)|\, v$ satisfies

$$w \leq Tw.$$

As above we conclude that subinvariant elements are invariant and hence $w \in F(T)$.

X.9 <u>Lemma</u>:
Let $k \in \mathbb{N}$ be fixed and choose $0 \leq v, w \in F(T)$.

(i) The following are equivalent:
  (a) $|T^n(Id-T^k)|\, v \downarrow w$.
  (b) $(T^n \wedge T^{n+k})v \uparrow v - \frac{1}{2} w$.

(ii) The following are equivalent:
  (c) $|T^n(Id-T^k)|\, v = 2v$
  (d) $(T^n \wedge T^{n+k})v = 0$.

Proof:

The assertions follow from the identity valid in vector lattices (Schaefer [1974], II.1.4, Cor.1):

$$(T^n \wedge T^{n+k})v = \frac{1}{2}\left((T^n + T^{n+k})v - |T^n - T^{n+k}|v\right)$$
$$= v - \frac{1}{2}\left|T^n(Id - T^n)\right|v. \quad ■$$

The next lemma contains a simple technical statement, which will be useful in the proof of our "zero-two" law.

X.10 Lemma:

For $k \in \mathbb{N}$ fixed, define $S_n := T^n \wedge T^{n+k}$ and suppose that $\lim_{n\to\infty} S_n v > 0$ for every $0 < v \in F(T)$. Then $\lim_{n\to\infty} S_n{}^m v > 0$ for every $m \in \mathbb{N}$ and $0 < v \in F(T)$. Moreover, the fixed space $F(T^r)$ is contained in $F(T^k)$ for every $r \in \mathbb{N}$.

Proof:

We know that $0 < \lim_{n\to\infty} S_n v =: v_1 \in F(T)$ and $0 < \lim_{n\to\infty} S_n v_1 =: v_2 \in F(T)$.
Therefore

$$S_n{}^2 v - v_2 = S_n{}^2 v - S_n v_1 + S_n v_1 - v_2$$

converges to zero, and the assertion is proved for $m = 2$. Repeating the same argument proves the first assertion for $m > 2$.
Now suppose that $F(T^r)$ is not contained in $F(T^k)$ for some $r \in \mathbb{N}$.
If we denote by $T_1$ the restriction of $T$ to the closed sublattice $F(T^r)$ this assumption implies $T_1{}^k \neq Id_{F(T^r)} = T_1{}^r$.
As in Rohlin's lemma (X.2.i) for some $m \in \mathbb{N}$, $k \neq 0 \bmod m$, one finds pairwise orthogonal positive functions $e_1, \dots, e_m \in F(T^r)$ such that $Te_i = e_{i+1}$ for every $i \bmod m$ and $k \neq 0 \bmod m$. Define $e := \sum_{i=1}^{m} e_i$ and observe that $Te = e = |e - 2e_i|$. Therefore

$$|T^n - T^{n+k}|e \geq |(T^n - T^{n+k})(e - 2e_i)| = 2\,|e_{i+n+k} - e_{i+n}|$$
$$= 2(e_{i+n+k} + e_{i+n}) \geq 2e_{i+n} \quad \text{for every } i \bmod m.$$

This implies $|T^n - T^{n+k}|e \geq 2 \cdot \sup\{e_i : 1 \leq i \leq m\} = 2e$.
By (App.X.9) we conclude that $S_n e = 0$ thereby contradicting the hypothesis of our lemma. ■

After these preparations we state the announced "0-2 law" due to Ornstein-Sucheston [1970] for irreducible operators and extended by Greiner-Nagel [1982] to positive operators on Banach lattices with order continuous norm.

X.11 Theorem ("0-2 law"):
Let $T$ be a bi-Markov operator on $L^1(X,\Sigma,\mu)$. For every $k\in\mathbb{N}$ there exists $1\!\!1_A\in F(T)$, $A\in\Sigma$, such that

(0) $\quad |T^n(Id - T^k)|\ 1\!\!1_{X\setminus A} \downarrow 0$ as $n\to\infty$,

and

(2) $\quad |T^n(Id - T^k)|\ 1\!\!1_A = 2\cdot 1\!\!1_A \quad$ for every $n\in\mathbb{N}$.

Remark:
Obviously, the characteristic function $1\!\!1_A$ of the above theorem depends on $k\in\mathbb{N}$ and generates a T-invariant projection band isomorphic to $L^1(A,\Sigma_A,\mu_A)$. Lemma (App.X.9) shows that the operators $T^n$ and $T^{n+k}$ restricted to this band are orthogonal for every $n\in\mathbb{N}$.

Proof:
Throughout the proof we keep $k\in\mathbb{N}$ fixed and set

$$S_n := T^n \wedge T^{n+k}.$$

Now choose $A\in\Sigma$ maximal such that $1\!\!1_A\in F(T)$ and $S_n 1\!\!1_A = 0$ for every $n\in\mathbb{N}$. For the following we may suppose that $A=\emptyset$, i.e.

$$\lim_{n\to\infty} S_n v > 0 \quad \text{and} \quad \lim_{n\to\infty} S_n^m v > 0$$

for every $0< v\in F(T)$ and $m\in\mathbb{N}$ (apply App.X.9 and App.X.10).
Then we consider

$$w := \inf_{n\in\mathbb{N}} |T^n(Id - T^k)|\, 1\!\!1 \in F(T)$$

and show that

$$w \le \frac{2}{\sqrt{m}}\cdot 1\!\!1$$

for every $m\in\mathbb{N}$.
Suppose to the contrary that there exists $m\in\mathbb{N}$ such that

$$w_1 := \left(w - \frac{2}{\sqrt{m}}\,1\!\!1\right)^+ > 0.$$

Since $F(T)$ is a sublattice of $L^1(\mu)$, $w_1$ is contained in $F(T)$ and there exists $n_o \in \mathbb{N}$ such that

$$(*) \qquad S^m_{n_o} w_1 > 0.$$

Finally we define an operator

$$U := T^r - S^m_{n_o}\left(\frac{Id+T^k}{2}\right)^m ,$$

where $r := m(n_o + k)$. This operator is positive since

$$S_{n_o}\left(\frac{Id+T^k}{2}\right) \leq \tfrac{1}{2}\left(T^{n_o+k} + T^{n_o+k}\right)$$

and therefore

$$S^m_{n_o}\left(\frac{Id+T^k}{2}\right)^m \leq S^{m-1}_{n_o}\, T^{n_o+k}\left(\frac{Id+T^k}{2}\right)^{m-1} \leq \cdots\cdots \leq T^{m(n_o+k)} .$$

Beginning with

$$\begin{aligned} T^{2r} &= \left(U + S^m_{n_o}\left(\frac{Id+T^k}{2}\right)^m\right)T^r \\ &= UT^r + S^m_{n_o}\, T^r\left(\frac{Id+T^k}{2}\right)^m \\ &= U^2 + \left(US^m_{n_o} + S^m_{n_o}\, T^r\right)\left(\frac{Id+T^k}{2}\right)^m \end{aligned}$$

one deduces that

$$T^{jr} = U^j + R\left(\frac{Id+T^k}{2}\right)^m$$

for $j \in \mathbb{N}$ and a certain positive operator $R$ depending on $j$ and satisfying

$$0 \leq R\mathbb{1} \leq U^j\mathbb{1} + R\mathbb{1} = T^{jr}\mathbb{1} = \mathbb{1} .$$

Consequently,

$$\begin{aligned} T^{jr}(Id - T^k) &= U^j(Id - T^k) + R\cdot 2^{-m} \sum_{\ell=0}^{m} \binom{m}{\ell} T^{k\ell}(Id - T^k) \\ &= U^j(Id - T^k) + R\cdot 2^{-m} \sum_{\ell=0}^{m+1} \left[\binom{m}{\ell} - \binom{m}{\ell-1}\right] T^{k\ell} . \end{aligned}$$

By taking the absolute values we obtain

$$\left|T^{jr}(Id - T^k)\right|\mathbb{1} \leq 2U^j\mathbb{1} + R\frac{2}{\sqrt{m}}\mathbb{1} \leq 2U^j\mathbb{1} + \frac{2}{\sqrt{m}}\mathbb{1} ,$$

where the estimate for $2^{-m} \sum_{\ell=0}^{m+1} \left|\binom{m}{\ell} - \binom{m}{\ell-1}\right|$ follows from Stirling's formula.

The definition of $w$ and the above computation shows that

$$w \leq \frac{2}{\sqrt{m}} 1\!\!1 + 2U^j 1\!\!1$$

for every $j \in \mathbb{N}$.

Let Q denote the band projection onto the T-invariant band in $L^1(X, \Sigma, \mu)$ generated by $w_1$. The component

$$v := Q1\!\!1$$

of $1\!\!1$ is contained in $F(T)$ and

$$w_1 = Qw_1 = (Qw - \frac{2}{\sqrt{m}} v)^+ = Qw - \frac{2}{\sqrt{m}} v .$$

The band projection $Q$ commutes first with $T^n$, then with $S_n$ which is dominated by $T^n$, and finally with $U$. Therefore

$$w_1 + \frac{2}{\sqrt{m}} v = Qw \leq \frac{2}{\sqrt{m}} v + 2U^j v$$

or

$$w_1 \leq 2U^j v$$

for every $j \in \mathbb{N}$. But $(U^j v)_{j \in \mathbb{N}}$ is a decreasing sequence since $v \geq v - S^m_{n_o} v = Uv$. Let $v_1$ be its limit. Then $v_1$ is fixed under $U$ and hence $T^r v_1 \geq v_1$. Again, $T' 1\!\!1 = 1\!\!1$ implies $T^r v_1 = v_1$ and therefore $T^k v_1 = v_1$ by the second part of Lemma (App.X.10). The invariance of $v_1$ by $U$, $T^r$ and $T^k$ yields $S^m_{n_o} v_1 = 0$. Since $w_1 \leq 2v_1$ and $0 \leq S^m_{n_o} w_1 \leq 2S^m_{n_o} v_1 = 0$ we end up with a contradiction to the inequality (*). ■

This "0-2 law" yields a decomposition of the bi-Markov operator $T$ into two parts with extremally different behavior. In particular, if $T$ is irreducible we have the impressive alternative that $T$ satisfies either $|T^n(Id - T^k)| 1\!\!1 \downarrow 0$ or $|T^n(Id - T^k)| 1 = 2 \cdot 1\!\!1$. These two properties may be interpreted using the probabilistic meaning of $T^n 1\!\!1_A(x)$ described at the beginning of this appendix.
In particular, for $k = 1$ we have the following alternatives (0) or (2):

X.12 <u>Probabilistic interpretation</u>:

(0) $\| T^n(Id - T)| \mathbb{1} \downarrow 0$ means by (App.X.9.i) that for almost every $x \in X$ and for $\varepsilon > 0$ there exists $n_o \in \mathbb{N}$ such that

$$T^n \wedge T^{n+1} \mathbb{1}(x) \geq 1 - \varepsilon$$

for $n \geq n_o$. By the formula determining $T^n \wedge T^{n+1}$ this is equivalent to

$$T^n \mathbb{1}_B(x) + T^{n+1} \mathbb{1}_{X \setminus B}(x) \geq 1 - \varepsilon$$

for every $B \in \Sigma$ and $n \geq n_o$. Since

$$T^n \mathbb{1}_B(x) + T^n \mathbb{1}_{B \setminus A}(x) = T^{n+1} \mathbb{1}_B(x) + T^{n+1} \mathbb{1}_{X \setminus B}(x) = 1$$

we deduce that

$$|T^{n+1} \mathbb{1}_B(x) - T^n \mathbb{1}_B(x)| \leq \varepsilon$$

for every $B \in \Sigma$ and $n \geq n_o$.

In probabilistic terms this means that for the corresponding Markov process when starting in $x$ and for every $B \in \Sigma$ the probabilities of being in $B$ at time $n$ or at time $n+1$ are almost equal (for $n$ sufficiently large).

(2) $(T^n \wedge T^{n+1})\mathbb{1} = 0$ for every $n \in \mathbb{N}$ implies by the formula defining the infimum of two positive operators, that for (almost) every $x \in X$, $n \in \mathbb{N}$, we can construct $B \in \Sigma$ such that

$$T^n \mathbb{1}_B(x) + T^{n+1} \mathbb{1}_{X \setminus B}(x) = 0$$

or $$T^n \mathbb{1}_B(x) = 0 \quad \text{and} \quad T^{n+1} \mathbb{1}_B(x) = 1.$$

By the same interpretation as above this means that $B$ distinguishes between the time $n$ and $n+1$, i.e. $T$ is "time determining" (compare Ornstein-Sucheston [1970]).

After this excursion into probability theory we will use the "0-2 law" for the proof of Theorem (App.X.8). But before doing so we sketch how the two things, i.e. the "0-2 law" and the strong convergence of the powers of a bi-Markov operator, are related:

Let $R \in \mathcal{L}(L^1(\mu))$ be a bi-Markov operator satisfying $F(R) = F(R^k)$ for every $k \in \mathbb{N}$.

Obviously, $R$ as any power $R^k$ is mean ergodic and we have

$$L^1(\mu) = F(R^k) \oplus \overline{(Id-R^k)L^1(\mu)} = F(R) \oplus \overline{(Id - R^k)L^1(\mu)}.$$

Therefore it suffices to show the strong convergence of $R^n$ on a total subset of $\overline{(Id-R^k)L^1(\mu)}$ for some $k\in\mathbb{N}$.
To that purpose take $0\leq f\leq 1\!\!1$ and observe that

$$|R^n(Id-R^k)|\,1\!\!1 \geq |R^n(Id-R^k)|f \geq |R^n(Id-R^k)f|.$$

Consequently, $R^n$ converges strongly on $L^1(\mu)$ as soon as $\lim_{n\to\infty} |R^n(Id-R^k)|\,1\!\!1 = 0$ for some $k\in\mathbb{N}$. This shows that a proof of the strong convergence theorem is achieved as soon as for some $k\in\mathbb{N}$ we can exclude the property (2) in (App.X.11).

This property (2) is a statement on the orthogonality of certain operators by (App.X.9.ii), and it is not at all clear how a spectral condition such as (App.X.8.a) could have some consequences in this direction. The following lemma bridges this gap.

X.13 Lemma:

Let $R$ denote a bi-Markov operator on $E = L^1(X,\Sigma,\mu)$.
Suppose that for every $n\in\mathbb{N}$ there exists a T-invariant projection band $E_n \neq \{0\}$ in E such that the operators

$$\{Id|_{E_n}, R|_{E_n}, \ldots, R^n|_{E_n}\}$$

are pairwise orthogonal. Then the spectrum $\sigma(R)$ contains the unit circle $\Gamma$.

Proof:
Assume that $\alpha\in\Gamma$ is not contained in $\sigma(R)$ and consider

$$S := \frac{1}{2}(Id + \bar{\alpha}R).$$

By the spectral mapping theorem, $r(S) < 1$ and therefore $\|S^{n_o}\| < 1$ for some $n_o\in\mathbb{N}$.
On the other hand, we apply the hypothesis of the lemma and assume without loss of generality that $E_{n_o} = E$, i.e. $\{Id, R, \ldots, R^{n_o}\}$ are pairwise orthogonal.

Therefore we obtain

$$|S^{n_o}| = 2^{-n_o} \Big| \sum_{i=0}^{n_o} \binom{n_o}{i} \overline{\alpha}^i R^i \Big|$$

$$\leq 2^{-n_o} \sum_{i=0}^{n_o} \binom{n_o}{i} R^i = \left(\frac{Id + R}{2}\right)^{n_o} .$$

Again it follows from the spectral mapping theorem that $1 \in \sigma(|S^{n_o}|)$, hence $1 \leq \| |S^{n_o}| \|$. But for operators on $L^1(\mu)$ we know that $\| S^{n_o} \| = \| |S^{n_o}| \|$ (Schaefer [1974], IV.1.5) contradicting the assumption $\alpha \in \rho(R)$. ∎

X.14 <u>Proof of the strong convergence theorem (App.X.8)</u>:

The peripheral spectrum of $T$ is cyclic (by Schaefer [1974], V.4.6) and finite (by hypothesis). Therefore

$$\sigma(R) \cap \Gamma = \{1\}$$

for $R := T^{n_o}$ and some $n_o \in \mathbb{N}$. As in the proof of the previous convergence theorems (App.X.6) and (App.X.7) it suffices to show that $R^n$ converges strongly as $n \to \infty$.

First of all we observe that $F(R) = F(R^k)$ for every $k \in \mathbb{N}$.

Next we apply the 0-2 law and find $1\!\!1_{A_k} \in F(R)$ such that

$$R^n \wedge R^{n+k} 1\!\!1_{A_k} = 0$$

and $$|R^n(Id - R^k)| 1\!\!1_{X \setminus A_k} \downarrow 0.$$

Denote by $E_k$ the R-invariant projection band generated by $1\!\!1_{A_k}$ and define

$$E^{(n)} := E_1 \cap E_2 \cap \ldots \cap E_n .$$

Then it is clear that

$$R^r|_{E^{(n)}} \text{ is orthogonal to } R^s|_{E^{(n)}}$$

for $0 \leq r < s \leq n$.

On the other hand, Lemma (App.X.13) implies the existence of $m_o \in \mathbb{N}$ such that for every non-trivial R-invariant projection band $B$ in $L^1(\mu)$ we can find $0 \leq r < s \leq m_o$ such that

$$R^r|_B \wedge R^s|_B > 0.$$

Combining both observations we obtain

$$E^{(m_o)} = \{0\}$$

or
$$E_1^{\perp} + E_2^{\perp} + \ldots + E_{m_o}^{\perp} = E.$$

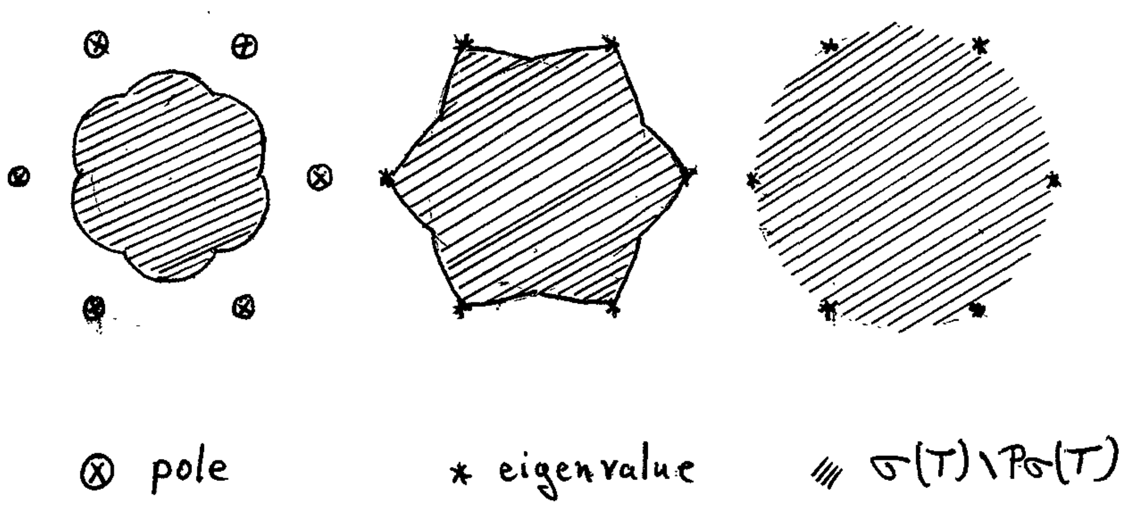
But the powers of $R$ restricted to $E_k^{\perp}$ converge by (App.X.11.0) and the considerations preceeding Lemma (App.X.13). Therefore, the proof of Theorem (App.X.8) is finished. ■

## X.15 Concluding Examples:

1.

For an irreducible bi-Markov operator $T$ on $L^1(X, \Sigma, \mu)$ the peripheral point spectrum is a subgroup of $\Gamma$ and $\sigma(T)$ is invariant under rotation by these eigenvalues:

Therefore the following spectra are prototypes of $\sigma(T)$ as appearing in the uniform, strong and weak convergence theorem:



2.

The spectrum of an operator $T_\varphi$ induced by an ergodic MDS $([0,1],\mathcal{B},m;\varphi)$, m the Lebesgue measure, always contains the unit circle (see VI.D.3), and the powers of $T_\varphi$ never converge in the strong operator topology (see IX.D.3). Therefore "weak convergence" is the best we can expect for $T_\varphi^n$. In addition, Property (0) of the 0-2 law holds for $\mathbb{1}_A = 0$ only, and we show that $\{T_\varphi^n : n \in \mathbb{N}_0\}$ are pairwise orthogonal: Let $n,m \in \mathbb{N}_0$ be fixed and assume $m < n$.

By Rohlin's lemma (X.2) and for every $k \in \mathbb{N}$ we obtain measurable sets $A_k \in [0,1]$ such that

(i) the sets $A_k, \varphi(A_k),\dots, \varphi^n(A_k)$ are pairwise disjoint, and

(ii) $m(\bigcup_{i=0}^{n} \varphi^i(A_k)) > 1 - \frac{1}{k}$ .

From (i) it follows immediately that also the sets

$$\varphi^{-i}(A_k),\ \varphi^{-i+1}(A_k),\dots,\ \varphi^{-i+n}(A_k)$$

are pairwise disjoint for every $i \in \{0,1,\dots,n\}$.

We therefore conclude

$$\begin{aligned} 0 \le (T_\varphi^m \wedge T_\varphi^n)\mathbb{1}_{\varphi^i(A_k)} &\le T_\varphi^m \mathbb{1}_{\varphi^i(A_k)} \wedge T_\varphi^n \mathbb{1}_{\varphi^i(A_k)} \\ &= \mathbb{1}_{\varphi^{i-m}(A_k)} \wedge \mathbb{1}_{\varphi^{i-n}(A_k)} \\ &= 0 \end{aligned}$$

for $i = 0,\dots,n$ and every $k \in \mathbb{N}$.

Finally, the inequalities in (ii) yield that the lattice ideal generated by the functions $\mathbb{1}_{\varphi^i(A_k)}$, $k \in \mathbb{N}$ and $i = 0,\dots,n$, is dense in $L^1(m)$. Consequently we obtain $T_\varphi^m \wedge T_\varphi^n = 0$ for $n,m \in \mathbb{N}_0$ and $n \neq m$.

3.

Uniform convergence is obtained if the bi-Markov operator (or some power) on $L^1(X,\Sigma,\mu)$ is compact or even only quasi-compact. Typical examples are kernel operators

$$Tf(x) := \int k(x,y)f(y)d\mu(y)$$

defined by a bounded kernel $0 \leq k \in L^\infty(X \times X, \mu \otimes \mu)$. Further compactness criteria for kernel operators may be found in Schaefer [1974], IV.10.

4.
Not every kernel operator on $L^1(X, \Sigma, \mu)$ is (quasi-)compact. But if it is irreducible and bi-Markov, then its peripheral point spectrum is a finite subgroup of $\Gamma$, hence condition (a) of the weak convergence theorem is satisfied. But surprisingly the much stronger conclusion (b) of the strong convergence theorem is valid. The proof of this result uses deep structure theorems on operators on Banach lattices and we refer to Greiner-Nagel [1982].

5.
The spectra of interesting Markov operators are computed in Vere-Jones [1963].

References: Bonsdorff [1980], Derriennic [1976], Drey [1971], Foguel [1976], [1979], Greiner-Nagel [1982], Lin [1982], Lin-Sine [1979], Luecke [1977], Ornstein-Sucheston [1970], Yoshida-Kakutani [1941].

## Appendix Y: Mean Ergodic Operator Semigroups

In our lectures the dynamics of a system has been described by a single mapping and its powers, but in many situations it is useful to consider systems in which the action is performed by an arbitrary semigroup of mappings. Many concepts and problems of ergodic theory can be extended to this context, but in this appendix we only discuss the generalizations of the mean ergodic theorem (see Lecture IV). Thereby our combined functional-analytic and semigroup-theoretical approach (see Lectures VII, VIII and IX) proves to be most useful.
The objects we are investigating in this appendix are a (real or complex) Banach space $E$ and a semigroup $\mathcal{S}$ of bounded linear operators on $E$ with adjoint semigroup $\mathcal{S}' = \{T' \in \mathcal{L}(E') : T \in \mathcal{S}\}$. On $\mathcal{L}(E)$ and therefore on $\mathcal{S}$ we consider the strong or weak operator topology, both making $\mathcal{S}$ into a semitopological semigroup (see VII.1).
The following is a very abstract version of the property appearing in the classical mean ergodic theorem (IV.1).

**Y.1** <u>Definition</u>:

The semigroup $\mathcal{S} \subset \mathcal{L}(E)$ is called <u>mean ergodic</u> if its closed convex hull

$$\overline{co}\,\mathcal{S}$$

in $\mathcal{L}_s(E)$ contains a zero element $P$, i.e. $TP = PT = P$ for every $T \in \overline{co}\,\mathcal{S}$.
Since the weak and strong operator topologies possess the same closed convex sets in $\mathcal{L}(E)$, we could take $\overline{co}\,\mathcal{S}$ in $\mathcal{L}_w(E)$ as well and would still obtain the same concept of mean ergodic semigroups. From the above definition we first draw some simple but important consequences. Then we prove equivalent characterizations leading to interesting examples.

Y.2 Proposition:

Let $\mathcal{S}$ be a mean ergodic semigroup in $\mathcal{L}(E)$ with zero element P. Then

(i) $Pf \in \overline{co}\,\mathcal{S}f$ and $P'f' \in \overline{co}\,\mathcal{S}'f'$ for every $f \in E$, $f' \in E'$ and the weak*- topology on $E'$.

(ii) P is a projection onto the fixed space $F := \{f \in E : Tf = f \text{ for all } T \in \mathcal{S}\}$. Its adjoint $P'$ projects $E'$ onto the dual fixed space $F' := \{f' \in E' : T'f' = f' \text{ for all } T' \in \mathcal{S}'\}$, and $(PE)'$ is (as a topological vector space) isomorphic to $P'E'$.

(iii) $P^{-1}(0)$ is the closed linear hull of $\{(Id - T)f : f \in E, T \in \mathcal{S}\}$.

Proof:

(i) The continuity of $T \mapsto Tf$ from $\mathcal{L}_s(E)$ into E implies that $Pf \in \overline{co}\,\{Tf : T \in \mathcal{S}\} = \overline{co}\,\mathcal{S}f$. Analogously, the continuity of $T \mapsto T'f'$ from $\mathcal{L}_w(E)$ into $(E', \sigma(E',E))$ shows that $P'f' \in \overline{co}\,\mathcal{S}'f'$.

(ii) $Pf \in \overline{co}\,\mathcal{S}f$ and $TP = PT = P^2 = P$ for all $T \in \mathcal{S}$ implies that P is a projection onto F. Similarly, one shows that $P'E' = F'$. The isomorphism of $(PE)'$ and $P'E'$ is proved in (App.B.7).

(iii) is left to the reader. ∎

It is surprising that for bounded semigroups the mean ergodicity can be characterized by a fixed point property.

Y.3 Theorem:

For a bounded semigroup $\mathcal{S}$ in $\mathcal{L}_s(E)$ the following conditions are equivalent:

(a) $\mathcal{S}$ is mean ergodic.

(b) $\overline{co}\,\mathcal{S}f$ contains a fixed point of $\mathcal{S}$ for every $f \in E$ and $\overline{co}^{\sigma^*}\mathcal{S}'f'$ contains a fixed point of $\mathcal{S}'$ for every $f' \in E'$.

Proof:

(a) $\Rightarrow$ (b): This has been proved in (App.Y.2.i).

(b) $\Rightarrow$ (a): $\overline{co}\,\mathcal{S}f \cap F$ is non-empty by the assumption on $\mathcal{S}$ and contains at most one element by the assumption on $\mathcal{S}'$.

Define $P : f \mapsto f_0 \in \overline{co}\,\mathcal{S}f \cap F$ for $f \in E$. We show that P is linear, i.e. homogenous and additive. While the first assertion is clear, we have to prove the additivity, i.e. $f_o \in \overline{co}\,\mathcal{S} f \cap F$ and $g_o \in \overline{co}\,\mathcal{S} g \cap F$ implies $f_o + g_o \in \overline{co}\,\mathcal{S}(f+g)$: For $\varepsilon > 0$ there exists $R \in co\,\mathcal{S}$ such that $\|Rf - f_o\| < \varepsilon$. Since $g_o \in \overline{co}\,\mathcal{S} Rg \subset \overline{co}\,\mathcal{S} g$ we find $S \in co\,\mathcal{S}$ such that $\|SRG - g_o\| < \varepsilon$. Together, this implies

$$\|SR(f+g) - (f_o + g_o)\| \leq \|SRf - f_o\| + \|SRg - g_o\| \leq \|S\|\varepsilon + \varepsilon .$$

Since P is obviously bounded we have found $P \in \mathcal{L}(E)$ satisfying $TP = PT = P$ for $T \in \mathcal{S}$. It remains to show that P is contained in $\overline{co}\,\mathcal{S}$, i.e. for $\varepsilon > 0$ and $\{f_1, \dots, f_n\} \subset E$ we have to find $S_n \in co\,\mathcal{S}$ such that $\|S_n f_i - Pf_i\| \leq \varepsilon$ for $i = 1, \dots, n$ : We proceed by induction and assume that there exists $S_{n-1} \in co\,\mathcal{S}$ satisfying

$\|S_{n-1} f_i - Pf_i\| \leq \varepsilon/c$ for $i = 1, \dots, n-1$ and $c := \sup\{\|T\|: T \in \mathcal{S}\}$.

Define $g := S_{n-1} f_n$.

By assumption there exists $T_n \in co\,\mathcal{S}$ such that $\|T_n g - Pg\| \leq \varepsilon$.

The operator $S_n := T_n \cdot S_{n-1}$ is contained in $co\,\mathcal{S}$, and the assertion follows from

$$S_n f_i - Pf_i = T_n(S_{n-1} f_i - Pf_i) \quad \text{for } i = 1, \dots, n-1,$$

and

$$S_n f_n - Pf_n = T_n g - Pg . \qquad \blacksquare$$

The above characterization has a first application to Banach spaces with uniformly convex norm. In such spaces every closed convex set contains a unique element having minimal norm (Dunford-Schwartz [1958], II.4.29). Therefore, if the operator semigroup $\mathcal{S}$ is contractive, this element in $\overline{co}\,\mathcal{S} f$ has to be fixed under $\mathcal{S}$. If the same holds in the dual space we obtain mean ergodicity of $\mathcal{S}$ and thereby retrieve a classical result due to Alaoglu-Birkhoff [1940] for Hilbert spaces.

Y.4 <u>Corollary</u>:

If E and $E'$ have uniformly convex norm (e.g. $E = L^p(\mu), 1 < p < \infty$) and $\mathcal{S} \subset \mathcal{L}(E)$ is a contractive semigroup, then $\mathcal{S}$ is mean ergodic.

In this application the particular norm structure of E and $E'$ implies mean ergodicity for arbitrary contractive semigroups. For more general Banach spaces we have to enforce our assumptions on the semigroup. The adequate conditions are of topological ("compactness") and algebraic ("amenability") nature.

Y.5 Definition:

A semitopological semigroup S is called left amenable, if the space $C_b(S)$ of all continuous bounded functions on S has a left-invariant mean, i.e. there exists $0 \leq \mu \in C_b(S)'$, $\langle 1\!\!1, \mu \rangle = 1$, such that $L_s' \mu = \mu$ for all left rotations $L_s f(t) = f(st)$ for $s, t \in S$ and $f \in C_b(S)$.

Right amenability is defined analogously using the right rotations $R_s$, and S is called amenable if $C_b(S)$ possesses a left- and right-invariant mean.

Amenable semigroups have been studied extensively, and we refer to Day [1969] and Greenleaf [1969] for further information. But before stating the main lemma containing the fixed point property of amenable operator semigroups we mention important classes of examples.

Y.6 Examples:

(1) Every abelian semigroup - if endowed with the discrete topology - is amenable. This may be proved using the Markov-Kakutani fixed point theorem (Schaefer [1974], III.7.12).

(2) Every compact group is amenable, since the Haar measure is an invariant mean.

(3) The free group with two generators is not amenable (Greenleaf [1969], 1.2.3).

Y.7 Lemma:

Let $\mathcal{S} \in \mathcal{L}(E)$ be a bounded semigroup.

(i) If $\mathcal{S}$ is right-amenable, then $\overline{\text{co}}\, \mathcal{S}' f'$ in $(E', \sigma(E', E))$ contains an $\mathcal{S}'$-fixed point for every $f' \in E'$.

(ii) If $\mathcal{S}$ is left-amenable, then $\overline{\text{co}}\, \mathcal{S} f$ in $(E'', \sigma(E'', E'))$ contains an $\mathcal{S}''$-fixed point for every $f \in E$.

Proof:

(i) Choose $f' \in E'$ and define by

$$f \mapsto \{T \mapsto \langle Tf, f' \rangle\}$$

a continuous linear map from E into $C_b(\mathcal{S})$, whose adjoint transforms the right-invariant mean $\mu \in C_b(\mathcal{S})'$ into an element $f_o' \in E'$. Since $\mu$ is contained in the weak$^*$-closed convex hull of the Dirac measures $\delta_T$, $T \in \mathcal{S}$, and since $\delta_T$ is transformed into $T'f'$ we obtain

$f_o' \in \overline{co}\,\mathcal{S}'f'$. A short calculation shows that $T'f_o' = f_o'$ for every $T \in \mathcal{S}$.

(ii) Now fix $f \in E$ and define by

$$f' \longmapsto \{T \longmapsto \langle Tf, f'\rangle\}$$

a continuous linear map from $(E', \beta(E',E))$ into $C_b(\mathcal{S})$, whose adjoint maps the left-invariant mean $\nu \in C_b(\mathcal{S})'$ into an element $f_o'' \in E''$. As above one shows that $f_o'' \in \overline{co}\,\mathcal{S}f$ in $(E'', \sigma(E'',E'))$ and $T''f_o'' = f_o''$ for every $T \in \mathcal{S}$. ∎

Recalling (Y.3.b) and applying (Y.7.i) we see that for right amenable bounded semigroups $\mathcal{S} \subset \mathcal{L}(E)$ a property assuring fixed points in $\overline{co}\,\mathcal{S}f$ implies mean ergodicity. But the following theorem shows that even a very weak separation property for the fixed spaces $F$ and $F'$ is sufficient.

Y.8 <u>Theorem</u>:
For a bounded right amenable semigroup $\mathcal{S} \subset \mathcal{L}(E)$ the following are equivalent:

(a) $\mathcal{S}$ is mean ergodic.

(b) $\overline{co}\,\mathcal{S}f \cap F$ is non-empty for every $f \in E$.

(c) $F$ separates $F'$.

Proof:
The implications (a) $\Rightarrow$ (b) $\Rightarrow$ (c) follow from (Y.3) and (Y.7).
(c) $\Rightarrow$ (a): By Lemma (Y.7.i) and by assumption, $\overline{co}\,\mathcal{S}'f' \cap F'$ contains exactly one element $f_o'$ for every $f' \in E'$. Define $P' : E' \longrightarrow F'$ by $P'f' := f_o'$ and show that $P'$ is linear: Let $f', g' \in E'$ and $P'(f'+g') \in \overline{co}\,\mathcal{S}'(f'+g')$. Since $P'f' + P'g' \in F'$ and $F$ separates $F'$ we conclude $P'f' + P'g' = P'(f'+g')$.
Next, we observe that $P'$ is continuous for $\sigma(E',E)$ and $\sigma(F',F)$ on $F'$. But $F'$ is $\sigma(E',E)$-closed in $E'$, hence $\sigma(E',E)$ and $\sigma(F',F)$ coincide on the equicontinuous sets of $F'$. Therefore, $P' \in \mathcal{L}(E')$ is $\sigma(E',E)$-continuous and has a pre-adjoint $P \in \mathcal{L}(E)$. If we assume that $Pf \notin \overline{co}\,\mathcal{S}f$ for some $f \in E$, then exists $f' \in E'$ such that

$\langle Pf,f'\rangle = \langle f,P'f'\rangle \notin \overline{co}\,\{\langle Tf,f'\rangle : T\in\mathcal{S}\}$ which is a contradiction to $P'f'\in\overline{co}\,\mathcal{S}'f'$. Therefore, we obtain $Pf\in\overline{co}\,\mathcal{S}f$ and $\mathcal{S}$ is mean ergodic by (Y.3.b). ∎

If the bounded semigroup $\mathcal{S}\subset\mathcal{L}(E)$ is actually amenable there exist fixed points in $E'$ and in $E''$. Therefore, $\mathcal{S}$ is mean ergodic as soon as we know that the fixed points in $E''$ are already contained in E. Weak compactness of $\mathcal{S}$ (or stronger: reflexivity of E) are the appropriate and natural assumptions.

Y.9 Corollary:
Let $\mathcal{S}\subset\mathcal{L}(E)$ be a bounded amenable semigroup. If $\mathcal{S}f$ is weakly compact for all f in a total subset of E, then $\mathcal{S}$ is mean ergodic.

Proof:
By (B.12) and (B.6) it follows that $\overline{co}\,\mathcal{S}f$ is weakly compact for all $f\in E$ and therefore coincides with its $\sigma(E'',E')$-closure in $E''$.
By (Y.7.ii) there exists an $\mathcal{S}$-fixed point in $\overline{co}\,\mathcal{S}f$ and the assertion follows from (Y.8.b).

Y.10 Examples:
(1) Every compact group $\mathcal{S}\subset\mathcal{L}_s(E)$ is mean ergodic.
(2) Every abelian semigroup $\mathcal{S}\subset\mathcal{L}(E)$ which is relatively compact for the weak operator topology is mean ergodic.
(3) Every bounded amenable semigroup $\mathcal{S}\subset\mathcal{L}(E)$, E a reflexive Banach space, is mean ergodic.
(4) Let E be a Banach lattice with order continuous norm (e.g. $E = L^p(\mu)$, $1\leq p<\infty$) and $\mathcal{S}$ a semigroup of positive contractions on E. If there exists a quasi-interior point $u\in E_+$ and a strictly positive linear form $\mu\in E'_+$ such that $Tu\leq u$ and $T'\mu\leq\mu$ for all $T\in\mathcal{S}$, then $\mathcal{S}$ is mean ergodic. See (Schaefer [1974], V.8.4).

All mean ergodic semigroups appearing in the examples above are relatively weakly compact. But in contrast to examples (1) - (3) where the algebraic extra-condition "amenable" had to be added, it is the topological nature of the underlying Banach space alone which implies mean ergodicity in (Y.4) as well as in Example (4). The main ingredient in the proof of (Y.10 (4)) is the construction of an associate semigroup $\mathcal{S}_2$ which is contractive on a certain Hilbert space (see Schaefer [1974], p.346) and therefore always mean ergodic by (Y.4).

Using a similar idea we are able to prove another mean ergodic theorem without making additional algebraic assumptions on the semigroup. Here the underlying Banach spaces are (non-commutative) $W^*$-algebras and their pre-duals. For the necessary terminology and basic results we refer to Pedersen [1979], Sakai [1971] and Stratila-Zsido [1979]. The main ingredient in the proof of the subsequent non-commutative mean ergodic theorem will be a non-commutative analogue of the canonical embedding

$$L^\infty(X,\Sigma,\mu) \longrightarrow L^2(X,\Sigma,\mu) \longrightarrow L^1(X,\Sigma,\mu),$$

$(X,\Sigma,\mu)$ a probability space. To that purpose consider a $W^*$-algebra $\mathcal{A}$ with pre-dual $\mathcal{A}_*$ and faithful (=strictly positive) state $\varphi \in \mathcal{A}_*$. Let $\Delta$ denote the modular operator (see Pedersen [1979], 8.13.14) on the GNS-Hilbert space $\mathcal{H}$ with canonical cyclic vector $\xi \in \mathcal{H}$ corresponding to $\varphi$.

As an appropriate injection $j_1 : \mathcal{A} \to \mathcal{H}$ we define

$$j_1(x) := \Delta^{1/4} x \xi \quad .$$

Then it can be shown (for the following, see Groh-Kümmerer [1982]) that the adjoint of $j_1$ defines an injection $j_2 : \mathcal{H} \to \mathcal{A}_*$ satisfying $j_2(\xi) = \varphi$. If $\mathcal{A}$ and $\mathcal{A}_*$ are ordered by their natural cones and if $\mathcal{H}$ is ordered by the (self dual) cone

$$\mathcal{P} := \overline{\{\Delta^{1/4} x \xi \ : x \in \mathcal{A}_+\}}$$

(see Stratila-Zsido [1979], 10.23) then the embeddings $j_1$ and $j_2$ enjoy the following fundamental properties:

(i) The map $j_1$ is an order isomorphism from the order interval $[0, 1]$ in $\mathcal{A}$ onto the order interval $[0,\xi]$ in $\mathcal{H}$. Analogously, $j_2$ is an order isomorphism, mapping $[0,\xi]$ onto $[0,\varphi]$ in $\mathcal{A}_*$.

(ii) The restriction of $j_2$ to $j_1(\mathcal{A}_1)$, $\mathcal{A}_1$ the unit ball of $\mathcal{A}$, is a homeomorphism for the weak (resp. norm) topologies on $\mathcal{H}$ and $\mathcal{A}_*$.

It should be noted that in the commutative case, i.e. if $\mathcal{A} = L^{\infty}(\mu)$ and $\varphi = \mu$, then $j_1$ and $j_2$ are indeed the canonical embeddings $L^{\infty}(\mu) \longrightarrow L^2(\mu) \longrightarrow L^1(\mu)$. Moreover, it is well known that in this situation every bi-Markov operator T on $L^1(\mu)$ can be restricted to a positive contraction on $L^{\infty}(\mu)$ and can be interpolated to a positive contraction on the intermediate space $L^2(\mu)$ (see Schaefer [1974], V.8.2). The above embeddings $j_1$ and $j_2$ allow an analogous construction:

Let T be a positive contraction on $\mathcal{A}_*$ satisfying $T\varphi \leq \varphi$ for the faithful state $\varphi \in \mathcal{A}_*$. Then

$$T_{\infty} := j^{-1} \circ T \circ j$$

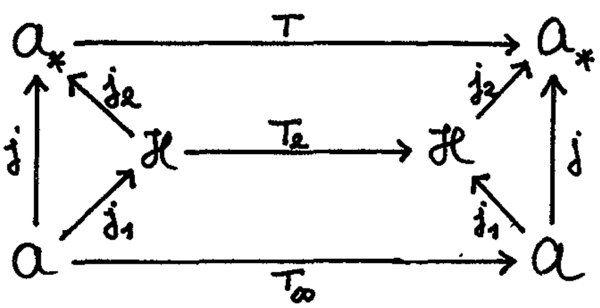
for $j := j_2 \circ j_1$ is a well-defined positive contraction on $\mathcal{A}$.

Moreover,

$$T_2 := j_1 \circ T_{\infty} \circ j_1^{-1}$$

yields a linear operator from the dense subspace $j_1\mathcal{A}$ into $\mathcal{H}$. By the non-commutative analogue of the Riesz convexity theorem (see Groh-Kümmerer [1982], 3.5) $T_2$ has a continuous extension to a contraction on $\mathcal{H}$ still denoted by $T_2$.

The following diagram and a look at the paper of Groh-Kümmerer [1982] helps to clarify the situation:




Using this interpolation method we are able to prove the announced mean ergodic theorem.

Y.11 Theorem (Groh-Kümmerer, 1982):
Let $E = \mathcal{A}_*$ be the pre-dual of a $W^*$-algebra $\mathcal{A}$ and assume that $\mathcal{S} \subset \mathcal{L}(E)$ is a semigroup of positive contractions. If $T\varphi \leq \varphi$ for some faithful state $\varphi \in \mathcal{A}_*$ and every $T \in \mathcal{S}$ then $\mathcal{S}$ is mean ergodic.

Proof:
We perform the construction described above and obtain the interpolated contraction semigroup

$$\mathcal{S}_2 := \{T_2 : T \in \mathcal{S}\}$$

on $\mathcal{H}$ which is mean ergodic by (Y.4). We denote by $P_2$ the corresponding projection. Since $j(\mathcal{A}_1)$ is invariant under $\mathcal{S}_2$ we obtain

$$P_2\eta \in \overline{co}\{T_2\eta : T \in \mathcal{S}\} \subset j_1(\mathcal{A}_1) \quad \text{for every} \quad \eta \in j_1(\mathcal{A}_1).$$

Moreover,

$$j_2(P_2\eta) \in j_2(\overline{co}\,\mathcal{S}_2\eta) \subset \overline{co}(\mathcal{S} j_2(\eta))$$

by property (ii) above. On the other hand it follows that

$$\| j_2(\eta)\| \leq \|\eta\| \leq c\ \|j_2(\eta)\|$$

for every $\eta \in j_1(\mathcal{A}_1)$ and some constant $c > 0$ independent of $\eta$. Therefore

$$\|j_2(P_2\eta)\| \leq \|P_2\eta\| \leq \|\eta\| \leq c \cdot \|j_2(\eta)\|,$$

and the mapping

$$j_2(\eta) \longrightarrow j_2(P_2\eta)$$

is norm continuous from $j(\mathcal{A})$ into $\mathcal{A}_*$. Since $j(\mathcal{A}) = \mathrm{lin} \bigcup_{n \in \mathbb{N}} [0, n\varphi]$ is dense in $\mathcal{A}_*$ we obtain a continuous extension $P \in \mathcal{L}(\mathcal{A}_*)$ satisfying $TP = PT = P$ for every $T \in \mathcal{S}$ and $P\psi \in \overline{co}\,\mathcal{S}\psi$ for every $\psi \in j(\mathcal{A})$.
As in the proof of Theorem (Y.3) it follows that $\mathcal{S}$ is mean ergodic. ∎

Since in most situations it is the $W^*$-algebra $\mathcal{A}$ and an operator semigroup on $\mathcal{A}$ - and not on $\mathcal{A}_*$ - which is of interest it may be useful to give a "dual" formulation of the above result.

Y.12 "Dual theorem":
Let $\mathcal{A}$ be a $W^*$-algebra and assume that $\mathcal{S} \subset \mathcal{L}(\mathcal{A})$ is a semigroup of positive, weak$^*$-continuous contractions. If there exists a $\mathcal{S}_*$-subinvariant faithful normal state on $\mathcal{A}$, then $\mathcal{S}$ is weak$^*$-mean

ergodic, i.e. there exists a weak* -continuous projection P from $\mathcal{A}$ onto the fixed space $F(\mathcal{J})$ such that $TP = PT = P$ for every $T \in \mathcal{J}$, and $Px \in \overline{co}\, \mathcal{J}x$ for every $x \in \mathcal{A}$ and the weak* topology on $\mathcal{A}$.

Obviously, Example (4) in (Y.10) applied to Banach lattices $L^1(X, \Sigma, \mu)$ is a particular case of Theorem (Y.11). Another application is to groups of *-automorphisms on $\mathcal{A}$ as discussed by Kovács-Szücs [1966]. Finally, if we consider the group of all inner automorphisms of a W* -algebra, we obtain the following characterization of "finite" W* -algebras (see Sakai [1971], 2.4.6).

Y.13 Corollary:
Let $\mathcal{A}$ be a W* -algebra possessing a strictly positive "trace" $\varphi \in \mathcal{A}_*$, i.e. $\varphi(x) = \varphi(uxu^*)$ for every $x \in \mathcal{A}$ and every unitary $u \in \mathcal{A}$. Then the group $\mathrm{Int}\,\mathcal{A}$ of all inner automorphisms is weak*- mean ergodic, and the corresponding projection maps $\mathcal{A}$ onto the center $Z(\mathcal{A}) := F(\mathrm{Int}\,\mathcal{A})$.

For a final application of our mean ergodic theory we return to the commutative situation and present a convergence result for positive contractions in $L^1(X, \Sigma, \mu)$, which is an abstract version of the classical martingale convergence theorem (see Ionescu Tulcea-Ionescu Tulcea [1969], II.5 and Appendix I).

Y.14 Application:
Let $(X, \Sigma, \mu)$ be a probability space and assume that $P_n \in \mathcal{L}(L^1(X, \Sigma, \mu))$ is a sequence of positive contractive projections leaving invariant the constant functions. If $P_n P_m = P_n$ for $n \leq m$ and $\bigcup_{n \in \mathbb{N}} P_n L^1(\mu)$ is dense in $L^1(\mu)$ then

$$P_n f \xrightarrow{\|\cdot\|} f$$

for every $f \in L^1(\mu)$.

Proof:
Instead of proving that $P_n$ converges in the strong operator topology to Id we show that

$$Q_n := Id - P_n$$

converges to zero. To that purpose consider the bounded abelian semigroup

$$\mathcal{S} := \{Q_n : n \in \mathbb{N}\}$$

which is relatively weakly compact since $\mathcal{S}[-1\!\!1, 1\!\!1] \subset [-1\!\!1, 1\!\!1]$ .

The fixed space of $\mathcal{S}$ is $\bigcap_{n \in \mathbb{N}} P_n^{-1}(0) = \{0\}$. Therefore, we know that $\mathcal{S}$ is mean ergodic with projection $P = 0$, i.e. there exist convex combinations

$$R_n := \sum_{i=1}^{n} c_{i,n} P_i \in \operatorname{co} \mathcal{S}$$

such that $R_n f$ converges to 0 for every $f \in L^1(\mu)$.

Now, keep $f \in L^1(\mu)$ fixed. For $k \in \mathbb{N}$ we find $n, m \in \mathbb{N}$ such that $\|R_n f\| \leq \frac{1}{2k}$ and $Q_m R_n = Q_m$, since $Q_i Q_j = Q_j$ for $i \leq j$.

From $\|Q_m\| \leq 2$ we conclude $\|Q_m f\| \leq \frac{1}{k}$ and find a subsequence $Q_{m_k} f$ converging to 0. Again from $Q_i Q_j = Q_j$ for $i \leq j$ follows the convergence of $(Q_n f)_{n \in \mathbb{N}}$ . ■

References: Alaoglu-Birkhoff [1940], Day [1969], Eberlein [1948], [1949], Greenleaf [1969], Groh [1984], Groh-Kümmerer [1982], Hiai-Sato [1977], Kümmerer-Nagel [1979], Nagel [1973], Sato [1978], [1979], Schaefer [1974] .

## Appendix Z: Ergodic Theory and Information

Our first aim in this appendix is to give an introduction into the basic ideas of information theory, starting with a rather intuitive concept of information (part 1) and ending up with a proof of the famous coding theorem of Shannon (part 2). In part 3 we prove essentially the theorem of McMillan which is needed in part 2.

We follow the outline of Khinchin 1957 , although this appendix is shorter and modernized from the mathematical point of view. Due to the systematic application of concepts and results from ergodic theory the proofs become more stringent and more general.

Finally, as a feedback this appendix may help the reader to a better understanding of the results of Lectures XI - XIII which are strongly motivated by information theory.

### part 1

The mathematical concept of "information" is intended to be a measurable quantity that specifies the amount of information in a given message (e.g. answer to a question). In the following we present a definition of "information" and motivate it by investigating its properties.

We assume that all possible answers $\mathcal{A}_1,\dots,\mathcal{A}_n$ to a question can be enumerated and (estimates of) their probabilities $p_1,\dots,p_n$, ($0 \leq p_i$ and $\sum_{i=1}^{n} p_i = 1$) are known,

Now we define the "information" of the scheme

$$\mathfrak{A} := \begin{pmatrix} \mathcal{A}_1, \dots, \mathcal{A}_n \\ p_1, \dots, p_n \end{pmatrix} .$$

Z.1 Definition:

$$I(\mathfrak{A}) := - \sum_{i=1}^{n} p_i \log p_i \ .$$

Since $I(\mathfrak{A})$ depends only on the probabilities $p_1,\ldots,p_n$ we write sometimes $I(p)$ where $p$ denotes the probability vector $(p_1,\ldots,p_n)$. Moreover, the logarithm "log" is taken to the base 2 and $x \log x := 0$ for $x = 0$.

Property 1:

$I(\mathfrak{A}) \geq 0$, and $I(\mathfrak{A}) = 0$ iff one of the answers $\mathcal{A}_i$ has probability $p_i = 1$.

Interpretation:

If we know in advance with probability one that the answer is going to be $\mathcal{A}_i$, we do not expect to obtain any information by asking the question.

Property 2:

For $\mathfrak{A} = \begin{pmatrix} \mathcal{A}_1,\ldots,\mathcal{A}_n \\ p_1,\ldots,p_n \end{pmatrix}$ and $\mathfrak{A}' = \begin{pmatrix} \mathcal{A}_0, \mathcal{A}_1,\ldots,\mathcal{A}_n \\ 0, p_1,\ldots,p_n \end{pmatrix}$

we have $I(\mathfrak{A}) = I(\mathfrak{A}')$.

Interpretation:

If one of the answers, i.e. $\mathcal{A}_0$, occurs with probability 0, we may as well leave it out of the scheme describing the question without changing the amount of information.

Property 3:

For fixed $n \in \mathbb{N}$ the information $I(\mathfrak{A})$ of the scheme $\mathfrak{A} = \begin{pmatrix} \mathcal{A}_1,\ldots,\mathcal{A}_n \\ p_1,\ldots,p_n \end{pmatrix}$ is maximal if and only if $p_i = \frac{1}{n}$ for $i = 1,\ldots,n$.

Proof:

$$I(p_1,\dots,p_n) - I(\tfrac{1}{n},\dots,\tfrac{1}{n}) = - \sum_{i=1}^{n} p_i \log p_i - \log n$$

$$= \sum_{i=1}^{n} p_i \log \frac{1}{np_i} = \sum_{i=1}^{n} \frac{p_i}{\ln 2} \ln \frac{1}{np_i}$$

$$\leq \sum_{i=1}^{n} \frac{p_i}{\ln 2} \left(\frac{1}{np_i} - 1\right) = \frac{1}{\ln 2} - \frac{1}{\ln 2} = 0 .$$

The inequality follows from $\ln x \leq x - 1$; equality holds iff $x = 1$, i.e. iff $\frac{1}{np_i} = 1$ for every i. ■

<u>Interpretation</u>:
Among all questions that admit n answers we expect the most information where all answers have the same probability.

<u>Property 4</u>:
The function $p \longmapsto I(p)$ is continuous on

$$S_n := \left\{ p = (p_1,\dots,p_n) : p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\} \quad \text{for all } n \in \mathbb{N}.$$

<u>Property 5</u>:
The function $p \longmapsto I(p)$ is concave on $S_n$ for every $n \in \mathbb{N}$, i.e.

$$I(\alpha p + (1-\alpha)q) \geq \alpha I(p) + (1-\alpha) I(q) \quad \text{for } \alpha \in [0,1].$$

Proof:
This follows immediately from the concavity of $p \longmapsto - p \log p$ on $\mathbb{R}$. ■

<u>Property 6</u>:

If $\mathfrak{A} = \begin{pmatrix} \mathcal{A}_1,\dots,\mathcal{A}_n \\ p_1,\dots,p_n \end{pmatrix}$ and $\mathfrak{A}' = \begin{pmatrix} \mathcal{A}_1',\dots,\mathcal{A}_k', \mathcal{A}_2,\dots,\mathcal{A}_n \\ \frac{p_1}{k},\dots,\frac{p_1}{k}, p_2,\dots,p_n \end{pmatrix}$

Then $I(\mathfrak{A}') = I(\mathfrak{A}) + p_1 I(\frac{1}{k},\dots,\frac{1}{k})$ .

Proof:

$$I(\mathfrak{A}) + p_1 I(\tfrac{1}{k},\dots,\tfrac{1}{k}) = -\sum_{i=1}^{n} p_i \log p_i + p_1 \log k$$

$$= -p_1(\log p_1 - \log k) - \sum_{i=2}^{n} p_i \log p_i$$

$$= -\sum_{i=1}^{k} \frac{p_1}{k} \log \frac{p_1}{k} - \sum_{i=2}^{n} p_i \log p_i = I(\mathfrak{A}') . \blacksquare$$

Interpretation:

The scheme $\mathfrak{A}'$ means that the answer $\mathcal{A}_1$ in $\mathfrak{A}$ is divided into k subcases $\mathcal{A}'_1,\dots,\mathcal{A}'_k$. One would expect that the information obtained by an answer to $\mathfrak{A}'$ is the same as the information provided by the same answer to $\mathfrak{A}$ if this answer is one of $\mathcal{A}_2,\dots,\mathcal{A}_n$. If the answer to $\mathfrak{A}$ is $\mathcal{A}_1$ (which happens with probability $p_1$) one gets additional information by asking $\mathfrak{A}'$ according to the possible subcases $\mathcal{A}'_1,\dots,\mathcal{A}'_k$. The amount of additional information is determined by the fact that all the k subcases of $\mathcal{A}_1$ have the same probability $\frac{1}{k}$. In summary, by asking $\mathfrak{A}'$ one gets the information $I(\mathfrak{A})$ of $\mathfrak{A}$ plus the additional information $I(\frac{1}{k},\dots,\frac{1}{k})$ with probability $p_1$.

Z.2 Theorem:

Any real valued function J on the set of all schemes satisfying properties 1,2,3,4 and 6 is of the form

$$J(\mathfrak{A}) = -c \sum_{i=1}^{n} p_i \ln p_i \quad \text{where} \quad \mathfrak{A} = \begin{pmatrix} \mathcal{A}_1,\dots,\mathcal{A}_n \\ p_1,\dots,p_n \end{pmatrix} ,$$

for some constant $c \in \mathbb{R}$.

For the proof we refer to Khinchin [1957].

The factor c is conventionally determined by the following normalization.

Property 7:

For $\mathfrak{A} = \begin{pmatrix} \mathcal{A}_1, \mathcal{A}_2 \\ \frac{1}{2}, \frac{1}{2} \end{pmatrix}$ we require that $I(\mathfrak{A}) = 1$.

Interpretation:

This normalization yields $c = (\ln 2)^{-1}$ in the above theorem and therefore $J(p) = -\sum_{i=1}^{2} p_i \frac{\ln p_i}{\ln 2} = -\sum_{i=1}^{2} p_i \log p_i$ , i.e. we obtain the base 2 in the logarithm.

In this case the information is said to be measured in bits. In the present lectures we adopted the policy not to specify the basis of the logarithm, i.e. we use the basis $e$ whenever it is convenient.

There is another mathematical framework in which one can describe the outcomes of an experiment, and which is more closely related to the mathematical setting of our lectures:

Let $X$ stand for the set of all possible "states of the world" $x$ relevant to the experiment with outcomes $\mathcal{A}_1, \ldots, \mathcal{A}_n$ . Then to each outcome $\mathcal{A}_i$ corresponds a subset $A_i$ of all states $x \in X$ that yield the outcome $\mathcal{A}_i$. In this notation, the experiment is described by a partition $\alpha = \{A_1, \ldots, A_n\}$ of the set $X$ . If we want to determine the information given by the experiment we assume a probability distribution $p$ on $X$, i.e. we have a probability space $(X, \Sigma, \rho)$, and we assume the partition $\alpha$ to be measurable, i.e. $\alpha \subseteq \Sigma$.

In this case the partition $\alpha$ corresponds to the scheme

$$\begin{pmatrix} A_1, \ldots\ldots, A_n \\ p(A_1), \ldots, p(A_n) \end{pmatrix}$$

and we define

$$I(\alpha) := -\sum_{A \in \alpha} p(A) \log p(A) .$$

Moreover for $A, B \in \Sigma$ one can define the conditional probability of $A$ given $B$ as

$$p(A|B) := \frac{p(A \cap B)}{p(B)} .$$

Using these conditional probabilities for a partition $\alpha = \{A_1,\dots,A_n\}$ and $B \in \Sigma$ we can consider the scheme

$$\begin{pmatrix} \mathcal{A}_1,\dots,\mathcal{A}_n \\ p(A_1|B),\dots,p(A_n|B) \end{pmatrix}$$

and the corresponding information

$$I(\alpha|B) := -\sum_{A\in\alpha} p(A|B)\log p(A|B),$$

which has to be interpreted as the information of the scheme

$$\begin{pmatrix} \mathcal{A}_1,\dots,\mathcal{A}_n \\ p(A_1),\dots,p(A_n) \end{pmatrix}$$

given that we know that the state $x$ is in $B$.

Z.3 Definition:
Given two partitions $\alpha, \beta$ then we define the conditional information of $\alpha$ given $\beta$ as

$$I(\alpha|\beta) := \sum_{B\in\beta} p(B)I(\alpha|B),$$

and interpret it as the average information of $\alpha$ given that we know in which element $B$ of $\beta$ the state $x$ happens to be.

Property 8 (see (XI.1):
$I(\beta) + I(\alpha|\beta) = I(\alpha\vee\beta)$.

A proof is given in (XI.D.3).

Interpretation:
The average inforamtion contained in the scheme described by $\alpha\vee\beta$ is equal to the average information contained in $\beta$ plus the average information of $\alpha$ given that we already know about $\beta$.

Recalling the Properties 1,...,8 above we realize that we have derived a reasonable way of measuring the information $I(\mathfrak{a})$ contained in a scheme $\mathfrak{a}$, resp. the information

$$I(\alpha) = h_p(\alpha)$$

contained in a partition $\alpha$ of a probability space $(X,\Sigma,p)$.

## part 2

We start by defining some ergodic-theoretical objects and by fixing some information-theoretical terminology:
A <u>source</u> (or <u>information source</u>) is a device that consecutively produces symbols from a finite set

$$L = \{\ell_1, \dots, \ell_k\},$$

called the <u>alphabet</u> of the source. In mathematical terms, a source will be (described by) a shift

$$(X, \Sigma, \mu; \tau)$$

where $X := L^{\mathbb{Z}}$ and $\mu([x_n = \ell_i])$ is interpreted as the probability that the source produces the <u>symbol</u> (or <u>letter</u>) $\ell_i$ at time $n$.

Let $\alpha = A_1, \dots, A_k$ with $A_i := [x_0 = \ell_i]$ denote the canonical generator of the above shift. Then the elements of $\alpha^n$ correspond to the <u>words</u> (or <u>blocks</u>) of length $n$ produced by the source starting at time 0. Thus $h_\mu(\alpha^n)$ can be interpreted as the average information contained in a word of length $n$ produced by the source. Consequently, $h_\mu(X;\tau) = \lim_{n\to\infty} \frac{1}{n} h_\mu(\alpha^n)$ is the average information per symbol produced by the source or the average information produced by the source in one time unit and will be called the <u>information rate</u> of the source.

A <u>channel</u> is a device that transmits sequences of symbols from its so-called <u>input alphabet</u> $L$ to sequences of symbols from its so-called <u>output alphabet</u> $M$. In order to model a channel we use again the shift $\tau$ on the <u>input space</u>

$$X := L^{\mathbb{Z}}$$

with product $\sigma$-algebra $\Sigma$ and on the <u>output space</u>

$$Y := M^{\mathbb{Z}}$$

with product $\sigma$-algebra $\mathcal{T}$. On the product $(X \times Y, \Sigma \times \mathcal{T})$, called the <u>channel space</u>, we still have the shift transformation if we identify $X \times Y$ with $(L \times M)^{\mathbb{Z}}$. With these notions, we define a channel as a <u>Markov kernel</u> from $(X, \Sigma)$ to $(Y, \mathcal{T})$, i.e. as a mapping

$$p : X \times \mathcal{T} \longrightarrow [0,1],$$

where $p(x,\cdot)$ is a probability measure on $(Y,\mathcal{T})$ and $p(\cdot,B)$ a measurable mapping from $X$ into $[0,1]$. Here $p(x,B)$ is interpreted as the probability that the output sequence is in $B \in \mathcal{T}$ given that the input sequence was $x \in X$.

Let $\mu$ be a fixed probability measure on $(X,\Sigma)$, called <u>input probability</u>. Then the Markov kernel $p$ yields a probability measure $p_\mu$ on $(X \times Y, \Sigma \times \mathcal{T})$ defined by

$$p_\mu(A \times B) := \int_X p(x,B) \mathbb{1}_A(x) d\mu$$

for $A \in \Sigma$, $B \in \mathcal{T}$.

In particular, we also denote by $p_\mu$ the <u>output probability</u> on $(Y,\mathcal{T})$ defined by $p$ and $\mu$, i.e.

$$p_\mu(B) := \int_X p(x,B) d\mu$$

for $B \in \mathcal{T}$.

Given two channels $p$ and $p'$ where $p$ has input alphabet $L$ and output alphabet $M$ and $p'$ has input alphabet $N$ and output alphabet $L$, we consider the space $Z := N^{\mathbb{Z}}$ with the product $\sigma$-algebra $\Upsilon$ and define the <u>compound channel</u> $p \circ p'$ by

$$p \circ p'(z,B) := \int_X p(x,B) dp'(z,x)$$

for $z \in Z$ and $B \in \mathcal{T}$.

A mapping $m : Z \to X$ can, of course, be regarded as a special channel, where for a given input $z \in Z$ there is no uncertainty about the output $x \in X$, i.e.

$$m(z,A) := \begin{cases} 0 & \text{if } m(z) \notin A \\ 1 & \text{if } m(z) \in A \end{cases} \quad \text{for } z \in Z \text{ and } A \in \Sigma.$$

Such channels will be called <u>deterministic</u>.

As indicated above we shall measure the information, passing through a channel, using the concepts of part 1:

Let $p$ be a channel with an input probability $\mu$. Then for the canonical generators $\alpha$ and $\beta$ of $(X,\Sigma,\mu;\tau)$ and $(Y,\mathcal{T},p_\mu;\tau)$ we may consider the following quantities:

$h_\mu(\alpha^n)$, i.e. the average information of an input sequence of length $n$,

$h_\mu(\beta^n) := h_{p_\mu}(\beta^n)$, i.e. the average information of an output sequence of length $n$,

$h_\mu(\alpha^n x \beta^n) := h_{p_\mu}(\alpha^n x \beta^n)$, i.e. the average informationof a pair of corresponding input and output sequences of length $n$.

Then we define the <u>transinformation</u>

$$t_n(\mu) := h_\mu(\alpha^n) - h_\mu(\alpha^n | \beta^n),$$

i.e. the average information in an input sequence of length $n$ minus the average information still contained in an input sequence of length $n$ given that we know the corresponding output sequence of length $n$. Or in other words: $t_n(\mu)$ is the average amount of information about an input sequence of length $n$ that is contained in the corresponding output sequence of length $n$.

Two extreme cases may help us to understand this terminology: If $h_\mu(\alpha^n|\beta^n) = 0$ then the input sequence can be determined (with probability 1) from the output sequence and therefore the whole information of the input sequence is transmitted through the channel, i.e. $t_n(\mu) = h_\mu(\alpha^n)$. If $h_\mu(\alpha^n|\beta^n) = h_\mu(\alpha^n)$, then the knowledge of the output sequence is of no use to determine the input sequence. Consequently, no information is transmitted through the channel and $t_n(\mu) = 0$.

After having fixed the basic terminology we need some particular properties of channels. The following is essential in order to apply our results on MDSs.

Z.4 <u>Definition</u>:
A source is called <u>stationary</u>, if $(X, \Sigma, \mu; \tau)$ is an MDS, i.e. the input probability $\mu$ is $\tau$-invariant. The channel $p$ is <u>stationary</u>, if $p(x,B) = p(\tau(x), \tau B)$ for $x \in X$ and $B \in \mathcal{T}$. Finally, a stationary channel $p$ is called <u>ergodic</u>, if $(X \times Y, \Sigma \times \mathcal{T}, p_\mu; \tau)$ is ergodic whenever $(X, \Sigma, \mu; \tau)$ is ergodic (compare Adler [1961]).

Concerning ergodicity, we remark that the converse implication is trivial, i.e. $(X \times Y, \Sigma \times \mathcal{T}, p_\mu ; \tau)$ ergodic implies $(X, \Sigma, \mu; \tau)$ and $(Y, \mathcal{T}, p_\mu ; \tau)$ to be ergodic.
It is easy to see that for a stationary input probability $\mu$ and a channel p the systems $(Y, \mathcal{T}, p_\mu ; \tau)$ and $(X \times Y, \Sigma \times \mathcal{T}, p_\mu ; \tau)$ are MDSs. In the following we assume channels and sources to be stationary. Then it follows from Property 8 in part 1 that

$$t_n = h_\mu(\alpha^n) - h_\mu(\alpha^n | \beta^n)$$

$$= h_\mu(\alpha^n) + h_\mu(\beta^n) - h_\mu(\alpha^n \times \beta^n)$$

$$= h_\mu(\beta^n) - h_\mu(\beta^n | \alpha^n)$$

and

$$0 \leq t_n \leq \min(h_\mu(\alpha^n), h_\mu(\beta^n)).$$

We define

$$t(\mu) := \lim_{n \to \infty} \frac{1}{n} t_n(\mu) = h_\mu(X, \tau) + h_\mu(Y, \tau) - h_\mu(X \times Y, \tau)$$

and the limit exists by (E.5).

The set $\mathcal{P}_\tau(X)$ of stationary input measures is non-empty and compact (compare App. S ). Therefore the following definition makes sense.

Z.5 Definition:

$c := \sup \{ t(\mu) : \mu \in \mathcal{P}(X; \tau) \}$ is called the information transmission capacity of the channel p .

Let us show how the capacity of simple channels can be calculated.

Z.6 Example:

A stationary channel is called memoryless if for

$$A = \bigcap_{i=0}^{n-1} \tau^{-i} A_i \in \alpha^n \quad \text{and} \quad B = \bigcap_{i=0}^{n-1} \tau^{-i} B_i \in \beta^n$$

we have

$$p(x, B) = \prod_{i=0}^{n-1} q(A_i, B_i) \qquad \text{for every } x \in A.$$

Here, $q : \alpha \times \beta \to [0,1]$ should be given by a stochastic matrix, i.e. $\sum_{B \in \beta} q(A, B) = 1$ for every $A \in \alpha$ .

This definition may be interpreted by saying that a memoryless channel transmits each symbol independently from former (or later) symbols.

It follows easily that for a memoryless channel

$$h_\mu(\beta^n | \alpha^n) = n \cdot h_\mu(\beta|\alpha) \quad \text{and} \quad h_\mu(\beta^n) \leq n \cdot h_\mu(\beta),$$

which are determined by the matrix $q$ and the vector $v := (\mu(A))_{A\in\alpha}$ . Therefore the optimal input probability $\bar{\mu}$ is given by a v-Bernoulli shift (where $h_\mu(\beta^n) = nh_\mu(\beta)$), and we obtain the capacity

$$\begin{aligned} c &= t(\bar{\mu}) = h_{\bar{\mu}}(\beta) - h_{\bar{\mu}}(\beta|\alpha) \\ &= -\sum_{B\in\beta} \Big(\sum_{A\in\alpha} q(A,B)\bar{\mu}(A)\Big) \log \Big(\sum_{A\in\alpha} q(A,B)\bar{\mu}(A)\Big) \\ &\quad + \sum_{B\in\beta} \sum_{A\in\alpha} q(A,B)\bar{\mu}(A) \log q(A,B) \\ &= \sum_{A,B} q(A,B)\bar{\mu}(A) \log \Big(q(A,B)\Big(\sum_{A'\in\alpha} q(A',B)\bar{\mu}(A')\Big)^{-1}\Big). \end{aligned}$$

Thus $c = \sup\limits_{v} \sum\limits_{i,j} q(i,j)v_i \log \big(q(i,j)\big(\sum\limits_{\ell} q(\ell,j)v_\ell\big)^{-1}\big)$,

and this equation can be used to find the optimal vector $v$ . Still this is not easy in general (see Ash 1965 ) but for example if

$$h_{\bar{\mu}}(\beta | A) = -\sum_{B\in\beta} q(A,B) \log q(A,B)$$

is the same for every $A \in \alpha$, then $v = \text{const.}$ is the optimal vector.

As a more concrete example we mention the so-called <u>binary symmetric channel</u> which is a memoryless channel with

$L = \{0,1\} = M$ and matrix $q = \begin{pmatrix} d & 1-d \\ 1-d & d \end{pmatrix}$ for some $0 \leq d \leq 1$, i.e.

$$p(x,[y_0 = 0]) = \begin{cases} d & \text{if } x_0 = 0 \\ 1-d & \text{if } x_0 = 1 \end{cases}$$

and of course

$$p(x,[y_0 = 1]) = \begin{cases} 1-d & \text{if } x_0 = 0 \\ d & \text{if } x_0 = 1 \end{cases} .$$

In this case the optimal vector is $v = (\frac{1}{2},\frac{1}{2})$ and we obtain $c = 1 + d \log d + (1-d) \log (1-d)$.

Various properties or types of channels are defined in the "classical" literature (e.g. Ash [1965], Feinstein [1958]), the most important ones being channels with finite (input-)memory. Here we introduce some more general properties of channels (see also Pfaffelhuber [1971], Gray and Ornstein [1979], Kieffer [1981]).

Z.7 Definition:
A channel $p$ with input space $(X,\Sigma)$ and output space $(Y,\mathcal{T})$ is called weakly continuous if the mapping $\psi: x \mapsto p_x$ from $X$ into the space $\mathcal{P}(Y)$ of all probability measures on $Y$ is continuous for the weak topology on $\mathcal{P}(Y)$.

Z.8 Lemma:
For a channel $p$ the following are equivalent:

(a) $p$ is weakly continuous.

(b) The mappings $\psi_B: x \mapsto p_x(B)$ from $X$ into $[0,1]$ are continuous for every $B \in \beta_{-n}^{n}$, $n \in \mathbb{N}$.

(c) For any $\varepsilon > 0$ and $n \in \mathbb{N}$ there is $m \in \mathbb{N}$ such that for $x, x' \in A \in \alpha_{-m}^{m}$ we have $|p_x(B) - p_{x'}(B)| < \varepsilon$ for every $B \in \sigma(\beta_{-n}^{n})$.

Proof:
(a) implies (b), since the sets $B \in \beta_{-n}^{n}$ are open and closed.
For the proof of the converse implication (b) $\Rightarrow$ (a) choose $g \in C(Y)$ and $\varepsilon > 0$. By the definition of the topology on $Y$ we find $n \in \mathbb{N}$ and a step function $g_n$ supported by elements of $\beta_{-n}^{n}$ such that $\|g - g_n\| < \varepsilon$.
Consequently $|p_x(g) - p_x(g_n)| \leq \int_Y |g_n - g|\, dp_x < \varepsilon$ for any $x \in X$.
Since the mappings $x \mapsto p_x(g_n)$ are continuous by hypothesis we obtain that $x \mapsto p_x(g)$ is continuous.
Finally the equivalence of (b) and (c) follows from the fact that $\beta_{-n}^{n}$ is finite and from the definition of the topology on $X$ and $\mathcal{P}(Y)$. ■

Z.9 <u>Proposition</u>:
Let p be a weakly continuous stationary channel. Then the capacity c equals

$$c_e := \sup \{ t(\mu) : \mu \in \mathcal{P}_\tau(X) \text{ and } \tau \text{ ergodic for } \mu \}.$$

Proof:
We have only to show that $c_e \geq c$: The measures $\mu \in \mathcal{P}_\tau$ for which $\tau$ is ergodic are exactly the extreme points of the compact convex set ( compare App. S.2). Now we use the following theorem of Bauer [1963]: "Let K be a compact convex set and let $f_1, f_2$ be two semicontinuous real valued functions on K such that $f_1$ is convex and $f_2$ is concave. If $f_1 \leq f_2$ holds on the extreme points of K then $f_1 \leq f_2$ holds on K ". As soon as we are able to apply this theorem to

$$f_1(\mu) := h_\mu(X;\tau) + h_\mu(Y;\tau),$$
$$f_2(\mu) := c_e + h_\mu(X \times Y;\tau),$$

we are done.

From (XII.12) we know that $f_1$ and $f_2$ are affine. The mappings $\mu \mapsto h_\mu(\alpha^n)$, $\mu \mapsto h_\mu(\beta^n)$ and $\mu \mapsto h_\mu(\alpha^n x \beta^n)$ are weakly contінous on $\mathcal{P}_\tau$ because p is weakly continuous.
Since $h_\mu(X;\tau) = \inf_{n \in \mathbb{N}} \frac{1}{n} h_\mu(\alpha^n)$ and similarily for $h_\mu(X;\tau)$ and $h_\mu(X \times Y;\tau)$, foth $f_1$ and $f_2$ are (upper) semicontinuous. ∎

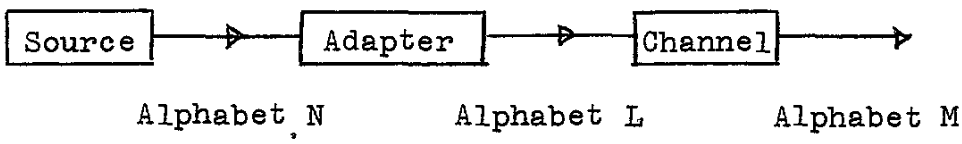
<u>Remark</u>:
It is possible to show this proposition under more general assumptions on p (see Parthasaraty [1961]). However, we need the weak continuity of p also for other purposes, and therefore we decided to present this simple proof that relies on the elementary theorem of Bauer [1963].

The condition (c) of Lemma (Z.8) can be interpreted as follows: the outcome of a sequence (from the $-n^{th}$ to the $n^{th}$ symbol) at the output of the channel does not depend much on input symbols that have occured or will occur at very distant times. In the following we shall need a slightly stronger version of (Z.8.c).

Z.10 Definition:

A channel p has asymptotically decreasing input dependence (for short: adid), if for any $\varepsilon > 0$ there is $m \in \mathbb{N}$ such that for any $n \in \mathbb{N}$ and any $x, x' \in A \in \alpha_{-n-m}^{n+m}$ and $B \in \sigma(\beta_{-n}^{n})$ we have $|p_x(B) - p_{x'}(B)| < \varepsilon$.

The commonly accepted definition of a channel with finite memory consists of two parts: one part concerns the finite input dependence and one part concerns the finite "internal memory" of the channel (e.g. Feinstein [1958]). It is fairly easy to see that the first part implies "adid" and the second part implies ergodicity. Therefore, any channel with finite input memory is ergodic and has "adid".
Now we proceed to formulate and prove Shannon's theorem.
Intuitively it says the following:
Given any channel with capacity c and a source $\mu$ with information rate $h_\mu < c$, then it is possible to adapt the source to the channel in such a way that one can retrieve "almost" the complete source sequences from the corresponding channel output sequences:




The adaption of the source to the channel input is done by means of block coding, i.e. by mapping blocks of a certain length n of symbols from the source alphabet to blocks of the same length of symbols from the channel input alphabet.

Z.11 Definition:

Consider two finite sets L, N, the corresponding product spaces $(X, \Sigma)$, $(Z, \Upsilon)$, and the canonical generators $\alpha, \gamma$. A mapping $f : \gamma^k \rightarrow \alpha^k$ defines a measurable mapping $m_f : Z \rightarrow X$ where

$m_f(x)$ is the unique element in $\bigcap_{i=-\infty}^{\infty} \tau^{-ki}(f(A_i))$, where $\tau^{ki}(x) \in A_i \in \alpha^k$. This mapping $m_f$ will be called a <u>block coding</u> of length $k$.

Now we should specify what is meant by "retrieving almost the complete source sequences from the corresponding channel output sequences". To that purpose we define the "guessing" of an input sequence of length $n$ from an output sequence of length $n$ by a mapping

$$g : \beta^n \longrightarrow \alpha^n ,$$

where $\alpha$ and $\beta$ are again the canonical generators. The guessing $g$ will be wrong if $x \in A \in \alpha^n$, $y \in B \in \beta^n$ but $g(B) \neq A$.

Therefore we define

$$E := \bigcup \{ A \times B : A \in \alpha^n , B \in \beta^n , g(B) \neq A \}$$

and call $p_\mu(E)$ the <u>probability of error</u> for the guessing $g$, the channel $p$ and the input probability $\mu$.

Then the theorem of Shannon says that for a certain class of channels $p$ with capacity $c$ and ergodic sources $\mu$ with rate $h_\mu < c$ and for any $\varepsilon > 0$ there is a coding

$$m_f : (Z,\mathbf{Y}) \longrightarrow (X, \Sigma)$$

between the source alphabet and the channel input alphabet and a guessing $g$ for the compound channel $p' = p \circ m_f$ such that the probability of error $p'_\mu(E) < \varepsilon$.

But before stating this theorem we prepare the basic combinatorial tools for its proof.

Z.12 <u>Definition</u>:
Let $p$ be a channel with input probability $\mu$ and let $\alpha, \beta$ be the canonical generators of the input and output space, respectively. A subset $\{A_1, \dots, A_k\}$ of $\alpha^n$ is called <u>$\lambda$-distinguishable</u>, if there is a partition $\beta' = \{B_1, \dots, B_k\} \leq \beta^n$ such that

$$p_\mu(B_i^c / A_i) = p_\mu(A_i \times B_i^c)/\mu(A_i) \leq \lambda \quad \text{for } i = 1, \dots, k .$$

Observe that a $\lambda$-distinguishable set $\{A_1,\ldots,A_k\} \subseteq \alpha^n$ may be used in order to construct a guessing $g : \beta^n \to \alpha^n$ with a low probability of error. Indeed by $g(B) := A_i$ for $B \subseteq B_i$ we obtain a guessing with

$$p_\mu(E) \leq \mu[x \notin \bigcup_{i=1}^k A_i] + p_\mu(\bigcup_{i=1}^k A_i \times B_i^c)$$

$$\leq \mu((\bigcup_{i=1}^k A_i)^c) + \sum_{i=1}^k \lambda \cdot \mu(A_i)$$

$$\leq \mu((\bigcup_{i=1}^k A_i)^c) + \lambda .$$

In the following we shall sometimes identify $A_i$ with $A_i \times Y$ or $B_i$ with $X \times B_i$, when no confusion seems possible.

Most types of channels (e.g. memoryless channels) defined in information theory are ergodic. For such channels we now state the theorem of McMillan (see part 3) exactly in the form we need it.

Z.13 Lemma:
Let $p$ be an ergodic channel and $\mu$ an ergodic input probability for $p$. Take the canonical generators $\alpha$, $\beta$ and $\alpha \times \beta$ on $X$, $Y$ and $X \times Y$ and define the functions $r_{on}$ as at the beginning of part 3. Then the functions

$$s_n(x,y) := r_{on}(x) + r_{on}(y) - r_{on}(x,y)$$

converge in $L^1(X \times Y, p_\mu)$ to $t(\mu) \cdot \mathbb{1}$.

Z.14 Main lemma (Feinstein, 1954):
Let $p$ be a channel with input probability $\mu$. Let $a, \lambda > 0, k \in \mathbb{N}$ and define

$$W := \{(x,y) \in X \times Y : s_n(x,y) > a\}.$$

Then there is a $\lambda$-distinguishable set in $\alpha^n$ with cardinality larger or equal to $2^{an}(\lambda - p_\mu(W^c))$.

Proof:

We construct a finite number of sets $B_i$ , which are unions of sets in $\beta^n$ and the corresponding sets $A_i$ with $p_\mu(B_i^c | A_i) \leq \lambda$ :

Successively, we take $A_1, A_2, A_3, \ldots$ such that

$$p_\mu(W \cap A_k \setminus \bigcup_{i=1}^{k-1} B_i) / \mu(A_k) \geq 1 - \lambda$$

and define $B_1, B_2, B_3 \ldots$ by

$$B_k := \{ y \in Y : (x,y) \in W,\ x \in A_k \text{ and } y \notin \bigcup_{i=1}^{k-1} B_i \}.$$

(The empty union is defined as $\emptyset$).

Observe that $B_1$ is a union of sets in $\beta^n$, since $s_n$ is constant on the sets of $\alpha^n \times \beta^n$ .

This procedure works for a number $t \geq 0$ of steps and we have to show that

$$t \geq 2^{an}(\lambda - p_\mu(W^c))$$

or $\quad (*) \quad p_\mu(W) \leq t\ 2^{-an} + 1 - \lambda$ .

For $B := \bigcup_{j=1}^{t} B_j$ we have $X \times B = \bigcup_{j=1}^{t} W\ A_j$ .

Now $p_\mu(W) \leq p_\mu(W \cap B) + p_\mu(W \cap B^c)$ and

$$p_\mu(W \cap B) \leq p_\mu(B) = \sum_{j=1}^{t} p_\mu(B_j) .$$

If $y \in B \in \beta^n$ and $B \subseteq B_j$ then $s_n(x,y) > a$ and $x \in A_j$ . Therefore

$- \log \mu(A_j) - \log p_\mu(B) + \log p_\mu(A_j \times B) > an$ .

Consequently,

$$p_\mu(B) < 2^{-an} p_\mu(A_j \times B) / \mu(A_j) \qquad \text{and}$$

$$\sum_{j=1}^{t} p_\mu(B_j) \leq \sum_{j=1}^{t} 2^{-an} p_\mu(A_j \times B_j) / \mu(A_j) \leq t \cdot 2^{-an} .$$

For the other term we have $p_\mu(W \cap B^c) \leq 1 - \lambda$ . Otherwise we could find $A \in \alpha^n$ such that

$$p_\mu(W \cap B^c \cap A) / \mu(A) \geq 1 - \lambda$$

and we could perform the procedure above more than $t$ times. Taking these estimates together, we obtain inequality ($*$). ∎

The main reason why we need channels with adid is that the notion of a $\lambda$-distinguishable set for a channel with finite input memory becomes independent of the input probability.

Z.15 Lemma:

Let $p$ be a channel, $\mu$ an input probability, and $\alpha, \beta$ the canonical generators. To a $\lambda$-distinguishable set $\{A_1,\dots,A_k\} \preceq \alpha^n$ corresponds a partition $\beta' = \{B_1,\dots,B_k\}$ such that $p_\mu(B_i^c / A_i) \le \lambda$.

With these notations the following holds:

(i) There are $x_i \in A_i$ such that $p(x_i, B_i^c) \le \lambda$.

(ii) If the channel $p$ has adid, then for any $\varepsilon > 0$ there are $A_i' \subseteq A_i$, $A_i' \in \alpha_{-m}^{m+n}$ such that $p(x, B_i^c) \le \lambda + \varepsilon$ for every $x \in A_i'$.

Proof:

(i) $p_\mu(B_i^c / A_i) \le \lambda$ means $p_\mu(B_i^c \times A_i) \le \lambda \cdot \mu(A_i)$.

If $p(x, B_i^c) > \lambda$ for every $x \in A_i$, then $p_\mu(B_i^c \times A_i) =$

$= \int p(x, B_i^c) 1\!\!1_{A_i}(x) d\mu(x) > \int \lambda \cdot 1\!\!1_{A_i}(x) d\mu(x) = \lambda \cdot \mu(A_i)$.

(ii) Define $A_i'$ by requiring $x_i \in A_i' \in \alpha_{-m}^{m+n}$. Since $B_i \in \beta' \preceq \beta^n$ and by Definition (Z.10) we have

$$p(x, B_i^c) \le p(x_i, B_i^c) + \varepsilon \le \lambda + \varepsilon \quad \text{for every } x \in A_i'.$$

Z.16 Theorem (Shannon, 1948):

Let $p$ be an ergodic channel with adid and capacity $c$ and let $(Z, \mathcal{Y}, \nu; \tau)$ be an ergodic source with canonical generator $\gamma$ on $Z := N^{\mathbb{Z}}$. We assume that $h_\nu(Z;\tau) < c$. For $\varepsilon > 0$, $k \in \mathbb{N}$ there exists $k \le n \in \mathbb{N}$, $m \in \mathbb{N}$ a block coding $m_f : Z \to X$ of length $n + 2m$ and a "guessing" $g : \beta^n \to \gamma^n$ such that the compound channel $p' = p \circ m_f$ has the following properties:

(i) $p'_{\nu}(E) \le \varepsilon$ , i.e. the probability of error is less than $\varepsilon$ .
(ii) $t'(\nu) \geq h_{\nu}(Z;\tau) - \varepsilon$ , i.e. up to $\varepsilon$ the total information of the source is transmitted through the compound channel.

Proof:
Let $0 < \delta < \frac{1}{4}(c - h_{\nu}(Z;\tau))$, and $\mu$ be an ergodic input probability for the channel $p$ with $t(\mu) \geq c - \delta$ (from Z.9). Choose $m$ so large that for any $n \in \mathbb{N}$, $x,x' \in A \in \alpha_{-m}^{n+m}$ and any $B \in \sigma(\beta^n)$ we have $|p_x(B) - p_{x'}(B)| < \delta$. Now take $n \geq m$ sufficiently large to obtain

$$\nu\left[|r_{o(n+m)}(Z) - h_{\nu}(Z;\tau)| < \delta\right] > 1 - \delta \quad \text{(use Z.22 in part 3)}$$

and

$$p_{\mu}\left[|s_n(x,y) - t(\mu)| < \delta\right] > 1 - \delta \qquad \text{(use Z.13)}$$

and

$$2^{-n\delta} < \delta \ , \ \frac{2m+1}{2m+n} < \delta .$$

For $W := \{(x,y) : s_n(x,y) > c - 2\delta\}$ we have $p_{\mu}(W^c) < \delta$.
Now we look for a $\lambda$-distinguishable set $F \lesssim \alpha^n$ having cardinality larger than $2^{n(c-3\delta)}$. By Feinstein's lemma (Z.14) this can be found as long as

$$2^{n(c-3\delta)} \le 2^{n(c-2\delta)}(\lambda - \delta), \text{ i.e. } \lambda \geq \delta + 2^{-n\delta}.$$

Therefore, we may take $\lambda = 2\delta$ .

Next we consider

$$Z_o := \{ z \in Z : r_{on}(z) < h_{\nu}(Z;\tau) + \delta \}$$

and observe that $\nu(Z_o^{\,c}) < \delta$ and

$$q_{on}(z) > 2^{-nr_{on}(z)} > 2^{-n(h_{\nu}(Z;\tau)+\delta)}$$

for $z \in Z_o$ . Since $r_{on}$ is constant on the sets of $\gamma^n$ , $Z_o$ is a union of sets in $\gamma^n$ , say

$Z_o$ $$Z_o = \bigcup_{C \in G} C$$

where $G \subseteq \gamma^n$ . For $z \in C \in G$ we have $q_{on}(z) = \nu(C)$ and therefore

$$1 > \nu(Z_o) = \sum_{C \in G} \nu(C) > |G| 2^{-n(h_{\nu}(Z;\tau)+\delta)} \text{ , or}$$

$$|G| < 2^{n(h_{\nu}(Z;\tau)+\delta)} < 2^{n(c-3\delta)} \le |F| .$$

By this last inequality we can find an invertible mapping $f_o : G \to F$. Since $F = \{A_1, \dots, A_k\}$ is $2\delta$-distinguishable, there is the corresponding partition $\beta' = \{B_1, \dots, B_k\} \preceq \beta^n$. For $C \in G$ we define $\beta_C := \{B \in \beta^n : B \subseteq B_i \text{ and } f_o(C) = A_i \in F\}$. Then we define $g(B) := C$ for $B \in \beta_C$ and extend this to a mapping $g : \beta^n \to \gamma^n$.

Now we can define the block coding $m_f$ :

For $A_i \in F$ there is by (Z.15) an $A_i' \in \alpha_{-m}^{m+n}$ such that $p(x, B_i^c) \leq 3\delta$ for every $x \in A_i'$. For every $C' \in \gamma_{-m}^{m+n}$ with $C' \subseteq C \in G$ and $f_o(C) = A_i$, we define $f(C') = A_i'$ and extend this to a mapping $f : \gamma_{-m}^{m+n} \to \alpha_{-m}^{m+n}$.

Finally we consider the compound channel $p'$ and calculate

$$
\begin{aligned}
p'_\nu(E) &= p'_\nu\left(\bigcup \{C \times B : C \in \gamma^n, B \in \beta^n, g(B) \neq C\}\right) \\
&\leq \nu(Z_o^c) + p'_\nu\left(\bigcup_{C \in G} \bigcup_{B \in \beta^n \setminus \beta_C} C \times B\right) \\
&= \nu(Z_o^c) + p'_\nu\left(\bigcup_{C \in G} [z \in C, y \notin B_i, f_o(C) = A_i]\right) \\
&\leq \nu(Z_o^c) + p'_\nu\left(\bigcup_{C \in G} [m_f(z) \in A_i', y \notin B_i, f_o(C) = A_i]\right) \\
&= \nu(Z_o^c) + p'_\nu\left(\bigcup_{A_i \in f_o(C)} [m_f(z) \in A_i', y \notin B_i]\right) \\
&\leq \nu(Z_o^c) + \sum_i \int p(x, B_i^c)\, \mathbb{1}_{A_i'}(x)\, d(\nu \circ m_f^{-1})(x) \\
&\leq \nu(Z_o^c) + \sum_i \int 3\delta \cdot \mathbb{1}_{A_i'}(x)\, d(\nu \circ m_f^{-1})(x) \\
&\leq 4\delta ,
\end{aligned}
$$

which proves (i) by choosing $4\delta \leq \varepsilon$.

As for (ii) we calculate $h_\nu(\gamma^n | \beta^n)$ by means of (i):

$$
\begin{aligned}
h_\nu(\gamma^n / \beta^n) &\leq h_\nu(\gamma^n \vee \{E, E^c\} \mid \beta^n) \\
&= h_\nu(\{E, E^c\} \mid \beta^n) + h_\nu(\gamma^n \mid \beta^n \vee \{E, E^c\}) \\
&\leq h_\nu(\{E, E^c\}) + h_\nu(\gamma^n / \beta^n \vee \{E, E^c\}) .
\end{aligned}
$$

For $C \in \gamma^n$ and $B \in \beta^n$ we have

$$p'_\nu (C|B \cap E^c) = \begin{cases} 0 & \text{if } C \neq g(B) \\ 1 & \text{if } C = g(B) . \end{cases}$$

Thus $h_\nu (\gamma^n | \beta^n \vee \{E, E^c\})$

$$= \sum_{B \in \beta^n} (p'_\nu (B \cap E)\, h_\nu (\gamma^n | B \cap E) + p'_\nu (B \cap E^c) h_\nu (\gamma^n | B \cap E^c))$$

$$\leq \sum_{B \in \beta^n} p'_\nu (B \cap E) \log |\gamma^n| + 0$$

$$\leq p'_\nu (E) \log |\gamma^n| .$$

Therefore

$$\frac{1}{\ell(2m+n)} (h_\nu (\gamma^{\ell(2m+n)} | \beta^{\ell(2m+n)})$$

$$\leq \frac{1}{2m+n} h_\nu (\gamma^{2m+n} | \beta^{2m+n})$$

$$\leq \frac{1}{2m+n} (h_\nu (\gamma^{2m}) + h_\nu (\gamma^n | \beta^n))$$

$$\leq \frac{1}{2m+n} (h_\nu (\gamma^{2m}) + h_\nu (\{E, E^c\}) + 4\delta n \log |\gamma| )$$

$$< \frac{1}{2m+n} ((2m \log |\gamma| + 1) + 4\delta \log |\gamma| )$$

$$\leq 5\delta \log |\gamma| .$$

Thus $t'(\nu) = h_\nu (Z; \tau) - \lim_{n \to \infty} \frac{1}{n} h_\nu (\gamma^n | \beta^n)$

$$\geq h_\nu (Z; \tau) - 5\delta \log |\gamma| . \quad \blacksquare$$

The theorem of Shannon essentially shows that it is possible to think of "information" as an incompressable measurable quantity. This is an important motive for the definition of a numerical measure for information (as in part 1 of this Appendix). In fact, the calculation of the numerical values of $h_\mu$ for a source $\mu$ and of the capacity $c$ for a channel $p$ allows definite practical predictions:
If $h_\mu < c$ it is possible to adapt the source to the channel in such a way that a source sequence (of length $n$ ) can be determined from the corresponding output sequences with a small probability of error. Thus the channel may be used to transmit the information provided by the source.

If $h_\mu > c$ , this is not possible. Thus there is no way of using the channel to transmit all the information provided by the source. This last statement still requires some proof.

Z.17 Proposition ("Converse" of Shannon's theorem):
Let $p$ be a channel with capacity $c$ and let $(Z, \Upsilon, \nu; \tau)$ be a source with $h_\nu(Z;\tau) > c$ . Then there exists $\varepsilon > 0$ such that there is no blocking coding $m_f$ and no guessing $g$ satisfying

(i) $p'_\nu(E) \leq \varepsilon$ for $p' = p \circ m_f$ ,

or

(ii) $t'(\nu) \geq h_\nu(Z;\tau) - \varepsilon$ .

Proof:
If we choose $0 < \varepsilon < h_\nu - c$ , it is clear that (ii) is impossible since otherwise

$$t'(\nu) \geq h_\nu(Z;\tau) - \varepsilon > c \geq c'$$

which is the capacity of the compound channel $p' = p \circ m_f$ and hence by definition at most equal to the capacity $c$ of the channel $p$ .

As for (i) we assume that for every $\varepsilon > 0$ it is possible to achieve $p'_\nu(E) \leq \varepsilon$ and use the argument in the proof of (Z.16ii) in order to obtain a contradiction. Indeed, we have shown the following:

$$h_\nu(\gamma^n | \beta^n) \leq h_\nu(\{E, E^c\}) + p'_\nu(E) \log |\gamma^n|$$

and therefore

$$\frac{1}{\ell n} h_\nu(\gamma^{\ell n} | \beta^{\ell n}) \leq \frac{1}{n} h_\nu(\gamma^n | \beta^n) \leq \frac{1}{n} h_\nu(\{E, E^c\}) + p'_\nu(E) \log |\gamma| .$$

Given $\varepsilon > 0$ it is now possible to achieve

$$t'(\nu) = h_\nu(\gamma;\tau) - \lim \frac{1}{\ell n} h_\nu(\gamma^{\ell n} | \beta^{\ell n})$$

$$\geq h_\nu(Z;\tau) - \frac{1}{n} h_\nu(\{E, E^c\}) + p'_\nu(E) \log |\gamma|$$

$$\geq h_\nu(Z;\tau) - \varepsilon$$

which contradicts (ii). ∎

Remark:

In our version of Shannon's theorem we have simply relied on the concepts from ergodic theory developped in this book, to find the most convenient assumptions. Although we have not aimed at utmost generality, we have achieved a version that contains most "Shannon theorems" in the literature. More general versions have recently been proved by Gray and Ornstein [1979], and Kieffer [1981]. Kieffer has managed to prove a "Shannon theorem" for weakly continuous ergodic channels; this seems to be the most general version.

part 3

In this final part we show a version of the theorem of McMillan [1953]. Let $(X, \Sigma, \mu; \varphi)$ denote an MDS and for a disjoint cover $\alpha \in \tilde{\Sigma}_d$ we consider the following closed subspaces of $E := L^1(X, \Sigma, \mu)$:

$$F_o := \langle \mathbb{1} \rangle,$$

$$F_n := \operatorname{lin} \{ \mathbb{1}_A : A \in \alpha_{-n}^{-1} \} \quad \text{and}$$

$$F := \overline{\bigcup_{n \geq 0} F_n} \quad .$$

As in (Y.14) one shows easily that for the conditional expectations (compare B.24)

$$P_n : E \longrightarrow F_n \quad \text{and} \quad P: E \longrightarrow F$$

we have strong operator convergence, i.e.

$$P_n f \xrightarrow{\| \|} Pf$$

for every $f \in E$.

In the following we take $n = 0,1,2,\dots,k \in \mathbb{N}$ and define

$$q_{nk} := \sum_{A \in \alpha^k} \mathbb{1}_A \cdot P_n \mathbb{1}_A \quad \text{and} \quad r_{nk} := -\frac{1}{k} \log q_{nk} \; .$$

Observe that

$$q_{ok} = \sum_{A \in \alpha^k} \mu(A) \mathbb{1}_A \quad , \; r_{ok} = -\frac{1}{k} \sum_{A \in \alpha^k} (\log \mu(A)) \cdot \mathbb{1}_A$$

and therefore

$$\int_X r_{ok} \, d\mu = -\frac{1}{k} \sum_{A \in \alpha^k} \mu(A) \log \mu(A) = \frac{1}{k} h_\mu(\alpha^k)$$

More generally, one obtains the following relations.

Z.18 Lemma:

(i) For $A \in \alpha^k$, $B \in \alpha_{-n}^{-1}$ and $x \in A \cap B$ we have

$q_{nk}(x) = \mu(A \cap B) / \mu(B)$ .

(ii) $\int r_{nk} \, d\mu = \frac{1}{k} h_\mu(\alpha^k | \alpha_{-n}^{-1})$ .

Proof:

(i) By definition,

$$q_{nk}(x) = P_n 1\!\!1_A(x) \qquad \text{for} \quad x \in A.$$

Since $x \in B$ and $P_n 1\!\!1_A$ is constant on $B$, we have

$$P_n 1\!\!1_A(x) = \frac{1}{\mu(B)} \langle P_n 1\!\!1_A, 1\!\!1_B \rangle$$
$$= \frac{1}{\mu(B)} \langle 1\!\!1_A, 1\!\!1_B \rangle = \frac{\mu(A \cap B)}{\mu(B)} .$$

(ii) By (i) we obtain

$$\int r_{nk} d\mu = \sum_{A \in \alpha^k} \sum_{B \in \alpha_{-n}^{-1}} \int r_{nk} 1\!\!1_{A \cap B} d\mu$$
$$= \sum_A \sum_B -\frac{1}{k} \mu(A \cap B) \log \frac{\mu(A \cap B)}{\mu(B)}$$
$$= \frac{1}{k} h_\mu(\alpha^k \mid \alpha_{-n}^{-1}) . \blacksquare$$

From the strong convergence of $P_n$ to $P$ it follows that the functions $q_{nk}$ norm converge to

$$q_k := \sum_{A \in \alpha^k} 1\!\!1_A P 1\!\!1_A$$

as $n \to \infty$. The analogous result for the functions $r_{nk}$ and

$$r_k := -\frac{1}{k} \log q_k$$

is less immediate.

Z.19 Lemma:

For every $k \in \mathbb{N}$, the sequence $(r_{nk})_{n \in \mathbb{N}}$ is contained in $L^1(X, \Sigma, \mu)$ and norm converges to $r_k$.

Proof:

The first assertion follows from (Z.18.ii).

Now take $A \in \alpha^k$, $B \in \alpha_{-n}^{-1}$ and $x \in A \cap B$. As observed above, we have

$$r_{nk}(x) = -\frac{1}{k} \log \frac{\mu(A \cap B)}{\mu(B)}$$

and therefore

$$2^{-k(t+1)}\mu(B) \leq \mu(A\ B) \leq 2^{-kt}\mu(B),$$

if $t \leq r_{nk}(x) \leq t+1$ .

Consequently

$$\int r_{nk}\, 1\!\!1_{[t \leq r_{nk} \leq t+1]}\, d\mu = \sum_{A \in \alpha^k} \sum_{B \in \alpha_{-n}^{-1}} \int_{A \cap B} r_{nk} 1\!\!1_{[t \leq r_{nk} \leq t+1]} d\mu$$

$$\leq \sum_{A} \sum_{B} 2^{-kt}(t+1)\mu(B)$$

$$\leq |\alpha^k| \cdot 2^{-kt}(t+1) .$$

As a technical aid we consider the functions

$$f_s : [0,1] \longrightarrow \mathbb{R}, \quad s \in \mathbb{R}_+ ,$$

defined by

$$f_s(x) := \inf \{-\log x, s\} , \ x \in [0,1].$$

By the boundness of $\left\{ \left| \frac{f_s(x) - f_s(x')}{x - x'} \right| : x,x' \in [0,1] \right\}$ it follows that $f_s \circ q_{nk}$ norm converges to $f_s \circ q_k$ as $n \longrightarrow \infty$ .

Therefore, the middle term on the right hand side of the estimate

$$\| r_{nk} - r_k \| \leq \| r_{nk} - \tfrac{1}{k} f_s \circ q_{nk} \| + \tfrac{1}{k} \| f_s \circ q_{nk} - f_s \circ q_k \| + \| \tfrac{1}{k} f_s \circ q_k - r_k \|$$

becomes small for $n$ sufficiently large. Also, the first and the last term can be made small (for $s$ sufficiently large) since we have shown above that

$$\int r_{nk}\, 1\!\!1_{[s \leq r_{nk}]}\, d\mu \leq \sum_{\substack{t \in \mathbb{N} \\ t \geq s}} |\alpha^k| \cdot 2^{-kt}(t+1) . \blacksquare$$

Before stating the main theorem we need one more technical result.

Z.20 Lemma:

$r_{ok} = \frac{1}{k} \sum_{i=0}^{k-1} T^i r_{i1}$ , where $T$ ($:= T_\varphi$) denotes the operator induced by $\varphi$ on $L^1(X, \Sigma, \mu)$.

Proof:

Choose $x \in A := A_o \cap \varphi^{-1}(A_1) \cap \dots \cap \varphi^{-k+1}(A_{k-1}) \in \alpha^k$.

Then $r_{o1}(x) = -\log q_{o1}(x) = -\log P_o \mathbb{1}_{A_o}(x) = -\log \mu(A_o)$ ,

$$r_{11}(\varphi(x)) = -\log q_{11}(\varphi(X)) = -\log \frac{\mu(A_1 \cap \varphi A_o)}{\mu(\varphi A_o)}$$

$$= -\log \frac{\mu(\varphi^{-1} A_1 \cap A_o)}{\mu(A_o)} ,$$

$$r_{21}(\varphi^2(x)) = -\log q_{21}(\varphi^2(x)) = -\log \frac{\mu(A_2 \cap \varphi A_1 \cap \varphi^2 A_o)}{\mu(\varphi A_1 \cap \varphi^2 A_o)}$$

$$= -\log \frac{\mu(\varphi^{-2} A_2 \cap \varphi^{-1} A_1 \cap A_o)}{\mu(\varphi^{-1} A_1 \cap A_o)} ,$$

and finally

$$r_{(k-1)1}(\varphi^{k-1}(x)) = \log \frac{\mu(\varphi^{-k+1} A_{k-1} \cap \dots \cap A_o)}{\mu(\varphi^{-k+2} A_{k-2} \cap \dots \cap A_o)} .$$

Thus we obtain

$$\frac{1}{k} \sum_{i=0}^{k-1} r_{i1}(\varphi^i(x)) = -\frac{1}{k} \log \mu(A) = -\frac{1}{k} \log P_o \mathbb{1}_A(x) = r_{ok}(x). \blacksquare$$

Z.21 <u>Theorem</u>:

Let $(X, \Sigma, \mu; \varphi)$ denote an MDS and $T$ $(:= T_\varphi)$ the induced operator on $L^1(X, \Sigma, \mu)$. Consider the functions $r_{nk}$ and $r_k$ defined above. Then

$$\lim_{k \to \infty} r_{ok} = r ,$$

where

$$r := \lim_{n \to \infty} T_n r_1$$

for the Cesaro means

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

Proof:

By the mean ergodic theorem (IV.6) it is clear that the limit $r$ exists and is T-invariant. But by (Z.20)

$$\| r_{ok} - T_n r_1 \| \leq \frac{1}{n} \sum_{i=0}^{n-1} \| r_{i1} \circ \varphi^i - r_1 \circ \varphi^i \|$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} \| r_{i1} - r_1 \|$$

which converges to zero by (Z.19). ■

Z.22 <u>Corollary</u> (McMillan, 1953):

If $(X, \Sigma, \mu; \varphi)$ is an ergodic MDS then $\lim_{k \to \infty} r_{ok} = h_\mu(X; \varphi) \cdot \mathbb{1}$ .

Proof:

By the ergodicity of $\varphi$ we have $\lim_{k \to \infty} r_{ok} = c \cdot \mathbb{1}$ and

$$c = \int \lim_{k \to \infty} r_{ok} \, d\mu = \lim_{k \to \infty} \int r_{ok} \, d\mu = \lim_{k \to \infty} \frac{1}{k} h_\mu(\alpha^k) d\mu = h_\mu(X, \varphi). \blacksquare$$

This last corollary has important applications in information theory as shown in part 2 of this Appendix.

Note that we have used the mean ergodic theorem and have shown $L^1$-convergence in (Z.21) and (Z.22), whereas usually a.e.-convergence is proved by means of Birkhoff's ergodic theorem and Doob's martingale theorem, which we did not need here.

References: Adler [1961], Ash [1965], Bauer [1963], Breiman [1957], Feinstein [1954], [1958], Gray and Ornstein [1979], Khinchin [1957], Kieffer [1981], McMillan [1953], Parthasarathy [1961], Pfaffelhuber [1971], Shannon [1948], Wolfowitz [1964].

R/1

# BIBLIOGRAPHY


Adler, R.L.

[1961] Ergodic and mixing properties of infinite memory channels. Proc. Amer. Math. Soc. 12, 924-930 (1961).

Adler, R.L., Konheim, A.G. and M.H. McAndrew

[1965] Topological entropy. Trans. Amer. Math. Soc. 114, 309-319 (1965).

Adler, R.L., Shields, P. and M.Smorodinsky

[1972] Irreducible Markov shifts. Ann. Math. Statistics 43, 1027 - 1029 (1972).

Adler, R.L. and B. Weiss

[1970] Similarity of automorphisms of the torus. Mem. Amer. Math. Soc. 98 (1970).

Akcoglu, M.A.

[1966] Pointwise ergodic theorems. Trans. Amer. Math. Soc. 125, 296-309 (1966).

[1975a] Positive contractions on $L^1$-spaces. Math. Z. 143, 1-13 (1975).

[1975b] A pointwise ergodic theorem in $L^p$-spaces. Can. J. Math. 27, 1075-1082 (1975).

[1979] Pointwise ergodic theorems in $L_p$ spaces. Ergodic Theory, Proceedings, Oberwolfach 1978. Lecture Notes Math. 729, p. 13-15 (1979).

Akcoglu, M.A., Huneke, J.P. and H. Rost

[1974] A counterexample to the Blum-Hanson theorem in general spaces. Pacific J. Math. 50, 305-308 (1974).

R/2


Akcoglu, M.A. and P.E. Kopp

[1977] Construction of dilations of positive $L_p$ contractions. Math. Z. 155, 119-127 (1977).

Akcoglu, M.A. and L. Sucheston

[1972] On operator convergence in Hilbert space and in Lebesgue space. Periodica Math. Hung. 2, 235-244 (1972).

[1975a] On convergence of iterates of positive contractions in $L_p$-spaces. J. Approx. Theory 13, 348-362 (1975).

[1975b] Weak convergence of positive contractions implies strong convergence of averages. Z. Wahrscheinlichkeitstheorie verw. Gebiete 32, 139-145 (1975).

[1977] Dilations of positive contractions on $L_p$ spaces. Canad. Math. Bull. 20, 285-292 (1977).

[1978] A ratio ergodic theorem for superadditive processes. Z. Wahrscheinlichkeitstheorie verw. Gebiete 44, 269-278 (1978).

Alaoglu, L. and G. Birkhoff

[1940] General ergodic theorems. Ann. Math. 41, 293-309 (1940).

Alfsen, E.

[1971] Compact Convex Sets and Boundary Integrals. Berlin-Heidelberg-New York: Springer 1971.

Alpern, S.

[1976] New proofs that weak mixing is generic. Inventiones math. 32, 263-279 (1976).

R/3

[1978] A topological analog of Halmos' conjugacy lemma. Inventiones math. 48, 1-6 (1978).

Ando, T.

[1968] Invariante Maße positiver Kontraktionen in C(X). Studia Math. 31, 173-187 (1968).

Anzai, H.

[1951a] On an example of a measure preserving transformation which is not conjugate to its inverse. Proc. Japan Acad. 27, 517-522 (1951).

[1951b] Ergodic skew-product transformations on the torus. Osaka Math. J. 3, 83-99 (1951).

Anzai, K.

[1977] On a mean ergodic theorem for several operators. Keio Engineering Reports 30, 59-68 (1977).

Arnold, V.I.

[1980] Gewöhnliche Differentialgleichungen. Berlin-Heidelberg-New York: Springer 1980.

Ash, R.

[1965] Information Theory. New York-London-Sydney: Interscience Publishers (John Wiley) 1965.

[1972] Real Analysis and Probability. New York-London: Academic Press 1972.

Atalla, R.E.
[1981] Markov operators and quasi-stonian spaces. Proc. Amer. Math. Soc. 82, 613-618 (1981).

Atalla, R. and R. Sine
[1976] Random Markov mappings. Math. Ann. 221, 195-199 (1976).

Avez. A. and V.J. Arnold
[1968] Ergodic Problems in Classical Mechanics. Benjamin (1968).

Axmann, O.
[1980] Struktur- und Ergodentheorie irreduzibler Operatoren auf Banachverbänden. Dissertation. Tübingen 1980.

Banach, S.
[1926] Sur la convergence presque partout de fonctionnelles linéaires. Bull. Sci. Math. France 50, 27-32 and 36-43 (1926).

Bartoszek, W.
[1980] Norm residuality of ergodic operators. Instytut Matematyki Politechniki Wrocławskiej; Raport Serii preprinty nr. 40 (1980).

Bauer, H.
[1972] Probability Theory and Elements of Measure Theory. New York: Holt, Rinehart and Winston, Inc. 1972 .

[1963] Kennzeichnung kompakter Simplexe mit abgeschlossener Extremalpunktmenge. Arch. Math. 14, 415-421 (1963).

Beck, A.
[1973] Recent Advances in Topological Dynamics. Lecture Notes in Math. 318. Berlin-Heidelberg-New York: Springer 1973.

Bellow, A.

[1975] An $L^p$-inequality with application to ergodic theory. Houston J. Math. 1, 153-159 (1975).

[1976] A problem in $L^p$-spaces. Lecture Notes in Math. 541 p. 381-388. Measure Theory, Oberwolfach 1975. Berlin-Heidelberg-New York: Springer 1976.

Bellow, A. and H. Fürstenberg

[1979] An application of number theory to ergodic theory and the construction of uniquely ergodic models. Israel J. Math. 33, 231-240 (1979).

Berglund, J.F. and K.H. Hofmann

[1967] Compact Semitopological Semigroups and Weakly Almost Periodic Functions. Lecture Notes in Math. 42. Berlin-Heidelberg-New York: Springer 1967.

Berglund, J.F., Junghenn, H.D. and P. Milnes

[1978] Compact right topological semigroups and generalizations of almost periodicity. Lecture Notes in Math. 663. Berlin-Heidelberg-New York: Springer 1978.

Billingsley, P.

[1965] Ergodic Theory and Information. Wiley 1965.

Birkhoff, G.D.

[1931] Proof of the ergodic theorem. Proc. Nat. Acad. Sci. USA 17, 656-660 (1931).

[1942] What is the ergodic theorem? Amer. Math. Soc. Monthly 49 , 222-226 (1942).

Birkhoff, G.D. and B.O. Koopman

[1932] Recent contributions to the ergodic theory. Proc. Nat. Acad. Sci.USA 18, 279-282 (1932).

Blum, J.R. and D.L. Hanson

[1960] On the mean ergodic theorem for subsequences. Bull. Amer. Math. Soc. 66, 308-311 (1960).

[1963] On the isomorphism problem for Bernoulli schemes. Bull. Amer. Math. Soc. 69, 221-223 (1963).

Boltzmann, L.

[1871] Einige allgemeine Sätze über Wärmegleichgewicht. Wiener Ber. 63, 679-711 (1871).

[1885] Über die Eigenschaften monozyklischer und anderer damit verwandter Systeme. J. reine angew. Math. 98, 68-94 (1885).

[1887] Über die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik. J. reine angew. Math. 100, 201-212 (1887).

Bonsdorff, H.

[1980] Characterizations of uniform recurrence for general Markov chains. Ann. Acad. Scie. Fennicae A, Diss. 32 (1980).

Boclé, J.

[1960] Sur la théorie ergodique. Ann. Inst. Fourier 10, 1-45 (1960).

K/7

Borel, E.

[1909] Les probabilités dénombrables et leurs applications arithmétiques. Rend. Circ. mat. Palermo 27, 247-271 (1909).

Bowen, R.

[1971] Entropy for group endomorphisms and homogeneous spaces. Trans. Amer. Math. Soc. 153, 401-414 (1971).

[1975] Equilibrium states and the Ergodic Theory of Anosov diffeomorphisms. Lecture Notes in Math. 470. Berlin-Heidelberg-New York: Springer 1974.

[1979] Invariant measures for Markov maps of the interval. Commun. Math. Phys. 69, 1-17 (1979).

Bratteli, O. and O.W. Robinson

[1979/80] Operator Algebras and Quantum Statistical Mechanics I and II. Berlin-Heidelberg-New York: Springer 1980.

Breiman, L.

[1957] The individual ergodic theorem of information theory. Ann. Math. Stat. 28, 809-811 (1957). Correction: Ann. Math. Stat. 31, 809-810 (1960).

[1977] On achieving channel capacity in finite-memory channels. Ill. J. Math. 4, 246-252 (1977).

Brown, J.R.

[1966] Approximation theorems for Markov operators. Pac. J. Math. 16, 13-23 (1966).

[1976] Ergodic Theory and Topological Dynamics. Acad. Press (1976).

[1973] Topological entropy for noncompact sets. Transactions Amer. Math. Soc. 184, 125-136 (1972).

R/8

Brown, A.L. and A. Page

[1970] Elements of Functional Analysis. London: Van Nostrand Reinhold Co. Ltd. 1970.

Brunel, A.

[1963] Sur un lemme ergodique voisin du lemme de E. Hopf, et sur une de ses applications. C.R.Acad.Sci. Paris, A, 256, 5481-5484 (1963).

[1970] New conditions for existence of invariant measures in ergodic theory. Lecture Notes in Math. 160, 7-17. Berlin-Heidelberg-New York: Springer 1970.

Brunel A. and D. Revuz

[1974] Quelques applications probabilistes de la quasi-compacité. Ann. Inst. Henri Poincaré, Sect. B, 10 , 301-337 (1974).

Brush, S.G.

[1971] Milestones in mathematical physics, proof of the impossibility of ergodic systems: The 1913 papers of Rosenthal and Plancherel. Transport Theory and Stat. Physics 1, 287-311 (1971).

Bryan, G.H. and J. Larmor

[1892] On our knowledge of thermodynamics. Rep. 61th meeting British Ass. Advancement Science, Cardiff 1891, 85-122 (1892).

Burburg, S.H.

[1876] On the second law of thermodynamics in connexion with the kinetic theory of gases. Philos. Magazine J. Science 1 (5th series), 61-67 (1876).

R/3

Burckel, R.B.

[1970] Weakly Almost Periodic Functions on Semigroups. New York-London-Paris: Gordon and Breach, 1970.

Burkholder, D.L.

[1962] Semi-Gaussian subspaces. Trans. Amer. Math. Soc. 104, 123-131 (1962).

Chacón, R.V.

[1962] Identification of the limit of operator averages. J. Math. Mech. 2, 961-968 (1962).

[1964] A class of linear transformations. Proc. Amer. Math. Soc. 15, 560-564 (1964).

R/10

Chacon, R.V. and U. Krengel
[1964] Linear modulus of a linear operator. Proc. Amer. Math. Soc. 15, 553-559 (1964).

Chacon, R.V. and D.S. Ornstein
[1960] A general ergodic theorem. Ill. J. Math. 4, 153-160 (1960).

Choksi, J.R. and S. Kakutani
[1979] Residuality of ergodic measurable transformations and of ergodic transformations which preserve an infinite measure. Indiana Univ. Math. J. 28, 453-469 (1979).

Cigler, J. and G. Helmberg
[1961] Neuere Entwicklungen der Theorie der Gleichverteilung. Jahresbericht d. DMV 64, 1-50 (1961).

Clausius, R.
[1872] Über den Zusammenhang des zweiten Hauptsatzes der mechanischen Wärmelehre mit dem Hamilton'schen Prinzip. Poggendorfs Annalen 146, 585-591 (1872).

Conze, J.P. and N. Dặng-Ngọc
[1978] Ergodic theorems for noncommutative dynamical systems. Invent. math. 46, 1-15 (1978).

Conze, J.P. and M. Keane
[1976] Thêorie Ergodique. Rennes 1973/74. Lecture Notes in Math. 532. Berlin-Heidelberg-New York: Springer 1976.

Cornfeld, I.P., Fomin, S.V. and J.G. Sinai
[1982] Ergodic Theory. Grundlehren math. Wiss. 245. Berlin-Heidelberg-New York: Springer 1982.

Dặng-Ngọc, N.
[1980] On the integrability of the maximal ergodic function. Proc. Amer. Math. Soc. 79, 565-570 (1980).

R/11


Davies, E.B.

[1981] One-parameter Semigroups. New York-London: Academic Press 1981.

Day, M.M.

[1969] Semigroups and amenability. in: Folley, K.W. (ed.): Semigroups p. 5-53. New York-London: Academic Press 1969.

Denker, M.

[1973] On strict ergodicity. Math. Z. 134, 231-253 (1973).

Denker, M., Grillenberger, C. and K. Sigmund

[1976] Ergodic Theory of Compact Spaces. Lecture Notes in Math. 527. Berlin-Heidelberg-New York: Springer 1976.

Denker, M. and K. Jacobs (ed.)

[1979] Ergodic Theory. Proc. Oberwolfach 1978. Lecture Notes in Math. 729. Berlin-Heidelberg-New York: Springer 1979.

Derriennic, Y.

[1976] Loi "zéro ou deux" pour les processus de Markov: Applications aux marches aléatoires. Ann. Inst. Henri Poincaré, Sect. B, 12, 111-129 (1976).

Dieudonné, J.

[1977] Panorama des mathématiques pures. Le choix bourbachique. A IV. Théorie ergodique. Paris: Gauthier-Villars 1977.

Dinaburg, E.I.

[1970] The relation between topological entropy and metric entropy. Soviet Math. Dokl. 11, 13-16 (1970).

Dugundji, J.

[1966] Topology. Boston: Allyn and Bacon, Inc. 1966.

P/12


Dunford, N. and J.T. Schwartz

[1956] Convergence almost averywhere of operator averages. J. Math. Mech. 5, 129-178 (1956).

[1958] Linear Operators, Part I. New York: Interscience 1958.

Dynkin, E.B.

[1965] Markov Processes. Berlin-Heidelberg-New York: Springer 1965.

Dynkin, E.B. and A.A. Juschkewitsch

[1969] Sätze und Aufgaben Über Markoffsche Prozesse. Berlin-Heidelberg-New York: Springer 1969.

Eberlein, W.F.

[1948] Abstract ergodic theorems. Proc. Nat. Acad. Scie. USA 34, 43-47 (1948).

[1949] Abstract ergodic theorems and weak almost periodic functions. Trans. Amer. Math. Soc. 67, 217-240 (1949).

Ehrenfest, P. and T.

[1911] Begriffliche Grundlagen der statistischen Auffassung. Encykl. Math. Wissenschaften IV, Artikel 32, Nr. 10. Leipzig 1911.

Ellis, R.

[1957] Locally compact transformation groups. Duke Math. J. 24, 119-125 (1957).

[1969] Lectures on Topological Dynamics. Benjamin 1969.

R/13

Emch, G.G.

[1972] Algebraic Methods in Statistical Mechanics and Quantum Field Theory. John Wiley & Sons 1972.

Falkowitz, M.

[1973] On finite invariant measures for Markov operators. Proc. Amer. Math. Soc. 38, 553-557 (1973).

Farquhar, I.E.

[1964] Ergodic Theory in Statistical Mechanics. John Wiley & Sons 1964.

Feinstein, A.

[1954] A new basic theorem of information theory. IRE Trans. Inform. Theory PGIT 4, 2-22 (1954).

[1958] Foundations of Information Theory. New York: McGraw-Hill (1958).

Feldman, M.B.

[1981] A proof of Lusin's Theorem. Amer. Math. Monthly 88, 191-192 (1981).

Feller, W.

[1957] An Introduction to Probability Theory and Its Applications. John Wiley & Sons 1957.

Foguel, S.R.

[1969] The Ergodic Theory of Markov Processes. van Nostrand Math. Studies 21 (1969).

[1976] More on the "zero-two" law. Proc. Amer. Math. Soc. 61, 262-264 (1976).

[1979] Harris operators. Israel J. Math. 33, 281-309 (1979).

[1980] Selected topics in the study of Markov operators. Carolina Lecture Series (1980).

R/14


Fong, H. and L. Sucheston

[1974] On a mixing property of operators in $L_p$-spaces. Z. Wahrscheinlichkeitstheorie verw. Gebiete 28, 165-171 (1974).

Fortet, R.

[1978] Condition de Doeblin et quasi-compacité. Ann. Inst. Henri Poincaré, Sect. B., 14, 379-390 (1978).

Friedman, N.A.

[1970] Introduction to Ergodic Theory. van Nostrand Math. Studies 29 (1970).

Friedman, N.A. and D.S. Ornstein

[1971] On isomorphism of weak Bernoulli transformations. Adv. in Math. 5, 365-394 (1971).

Fuchs,

[1970] Infinite Abelian Groups I. New York-London: Academic Press 1970.

Furstenberg, H.

[1977] Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. J. Analyse Math. 31, 204-256 (1977).

[1981a] Recurrence in Ergodic Theory and Combinatorial Number Theory. Princeton, New Jersey: Princeton University Press 1981.

[1981b] Poincaré Recurrence and Number Theory. Bull.Amer.Math.Soc. 5, 211 - 234 (1981).

R/15

Furstenberg, H., Y. Katznelson and D. Ornstein
[1982] The ergodic theoretical proof of Szemerédi's theorem. Bull. Amer. Math. Soc. 7, 527-552 (1982)

Furstenberg, H. and Y. Katznelson
[1978] An ergodic Szemerêdi theorem for commuting transformations. J. Analyse Math. 34, 275-291 (1978).

Furstenberg, H. and B. Weiss
[1978] Topological dynamics and combinatorial number theory. J. Analyse Math. 34, 61-85 (1978).

Gait, J. and S. Koo
[1972] Averages of functions and ergodic measures in F-spaces. Math. Systems Theory 6, 23-25 (1972).

Gallavotti, G. and D.S. Ornstein
[1974] Billiards and Bernoulli schemes. Comm. Math. Phys. 38, 83-101 (1974).

Garsia, A.
[1965] A simple proof of E. Hopf's maximal ergodic theorem. J. Math. Mech. 14, 381-382 (1965).
[1970] Topics in Almost Everywhere Convergence. Chicago: Markham Publ. Comp. 1970.

Glicksberg, I. and K. de Leeuw
[1959] Almost periodic compactifications. Bull. Amer. Math. Soc. 65, 134-139 (1959).
[1961] Applications of almost periodic compactifications. Acta Math. 105, 63-97 (1961).

Gallavotti, G.
[1975] Lectures on the billiard. in: Dynamical Systems, Theory and Applications. Lecture Notes Physics 38, 112-143. Berlin-Heidelberg-New York: Springer 1975.
[1981] Aspetti delle teotia ergodica, qualitativa e statistica del moto. Pitagora Editrice, Bologna.

R/16

Gnedenko, B.W.
[1962] Lehrbuch der Wahrscheinlichkeitsrechnung. Berlin: Akademieverlag 1962.

Goodman, T.N.T.
[1971] Relating topological entropy and measure entropy. Bull. London Math. Soc. 3, 176-180 (1971).

Goodwyn, L.W.
[1969] Topological entropy bounds measure-theoretic entropy. Proc. Amer. Math. Soc. 23, 679-688 (1969).

Gottschalk, W.H. and G.A. Hedlund
[1955] Topological Dynamics. Amer. Math. Soc. Colloq. Publ. 36. Providence 1955, reprinted 1968.

Greenleaf, F.P.
[1969] Invariant Means on Topological Groups and Their Applications. van Nostrand Math. Studies 16 (1969).

Greiner, G. and R. Nagel
[1982] La loi "zero ou deux" et ses conséquences pour le comportement asymptotique des opérateurs positifs. J. Math. pures appl.

Gray, R.M. and D.S. Ornstein
[1979] Block Coding for discrete stationary $\bar{d}$-continuous noisy channels. IEEE Trans. on Inf. Theory IT 25, 292-306 (1979).

R/17


Groh, U. and B. Kümmerer

[1982] Bibounded operators on $W^*$-algebras. Math. Scand.

Haezendonk, J.

[1973] Abstract Lebesgue-Rohlin spaces. Bull. Soc. Math. Belgique 25, 243-258 (1973).

Hajian, A.B. and Y. Ito

[1967] Conservative positive contractions in $L^1$. Proc. Fifth Berkeley Symp., Vol. 2, part 2, 361-374 (1967).

[1978] Transformations that do not accept a finite invariant measure. Bull. Amer. Math. Soc. 84, 417-427 (1978).

Hajian, A.B. and S. Kakutani

[1964] Weakly wandering sets and invariant measures. Trans. Amer. Math. Soc. 110, 136-151 (1964).

Halmos, P.R.

[1944a] Approximation theorems for measure preserving transformations. Trans. Amer. Math. Soc. 55, 1-18 (1944).

[1949] Measurable transformations. Bull. Amer. Math. Soc. 55, 1015-1034 (1949).

[1950] Measure Theory. Berlin-Heidelberg-New York: Springer 1950

[1956] Lectures on Ergodic Theory. Chelsea 1956.

[1961] Recent progress in ergodic theory. Bull. Amer. Math. Soc. 67, 70-80 (1961).

[1974] Lectures on Boolean Algebras. Berlin-Heidelberg-New York: Springer 1974.

[1944b] In general a measure-preserving transformation is mixing. Ann. of Math. 45, 786-792 (1944).

R/18


Halmos, P.R. and J. von Neumann

[1942] Operator methods in classical mechanics II. Ann. Math. 43, 332-350 (1942).

Halmos, P R and H Samelson

[1942] On monothetic groups. Proc. Nat. Acad. Sciences USA 28, 254-258 (1942).

Hansel, G.

[1974] Strict uniformity in ergodic theory. Math. Z. 135, 221-248 (1974).

Hansel, G. and J.P. Raoult

[1973] Ergodicity, uniformity and unique ergodicity. Indiana Univ. Math. J. 23, 221-237 (1973).

Helson, H. and W. Parry

[1978] Cocycles and spectra. Arkiv Mat. 16, 195-206 (1978).

Hewitt, E. and K.A. Ross

[1963] Abstract Harmonic Analysis. Vol. I. 1st.ed. 1963, 2nd ed. 1979. Berlin-Heidelberg-New York: Springer 1963.

Hiai, F. and R. Sato

[1977] Mean ergodic theorems for semigroups of positive linear operators. J. Math. Soc. Japan 29, 123-134 (1977).

Hille, E. and R.S. Phillips

[1957] Functional Analysis and Semi-groups. 2nd ed. Providence, R.I. 1957.

Hlawka, E.

[1979] Theorie der Gleichverteilung. Mannheim-Wien-Zürich: Bibliographisches Institut 1979.

Helmer, D.

[1980] Joint continuity of affine semigroup actions. Semigroup Forum 21, 153-165 (1980).

Hofbauer, F. and Keller, G.

[1982] Ergodic Properties of Invariant Measures for Piecewise Monotonic Transformations. Math.Z. 180, 119 - 140 (1982) .

R/19

Hofmann, K.H. and P.S. Mostert

[1966] Elements of Compact Semigroups. Columbus, Ohio: Charles E. Merrill Books, Inc. 1966.

Hopf, E.

[1937] Ergodentheorie. Berlin-Heidelberg-New York: Springer 1970.

[1954] The general temporally discrete Markov process. J. Rat. Mech. Anal. 3, 13-45 (1954).

[1961] On the ergodic theorem for positive linear operators. J. Math. Mech. 205, 101-106 (1961).

Horowitz, S.

[1972] Transition probabilities and contractions on $L_\infty$. Z. Wahrscheinlichkeitstheorie verw. Gebiete 24, 263-274 (1972).

Ionescu-Tulcea, A.

[1964] Ergodic properties of isometries in $L_p$-spaces, $1 < p < \infty$. Bull. Amer. Math. Soc. 70, 366-371 (1964).

Ionescu-Tulcea, A. and C.

[1963] Abstract ergodic theorems. Trans. Amer. Math. Soc. 107, 107-124 (1963).

[1969] Topics in the Theory of Lifting. Berlin-Heidelberg-New York: Springer 1969.

Ito, Y.

[1965] Uniform integrability and the pointwise ergodic theorem. Proc. Amer. Math. Soc. 16, 222-227 (1965).

R/20


Iwanik, A.

[1980] Approximation theorems for stochastic operators. Indiana Univ. Math. J. 29, 415-425 (1980).

Jacobs, K.

[1956] Ergodentheorie und festperiodische Funktionen und Halbgruppen. Math.Z.64, 293 - 338 (1956).

[1960] Neuere Methoden und Ergebnisse der Ergodentheorie. Berlin-Heidelberg-New York: Springer 1960.

[1963] Lecture Notes on Ergodic Theory. Aarhus University 1963.

[1965] Einige neuere Ergebnisse der Ergodentheorie. Jahresbericht Deutsche Math. Verein. 67, 143-182 (1965).

[1972a] Poincarés Wiederkehrsatz. Selecta Math. IV, p. 31-56. Heidelberger Tachenbücher. Berlin-Heidelberg-New York: Springer 1972 .

[1972b] Gleichverteilung mod 1. Selecta Math. IV, p. 57-93. Heidelberger Taschenbücher 98. Berlin-Heidelberg-New York: Springer 1972.

Jewett, R.J.

[1970] The prevalence of uniquely ergodic systems. J. Math. Mech. 19, 717-729 (1970).

Jones, L.K.

[1971] A mean ergodic theorem for weakly mixing operators. Advances Math. 7, 211-216 (1971).

[1972] An elementary lemma on sequences of integers and its applications to functional analysis. Math. Z. 126, 299-307 (1972).

[1978] Measure and Integral. New York-San Francisco-London: Academic Press 1978.

[1983] Arithmetische Progressionen. Jahresberichte d. Dt. Math.-Verein. 85, 55-65 (1983)

Jones, L.K. and M. Lin

[1976] Ergodic theorems of weak mixing type. Proc. Amer. Math. Soc. 57, 50-52 (1976).

[1980] Unimodular eigenvalues and weak mixing. J. Functional Anal. 35, 42-48 (1980).

Jones, R.L.

[1977] Inequalities for the ergodic maximal function. Studia Math. 60, 111-129 (1977).

del Junco, A. and J. Rosenblatt

[1979] Counterexamples in ergodic theory and number theory. Math. Ann. 245, 185-197 (1979).

Kac, M.

[1947] On the notion of recurrence in discrete stochastic processes. Bull. Amer. Math. Soc. 53, 1002-1010 (1947).

Kakutani, S.

[1950] Ergodic Theory. Proc. Int. Congress Math. Cambridge 128-142 (1950).

Kakutani, S. and K. Petersen

[1981] The speed of convergence in the ergodic theorem. Mh. Math. 91, 11-18 (1981).

Karlin, S.

[1959] Positive operators. J. Math. Mech. 8, 907-937 (1959).

Katok, A.B.

[1980] Lyapunov Exponents, Entropy and Periodic Orbits for Diffeomorphisms. Publ. Math. I.H.E.S. 51, 137-173 (1980).

[1981] Ergodic Theory and Dynamical Systems I. Proc. Special Year, Maryland 1979/80. Birkhäuser 1981.

Katok, A.B., Sinai, Ya.G. and A.M. Stepin

[1977] Theory of dynamical systems and general transformation groups with invariant measure. J. Soviet Math. 77, 974-1065 (1977).

Katok, A.B. and A.M. Stepin

[1967] Approximations in ergodic theory. Russian Math. Surveys 22, 77-102 (1967).

[1970] Metric properties of measure-preserving homeomorphisms. Russian Math. Surveys 25, 191-220 (1970).

Katznelson, Y.

[1971] Ergodic automorphisms of $T^n$ are Bernoulli shifts. Israel J. Math. 10, 186-195 (1971).

[1976] An Introduction to Harmonic Analysis. New York: Dover Publications, Inc. 1976.

Keane, M. and M. Smorodinsky

[1979] Bernoulli schemes of the same entropy are finitarily isomorphic. Ann. Math. 109, 397-406 (1979).

Kern, M., Nagel, R. and G. Palm

[1977] Dilations of positive operators: construction and ergodic theory. Math. Z. 156, 265-277 (1977).

Keynes, H.B. and J.B. Robertson

[1968] On ergodicity and mixing in topological transformation groups. Duke Math. J. 35, 809-819 (1968).

[1969] Generators for topological entropy and expansivness. Math. Systems Theory 3, 51-59 (1969(.

Krengel, U.

[1963] Über den Absolutbetrag stetiger linearer Operatoren und seine Anwendung auf ergodische Zerlegungen. Math. Scand. 13, 151-187 (1963).

[1966] On the global limit behaviour of Markov chains and of the general non-singular Markov processes. Z. Wahrscheinlichkeitstheorie verw. Gebiete 6, 302-316 (1966).

[1972] Weakly wandering vectors and weakly independent partitions. Trans. Amer. Math. Soc. 164, 199-226 (1972).

[1977] Recent progress on ergodic theorems. Soc. Math. France, Astérisque 50, 151-192 (1977).

[1978] On the speed of convergence in the ergodic theorem. Monatshefte Math. 86, 3-6 (1978).

[1983] Ergodic Theorems. de Gruyter-Verlag.

Krengel, U. and L. Sucheston

[1969] On mixing in infinite measure spaces. Z. Wahrscheinlichkeitstheorie verw. Gebiete 13, 150-164 (1969).

Krieger, W.

[1970] On entropy and generators of measure-preserving transformations. Trans. Amer. Math. Soc. 149, 453-464 (1970).

[1972] On unique ergodicity. Proc. Sixth Berkeley Symposium. Math. Stat. Prob. 1970, Vol. II. Univ. California Press, 1972, p. 327-345.

Krylov, N. and N. Bogoljubov

[1937] La théorie générale de la mesure dans son application à l'étude des systèmes de la mécanique non linéaire. Ann. Math. 38, 65-113 (1937).

Kümmerer, B.

[1978] A non-commutative individual ergodic theorem. Inventiones math. 46, 139-145 (1978).

R/25

Kümmerer, B. and R. Nagel

[1979] Mean ergodic semigroups on $W^*$-algebras. Acta Sci. Math. 41, 151-159 (1979).

Kushnirenko, A.G.

[1967] Metric invariants of entropy type. Russian Math. Surveys 22, 53-61 (1967).

Lamperti, J.

[1977] Stochastic Processes. A Survey of the Mathematical Theory. Applied Mathematical Sciences 23. Berlin-Heidelberg-New York: Springer 1977.

Lance, C.

[1976] Ergodic theorems for convex sets and operator algebras. Inventiones math. 37, 201-214 (1976).

Lasota, A.

[1980] A fixed point theorem and its application in ergodic theory. Tôhoku Math. J. 32, 567-575 (1980).

Lasota, A. and J. Yorke

[1973] On the existence of invariant measures for piecewise monotonic transformations. Trans. Amer. Math. Soc. 186, 481-488 (1973).

Lawson, J.D.

[1974] Joint continuity in semitopological semigroups. Ill. J. Math. 18, 275-285 (1974).

[1976] Additional notes on continuity in semitopological semigroups. Semigroup Forum 12, 265-280 (1976).

R/26


Lin, M.

[1971] Mixing for Markov operators. Z. Wahrscheinlichkeitstheorie verw. Gebiete 19, 231-242 (1971).

[1974] On the uniform ergodic theorem. Proc. Amer. Math. Soc. 43 (2), 337-340 (1974).

[1975] Quasi-compactness and uniform ergodicity of Markov operators. Ann. Inst. Henri Poincaré, Section B, 10, 345-354 (1975).

[1978] Quasi-compactness and uniform ergodicity of positive operators. Israel J. Math. 29, 309-311 (1978).

[1981] On weakly mixing Markov operators and non-singular transformations. Z. Wahrscheinlichkeitstheorie verw. Gebiete 55, 231-236 (1981).

[1982] On the "zero-two" law for conservative Markov processes.

Lin, M. and R. Sine

[1979] A spectral condition for strong convergence of Markov operators. Z. Wahrscheinlichkeitstheorie verw. Gebiete 47, 27-29 (1979).

Lotz, H.P.

[1968] Über das Spektrum positiver Operatoren. Math. Z. 108, 15-32 (1968).

[1981] Uniform ergodic theorems for Markov operators on C(X). Math. Z. 178, 145-156 (1981).

Mackey, G.W.

[1974] Ergodic theory and its significance for statistical mechanics and probability theory. Advances Math. 12, 178-268 (1974).

Lotz, H.P. and Schaefer, H.H.

[1968] Über einen Satz von F. Niiro und J. Sawashima. Math. Z. 108, 33-36 (1968)

Ludwig, G.

[1976] Einführung in die Grundlagen der Theoretischen Physik, Vol. 3: Quantenmechanik. Vieweg & Sohn, Braunschweig 1976.

Luecke, G.R.

[1977] Norm Convergence of $T^n$. Can.J.Math. 29, 1340 - 1344 (1977).

R/27


Martin, N.F.G. and J.W. England

[1981] Mathematical Theory of Entropy. Encyclopedia of Math. and its Appl. Vol. 12. Addison-Wesley Publ. Comp. 1981.

Maxwell, J.C.

[1879] On Boltzmanns theorem on the average distribution of energy in a system of material points. Trans. Cambridge Phil. Soc. 12, pp. 547 (1879).

McMillan, B.

[1953] The basic theorems of information theory. Ann. Math. Stat. 24, 196-219 (1953).

Meshalkin, L.D.

[1959] A case of isomorphism of Bernoulli schemes. Dokl. Akad. Nauk SSSR 128, 41-44 (1959).

Misiurewicz, M.

[1976] A short proof of the variational principle for a $\mathbb{Z}_+^{\mathbb{N}}$ action on a compact space. Astérisque 40, 147-157 (1976).

Nagel, R.

[1973a] Mittelergodische Halbgruppen linearer Operatoren. Ann. Inst. Fourier 23,75-87 (1973).

[1973b] Ordnungsstetigkeit in Banachverbänden. Manuscripta Math. 9, 9-27 (1973).

[1974] Ergodic and mixing properties of linear operators. Proc. Royal Irish Acad. 74, 245-261 (1974).

Moser, J. (ed.)

[1975] Dynamical systems, Theory and Applications. Lecture Notes Physics 38. Berlin-Heidelberg-New York: Springer 1975.

Nagel, R. and G. Palm

[1982] Lattice dilations of positive contractions on $L^p$-spaces. Canad. Math. Bull.

Nagel, R. and M. Wolff

[1972] Abstract dynamical systems with an application to operators with discrete spectrum. Arch. Math. 23, 170-176 (1972).

Nakano, K. and T. Shimogaki

[1962] A note on the cut extension of C-spaces. Proc. Japan Acad. 38, 473-477 (1962).

Nemyckii, V. and V. Stepanov

[1960] Qualitative Theory of Differential Equations. Princeton,

v. Neumann, J.

[1932a] Zur Operatorenmethode in der klassischen Mechanik. Ann. of Math. (2) 33, 587-642 (1932).

[1932b] Proof of the quasi-ergodic hypothesis. Proc. Nat. Acad. Sci. USA 18, 70-82 (1932).

[1932c] Einige Sätze über meßbare Abbildungen. Ann. Math. 33, 574-586 (1932).

[1932d] Physical applications of the ergodic hypothesis. Proc. Nat. Acad. Sci. USA 18, 263-266 (1932).

Neveu, J.

[1964] Bases Mathématiques du Calcul des Probabilités. Paris: Masson et $C^{ie}$ 1964.

[1967] Existence of bounded invariant measures in ergodic theory. Proc. Fifth Berkeley Symp. Vol. 2, part 2, 461-472 (1967).

Namioka, I.

[1974] Separate continuity and joint continuity. Pacific J. Math. 51, 515-531 (1974).

R/29

Olevskiĭ, A.M.

[1975] Fourier Series with Respect to General Orthogonal Systems. Ergebnisse der Math. Band 86. Berlin-Heidelberg-New York: Springer 1975.

Orey, S.

[1971] Limit Theorems for Markov Chain Transition Probabilities. van Nostrand Math. Studies 34 (1971).

Ornstein, D.S.

[1970] Bernoulli shifts with the same entropy are isomorphic. Advances Math. 4, 337-352 (1970).

[1971] Some new results in the Kolmogorov-Sinai theory of entropy and ergodic theory. Bull. Amer. Math. Soc. 77, 878-890 (1971).

[1971] A remark on the Birkhoff ergodic theorem. Illinois J. Math. 15, 77-79 (1971).

[1973] An example of a Kolmogorov automorphism that is not a Bernoulli shift. Advances Math. 10, 49-62 (1973).

[1974] Ergodic Theory, Randomness, and Dynamical Systems. Yale Math. Monographs 5, Yale Univ. Press 1974.

[1975] What does it mean for a mechanical system to be isomorphic to the Bernoulli flow. in: Dynamical Systems, Theory and Applications. Lecture Notes Physics 38, 209-223. Berlin-Heidelberg-New York: Springer 1975.

Ornstein, D.S. and P.C. Shields

[1973] Mixing Markov Shifts of Kernel Type are Bernoulli. Adv. in Math. 10, 143-146 (1973).

Ornstein, D.S. and L. Sucheston

[1970] An operator theorem on $L_1$-convergence to zero with applications to Markov kernels. Ann. Math. Statist. 41, 1631-1639 (1970).

K/30


Ornstein, D.S. and B. Weiss

[1973] Geodesic flows are Bernoullian. Israel J. Math. 14, 184-198 (1973).

Oxtoby, J.C.

[1952] Ergodic sets. Bull. Amer. Math. Soc. 58, 116-136 (1952).

Palm, G.

[1976a] A common generalization of topological and measure-theoretic entropy. Soc. Math. de France, Astêrisque 40, 159-165 (1976).

[1976b] Entropie und Erzeuger in dynamischen Verbänden. Z. Wahrscheinlichkeitstheorie verw. Gebiete 36, 27-45 (1976).

Parry, W.

[1969] Entropy and Generators in Ergodic Theory. Benjamin 1969.

[1981] Topics in Ergodic Theory. Cambridge University Press 1981.

Parry, W and S. Tuncel

[1982] Classification Problems in Ergodic Theory Cambridge University Press 1982.

Parry, W. and P. Walters

[1972] Endomorphisms of a Lebesgue space. Bull. Amer. Math. Soc. 78, 272-276 (1972).

Parthasarathy, K.R.

[1961] On the integral representation of the rate of transmission of a stationary channel. Ill. J. Math. 2, 299-305 (1961).

Pedersen, G.K.

[1979] $C^*$-Algebras and their Automorphism Groups. Academic Press 1979.

Peressini, A.L.

[1967] Ordered Topological Vector Spaces. New York-Evanston-London: Harper and Row 1967.

Peller, V.

[1978] L'inegalitê de von Neumann, la dilatation isomêtrique et l'approximation par isomêtries dans $L^p$. C.R. Acad. Sc. Paris, 287, 311-314 (1978).

Pfaffelhuber, E.

[1971] Channels with asymptotically decreasing memory and anticipation. IEEE Trans. Inform. Theory IT 17, 379-385 (1971).

Phelps, R.R.

[1966] Lectures on Choquet's Theorem. van Nostrand Math. Studies 7 (1966).

Pianigiani, U.

[1979] On the existence of invariant measures. Proc. Third Inter. Conf. Univ. Texas, Arlington 1978, p. 299-307. Academic Press 1979.

Pietsch, A.

[1972] Nuclear Locally Convex Spaces. Ergebnisse der Math. und ihrer Grenzgebiete, Bd. 66. Berlin-Heidelberg-New York: Springer 1972.

Poincaré, H.

[1890] Sur le problème des trois corps et les équations de la dynamique. Acta math. 13, 1-270 (1890).

[1899] Les méthodes nouvelles de la mécanique céleste, tom. III, chap. XXVI "Stabilité à la Poisson", Paris: Gauthiers-Villars 1899.

Pontrjagin, L.S. and L.G. Schnirelmann

[1932] Sur une proprieté métrique de la dimension. Ann. Math. 33, 152-162, (1932).

Revuz, D.

[1975] Markov Chains. North Holland 1975.

Riečan, B.

[1978] A note on Lebesgue spaces. Comm. Math. Univ. Carolinae 19, 783-787 (1978).

Raimi, R.A.

[1964] Minimal sets and ergodic measures in $\beta\mathbb{N} \setminus \mathbb{N}$. Bull. Amer. Math. Soc. 70, 711-712 (1964).

Rohlin, V.A.
[1952] On the fundamental ideas of measure theory. Amer. Math. Soc. Translation 71, (1952).
[1959] On the entropy of a metric automorphism. Dokl. Akad. Nauk SSSR 124, 980-983 (1959).
[1960] New progress in the theory of transformations with invariant measure. Russ. Math. Surveys 15, 1-22 (1960).
[1961] Exact endomorphisms of a Lebesgue space. Amer. math. Soc. Transl. 39, 1-36 (1964).
[1966] Selected topics from the metric theory of dynamical systems. Amer. Math. Soc. Transl. 2, 49, 171-209 (1966).
[1967] Lectures on the entropy theory of measure-preserving transformations. Russian Math. Surveys 22, 1-52 (1967).

Rohlin, V.A. and Y.G. Sinai
[1961] Construction and properties of invariant measurable partitions. Soviet Math. Dokl. 2, 1611-1614 (1961).

Rostand, E.
[1897] Cyrano de Bergerac. edition: Livre de poche, Paris

Rudin, W.
[1958] Averages of continuous functions on compact spaces. Duke Math. J. 25, 197-204 (1958).

Ruelle, D.
[1969] Statistical Mechanics: Rigorous Results. Reading, Mass.: W.A. Benjamin, Inc. 1969.

Ruppert, W.
[1973] Rechtstopologische Halbgruppen. J.reine angew. Math. 261, 123 - 133 (1973).

Sakai, S.
[1971] C*-Algebras and W*-Algebras. Ergebnisse Math. Bd. 60. Berlin-Heidelberg-New York: Springer 1971.

Rota, G.-C. (ed)
[1978] Studies in Probability and Ergodic Theory. Adv.Math., Suppl.Studies, Vol.2 (1978). Academic Press, New York -San Francisco - London.

Sato, R.

[1975a] On mean ergodic theorems for positive operators in Lebesgue spaces. J. Math. Soc. Japan 27, 207-212 (1975).

[1975b] On the existence of positive invariant functions for semigroups of operators. Tôhoku Math. J. 27, 187-196 (1975).

[1977a] An individual ergodic theorem. Proc. Amer. Math. Soc. 65 , p. 235-241 (1977).

[1977b] Mean ergodic theorems for semigroups of positive linear operators. J. Math. Soc. Japan 29, 123-134 (1977).

[1977c] Invariant measures for ergodic semigroups of operators. Pacific J. Math. 71, 173-192 (1977).

[1978] On abstract mean ergodic theorems. Tôhoku Math. J. 30, 575-581 (1978).

[1979] On abstract mean ergodic theorems II. Math. J. Okayama Univ. 21, 141-147 (1979).

[1980 ] On the Blum-Hanson theorem in $L_p$. Math. J. Okayama Univ. 22, 27-32 (1980).

Schaefer, H.H.

[1966] Topological Vector Spaces. 1st and 2nd print, New York-London: Macmillan 1966/67. 3rd print, Berlin-Heidelberg-New York: Springer 1971.

[1967/68] Invariant ideals of positive operators in C(X), I and II. Ill. J. Math. 11, 703-715 and 12, 525-538

[1974] Banach Lattices and Positive Operators. Berlin-Heidelberg-New York: Springer 1974.

Schaefer, H.H., Wolff, M. and W. Arendt

[1978] On lattice isomorphisms with positive real spectrum and groups of positive operators. Math. Z. 164, 115-123 (1978).

Seever, G.L.

[1968] Measures on F-spaces. Trans. Amer. Math. Soc. 133, 267-280 (1960).

Sell, G.R.

[1971] Topological Dynamics and Ordinary Differential Equations. London: van Nostrand Reinhold Math. Studies 33 (1971).

Semadeni, Z.

[1971] Banach spaces of continuous functions. Warsaw: Polish Scientific Publishers 1971.

Shannon, C.

[1948] A mathematical theory of communication. Bell. System Tech. J. 27, 379-423, 623-656 (1948).

Shields, P.C.

[1973] Cutting and stacking of intervals. Math. Systems Theory 7, 1-4 (1973).

[1974] The Theory of Bernoulli Shifts. Chicago, London: The University of Chicago Press 1974.

Sinai, J.G.

[1959] On the concept of entropy for dynamical systems. Dokl. Akad. Nauk. SSSR 124, 768-771 (1959) (Russian).

[1963] Probabilistic ideas in ergodic theory. Proc. Inter. Congress Math., Stockholm 1962, engl. translation in: Amer. Math. Soc. Transl., Ser. 2, 31, 62-84 (1963).

[1963] On the foundations of the ergodic hypothesis for a dynamical system of statistical mechanics. Soviet Math. Dokl. 4, 1818-1822 (1963).

[1970a] Theory of Dynamical Systems. Part I: Ergodic Theory. Aarhus University 1970.

[1970b] Dynamical systems with elastic reflections. Russ. Math. Surveys 25, 137-189 (1970).

[1977] Introduction to Ergodic Theory. Princeton Univ. Press 1977.

Sine, R.

[1975] Convex combinations of uniformly mean stable Markov operators. Proc. Amer. Math. Soc. 51, 123-126 (1975).

[1976] A note on the ergodic properties of homeomorphisms. Proc. Amer. Math. Soc. 57, 169-172 (1976).

Smale, S.

[1967] Differentiable dynamical systems. Bull. Amer. Math. Soc. 73, 747-817 (1967).

[1980] The Mathematics of Time. Essays on Dynamical Systems, Economic Processes, and Related Topics. Berlin-Heidelberg-New York: Springer 1980.

Smorodinsky, M.

[1971] A partition on a Bernoulli shift which is not 'weakly Bernoulli'. Math. Systems Theory 5, 201-203 (1971).

[1971] Ergodic Theory, Entropy. Lecture Notes in Math. 214. Berlin-Heidelberg-New York: Springer 1971.

Sucheston, L. (ed.)

[1970] Contributions to Ergodic Theory and Probability. Lecture Notes in Math. 160. Berlin-Heidelberg-New York: Springer 1970.

Stepin, A.M.

[1971] Spectra of dynamical systems. Actes Congrès Inter. Math. (Nice 1970), p. 941-946. Paris: Gauthier-Villars 1971.

R/36


Streater, R.F. and A.S. Wightman

[1964] PCT, Spin and Statistics, and All That. New York-Amsterdam: W.A. Benjamin Inc. 1964.

Szily, M.C.

[1872] On Hamilton's principle and the second proposition of the mechanical theory of heat. Philos. Magazine J. Science 43 (4), 339-343 (1872).

[1873] Das dynamische Prinzip von Hamilton in der Thermodynamik. Poggendorfs Annalen 149, 74-86 (1973).

[1876] The second proposition of the mechanical theory of heat deduced from the first. Philos. Magazine J. Science 1 (5), 22-31 (1876).

Sz.-Nagy, B. and C. Foias

[1970] Harmonic Analysis of Operators on Hilbert Space. Amsterdam: North Holland 1970.

Takahashi, W.

[1971] Invariant functions for amenable semigroups of positive contractions on $L^1$. Kodai Math. Sem. Rep. 23, 131-143 (1971).

Talagrand, M.

[1979] Deux generalisations d'un thêorême de J. Namioka. Pacific J. Math. 81, 239-251 (1979).

Totoki, H.

[1969] Ergodic Theory. Lecture Notes, Aarhus University 1969.

Troallic, J.P.

[1979] Espaces fonctionnelles et thêorêmes de J. Namioka. Bull. Soc. math. France 107, 127-137 (1979).

Vere-Jones, D.

[1963] On the spectra of some linear operators associated with queueing systems. Z. Wahrscheinlichkeitstheorie verw. Gebiete 2, 12-21 (1963).

R/37

Vershik, A.M. and S.A. Yuzvinskii

[1970] Dynamical systems with invariant measures. Progress Math. 8, 151-215 (1970).

Walters, P.

[1975] Ergodic Theory - Introductory Lectures. Lecture Notes in Math. 458. Berlin-Heidelberg-New York: Springer 1975.

[1982] An Introduction to Ergodic Theory. Graduate Texts in Math. 79. Berlin-Heidelberg-New York: Springer 1982.

Weiss, B.

[1972] The isomorphism problem in ergodic theory. Bull. Amer. Math. Soc. 78, 668-684 (1972).

v. Weizsäcker, H.

[1974] Sublineare Abbildungen und ein Konvergenzsatz von Banach. Math. Ann. 212, 165-171 (1974).

Weyl, H.

[1916] Über die Gleichverteilung von Zahlen mod 1. Math. Ann. 77, 313-352 (1916).

Wiener, N.

[1939] The ergodic theorem. Duke Math. J. 5, 1-18 (1939).

Witz, K.G.

[1964] Applications of a compactification for bounded operator semigroups. Ill. J. Math. 8, 685-696 (1964).

Wolfowitz, J.

[1964] Coding Theorems of Information Theory. Berlin-Heidelberg-New York: Springer 1964.

[1967] The moments of recurrence time. Proc. Amer. Math. Soc. 18, 613-614 (1967).

Yeadon, F.J.

[1977] Ergodic theorems for semifinite von Neumann algebras I. J. London Math. Soc. 16, 326-332 (1977).

[1980] Ito's theorem and the ratio theorem. Bull. London Math Soc. 12, 279-288 (1980).

Yoshida, K.

[1940] An abstract treatment of the individual ergodic theorems. Proc. Imp. Acad. Tokyo 16, 280-284 (1940).

Yoshida, K. and S. Kakutani

[1941] Operator-theoretical treatment of Markov process and mean ergodic theorem. Ann. Math. 42, 188-228 (1941).

Wright, F.B.

[1963] Ergodic Theory. Proc. Int. Symp. Tulane Univ. New Orleans 1961. New York: Academic Press 1963.

## Index