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

Operator Theoretic Aspects of Ergodic Theory

July 21, 2012

Springer

Contents

Preface ix

1 What is Ergodic Theory? 1

2 Topological Dynamical Systems 7
2.1 Some Basic Examples 7
2.2 Basic Constructions 11
2.3 Topological Transitivity 17
2.4 Transitivity of Subshifts 21
Exercises 23

3 Minimality and Recurrence 27
3.1 Minimality 28
3.2 Topological Recurrence 30
3.3 Recurrence in Extensions 32
Exercises 35

4 The C^*-algebra $\mathrm{C}(K)$ and the Koopman Operator 37
4.1 Continuous Functions on Compact Spaces 38
4.2 The Space $\mathrm{C}(K)$ as a Commutative C^*-Algebra 43
4.3 The Koopman Operator 45
4.4 The Theorem of Gelfand–Naimark 49
Supplement: Proof of the Gelfand–Naimark Theorem 50
Exercises 56

5 Measure-Preserving Systems 59
5.1 Examples 61
5.2 Measures on Compact Spaces 66
5.3 Haar Measure and Rotations 70
Exercises 71

6 Recurrence and Ergodicity 75
6.1 The Measure Algebra and Invertible MDS 75
6.2 Recurrence 77
6.3 Ergodicity 82
Supplement: Induced Transformation and Kakutani–Rokhlin Lemma 85
Exercises 88

7 The Banach Lattice L^p and the Koopman Operator 91
7.1 Banach Lattices and Positive Operators 92
7.2 The Space $L^p(X)$ as a Banach Lattice 97
7.3 The Koopman Operator and Ergodicity 99
Supplement: Interplay between Lattice and Algebra Structure 102
Exercises 104

8 The Mean Ergodic Theorem 107
8.1 Von Neumann's Mean Ergodic Theorem 108
8.2 The Fixed Space and Ergodicity 111
8.3 Perron's Theorem and the Ergodicity of Markov Shifts 114
8.4 Mean Ergodic Operators 116
Supplement: Operations with Mean Ergodic Operators 120
Exercises 123

9 Mixing Dynamical Systems 125
9.1 Strong Mixing 126
9.2 Weakly Mixing Systems 131
9.3 Weak Mixing of All Orders 138
Exercises 143

10 Mean Ergodic Operators on $C(K)$ 147
10.1 Invariant Measures 147
10.2 Uniquely and Strictly Ergodic Systems 150
10.3 Mean Ergodicity of Group Rotations 152
10.4 Furstenberg's Theorem on Group Extensions 154
10.5 Application: Equidistribution 157
Exercises 160

11 The Pointwise Ergodic Theorem 163
11.1 Pointwise Ergodic Operators 164
11.2 Banach's Principle and Maximal Inequalities 166
11.3 Applications 170
Exercises 173

12 Isomorphisms and Topological Models 175
12.1 Point Isomorphisms and Factor Maps 175
12.2 Algebra Isomorphisms of Measure Preserving Systems 179
12.3 Topological Models 183
12.4 The Stone Representation 186
12.5 Mean Ergodic Subalgebras 188
Exercises 191

13 Markov Operators 193
13.1 Examples and Basic Properties 194
13.2 Embeddings, Factor Maps and Isomorphisms 196
13.3 Markov Projections 199
13.4 Factors and their Models 203
Supplement: Operator Theory on Hilbert Space 206
Exercises 210

14 Compact Semigroups and Groups 211
14.1 Compact Semigroups 211
14.2 Compact Groups 217
14.3 The Character Group 219
14.4 The Pontryagin Duality Theorem 223
14.5 Application: Ergodic Rotations and Kronecker's Theorem 225
Exercises 227

15 Topological Dynamics Revisited 229
15.1 The Čech–Stone Compactification 229
15.2 The Space of Ultrafilters 229
15.3 Multiple Recurrence 229
15.4 Applications to Ramsey Theory 229
Exercises 229

16 The Jacobs–de Leeuw–Glicksberg Decomposition 231
16.1 Weakly Compact Operator Semigroups 232
16.2 The Jacobs–de Leeuw–Glicksberg Decomposition 237
16.3 The Almost Weakly Stable Part 240
16.4 Compact Group Actions on Banach Spaces 244
16.5 The Reversible Part Revisited 250
Supplement: Unitary Representations of Compact Groups 252
Exercises 257

17 Dynamical Systems with Discrete Spectrum 261
17.1 Operators with Discrete Spectrum 261
17.2 Monothetic Groups 263
17.3 Dynamical Systems with Discrete Spectrum 265
17.4 Examples 269
Exercises 272

18 A Glimpse at Arithmetic Progressions 275
18.1 The Furstenberg Correspondence Principle 276
18.2 A Multiple Ergodic Theorem 279
18.3 The Furstenberg–Sárközy Theorem 283
Exercises 286

19 Joinings 287
19.1 Dynamical Systems and their Joinings 287
19.2 Inductive Limits and the Minimal Invertible Extension 287
19.3 Relatively Independent Joinings 287
Supplement: Tensor Products and Projective Limits 287
Exercises 287

20 The Host–Kra–Tao Theorem 289
20.1 Idempotent Classes 289
20.2 The Existence of Sated Extensions 289
20.3 The Furstenberg Self-Joining 289
20.4 Proof of the Host–Kra–Tao Theorem 289
Exercises 289

21 More Ergodic Theorems 291
21.1 Polynomial Ergodic Theorems 291
21.2 Weighted Ergodic Theorems 291
21.3 Wiener–Wintner Theorems 291
21.4 Even More Ergodic Theorems 291
Exercises 291

A Topology 293

B Measure and Integration Theory 305

C Functional Analysis 319

D The Riesz Representation Theorem 333

E Theorems of Eberlein, Grothendieck, and Ellis 345

Bibliography 357

Index 371

Symbol Index 379

Preface

Ergodic Theory has its roots in Maxwell's and Boltzmann's kinetic theory of gases and was born as a mathematical theory around 1930 by the groundbreaking works of von Neumann and Birkhoff. In the 1970s, Furstenberg showed how to translate questions of combinatorial number theory into ergodic theory. This inspired a new line of research, which ultimately led to stunning recent results of Host and Kra, Green and Tao and many others.

In its 80 years of existence, ergodic theory has developed into a highly sophisticated field that builts heavily on methods and results from many other mathematical disciplines, e.g., probability theory, combinatorics, group theory, topology and set theory, even mathematical logic. Right from the beginning, also *operator theory* (which for the sake of simplicity we use synonymously to "functional analysis" here) played a central role. To wit, the main protagonist of the seminal papers of von Neumann [von Neumann (1932b)] and Birkhoff [Birkhoff (1931)] is the so-called *Koopman operator*

$$T : \mathrm{L}^2(\mathrm{X}) \to \mathrm{L}^2(\mathrm{X}) \qquad (Tf)(x) := f(\varphi(x)), \tag{0.1}$$

where $\mathrm{X} = (X, \Sigma, \mu)$ is the underlying probability space, and $\varphi : X \to X$ is the measure-preserving dynamics. (See Chapter 1 for more details.) By passing from the state space dynamics φ to the (linear) operator T, one linearises the situation and can then employ all the methods and tools from operator theory, like, e.g., spectral theory, duality theory, harmonic analysis.

However, this is not a one-way street. Results and problems from ergodic theory, once formulated in operator theoretic terms, tend to emancipate from their parental home and to lead their own life in functional analysis, with sometimes stunning results of wide applicability (like the mean ergodic theorem, see Chapter 8).

We, as functional analysts, are fascinated by this interplay, and the present book is the result of this fascination. With it we offer the reader a systematic introduction into the operator theoretic aspects of ergodic theory, and show some of their surprising applications.

As our focus is on operator theoretic *aspects*, we have sometimes taken the liberty to develop the functional analysis beyond its immediate connection with ergodic theory, and hence to leave out topics in ergodic theory where functional analysis has (yet?) not so much to say. (An example of such a missing topic is *entropy*.) These omissions, however, should be outweighed by some inclusions, such as a detailed account of the Jacobs–deLeeuw–Glicksberg splitting theory and a proof of the Host–Kra–Tao theorem on multiple ergodic averages. (See below for a more detailed description of the contents.) Readers who want to see other parts of and trends in ergodic theory we refer to one of the excellent recent monographs such as [Glasner (2003)], [Kalikow and McCutcheon (2010)], and [Einsiedler and Ward (2011)].

This book can serve quite different purposes. First, we do not expect any prior encounter with ergodic theory, so it can be used as a basis for an introductory course into the subject. We certainly require familiarity with standard functional analysis, but whenever there were doubts about what may be considered "standard", we included full proofs.[1] Second, it can serve someone coming from functional analysis to make the first steps into ergodic theory; and third, it can help someone from ergodic theory to learn about the many operator theoretic aspects of his own discipline. Finally—one great hope of ours—it may serve as a foundation of future research, leading towards new and yet unknown connections between ergodic and operator theory.

A Short Synopsis

The following is a brief overview about the contents of the book, highlighting some of its distinguishing features. As said above, no prior knowledge of ergodic theory is required. Instead, we expect the reader to have some background in topology, measure theory and functional analysis, most of which is collected in Appendices A, B, and C.

In Chapter 1 entitled "What is Ergodic Theory?" we give a brief and intuitive introduction into the subject. The mathematical theory then starts in Chapters 2 and 3 with the case of *topological dynamical systems*. We present the basic definitions (transitivity, minimality and recurrence) and cover the standard examples, constructions and theorems, like for instance the Birkhoff recurrence theorem.

Operator theory appears first in Chapter 4 when we introduce as in (0.1) the *Koopman operator* T on the Banach space $\mathrm{C}(K)$ induced by a topological dynamical system $(K;\varphi)$. In order to be able to characterise properties of the (nonlinear) dynamical system through properties of this linear operator T, a careful analysis of the space $\mathrm{C}(K)$ and the operators thereon is needed. Therefore we cover the ba-

[1] This concerns the Gelfand–Naimark theorem, the Stone–Weierstrass theorem, Pontryagin duality and the Peter-Weyl theorem (main text) and the Riesz–Representation theorem, Eberlein's, Grothendieck's and Krein-Šmulian's theorems on weak compactness, Ellis' theorem and the existence of the Haar measure (appendices).

sic results about such spaces (Urysohn's lemma, Tietze's and the Stone–Weierstrass theorem). Then we emphasise its structure as a Banach algebra and give a proof of the classical Gelfand–Naimark theorem, which allows to represent each commutative C^*-algebra as a space $\mathrm{C}(K)$. The Koopman operators are then identified as the morphisms of such algebras.

In Chapter 5 we introduce *measure-preserving dynamical systems* and cover standard examples and constructions. In particular, we discuss the relation of measures on a compact space K with functionals on the Banach space $\mathrm{C}(K)$. (However, the proof of the Riesz representation theorem is deferred to Appendix D.) The classical topics of recurrence and ergodicity as the most basic properties of measure-preserving systems are discussed in Chapter 6.

Subsequently, in Chapter 7 we turn to the corresponding operator theory. As in the topological case, a measure-preserving map φ on the probability space $\mathrm{X} = (X, \Sigma, \mu)$ induces a *Koopman operator* T on each space $\mathrm{L}^p(\mathrm{X})$ as in (0.1). While in the topological situation we look at the space $\mathrm{C}(K)$ as a Banach algebra and at the Koopman operator as an algebra homomorphism, in the measure-theoretic context the corresponding spaces are *Banach lattices* and the Koopman operators are *lattice homomorphisms*. Consequently, we include a short introduction into abstract Banach lattices and their morphisms. Finally, we characterise the ergodicity of a measure-preserving dynamical system by the fixed space or irreducibility of the Koopman operator.

After these preparations, we discuss the most central operator theoretic results in ergodic theory, von Neumann's *mean ergodic theorem* (Chapter 8) and Birkhoff's *pointwise ergodic theorem* (Chapter 11). The former is placed in the more general context of *mean ergodic operators* and in Chapter 10 we discuss this concept for Koopman operators of topological dynamical systems. Here the classical results of Krylov–Bogoliubov about the existence of invariant measures is proved, the concepts of *unique* and *strict ergodicity* are introduced and exemplified with Furstenberg's theorem on group extensions.

In between the discussion of the ergodic theorems, in Chapter 9, we introduce the concepts of strongly and weakly *mixing* systems. The topic has again a strong operator theoretic flavour, as the different types of mixing are characterised by different asymptotic behaviour of the powers T^n as $n \to \infty$ of the Koopman operator. In a sense, the results of this chapter are incomplete since they are developed without using the fact that orbits of the Koopman operator are relatively weakly compact. The complete truth about weakly mixing systems is revealed only in Chapter 16, when this compactness is studied more thoroughly (see below).

Next, in Chapter 12 we consider different concepts of an "isomorphism"—point isomorphism, measure algebra isomorphism and Markov isomorphism—of measure-preserving systems. In our approach, the latter is the most important, while underlying state space maps are rather insignificant. This includes in particular the dynamics φ itself; and indeed, by virtue of the Gelfand–Naimark theorem it is shown that each measure-preserving system has many *topological models*. One canonical model, the *Stone representation* is discussed in detail.

In Chapter 13 we introduce the class of *Markov operators*, which plays a central role in later chapters. We discuss different types of Markov operators (embeddings, factor maps, Markov projections) and introduce the related concept of a *factor* of a measure-preserving system.

Compact semigroups and compact groups (Chapter 14) occur, e.g., when considering the semigroup $\{T^n : n \in \mathbb{N}_0\}$ and its closures in appropriate operator topologies. Since the structure of these (semi)groups is essential for an understanding of dynamical systems, we present the essentials of Pontrjagin's duality theory for compact Abelian groups. This chapter is accompanied by the results in Appendix E, where the existence of the Haar measure and Ellis' theorem for compact semitopological groups is proved in its full generality.

In Chapter 15 we approach the Čech–Stone compactification of a (discrete) semigroup via the Gelfand–Naimark theorem and return to topological dynamics by showing some less classical results, like the one of Birkhoff about multiple recurrence. Here we encounter the first applications of dynamical systems to combinatorics and prove the theorems of van der Waerden, Hindman and Hales–Jewett.

In Chapter 16 we develop a powerful tool in the study of the asymptotic behaviour of semigroup actions on Banach spaces, the *Jacobs–deLeeuw–Glicksberg* (JdLG-)decomposition. Applied to the semigroup generated by a Markov operator T it yields an orthogonal splitting of the corresponding L^2-space into the so-called "reversible" and the "almost weakly stable" part. The former is the range of a Markov projection and hence a factor, and the operator generates a compact group on it. The latter is characterised by a convergence to 0 (in some sense) of the powers of T.

Applied to the Koopman operator of a measure-preserving system, the reversible part is the so-called *Kronecker factor*. It turns out that this factor is trivial if and only if the system is weakly mixing. On the other hand, this factor is the whole system if and only if the Koopman operator has *discrete spectrum* in which case the system is (Markov) isomorphic to a rotation on a compact monothetic group (Halmos–von Neumann theorem, Chapter 17).

In Chapter 18 we employ the JdLG-decomposition to establish the existence of arithmetic progressions of length 3 in certain subsets of $\mathbb{N}$, thus proving the first nontrivial case of Szemerédi's theorem on arithmetic progressions.

In Chapter 19 we leave the classical setup of ergodic theory towards general measure-preserving (countable discrete) semigroup actions on L^1-spaces. Moreover, we introduce the concept of *joining* of such systems, discuss the connection with Markov operators, present the standard types (graph joining, relatively independent joining), and prove the *relative independence theorem*.

These notions and results are then applied in Chapter 20 where we present a proof (due to Austin and de la Rue) of the *Host–Kra–Tao theorem* on multiple ergodic averages.

In the final Chapter 21 more ergodic theorems lead the reader to a less classical area, and actually to a field of active research.

History

The text of this book has a long and complicated history. In the early 1960s, Helmut H. Schaefer founded his Tübingen school systematically investigating Banach lattices and positive operators [Schaefer (1974)]. Growing out of that school, Rainer Nagel in the early 1970's developed in a series of lectures an abstract "ergodic theory on Banach lattices" in order to unify topological dynamics with the ergodic theory of measure-preserving systems. This approach was pursued subseqently together with several doctoral students among whom Günther Palm who succeeded in [Palm (1976)] to unify the topological and measure-theoretic entropy theories. This and other results, e.g., on discrete spectrum [Nagel and Wolff (1972)], mean ergodic semigroups [Nagel (1973)] and dilations of positive operators [Kern et al. (1977)] led eventually to a manuscript by Roland Derndinger, Rainer Nagel and Günther Palm entitled "Ergodic Theory in the Perspective of Functional Analysis", ready for publication around 1980.

However, the "Zeitgeist" seemed to be against this approach. Ergodic theorists at the time were fascinated by other topics, like the isomorphism problem and the impact the concept of entropy had made on it. For this reason, the publication of the manuscript was delayed for several years. In 1987, when the manuscript was finally accepted by Springer's Lecture Notes in Mathematics, time had passed over it, and none of the authors was willing or able to do the final editing. The book project was buried and the manuscript remained in an unpublished preprint form [Derndinger et al. (1987)].

Then, some twenty years later and inspired by the survey articles by Bryna Kra [Kra (2006), Kra (2007)] and Terence Tao [Tao (2007)] on the Green–Tao theorem, Rainer Nagel took up the topic again. He quickly motivated two of his former doctoral students (T.E., B.F.) and a former master student (M.H.) for the project. However, it was clear that the old manuscript could only serve as an important source, but a totally new text would have had to be written. During the academic year 2008/09 the authors organised an international internet seminar under the title *"Ergodic Theory – An Operator Theoretic Approach"* and wrote the corresponding lecture notes. Over the last 3 years, these notes were expanded considerably, rearranged and rewritten several times. Until, finally, they became the book that we now proudly present to the mathematical public.

Acknowledgements

Many people have contributed, directly or indirectly, to the completion of the present work. During the Internetseminar 2008/09 we profitted enourmously from the participants, not only by their mathematical comments but also by their enthusiastic encouragement. This support continued during the subsequent years in which a part of the material was taken as a basis for lecture courses or seminars at a master's level. We thank all who were involved in this process, particularly Omar Aboura, Abramo Bertucco, Joanna Kulaga, Heinrich Küttler, Daniel Maier, Carlo Monta-

nari, Felix Pogorzelski, Manfred Sauter, Pavel Zorin-Kranich. We hope that they will enjoy this final version.

We thank the *Mathematisches Forschungsinstitut Oberwolfach* for hosting us in a "research in quadruples"-week in April 2012.

B.F. acknowledges the support of the János Bolyai Research Fellowship of the Hungarian Academy of Sciences and of the Hungarian Research Fund (OTKA-100461).

Special thanks go to our colleagues Roland Derndinger, Mariusz Lemańczyk, Jürgen Voigt and Hendrik Vogt.

Amsterdam, Budapest, Delft and Tübingen, July 2012

Tanja Eisner
Bálint Farkas
Markus Haase
Rainer Nagel

Chapter 1
What is Ergodic Theory?

Ergodic Theory is not one of the classical mathematical disciplines and its name, in contrast to, e.g., number theory, does not explain its subject. However, its origin can be described quite precisely.

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

> "Monoden, welche nur durch die Gleichung der lebendigen Kraft beschränkt sind, will ich als **Ergoden** bezeichnen."[2]

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

> "... haben Boltzmann und Maxwell eine Klasse von mechanischen Systemen durch die folgende Forderung definiert:
> Die einzelne ungestörte Bewegung des Systems führt bei unbegrenzter Fortsetzung schließlich durch jeden Phasenpunkt hindurch, der mit der mitgegebenen Totalenergie verträglich ist. – Ein mechanisches System, das diese Forderung erfüllt, nennt Boltzmann ein **ergodisches System**." [3]

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

Before we, too, undertake this journey, let us explain the original problem without going too deep into the underlying physics. We start with an (ideal) gas contained in

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

[2] "Monodes which are restricted only by the equations of the living power, I shall call Ergodes."

[3] "...Boltzmann and Maxwell have defined a class of mechanical systems by the following claim: By unlimited continuation the single undisturbed motion passes through every state which is compatible with the given total energy. Boltzmann calls a mechanical system which fulfils this claim an ergodic system."

a box and represented by d (frictionless) moving particles. Each particle is described by six coordinates (three for position, three for velocity), so the situation of the gas (better: the *state* of the system) is given by a point $x \in \mathbb{R}^{6d}$. Clearly, not all points in $\mathbb{R}^{6d}$ can be attained by our gas in the box, so we restrict our considerations to the set X of all *possible* states and call this set the *state space* of our system. We now observe that our system changes while time is running, i.e., the particles are moving (in the box) and therefore a given state (= point in X) also "moves" (in X). This motion (in the box, therefore in X) is governed by Newton's laws of mechanics and then by Hamilton's differential equations. The solutions to these equations determine a map

$$\varphi : X \to X$$

in the following way: If our system, at time $t = 0$, is in the state $x_0 \in X$, then at time $t = 1$ it will be in a new state x_1, and we define φ by $\varphi(x_0) := x_1$. As a consequence, at time $t = 2$ the state x_0 becomes

$$x_2 := \varphi(x_1) = \varphi^2(x_0)$$

and

$$x_n := \varphi^n(x_0)$$

at time $t = n \in \mathbb{N}$. The so-obtained set $\{\varphi^n(x_0) \mid n \in \mathbb{N}_0\}$ of states is called the **orbit** of x_0. In this way, the physical motion of the system of particles becomes a "motion" of the "points" in the state space. The motion of all states within one time unit is given by the map φ. For these objects we introduce the following terminology.

Definition 1.1. A pair (X, φ) consisting of a state space X and a map $\varphi : X \to X$ is called a **dynamical system**. [4]

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

First objection. In the physical situation it is not possible to determine exactly the given initial state $x_0 \in X$ of the system or any of its later states $\varphi^n(x_0)$.

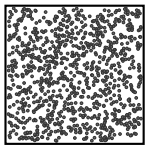


Fig. 1.1 Try to determine the exact state of the system for only $d = 1000$ gas particles.

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

To overcome this objection we introduce "observables", i.e., functions $f : X \to \mathbb{R}$ assigning to each state $x \in X$ the value $f(x)$ of a measurement, for instance of the temperature. The motion in time ("evolution") of the states described by the map $\varphi : X \to X$ is then reflected by a map T_φ of the observables defined as

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

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

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

So instead of looking at the orbits $\{\varphi^n(x_0) : n \in \mathbb{N}_0\}$ of the state map φ we study the orbit $\{(T_\varphi)^n f : n \in \mathbb{N}_0\}$ of an observable f under the linear operator T_φ. This allows one to use operator theoretic tools and is the leitmotiv of this book.

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

Second objection. The motion of our system happens so quickly that we will neither be able to determine the states

$$x_0, \varphi(x_0), \varphi^2(x_0), \dots$$

nor their measurements

$$f(x_0), T_\varphi f(x_0), T_\varphi^2 f(x_0), \dots$$

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

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

$$\frac{1}{N}\sum_{n=0}^{N-1} T_\varphi^n f(x_0) = \frac{1}{N}\sum_{n=0}^{N-1} f(\varphi^n(x_0))$$

and their limit

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T_\varphi^n f(x_0),$$

called the **time mean** (of the observable f in the state x_0). This appears to be a good idea, but it is still based on the knowledge of the states $\varphi^n(x_0)$ and their measurements $f(\varphi^n(x_0))$, so the first objection remains valid. At this point, Boltzmann asked for more.

Third objection. The time mean $\lim_N \frac{1}{N}\sum_{n=0}^{N-1} f(\varphi^n(x_0))$ should not depend on the initial state $x_0 \in X$.

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

Ergodic Hypothesis. *For each initial state $x_0 \in X$ and each (reasonable) observable $f : X \to \mathbb{R}$, it is true that "time mean equals space mean", i.e.,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\varphi^n(x_0)) = \int_X f \mathrm{d}\mu.$$

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

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

Historically, these goals could be achieved only after the necessary tools had been created. Fortunately, between 1900 and 1930 new mathematical theories emerged such as topology, measure theory and functional analysis. Using the tools from these new theories, J. von Neumann and G. D. Birkhoff proved in 1931 the so-called *mean ergodic theorem* [von Neumann (1932b)] and the *individual ergodic theorem* [Birkhoff (1931)] and thereby established ergodic theory as a new and independent mathematical discipline, see [Bergelson (2004)].

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

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

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

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

The Green–Tao theorem had a precursor first proved by Szemerédi in a combinatorial way but eventually given a purely ergodic theoretic proof by H. Furstenberg.

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

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

then it contains arbitrarily long arithmetic progressions.

A complete proof of this theorem remains beyond the reach of this book. However, we shall provide some of the necessary tools.

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

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

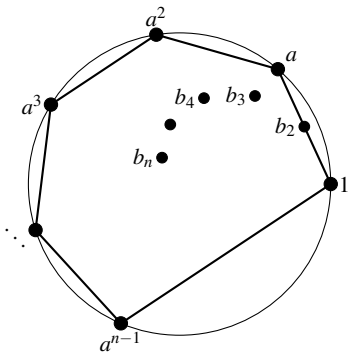


Fig. 1.2 Polygon

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

Chapter 2
Topological Dynamical Systems

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

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

Definition 2.1. A **topological dynamical system** (TDS)is a pair $(K;\varphi)$, where K is a nonempty compact space and $\varphi : K \to K$ is continuous. A TDS $(K;\varphi)$ is **surjective** if φ is surjective, and the TDS is called **invertible** if φ is invertible, i.e., a homeomorphism.

An invertible TDS $(K;\varphi)$ defines two "one-sided" TDSs, namely the **forward system** $(K;\varphi)$ and the **backward system** $(K;\varphi^{-1})$.

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

2.1 Some Basic Examples

First we list some basic examples of dynamical systems.

Example 2.2 (Finite State Space). Take $d \in \mathbb{N}$ and consider the finite set $K := \{1,\dots,d\}$ with the discrete topology. Then K is compact and every map $\varphi : K \to K$

is continuous. The TDS $(K;\varphi)$ is invertible if and only if φ is a permutation of the elements of K.

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

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

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

Then

$$|[Tx]_i| = \left|\sum_{j=1}^d t_{ij}x_j\right| \le \sum_{j=1}^d t_{ij}\left|x_j\right| \le \|x\|_\infty \sum_{j=1}^d t_{ij} \le \|x\|_\infty$$

for every $j=1,\dots,d$. Hence T is contractive.

Example 2.4 (Shift). Take $k\in\mathbb{N}$ and consider the set

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

of infinite sequences within the set $\{0,\dots,k-1\}$. In this context, the set $\{0,\dots,k-1\}$ is often called an **alphabet**, its elements being the **letters**. So the elements of K are the infinite **words** composed of these letters. Endowed with the discrete metric the alphabet is a compact metric space. Hence, by Tychonoff's theorem K is compact when endowed with the product topology (see Section A.5 and Exercise 1). On K we consider the (left) **shift** τ defined by

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

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

Example 2.5 (The Cantor System). The **Cantor set** is

$$C=\left\{x\in[0,1] : x=\sum_{j=1}^\infty \tfrac{a_j}{3^j},\ a_j\in\{0,2\}\right\}, \tag{2.1}$$

cf. Appendix A.8. As a closed subset of the unit interval, the Cantor set C is compact. Consider on C the mapping

$$\varphi(x)=\begin{cases} 3x & 0\le x\le \frac{1}{3},\\ 3x-2 & \frac{2}{3}\le x\le 1.\end{cases}$$

The continuity of φ is clear, and a close inspection using (2.1) reveals that φ maps C to itself. Hence $(C;\varphi)$ is a TDS, called the **Cantor system**.

Example 2.6 (Translation mod 1). Consider the interval $K := [0,1)$ and define

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

By Exercise 2, d is a metric on K, continuous with respect to the standard one, and turning K into a compact metric space. For a real number $x \in \mathbb{R}$ we write

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

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

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

By Exercise 2, φ is continuous with respect to the metric d, hence it gives rise to a TDS on $[0,1)$. We shall abbreviate this TDS by $([0,1);\alpha)$.

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

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

Since φ is obviously continuous, it gives rise to an invertible TDS defined on $\mathbb{T}$ called the **rotation** by a. We shall denote this TDS by $(\mathbb{T};a)$.

Example 2.8 (Rotation on Compact Groups). The previous example is an instance of the following general set-up. A **topological group** is a group $(G,\cdot)$ endowed with a topology such that the maps

$$(g,h) \mapsto g \cdot h \qquad\qquad G \times G \to G$$

and

$$g \mapsto g^{-1} \qquad\qquad G \to G$$

are continuous. A topological group is a **compact group** if G is compact.

Given a compact group G, the **left rotation** by $a \in G$ is the mapping

$$\varphi_a : G \to G, \qquad \varphi_a(g) := a \cdot g.$$

Then $(G;\varphi_a)$ is an invertible TDS, which we henceforth shall denote by $(G;a)$. Similarly, the **right rotation** by $a \in G$ is

$$\rho_a : G \to G, \qquad \rho_a(g) := g \cdot a$$

and $(G;\rho_a)$ is an invertible TDS.

If the group is Abelian, then left and right rotations are identical and one often speaks of *translation* instead of rotation. The torus $\mathbb{T}$ is our most important example of a compact Abelian group.

Example 2.9 (Group Rotation on $\mathbb{R}/\mathbb{Z}$). Consider the additive group $\mathbb{R}$ with the standard topology and its closed subgroup $\mathbb{Z}$. We consider the factor group $\mathbb{R}/\mathbb{Z}$

together with the canonical homomorphism

$$q(x) := x + \mathbb{Z}, \qquad \mathbb{R} \to \mathbb{R}/\mathbb{Z}.$$

If we endow $\mathbb{R}/\mathbb{Z}$ with the quotient topology with respect to q, $\mathbb{R}/\mathbb{Z}$ becomes a topological group. It is even compact, because $\mathbb{R}/\mathbb{Z} = q([0,1])$, cf. Proposition A.4. For $\alpha \in \mathbb{R}$ we define $\varphi(x+\mathbb{Z}) = \alpha + x + \mathbb{Z}$. Then φ is continuous, and we obtain a TDS, which we denote by $(\mathbb{R}/\mathbb{Z}; \alpha)$.

The next example is not a group rotation, but quite similar to the previous example.

Example 2.10 (The Heisenberg Group). The non-Abelian group G of upper triangular real matrices with ones on the diagonal, i.e.,

$$G = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{R} \right\},$$

is called the **Heisenberg group**. We endow it with the usual topology of $\mathbb{R}^3$. Its elements with integer entries form a subgroup

$$H = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} : a, b, c \in \mathbb{Z} \right\}.$$

Then G is a topological group, not compact, and H is closed and discrete in G. However, note that H is not a normal subgroup. The set G/H of *left cosets* becomes a Hausdorff topological space if endowed with the quotient topology under the canonical map

$$q(g) := gH, \qquad G \to G/H.$$

The set

$$K = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in [0,1] \right\}$$

is compact, and one has $G = KH$ (Exercise 3). Hence $G/H = q(K)$ is compact. For

$$a = \begin{pmatrix} 1 & \alpha & \gamma \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{pmatrix} \in G,$$

the left rotation by a on G/H is defined by $gH \mapsto agH$. In this way we obtain an invertible TDS, denoted by $(G/H; a)$. (See also Example 2.16.)

Example 2.11 (Dyadic Adding Machine). For $n \in \mathbb{N}$ consider the cyclic group $\mathbb{Z}_{2^n} := \mathbb{Z}/2^n\mathbb{Z}$, endowed with the discrete topology. The factor maps

$$\pi_{ij} : \mathbb{Z}_{2^j} \to \mathbb{Z}_{2^i}, \qquad \pi_{ij}(x + 2^j\mathbb{Z}) := x + 2^i\mathbb{Z} \qquad (i \le j)$$

are trivially continuous, and satisfy the relations

$$\pi_{ii} = \mathrm{id} \quad \text{and} \quad \pi_{ij} \circ \pi_{jk} = \pi_{ik} \qquad (i \le j \le k).$$

The topological product space $G := \prod_{n\in\mathbb{N}} \mathbb{Z}_{2^n}$ is a compact Abelian group by Tychonoff's Theorem A.5. Since each π_{ij} is a group homomorphism, the set

$$\mathbb{A}_2 = \left\{ x = (x_n)_{n\in\mathbb{N}} \in \prod\nolimits_{n\in\mathbb{N}} \mathbb{Z}_{2^n} \,:\, x_i = \pi_{ij}(x_j) \ \text{ for all } i \le j \right\}$$

is a closed subgroup of G, called the group of **dyadic integers**. Being closed in the compact group G, $\mathbb{A}_2$ is compact. (This construction is an instance of what is called an **inverse** or **projective limit**.) Finally, consider the element

$$\mathbf{1} := (1 + 2\mathbb{Z}, 1 + 4\mathbb{Z}, 1 + 8\mathbb{Z}, \dots) \in \mathbb{A}_2.$$

The translation TDS $(\mathbb{A}_2, \mathbf{1})$ is called the **dyadic adding machine**.

2.2 Basic Constructions

In the present section we shall make precise what it means for two TDSs to be "essentially equal". Then we shall present some techniques for constructing "new" TDSs from "olds" ones, and review our list of examples from the previous section in the light of these constructions.

1. Homomorphisms

A **homomorphism** between two TDSs $(K_1; \varphi_1)$, $(K_2; \varphi_2)$ is a continuous map $\Psi : K_1 \to K_2$ such that $\Psi \circ \varphi_1 = \varphi_2 \circ \Psi$, i.e., the diagram

$$\begin{array}{ccc} K_1 & \xrightarrow{\varphi_1} & K_1 \\ {\scriptstyle\Psi}\downarrow & & \downarrow{\scriptstyle\Psi} \\ K_2 & \xrightarrow{\varphi_2} & K_2 \end{array}$$

is commutative. We indicate this by writing simply

$$\Psi : (K_1; \varphi_1) \to (K_2; \varphi_2).$$

A homomorphism $\Psi : (K_1; \varphi_1) \to (K_2; \varphi_2)$ is a **factor map** if it is surjective. Then, $(K_2; \varphi_2)$ is called a **factor** of $(K_1; \varphi_1)$ and $(K_1; \varphi_1)$ is called an **extension** of $(K_2; \varphi_2)$.

If Ψ is bijective, then it is called an **isomorphism** (or **conjugation**). Two TDSs $(K_1;\varphi_1)$, $(K_2;\varphi_2)$ are called **isomorphic** (or **conjugate**) if there is an isomorphism between them. An isomorphism $\Psi : (K;\varphi) \to (K;\varphi)$ is called an **automorphism**, and the set of automorphisms of a TDS $(K;\varphi)$ is denoted by $\mathrm{Aut}\,(K;\varphi)$. This is clearly a group with respect to composition.

Example 2.12 (Right Rotations as Automorphism). Consider a left group rotation TDS $(G;a)$. Then any right rotation ρ_h, $h \in G$, is an automorphism of $(G;a)$:

$$\rho_h(a\cdot g) = (ag)h = a(gh) = a\cdot(\rho_h(g)) \qquad (g \in G).$$

This yields an injective homomorphism $\rho : G \to \mathrm{Aut}\,(G;a)$ of groups.

Moreover, the inversion map $\Psi(g) := g^{-1}$ is an isomorphism of the TDSs $(G;a)$ and $(G;\rho_{a^{-1}})$ since

$$\rho_{a^{-1}}(\Psi(g)) = g^{-1}a^{-1} = (ag)^{-1} = \Psi(a\cdot g) \qquad (g \in G).$$

Example 2.13 (Rotation and Translation). For $\alpha \in [0,1)$ let $a := \mathrm{e}^{2\pi\mathrm{i}\alpha}$. Then the three TDSs

1) $([0,1);\alpha)$ from Example 2.6,
2) $(\mathbb{T};a)$ from Example 2.7,
3) and $(\mathbb{R}/\mathbb{Z};\alpha)$ from Example 2.9

are all isomorphic.

Proof. The (well defined!) map $\Phi : \mathbb{R}/\mathbb{Z} \to \mathbb{T}, \quad x+\mathbb{Z} \mapsto \mathrm{e}^{2\pi\mathrm{i}x}$ is a group isomorphism. It is continuous, since $\Phi\circ q$ is, and $q : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ is a quotient map (Appendix A.4). Since $\varphi_a \circ \Phi = \Phi \circ \varphi_\alpha$, Φ is an isomorphism of the TDSs 2) and 3). The isomorphy of 1) and 2) is treated in Exercise 4. □

Example 2.14 (Shift $\cong$ Cantor System). The two TDSs

1) $(C;\varphi)$ from Example 2.5 (Cantor system)
2) $(\mathscr{W}_2^+;\tau)$ from Example 2.4 (shift on two letters)

are isomorphic (Exercise 5).

Remark 2.15 (Factors). Let $(K;\varphi)$ be a TDS. An equivalence relation $\sim$ on K is called a **φ-congruence** if $\sim$ is compatible with the action of φ, i.e.,

$$x \sim y \quad \Rightarrow \quad \varphi(x) \sim \varphi(y).$$

Let $L := K/\!\sim$ be the space of equivalence classes with respect to $\sim$. Then L is a compact space by endowing it with the quotient topology with respect to the canonical surjection $q : K \to K/\!\sim$, $q(x) = [x]_\sim$, the equivalence class of $x \in K$. Moreover, a dynamics is induced on L via

$$\psi([x]_\sim) := [\varphi(x)]_\sim \qquad (x \in K),$$

well-defined because φ is a congruence. Hence we obtain a new TDS $(L;\psi)$ and $q:(K;\varphi)\to(L;\psi)$ is a factor map.

Conversely, suppose that $\pi:(K;\varphi)\to(L;\psi)$ is a factor map between two TDSs. Then $\sim$, defined by $x\sim y$ if and only if $\pi(x)=\pi(y)$, is a φ-congruence. The map

$$\Phi:(K/_\sim;\varphi)\to(L;\psi),\qquad \Phi([x]_\sim):=\pi(x)$$

is an isomorphism of TDSs.

Example 2.16 (Homogeneous Systems). Consider a group rotation $(G;a)$, and let $H\le G$ be a closed subgroup of G. The equivalence relation

$$x\sim y \quad\Leftrightarrow\quad y^{-1}x\in H$$

is a congruence, since $(ay)^{-1}(ax)=x^{-1}a^{-1}ax=y^{-1}x$. The set of corresponding equivalence classes is

$$G/_H:=\{gH\,:\,g\in G\}$$

and the induced dynamics on it is given by $gH\mapsto agH$. The so defined TDS is denoted by $(G/_H;a)$ and is called a **homogeneous system**. Example 2.10 above is an instance of a homogeneous system.

Example 2.17 (Group Factors). Let $(K;\varphi)$ be a TDS, and let $H\subseteq \mathrm{Aut}\,(K;\varphi)$ be a subgroup of $\mathrm{Aut}\,(K;\varphi)$. Consider the equivalence relation

$$x\sim_H y \quad\Leftrightarrow\quad \exists h\in H:\ h(x)=y$$

on K. The set of equivalence classes is denoted by $K/_H$. Since all $h\in H$ act as automorphism of $(K;\varphi)$, this is a φ-congruence, hence constitutes a factor

$$q:(K;\varphi)\to(K/_G;\varphi)$$

with $\varphi([x]_H):=[\varphi(x)]_H$ by abuse of language. A homogeneous system is a special case of a group factor (Exercise 6).

2. Products, Skew-Products and Group Extensions

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

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

It is invertible if and only if both $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ are invertible. Iterating this construction we obtain finite products of TDSs as in the following example.

Example 2.18 (d-Torus). The **d-torus** is the d-fold direct product $G:=\mathbb{T}\times\cdots\times\mathbb{T}=\mathbb{T}^d$ of the one-dimensional torus. It is a compact topological group. The rotation

by $a = (a_1, a_2, \ldots, a_d) \in G$ is the d-fold product of the one-dimensional rotations $(\mathbb{T}; a_i)$, $i = 1, \ldots, d$.

The construction of products is not restricted to a finite number of factors. If (K_ι, φ_ι) is a TDS for each ι from some index set I, then by Tychonoff's Theorem A.5

$$K := \prod_{\iota \in I} K_\iota, \quad \varphi(x) = (\varphi_\iota(x_\iota))_{\iota \in I}$$

is a TDS, called the **product** of the collection $((K_\iota, \varphi_\iota))_{\iota \in I}$. The canonical projections

$$\pi_\iota : (K, \varphi) \to (K_\iota, \varphi_\iota), \qquad \pi(x) := x_\iota \qquad (\iota \in I)$$

are factor maps.

A product of two TDSs is a special case of the following construction. Let $(K; \varphi)$ be a TDS, let L be a compact space, and let

$$\Phi : K \times L \to L$$

be continuous. The mapping

$$\psi : K \times L \to K \times L, \qquad \psi(x, y) := (\varphi(x), \Phi(x, y))$$

is continuous, hence we obtain a TDS $(K \times L; \psi)$, called the **skew product** of $(K; \varphi)$ with L along Φ.

Remarks 2.19. Let $(K \times L; \psi)$ be a skew product along Φ as above.

1) The projection

$$\pi : (K \times L; \psi) \to (K; \varphi) \qquad \pi(x, y) = x$$

is a factor map.

2) If $\Phi(x, y)$ is independent of $x \in K$, the skew product is just an ordinary product as defined above.

3) We abbreviate $\Phi_x := \Phi(x, \cdot)$. If φ and each mapping Φ_x, $x \in K$, is invertible, then ψ is invertible with

$$\psi^{-1}(x, y) = (\varphi^{-1}(x), \Phi^{-1}_{\varphi^{-1}(x)}(y)).$$

4) If we iterate ψ on (x, y) we obtain

$$\psi^n(x, y) = (\varphi^n(x), \Phi_{\varphi^{n-1}(x)} \ldots \Phi_{\varphi(x)} \Phi_x(y)),$$

i.e., $\psi^n(x, y) = (\varphi^n(x), \Phi_x^{[n]}(y))$ with $\Phi_x^{[n]} := \Phi_{\varphi^{n-1}(x)} \circ \ldots \circ \Phi_{\varphi(x)} \circ \Phi_x$. The mapping

$$\widehat{\Phi} : \mathbb{N}_0 \times K \to L^L, \qquad \widehat{\Phi}(n, x) := \Phi_x^{[n]}$$

is called the **cocycle** of the skew product. It satisfies the relations

$$\Phi_x^{[0]} = \mathrm{id}, \quad \Phi_x^{[m+n]} = \Phi_{\varphi^n(x)}^{[m]} \circ \Phi_x^{[n]} \qquad (n,m \in \mathbb{N}_0, x \in K).$$

A **group extension** is a particular instance of a skew product, where the second factor is a compact group G, and $\Phi : K \times G \to G$ is given by left rotations. That means that we have (by abuse of language)

$$\Phi(x,g) = \Phi(x) \cdot g \qquad (x \in K, g \in G)$$

where $\Phi : K \to G$ is continuous. By Remark 2.19.2 a group extension is invertible if its first factor $(K;\varphi)$ is invertible.

Example 2.20 (Skew Rotation). Let $\Phi := \mathrm{I} : \mathbb{T} \to \mathbb{T}$ be the identity map. Then for the corresponding group extension of a rotation $(\mathbb{T};a)$ along Φ we have

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

This TDS is called the **skew rotation** by $a \in \mathbb{T}$.

Remark 2.21. Let $H = K \times G$ be a group extension of $(K;\varphi)$ along $\Phi : K \to G$. For $h \in G$ we take $\rho_h : H \to H$, $\rho_h(x,g) := (x,gh)$ to be the right multiplication with h in the second coordinate. Then $\rho_h \in \mathrm{Aut}\,(H;\psi)$; indeed,

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

for all $(x,g) \in H$. Hence $\rho : G \to \mathrm{Aut}\,(K \times G;\psi)$ is a group homomorphism.

3. Subsystems and Unions

Let $(K;\varphi)$ be a TDS. A subset $A \subseteq K$ is called **invariant** if $\varphi(A) \subseteq A$, **stable** if $\varphi(A) = A$, and **bi-invariant** if $A = \varphi^{-1}(A)$. If we want to stress the dependence of these notions on φ, we say invariant *under* φ or φ-invariant.

Lemma 2.22. *Let $(K;\varphi)$ be a TDS and $A \subseteq K$. Then the following assertions hold.*

a) *A bi-invariant or stable $\Rightarrow$ A invariant.*

b) *A stable and φ injective $\Rightarrow$ A bi-invariant.*

c) *A bi-invariant and φ surjective $\Rightarrow$ A stable.*

Proof. This is Exercise 7. □

Every nonempty invariant closed subset $A \subseteq K$ gives rise to a **subsystem** by restricting φ to A. For simplicity we write φ again in place of $\varphi|_A$, and so the subsystem is denoted by $(A;\varphi)$. Note that even if the original system is invertible, the subsystem need not be invertible any more unless A is bi-invariant (=stable in this case). We record the following for later use.

Lemma 2.23. *Suppose that $(K;\varphi)$ is a TDS and $\emptyset \neq A \subseteq K$ is closed and invariant. Then there is a nonempty, closed set $B \subseteq A$ such that $\varphi(B) = B$.*

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

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

All these sets are compact and nonempty, since A is closed and φ is continuous. Thus the set $B := \bigcap_{n\in\mathbb{N}} \varphi^n(A)$ is again nonempty and compact, and satisfies $\varphi(B) \subseteq B$. In order to prove $\varphi(B) = B$, let $x \in B$ be arbitrary. Then for each $n \in \mathbb{N}$ we have $x \in \varphi^n(A)$, i.e., $\varphi^{-1}\{x\} \cap \varphi^{n-1}(A) \neq \emptyset$. Hence these sets form a decreasing sequence of compact nonempty sets, and therefore their intersection $\varphi^{-1}\{x\} \cap B$ is not empty either. It follows that $x \in \varphi(B)$ as desired. □

From this lemma the following results is immediate.

Corollary 2.24 (Maximal Surjective Subsystem). *For a TDS $(K;\varphi)$ let*

$$K_s := \bigcap\nolimits_{n\geq 0} \varphi^n(K).$$

Then $(K_s;\varphi)$ is the unique maximal surjective subsystem of $(K;\varphi)$.

Let $(K_1;\varphi_1)$, $(K_2;\varphi_2)$ be two TDSs. Then there is a unique topology on $K := (K_1 \times \{1\}) \cup (K_2 \times \{2\})$ such that the canonical embeddings

$$J_n : K_n \to K, \qquad J_n(x) := (x,n) \qquad (n = 1,2)$$

become homeomorphisms onto open subsets of K. (This topology is the inductive topology defined by the mappings J_1, J_2, cf. Appendix A.4.) It is common to identify the original sets K_1, K_2 with their images within K. Doing this, we can write $K = K_1 \cup K_2$ and define the map

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

Then $(K;\varphi)$ is a TDS, called the (disjoint) **union** of the original TDSs, and it contains the original systems as subsystems. It is invertible if and only if $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ are both invertible. In contrast to products, disjoint unions can only be formed out of *finite* collections of given TDSs.

Example 2.25 (Subshifts). As in Example 2.4, consider the shift $(\mathscr{W}_k^+;\tau)$ on the alphabet $\{0,1,\ldots,k-1\}$. We determine its subsystems, called **subshifts**. For this purpose we need some further notions. An n-**block** of an infinite word $x \in \mathscr{W}_k^+$ is a finite sequence $y \in \{0,1,\ldots,k-1\}^n$ which occurs in x at some position. Take now an arbitrary set

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

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

$$\mathscr{W}_k^B := \big\{x \in \mathscr{W}_k^+ : \text{no block of } x \text{ is contained in } B\big\}.$$

It is easy to see that $\mathscr{W}_k^B$ is a closed τ-invariant subset, hence, if nonempty, it gives rise to a subsystem $(\mathscr{W}_k^B;\tau)$. We claim that *each subsystem of* $(\mathscr{W}_k^+;\tau)$ *arises in this way.*

To prove this claim, let $(F;\tau)$ be a subsystem and consider the set B of finite sequences that are not present in any of the words in F, i.e.

$$B := \big\{y : y \text{ is a finite sequence and not contained in any } x \in F\big\}.$$

We have $F \subseteq \mathscr{W}_k^B$ by definition, and we claim that actually $F = \mathscr{W}_K^B$. To this end let $x \notin F$. Then there is an open cylinder $U \subseteq \mathscr{W}_k^+$ with $x \in U$ and $U \cap F = \emptyset$. By taking a possibly smaller cylinder we may suppose that U has the form

$$U = \big\{z \in \mathscr{W}_k^+ : z_0 = x_0, z_1 = x_1, \ldots, z_n = x_n\big\}$$

for some $n \in \mathbb{N}_0$. Since $F \cap U = \emptyset$ and F is shift invariant, the block $y := (x_0, \ldots, x_n)$ does not occur in any of the words in F. Hence $y \in B$, so $x \notin \mathscr{W}_k^B$.

Particularly important are those subshifts $(F;\tau)$ that arise from a *finite* set B as $F = \mathscr{W}_k^B$. These are called **subshifts of finite type**. The subshift is called of **order** n if there is an excluded block-system B containing only sequences not longer than n, i.e., $B \subseteq \bigcup_{i \le n} \{0,1,\ldots,k-1\}^i$. In this case, by extending shorter blocks in all possible ways, one may suppose that all blocks in B have length *exactly* n.

Of course, all these notions make sense and all these results remain valid for the invertible case, i.e., for subshifts of $(\mathscr{W}_k;\tau)$.

2.3 Topological Transitivity

Investigating a TDS means to ask questions like: How does a particular state of the system evolve in time? How does φ mix the points of K as it is applied over and over again? Will two points that are close to each other initially, stay close even after a long time? Will a point return to its original position, at least very near to it? Will a certain point x never leave a certain region or will it come arbitrarily close to any other given point of K?

In order to study such questions we define the **forward orbit** of $x \in K$ as

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

and, if the TDS is invertible, the **(total) orbit** of $x \in K$ as

$$\mathrm{orb}(x) := \{\varphi^n(x) : n \in \mathbb{Z}\}.$$

And we shall write

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

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

Definition 2.26. Let $(K;\varphi)$ be a TDS. A point $x \in K$ is called **forward transitive** if its forward orbit $\mathrm{orb}_+(x)$ is dense in K. If there is at least one forward transitive point, then the TDS $(K;\varphi)$ is called (topologically) **forward transitive**.

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

Example 2.27. Let $K := \mathbb{Z} \cup \{\pm\infty\}$ be the **two-point compactification** of $\mathbb{Z}$, and define

$$\varphi(n) := \begin{cases} n+1 & \text{if } n \in \mathbb{Z}, \\ n & \text{if } n = \pm\infty. \end{cases}$$

Then $(K;\varphi)$ is an invertible TDS, each point $n \in \mathbb{Z}$ is transitive but no point of K is forward nor backward transitive.

Remarks 2.28. 1) We sometimes say just "transitive" in place of "topologically transitive". The reason is that algebraic transitivity—the fact that a point eventually reaches *exactly* every other point—is rare in topological dynamics and does not play any role in this theory.

2) The distinction between forward transitivity and (two-sided) transitivity is only meaningful in invertible systems. If a system under consideration is not invertible we may therefore drop the word "forward" without causing confusion.

A point $x \in K$ is forward transitive if we can reach, at least approximately, any other point in K after some time. The following result tells us that this is a mixing-property: Two arbitrary open regions are mixed with each other under the action of φ after finitely many steps.

Proposition 2.29. *Let $(K;\varphi)$ be a TDS and consider the following assertions.*

(i) *$(K;\varphi)$ is forward transitive, i.e., there is a point $x \in K$ with $\overline{\mathrm{orb}}_+(x) = K$.*

(ii) *For all open sets $\emptyset \neq U, V$ in K there is $n \in \mathbb{N}$ with $\varphi^n(U) \cap V \neq \emptyset$.*

(iii) *For all open sets $\emptyset \neq U, V$ in K there is $n \in \mathbb{N}$ with $\varphi^{-n}(U) \cap V \neq \emptyset$.*

Then (ii) *and* (iii) *are equivalent,* (ii) *implies* (i) *if K is metrisable, and* (i) *implies* (ii) *if K has no isolated points.*

Proof. The proof of the equivalence of (ii) and (iii) is left to the reader.

(i)$\Rightarrow$(ii): Suppose that $x \in K$ has dense forward orbit and that U, V are nonempty open subsets of K. Then certainly $\varphi^k(x) \in U$ for some $k \in \mathbb{N}_0$. Consider the open set

$W := V \setminus \{x, \varphi(x), \dots, \varphi^k(x)\}$. If K has no isolated points, W cannot be empty, and hence $\varphi^m(x) \in W$ for some $m > k$. Now, $\varphi^m(x) = \varphi^{m-k}(\varphi^k(x)) \in \varphi^{m-k}(U) \cap V$.

(iii)$\Rightarrow$(i): If K is metrisable, there is a countable base $\{U_n : n \in \mathbb{N}\}$ for the topology on K (see Section A.7). For each $n \in \mathbb{N}$ consider the open set

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

By assumption b), G_n intersects nontrivially every nonempty open set, and hence is dense in K. By Baire's Category Theorem A.10 the set $\bigcap_{n\in\mathbb{N}} G_n$ is nonempty (it is even dense). Any point in this intersection has dense forward orbit. □

Remarks 2.30. 1) In general, (i) does not imply (ii) even if K is metrisable. Take, e.g., $K := \mathbb{N} \cup \{\infty\}$, and $\varphi : K \to K$, $\varphi(n) = n+1$, $\varphi(\infty) = \varphi(\infty)$. The point $1 \in K$ has dense forward orbit, but for $U = \{2\}$, $V = \{1\}$ condition b) fails to hold.

2) The proof of Proposition 2.29 yields even more: if K is metrisable without isolated points, then the set of points with dense forward orbit is either empty or a dense G_δ, see Appendix A.9.

There is an analogous statement for transitivity of invertible systems. It is proved almost verbatim as Proposition 2.29.

Proposition 2.31. *Let $(K;\varphi)$ be an invertible TDS, with K metrisable. Then the following properties are equivalent.*

(i) *$(K;\varphi)$ is topologically transitive, i.e., there is a point $x \in K$ with dense orbit.*

(ii) *For all $\emptyset \neq U, V$ open sets in K there is $n \in \mathbb{Z}$ with $\varphi^n(U) \cap V \neq \emptyset$.*

Let us turn to our central examples.

Theorem 2.32 (Rotation Systems I). *Let $(G;a)$ be a left rotation TDS. Then the following statements are equivalent.*

(i) *$(G;a)$ is topologically forward transitive.*

(ii) *Every point of G has dense forward orbit.*

(iii) *$(G;a)$ is topologically transitive.*

(iv) *Every point of G has dense total orbit.*

Proof. Since every right rotation $\rho_h : g \mapsto gh$ is an automorphism of $(G;a)$, we have

$$\rho_h(\mathrm{orb}_+(g)) = \mathrm{orb}_+(gh) \quad \text{and} \quad \rho_h(\mathrm{orb}(g)) = \mathrm{orb}(gh)$$

for every $g, h \in G$. Taking closures we obtain the equivalences (i)$\Leftrightarrow$(ii) and (iii)$\Leftrightarrow$(iv). Now clearly (ii) implies (iv). In order to prove the converse, fix $g \in G$ and consider the nonempty invariant closed set

$$A := \overline{\mathrm{orb}}_+(g) = \overline{\{a^n g : n \geq 0\}}.$$

By Lemma 2.23 there is a nonempty closed set $B \subseteq A$ such that $aB = B$. Since B is nonempty, fix $h \in B$. Then $\mathrm{orb}(h) \subseteq B$ and hence $\overline{\mathrm{orb}}(h) \subseteq B$. By (iv), this implies that $G \subseteq B \subseteq A$, and thus $A = G$. □

A corollary of this result is that a topologically transitive group rotation $(G;a)$ does not admit any nontrivial subsystem. We exploit this fact in the following examples.

Example 2.33 (Kronecker's Theorem, Simple Case). *The rotation $(\mathbb{T};a)$ is topologically transitive if and only if $a \in \mathbb{T}$ is not a root of unity.*

Proof. If $a^{n_0} = 1$ for some $n_0 \in \mathbb{N}$, then $\{z \in \mathbb{T} : z^{n_0} = 1\}$ is closed and φ-invariant, so by Theorem 2.32, $(\mathbb{T};a)$ is not transitive.

For the converse, suppose that a is not a root of unity and take $\varepsilon > 0$. By compactness, the sequence $(a^n)_{n\geq 0}$ has a convergent subsequence, and so there exist $l < k \in \mathbb{N}$ such that $\left|1 - a^{k-l}\right| = \left|a^l - a^k\right| < \varepsilon$. By hypothesis one has $b := a^{k-l} \neq 1$, and since $|1-b| < \varepsilon$ every $z \in \mathbb{T}$ lies in ε-distance to some positive power of b. But $\{b^n : n \geq 0\} \subseteq \mathrm{orb}_+(1)$, and the proof is complete. □

If $a = \mathrm{e}^{2\pi \mathrm{i} x}$ then a is a root of unity if and only if x is a rational number. In this case we call the associated rotation system $(\mathbb{T},a)$ a **rational rotation**, otherwise an **irrational rotation**. Hence the rotation system $(\mathbb{T},a)$ is topologically transitive if and only if a is irrational.

Example 2.34. The product of two topologically transitive systems need not be topologically transitive. Consider $a_1 = \mathrm{e}^{\mathrm{i}}$, $a_2 = \mathrm{e}^{2\mathrm{i}}$, and the product system $(\mathbb{T}^2;(a_1,a_2))$. Then $M = \{(x,y) \in \mathbb{T}^2 : x^2 = y\}$ is a nontrivial, closed invariant set.

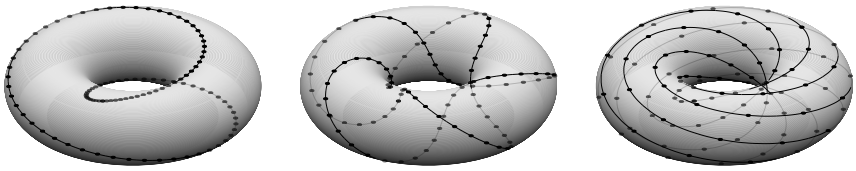


Fig. 2.1 The first 100 iterates of a point under the rotation in Example 2.34 and its orbit closure, and the same for the rotation ratios 2 : 5 and 8 : 5.

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

Theorem 2.35 (Kronecker). *Let $a = (a_1,\ldots,a_d) \in \mathbb{T}^d$. Then the rotation TDS $(\mathbb{T}^d;a)$ is topologically transitive if and only if $a_1,a_2,\ldots,a_d$ are linearly independent in the $\mathbb{Z}$-module $\mathbb{T}$. This means that if $a_1^{k_1}a_2^{k_2}\cdots a_d^{k_d} = 1$ for $k_1,k_2,\ldots,k_d \in \mathbb{Z}$, then $k_1 = k_2 = \cdots = k_d = 0$.*

We do not give the proof here, but shall return to it later in Chapter 14. Instead, we conclude this chapter by characterising topological transitivity of certain subshifts.

2.4 Transitivity of Subshifts

To a subshift $(F;\tau)$ of $(\mathscr{W}_K^+;\tau)$ of order 2 (Example 2.25) we associate its **transition matrix** $A=(a_{ij})_{i,j=0}^{k-1}$ by

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

The excluded blocks (i,j) of length 2 are exactly those with $a_{ij}=0$. Hence the positive matrix A describes the subshift completely.

In subshifts of order 2 we can "glue" words together as we shall see in the proof of the next lemma.

Lemma 2.36. *For $n\in\mathbb{N}$ one has $[A^n]_{ij}>0$ if and only if there is $x\in F$ with $x_0=i$ and $x_n=j$.*

Proof. We argue by induction on $n\in\mathbb{N}$, the case $n=1$ being clear by the definition of the transition matrix A. Suppose that this claim is proved for $n\geq 1$. The inequality $[A^{n+1}]_{ij}>0$ holds if and only if there is m such that

$$[A^n]_{im}>0 \quad \text{and} \quad A_{mj}>0.$$

By the induction hypothesis this means that there are $x,y\in F$ with $x_0=i$, $x_n=m$ and $y_0=m$, $y_1=j$.

So if by assumption $[A^{n+1}]_{ij}>0$, then we can define $z\in\mathscr{W}_k^+$ by

$$z_s:=\begin{cases}x_s & \text{if } s<n,\\ m & \text{if } s=n,\\ y_{s-n} & \text{if } s>n.\end{cases}$$

Then actually $z\in F$, because z does not contain any excluded block of length 2 by construction. Hence one direction of the claim is proved.

For the converse direction assume that there is $x\in F$ and $n\in\mathbb{N}$ with $x_0=i$ and $x_{n+1}=j$. Set $y=\tau^n(x)$, then $y_0=x_n=:m$ and $y_1=j$, as desired. □

A $k\times k$-matrix A with positive entries which has the property that for all $i,j\in\{0,\dots,k-1\}$ there is some $n\in\mathbb{N}$ with $[A^n]_{ij}>0$ is called **irreducible**. This property of transition matrices characterises forward transitivity.

Proposition 2.37. *Let $(F;\tau)$ be a subshift of order 2 and suppose that every letter occurs in some word in F. Consider the next assertions.*

(i) *The transition matrix A of $(F;\tau)$ is irreducible.*

(ii) *$(F;\tau)$ is forward transitive.*

Then (i) implies (ii), and if F does not have isolated points or if the shift is two-sided, then (ii) implies (i).

Proof. (i)$\Rightarrow$(ii): Since F is metrisable, we can apply Proposition 2.31. It suffices to consider open sets U and V intersecting F that are of the form

$$U = \{x : x_0 = u_0, \dots, x_n = u_n\} \quad \text{and} \quad V = \{x : x_0 = v_0, \dots, x_m = v_m\}$$

for some $n, m \in \mathbb{N}_0$, $u_0, \dots, u_n, v_0, \dots, v_m \in \{0, \dots, k-1\}$. Then we have to show $F \cap U \cap \tau^j(V) \neq \emptyset$ for some $j \in \mathbb{N}$. There is $x \in U \cap F$ and $y \in V \cap F$, and by assumption there are $N \in \mathbb{N}$, $z \in F$ with $z_0 = u_n$ and $z_N = v_0$. Now define w by

$$w_s := \begin{cases} x_s & \text{if } s < n, \\ u_n & \text{if } s = n, \\ z_{s-n} & \text{if } n < s < n+N, \\ v_0 & \text{if } s = n+N, \\ y_{s-n-N} & \text{if } s > n+N. \end{cases}$$

By construction w does not contain any excluded word of length 2, so $w \in F$, and of course $w \in U$, $\tau^{n+N}(w) \in V$.

(ii)$\Rightarrow$(i): Suppose F has no isolated points. If A is not irreducible, then there are $i, j \in \{0, \dots, k-1\}$ such that $[A^n]_{ij} = 0$ for all $n \in \mathbb{N}$. By Lemma 2.36 this means that there is no word in F of the form $(i, \cdots, j, \cdots)$. Consider the open sets

$$U = \{x : x_0 = i\} \quad \text{and} \quad V = \{x : x_0 = j\}$$

both of which intersect F since i and j occur in some word in F and F is shift invariant. However, for no $N \in \mathbb{N}$ can $\tau^N(V)$ intersect $U \cap F$, so by Proposition 2.31 the subshift cannot be topologically transitive.

Suppose now that $(F;\tau)$ is a two-sided shift, $y \in F$ has dense forward orbit and $x \in F$ is isolated. Then $\tau^n(y) = x$ for some $n \geq 0$. Since τ is a homeomorphism, y is isolated, too, and hence by compactness must have finite orbit. Consequently, F is finite and τ is a cyclic permutation of F. It is then easy to see that A is a cyclic permutation matrix, whence irreducible. □

For noninvertible shifts the statements (i) and (ii) are in general not equivalent.

Example 2.38. Consider the subshift $\mathscr{W}_3^B$ of $\mathscr{W}_3^+$ defined by the excluded blocks $B = \{(0,1), (0,2), (1,0), (1,1), (2,1), (2,2)\}$, i.e.,

$$F = \{(0,0,0,\dots,0,\dots), (1,2,0,\dots,0,\dots), (2,0,0,\dots,0,\dots)\}.$$

The $(F;\tau)$ has the transition matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Clearly, $(F;\tau)$ is transitive, but—as a moment's thought shows—A is not irreducible.

Even for two-sided subshifts, transitivity does not imply that its transition matrix is irreducible.

Example 2.39. Consider the subshift $F := \mathscr{W}_2^B$ of $\mathscr{W}_2$ given by the excluded blocks $B = \{(1,0)\}$, i.e., the elements of F are of the form

$$(\dots,0,\dots,0,\dots),\ (\dots,1,\dots,1,\dots),\ (\dots,0,\dots,0,1,\dots,1,\dots).$$

The transition matrix

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is clearly not irreducible, whereas $(F;\tau)$ is transitive (of course, not *forward* transitive!).

For more on subshifts of finite type we refer to Sec. 17 of [Denker et al. (1976)].

Exercises

1. Fix $k \in \mathbb{N}$ and consider $\mathscr{W}_k^+ = \{0,1,\dots,k-1\}^{\mathbb{N}_0}$ with its natural compact (product) topology as in Example 2.4. For words $x = (x_j)_{j\geq 0}$ and $y = (y_j)_{j\geq 0}$ in $\mathscr{W}_k^+$ define

$$\nu(x,y) := \min\{j \geq 0 : x_j \neq y_j\} \in \mathbb{N}_0 \cup \{\infty\}.$$

Show that $d(x,y) := \mathrm{e}^{-\nu(x,y)}$ is a metric that induces the product topology on $\mathscr{W}_k^+$. Show further that no metric on $\mathscr{W}_k^+$ inducing its topology can turn the shift τ into an isometry. How about the two-sided shift?

2. Prove that the function $d(x,y) := \min(|x-y|, 1-|x-y|)$ is a metric on $[0,1)$ which turns it into a compact space. For $\alpha \in [0,1)$ show that the addition with $\alpha \pmod 1$ is continuous with respect to the metric d. See also Example 2.6.

3. Let G be the Heisenberg group and H, K as in Example 2.10. Prove that $G = KH$.

4. Let $\alpha \in [0,1)$ and $a := \mathrm{e}^{2\pi\mathrm{i}\alpha} \in \mathbb{T}$. Show that the map

$$\Psi : ([0,1),\alpha) \to (\mathbb{T};a), \qquad \Psi(x) := \mathrm{e}^{2\pi\mathrm{i}x}$$

is an isomorphism of TDSs, cf. Example 2.13.

5. Let $(C;\varphi)$ be the Cantor system (Example 2.5). Show that the map

$$\Psi : (\mathscr{W}_2^+;\tau) \to (C;\varphi), \qquad \Psi(x) := \sum_{j=1}^{\infty} \frac{2x_{j-1}}{3^j}$$

is an isomorphism of TDSs, cf. Example 2.14.

6. Let $(G;a)$ be a group rotation and let $H \leq G$ be a closed subgroup of G. Show that the homogeneous system $(G/H;a)$ is a group factor of $(G;a)$, cf. Example 2.17.

7. Prove Lemma 2.22.

8 (Dyadic Adding Machine). Let $K := \{0,1\}^{\mathbb{N}_0}$ with the product topology, and define $\Psi : K \to \mathbb{A}_2$ by

$$\Psi((x_n)_{n\geq 0}) := (x_0 + \mathbb{Z}_2, x_0 + 2x_1 + \mathbb{Z}_4, x_0 + 2x_1 + 4x_2 + \mathbb{Z}_8, \dots\dots).$$

Show that Ψ is a homeomorphism. Now let φ be the unique dynamics on K that turns $\Psi : (K;\varphi) \to (\mathbb{A}_2;\mathbf{1})$ into an isomorphism. Describe the action of φ on a $\{0,1\}$-sequence $x \in K$.

9. Let $(K;\varphi)$ be a TDS.

a) Show that if $A \subseteq K$ is invariant/stable, then $\overline{A}$ is invariant/stable, too.

b) Show that the intersection of arbitrary many (bi-)invariant subsets of K is again (bi-)invariant.

c) Let $\Psi : (K;\varphi) \to (L;\psi)$ be a homomorphism of TDSs. Show that if $A \subseteq K$ is invariant/stable, then so is $\Psi(A) \subseteq L$.

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

11. Consider the **tent map** $\varphi(x) := 1 - |2x-1|$, $x \in [0,1]$. Prove that for every nonempty closed sub-interval I of $[0,1]$, there is $n \in \mathbb{N}$ with $\varphi^n(I) = [0,1]$. Show that $([0,1];\varphi)$ is transitive.

12. Consider the function $\psi : [0,1] \to [0,1]$, $\psi(x) := 4x(1-x)$. Show that $([0,1];\psi)$ is isomorphic to the tent map system from Exercise 11. (Hint: Use $\Psi(x) := \sin(\pi x/2)^2$ as an isomorphism.)

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

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

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

14. Prove that for a Hausdorff topological group G and a closed subgroup H the set of left cosets G/H becomes a Hausdorff topological space under the quotient map $q : G \mapsto G/H$, $q(g) = gH$. If for some compact set $K \subseteq G$ the equality $KH = G$ holds, then G/H is compact. If moreover H is a closed, normal subgroup, then G/H is a compact group.

(This is an abstract version of Examples 2.9 and 2.10.)

15. a) Describe all subshifts of $\mathscr{W}_2^+$ of order 2 and determine their transition matrices. Which of them are forward transitive?

b) Show that Proposition 2.37 does not remain true for two-sided subshifts if we replace forward transitivity by transitivity.

Chapter 3
Minimality and Recurrence

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

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

and then

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

Roughly speaking, these quotations mean the following: If a TDS does not contain a nontrivial invariant set, then each of its points returns arbitrarily near to itself under the dynamical action infinitely often.

Our aim is to prove these results, today known as Birkhoff's recurrence theorem, and to study the connected notions. To start with, let us recall from Chapter 2 that a subset $A \subseteq K$ of a TDS $(K;\varphi)$ is *invariant* if $\varphi(A) \subseteq A$. If A is invariant and nonempty, then it gives rise to the *subsystem* $(A;\varphi|_A)$.

By Exercise 9, the closure of an invariant set is invariant and the intersection of arbitrarily many invariant sets is again invariant. It is also easy to see that for every point $x \in K$ the forward orbit $\mathrm{orb}_+(x)$ and its closure are both invariant. So there are many possibilities to construct subsystems: Simply pick a point $x \in K$ and consider $F := \overline{\mathrm{orb}}_+(x)$. Then $(F;\varphi|_F)$ is a subsystem, which coincides with $(K;\varphi)$ whenever $x \in K$ is a *transitive point*. Can we always find nontransitive points, hence nontrivial subsystems?

3.1 Minimality

Systems without any subsystems play a special role for recurrence, so let us give them a special name.

Definition 3.1. A TDS $(K;\varphi)$ is called **minimal** if there are no nontrivial closed φ-invariant sets in K. This means that whenever $A \subseteq K$ is closed and $\varphi(A) \subseteq A$, then $A = \emptyset$ or $A = K$.

Remarks 3.2. 1) Given any TDS, its subsystems (= nonempty, closed, φ-invariant sets) are ordered by set inclusion. Clearly, a subsystem is minimal in this order if and only if it is a minimal TDS. Therefore we call such a subsystem a **minimal subsystem**.

2) By Lemma 2.23, if $(K;\varphi)$ is minimal, then $\varphi(K) = K$.

3) Even if a TDS is invertible, its subsystems need not be. However, by 2) every minimal subsystem of an invertible TDS is invertible. More precisely, it follows from Lemma 2.23 that an invertible TDS is minimal if and only if it has no nontrivial closed *bi-invariant* subsets.

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

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

6) More generally, if $\pi : (K;\varphi) \to (L;\psi)$ is a factor map, and if $(K;\varphi)$ is minimal, then so is $(L;\psi)$ (Exercise 1). In particular, if a product TDS is minimal so is each of its components. The converse is not true, see Example 2.34.

Minimality can be characterised in different ways.

Proposition 3.3. *For a TDS $(K;\varphi)$ the following assertions are equivalent.*

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

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

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

In particular, every minimal system is topologically forward transitive.

Proof. This is Exercise 2. □

Proposition 3.3 allows to extend the list of equivalences for rotation systems given in Theorem 2.32.

Theorem 3.4 (Rotation Systems II). *For a rotation system $(G;a)$ the following assertions are equivalent to any of the properties* (i)–(iv) of Theorem 2.32.

(v) The TDS $(G;a)$ is minimal.

(vi) $\{a^n : n \geq 0\}$ is dense in G.

(vii) $\{a^n : n \in \mathbb{Z}\}$ is dense in G.

Whenever a compact space K has at least two points, the TDS $(K;\mathrm{id})$ is not minimal. Also, the full shift $(\mathscr{W}_k^+;\tau)$ on words from a k-letter alphabet for $k \geq 2$ is nnot minimal. Indeed, every constant sequence is shift invariant and constitutes a one-point minimal subsystem.

It is actually a general fact that one always finds a subsystem that is minimal.

Theorem 3.5. *Every TDS $(K;\varphi)$ has at least one minimal subsystem.*

Proof. Let $\mathscr{M}$ be the family of all nonempty closed φ-invariant subsets of K. Then, of course, $K \in \mathscr{M}$, so $\mathscr{M}$ is nonempty. Further, $\mathscr{M}$ is ordered by set inclusion. Given a chain $\mathscr{C} \subseteq \mathscr{M}$ the set $C := \bigcap_{A\in\mathscr{C}} A$ is not empty (by compactness) and hence a lower bound for the chain $\mathscr{C}$. Zorn's lemma yields a minimal element F in $\mathscr{M}$. By Remark 3.2.1 above, $(F;\varphi|_F)$ is a minimal TDS. □

Besides for compact groups there is another important class of TDSs for which minimality and transitivity coincide. A TDS $(K;\varphi)$ is called **isometric** if there is a metric d inducing the topology of K such that φ is an isometry with respect to d.

Proposition 3.6. *An isometric TDS $(K;\varphi)$ is minimal if and only if it is topologically transitive.*

Proof. Suppose that $x_0 \in K$ has dense forward orbit and pick $y \in K$. By Proposition 3.3 it suffices to prove that $x_0 \in \overline{\mathrm{orb}}_+(y)$. Let $\varepsilon > 0$ be arbitrary. Then there is $m \in \mathbb{N}_0$ such that $d(\varphi^m(x_0), y) \leq \varepsilon$. The sequence $(\varphi^{mn}(x_0))_{n\in\mathbb{N}}$ has a convergent subsequence $(\varphi^{mn_k}(x_0))_{k\in\mathbb{N}}$. Using that φ is an isometry we obtain

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

This yields that there is $n \geq m$ with $d(x_0, \varphi^n(x_0)) < \varepsilon$, and therefore

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

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

Corollary 3.7 ("Structure Theorem" for Isometric Systems). *An isometric system is a possibly infinite disjoint union of minimal subsystems.*

Proof. By Remark 3.2.4, different minimal subsystems must be disjoint. Hence the statement is equivalent to saying that every point in K is contained in a minimal system. But for any $x \in K$ the system $(\overline{\mathrm{orb}}_+(x);\varphi)$ is topologically transitive, hence minimal by Proposition 3.6. □

Example 3.8. Consider the rotation by $a \in \mathbb{T}$ of the closed unit disc

$$\overline{\mathbb{D}} := \{z \in \mathbb{C} : |z| \leq 1\}, \quad \varphi(z) := a \cdot z.$$

The system $(\overline{\mathbb{D}}; \varphi)$ is isometric but not minimal. Now suppose that $a \in \mathbb{T}$ is not a root of unity. Then $\overline{\mathbb{D}}$ is the disjoint union $\overline{\mathbb{D}} = \bigcup_{r\in[0,1]} r\mathbb{T}$ and each $r\mathbb{T}$ is a minimal subsystem. The case that e is a root of unity is left as Exercise 3.

The shift system $(\mathscr{W}_k^+, \tau)$ is an example with points that are not contained in a minimal subsystem, see Example 3.10 below.

3.2 Topological Recurrence

For a rotation on the circle $\mathbb{T}$ we have observed in Example 2.33 two mutually exclusive phenomena:

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

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

Definition 3.9. Let $(K; \varphi)$ be a TDS. A point $x \in K$ is called

a) **recurrent** if for every open neighbourhood U of x, there is $m \in \mathbb{N}$ such that $\varphi^m(x) \in U$.

b) **almost periodic** if for every open neighbourhood U of x, the set of **return times**

$$\{m \in \mathbb{N} : \varphi^m(x) \in U\}$$

has bounded gaps. A subset $M \subseteq \mathbb{N}$ has **bounded gaps** (= **syndetic**, or **relatively dense**) if there is $N \in \mathbb{N}$ such that $M \cap [n, n+N] \neq \emptyset$ for every $n \in \mathbb{N}$.

c) **periodic** if there is $m \in \mathbb{N}$ such that $\varphi^m(x) = x$.

All these properties are invariant under automorphisms. Clearly, a periodic point is almost periodic, and an almost periodic point is recurrent. None of the converse implications is true in general (see Example 3.10 below). Further, a recurrent point is even **infinitely recurrent**, i.e. it returns infinitely often to all of its neighbourhoods (see Exercise 4).

For all these properties of a point $x \in K$ only the subsystem $\overline{\mathrm{orb}}_+(x)$ is relevant, whence we may suppose right away that the system is topologically transitive and x is a transitive point. In this context it is very helpful to use the notation

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

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

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

Example 3.10. Consider the full shift $(\mathscr{W}_k^+;\tau)$ in a k-alphabet.

a) A point $x \in \mathscr{W}_k^+$ is recurrent if every finite block of x occurs infinitely often in x (Exercise 5).

b) The point x is almost periodic if it is recurrent and the gaps between two occurrences of a given block y of x is bounded (Exercise 5).

c) By a) and b) a recurrent, but not almost periodic point is easy to find: Consider—for the sake of simplicity—$k = 2$, and enumerate the words formed from the alphabet $\{0,1\}$ according to a lexicographical ordering. Now write these words into one infinite word $x \in \mathscr{W}_2^+$:

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

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

d) A concrete example for a nonperiodic but almost periodic point is given in Exercise 12.

According to Example 2.33, all points in the rotation system $(\mathbb{T};a)$ are recurrent. This happens also in minimal systems: if K is minimal and $x \in K$, then $\overline{\mathrm{orb}}_{>0}(x)$ is a nontrivial, closed invariant set, hence it coincides with K and thus contains x. One can push this a little further.

Theorem 3.11. *Let $(K;\varphi)$ be a TDS and $x \in K$. Then the following assertions are equivalent.*

(i) *x is almost periodic.*

(ii) *$(\overline{\mathrm{orb}}_+(x);\varphi)$ is minimal.*

(iii) *x is contained in a minimal subsystem of $(K;\varphi)$.*

Proof. The equivalence of (ii) and (iii) is simple. Suppose that (iii) holds. Then we may suppose without loss of generality that $(K;\varphi)$ is minimal. Let $U \subseteq K$ be a nonempty open neighbourhood of x. By Proposition 3.3, $\bigcup_{n\in\mathbb{N}_0} \varphi^{-n}(U) = K$. By compactness there is $m \in \mathbb{N}$ such that $K = \bigcup_{j=0}^{m} \varphi^{-j}(U)$. For each $n \in \mathbb{N}$ we have $\varphi^n(x) \in \varphi^{-j}(U)$, i.e, $\varphi^{n+j}(x) \in U$ for some $0 \le j \le m$. So the set of return times $\{k \in \mathbb{N}_0 : \varphi^k(x) \in U\}$ has gaps of length at most m.

Conversely, suppose that (i) holds and let $F := \overline{\mathrm{orb}}_+(x)$. Take $y \in F$ and a closed neighbourhood U of x, cf. Lemma A.3 in Appendix A. Since x is almost periodic,

the set $\{n \in \mathbb{N} : \varphi^n(x) \in U\}$ has gaps with maximal length $m \in \mathbb{N}$, say. This means that

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

and hence $y \in \overline{\mathrm{orb}}_+(x) \subseteq \bigcup_{i=1}^{m+1} \varphi^{-i}(U)$. This implies $\varphi^i(y) \in U$ for some $1 \leq i \leq m+1$, and therefore $x \in \overline{\mathrm{orb}}_+(y)$. Hence (ii) follows by virtue of Proposition 3.3. □

Let us draw some immediate conclusions from this characterisations.

Proposition 3.12. a) *If a TDS contains a forward transitive and almost periodic point, then it is minimal.*

b) *In a minimal TDS every point is almost periodic.*

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

d) *In a group rotation system $(G;a)$ every point is almost periodic.*

e) *In a homogeneous system $(G/H;a)$ every point is almost periodic.*

Proof. a) and b) are immediate from Theorem 3.11, and c) follows from Theorem 3.11 together with Corollary 3.7.

For the proof of d), note that since right rotations are automorphisms of $(G;a)$ and 1 can be moved via a right rotation to any other point of G, it suffices to prove that 1 is almost periodic. Define $H := \mathrm{cl}\{a^n : n \in \mathbb{Z}\}$. This is a compact subgroup containing a, and hence the rotation system $(H;a)$ is a subsystem of $(G;a)$. It is minimal by Theorem 3.4. Hence, by Theorem 3.11, $1 \in H$ is almost periodic.

Assertion e) follows from d) and Exercise 6. □

Example 3.13. *Let $\alpha \in [0,1)$ and $I \subseteq [0,1)$ be an interval containing α in its interior. Then the set*

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

has bounded gaps.

Proof. Consider $([0,1);\alpha)$, the translation mod 1 by α (Example 2.6). It is isomorphic to the group rotation $(\mathbb{T};e^{2\pi i\alpha})$, so the claim follows from Proposition 3.12. □

Finally, combining Theorem 3.11 with Theorem 3.5 on the existence of minimal subsystems yields the following famous theorem.

Theorem 3.14 (Birkhoff). *Each TDS contains at least one almost periodic, hence recurrent point.*

3.3 Recurrence in Extensions

Given a compact group G, a TDS $(K;\varphi)$ and a continuous function $\Phi : K \to G$ we consider as in Section 2.2 on page 15 the group extension $(H;\psi)$ with

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

Proposition 3.15. *Let $(K;\varphi)$ be a TDS, G a compact group and $(H;\psi)$ the group extension along some $\Phi : K \to G$. If $x_0 \in K$ is a recurrent point, then $(x_0,g) \in H$ is recurrent in H for all $g \in G$.*

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

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

Since the projection $\pi : H \to K$ is continuous and H is compact, we have $\pi(\overline{\mathrm{orb}}_{>0}(x_0,1)) = \overline{\mathrm{orb}}_{>0}(x_0)$. The point x_0 is recurrent by assumption, so $x_0 \in \overline{\mathrm{orb}}_{>0}(x_0)$. Therefore, for some $h \in G$ we have $(x_0,h) \in \overline{\mathrm{orb}}_{>0}(x_0,1)$. Multiplying from the right by h in the second coordinate we obtain by continuity that

$$(x_0,h^2) \in \overline{\mathrm{orb}}_{>0}(x_0,h) \subseteq \overline{\mathrm{orb}}_{>0}(x_0,1).$$

Inductively this leads to $(x_0,h^n) \in \overline{\mathrm{orb}}_{>0}(x_0,1)$ for all $n \in \mathbb{N}$. By Proposition 3.12 we know that 1 is recurrent in $(G;h)$, so if V is a neighbourhood of 1, then $h^n \in V$ for some $n \in \mathbb{N}$. This means that $(x_0,h^n) \in U \times V$ for any neighbourhood U of x_0. Thus

$$(x_0,1) \in \overline{\{(x_0,h),(x_0,h^2),\dots\}} \subseteq \overline{\mathrm{orb}}_{>0}(x_0,1).$$

This means that $(x_0,1)$ is a recurrent point in $(H;\psi)$. □

An analogous result is true for almost periodic points.

Proposition 3.16. *Let $(K;\varphi)$ be a TDS, G a compact group and $(H;\psi)$ the group extension along $\Phi : K \to G$. If $x_0 \in K$ is an almost periodic point, then $(x_0,g) \in H$ is almost periodic in H for all $g \in G$.*

Proof. As before, it suffices to prove that (x_0,h) is almost-periodic for *one* $h \in G$. The set $\overline{\mathrm{orb}}_+(x_0)$ is minimal by Theorem 3.11, so by passing to a subsystem we can assume that $(K;\varphi)$ is minimal. Now let $(H';\psi)$ be a minimal subsystem in $(H;\psi)$ (Theorem 3.5). The projection $\pi : H \to K$ is a homomorphism from $(H;\psi)$ to $(K;\varphi)$, so the image $\pi(H')$ is a φ-invariant subset in K, and therefore must be equal to K. Let x_0 be an almost periodic point in $(K;\varphi)$, and $h \in G$ such that $(x_0,h) \in H'$. Then, by Theorem 3.11 (x_0,h) is almost periodic. □

We apply the resulst from above to the problem of *Diophantine approximation*. It is elementary that any real number can be approximated by rational numbers, but how well and, at the same time, how "simple" (keeping the denominator small) this approximation can be, is a hard question. The most basic answer is Dirichlet's theorem, see Exercise 11. Using the recurrence result from above, we can give a more sophisticated answer.

Corollary 3.17. *Let $\alpha \in \mathbb{R}$ and $\varepsilon > 0$ be given. Then there exists $n \in \mathbb{N}$, $k \in \mathbb{Z}$ such that*

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

Proof. Consider the TDS $([0,1);\alpha)$ from Example 2.6, and recall that, endowed with the appropriate metric and with addition modulo 1, it is a compact group homeomorphic to $\mathbb{T}$. We consider a group extension similar to Example 2.20. Let

$$\Phi : [0,1) \to [0,1), \qquad \Phi(x) = 2x + \alpha \pmod 1$$

and consider the group extension of $[0,1)$ along Φ. That means

$$H := [0,1) \times [0,1), \qquad \psi(x,y) = (x+\alpha, 2x+\alpha+y) \pmod 1.$$

Then by Proposition 3.15 $(0,0)$ is a recurrent point in $(H;\psi)$. By induction we obtain

$$\psi(0,0) = (\alpha,\alpha), \quad \psi^2(0,0) = \psi(\alpha,\alpha) = (2\alpha,4\alpha), \ldots, \psi^n(0,0) = (n\alpha, n^2\alpha).$$

The recurrence of $(0,0)$ implies that for any $\varepsilon > 0$ we have $d(0, n^2\alpha - \lfloor n^2\alpha \rfloor) < \varepsilon$ for some $n \in \mathbb{N}$, hence the assertion follows. □

By the same technique, i.e., using the recurrence of points in appropriate group extensions, we obtain the following result.

Proposition 3.18. *Let $p \in \mathbb{R}[x]$ be a polynomial of degree $k \in \mathbb{N}$ with $p(0) = 0$. Then for every $\varepsilon > 0$ there is $n \in \mathbb{N}$ and $m \in \mathbb{N}_0$ with*

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

Proof. Start from a polynomial $p(x)$ of degree $\deg p = k$ and define

$$p_k(x) := p(x), \quad p_{k-i}(x) := p_{k-i+1}(x+1) - p_{k-i+1}(x) \qquad (i = 1, \ldots, k).$$

Then each polynomial p_i has degree i. So p_0 is a constant α. Consider the TDS $([0,1);\alpha)$ (Example 2.6) and the following tower of group extensions. Set $H_1 := [0,1)$, $\psi_1(x) := x + \alpha \pmod 1$, and for $2 \leq i \leq k$ define

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

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

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

We can apply Proposition 3.15, and obtain by starting from a recurrent point x_1 in $(H_1;\psi_1)$ that every point $(x_1, x_2, \ldots, x_i)$ in $(H_i;\psi_i)$, $1 \leq i \leq k$, is recurrent. In particular, the point $p_1(0)$ just as any other point in $[0,1)$ is recurrent in the TDS $([0,1);\alpha)$, hence so is the point $(p_1(0), p_2(0), \ldots, p_k(0)) \in H_k$. By the definition of ψ_k we have

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

and, analogously, $\psi_k^n\big((p_1(0),p_2(0),\dots,p_k(0))\big) = \big(p_1(n),p_2(n),\dots,p_k(n)\big)$. By looking at the last component we see that for some $n \in \mathbb{N}$ the point $p_k(n) = p(n)$ comes arbitrarily close to $p_k(0) = p(0) = 0$. The assertion is proved. □

We refer to [Furstenberg (1981)] for more results on Diophantine approximations that can be obtained by means of simple arguments using topological recurrence. After having developed more sophisticated tools, we shall see more applications of dynamical systems to number theory in Chapters 10, 11 and 18.

Exercises

1. Let $\pi : (K;\varphi) \to (L;\psi)$ be a homomorphism of TDSs. Prove that if $B \subseteq L$ is a nonempty, closed ψ-invariant subset of L, then $\pi^{-1}(B)$ is a nonempty, closed φ-invariant subset of K. Conclude that if $(K;\varphi)$ is minimal, then so is $(L;\psi)$.

2. Prove Proposition 3.3.

3. Consider as in Example 3.8 the system $(\mathbb{D},\varphi)$, where $\varphi(z) = az$ is rotation by $a \in \mathbb{T}$. Describe the minimal subsystems in the case that a is a root of unity.

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

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

5. Prove the statements a) and b) from Example 3.10.

6. Let $\pi : (K;\varphi) \to (L;\psi)$ be a homomorphism of TDSs. Show that if $x \in K$ is periodic/recurrent/almost periodic, then $\pi(x) \in L$ is periodic/recurrent/almost periodic as well.

7. Consider the **tent map** $\varphi(x) := 1 - |2x-1|$, $x \in [0,1]$. By Exercise 9 the system $([0,1];\varphi)$ is topologically transitive. Is this TDS also minimal?

8. Show that the dyadic adding machine $(\mathbb{A}_2;\mathbf{1})$ is minimal, see Example 2.11 and Exercise 8.

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

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

11 (Dirichlet's Theorem). Let $\alpha \in \mathbb{R}$ be irrational and $n \in \mathbb{N}$ be fixed. Show that there are $p_n \in \mathbb{Z}$, $q_n \in \mathbb{N}$ such that $1 \leq q_n \leq n$ and

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

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

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

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

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

Chapter 4
The C^*-algebra $\mathrm{C}(K)$ and the Koopman Operator

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

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

for scalar-valued functions f on K, and we shall look at the functional analytic and operator theoretic aspects of TDSs. On the space of all such functions there is a rich structure: One can add, multiply, take absolute values or the complex conjugate. All these operations are defined pointwise, that is, we have

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

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

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

for all $f, g : K \to \mathbb{C}$, $\lambda \in \mathbb{C}$. In particular, the operator T is linear, and, denoting by $\mathbf{1}$ the function taking everywhere the value 1, satisfies $T\mathbf{1} = \mathbf{1}$.

Now, since φ is continuous, the operator T_φ leaves invariant the space $\mathrm{C}(K)$ of all complex-valued *continuous* functions on K; i.e., $f \circ \varphi$ is continuous whenever f is. Our major goal in this chapter is to show that the Koopman operator T_φ on $\mathrm{C}(K)$ actually *contains all information* about φ, that is, φ can be recovered from T_φ. In order to achieve this, we need a more detailed look on the Banach space $\mathrm{C}(K)$.

4.1 Continuous Functions on Compact Spaces

In this section we let $K \neq \emptyset$ be a nonempty compact topological space and consider the vector space $\mathrm{C}(K)$ of all continuous $\mathbb{C}$-valued functions on K. Since a continuous image of a compact space is compact, each $f \in \mathrm{C}(K)$ is bounded, hence

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

is finite. The map $\|\cdot\|_\infty$ is a norm on $\mathrm{C}(K)$ turning it into a complex *Banach space* (see also Appendix A.6). In our study of $\mathrm{C}(K)$ we shall rely on three classical theorems: Urysohn's Lemma, Tietze's Extension Theorem and the Stone–Weierstraß Theorem.

Urysohn's Lemma

Given a function $f : K \to \mathbb{R}$ and a number $r \in \mathbb{R}$, we introduce the notation

$$[f > r] := \{x \in K : f(x) > r\}$$

and, analogously, $[f < r], [f \geq r], [f \leq r]$ and $[f = r]$. The first lemma tells that in a compact space (by definition Hausdorff) two disjoint closed sets can be separated by disjoint open sets. This property of a topological space is called **normality**.

Lemma 4.1. *Let A, B be disjoint closed subsets of a compact space K. Then there are disjoint open sets $U, V \subseteq K$ with $A \subseteq U$ and $B \subseteq V$; or, equivalently, $A \subseteq U$ and $\overline{U} \cap B = \emptyset$.*

Proof. Let $x \in A$ be fixed. For every $y \in B$ there are disjoint open neighbourhoods $U(x,y)$ of x and $V(x,y)$ of y. Finitely many of the $V(x,y)$ cover B by compactness, i.e., $B \subseteq V(x,y_1) \cup \cdots \cup V(x,y_k) =: V(x)$. Set $U(x) := U(x,y_1) \cap \cdots \cap U(x,y_k)$, which is a open neighbourhood of x, disjoint from $V(x)$. Again by compactness finitely many $U(x)$ cover A, i.e., $A \subseteq U(x_1) \cup \cdots \cup U(x_n) =: U$. Set $V := V(x_1) \cap \cdots \cap V(x_n)$. Then U and V are disjoint open sets with $A \subseteq U$ and $B \subseteq V$. □

Lemma 4.2 (Urysohn). *Let A, B be disjoint closed subsets of a compact space K. Then there exists a continuous function $f : K \to [0,1]$ such that $A \subseteq [f = 0]$ and $B \subseteq [f = 1]$.*

Proof. We first construct for each dyadic rational $r = n/2^m$, $0 \leq n \leq 2^m$, an open set $U(r)$ with

$$A \subseteq U(r) \subseteq \overline{U(r)} \subseteq U(s) \subseteq \overline{U(s)} \subseteq B^{\mathrm{c}} \text{ for } r < s \text{ dyadic rationals in } (0,1).$$

We start with $U(0) := \emptyset$ and $U(1) := K$. By Lemma 4.1 there is some U with $A \subseteq U$ and $\overline{U} \cap B = \emptyset$. Set $U(1/2) = U$. Suppose that for some $m \geq 1$, all the sets $U(n/2^m)$ with $n = 0, \ldots, 2^m$ are already defined. Let $k \in [1, 2^{m+1}] \cap \mathbb{N}$ be odd. Lemma 4.1

yields open sets $U(k/2^{m+1})$, when this lemma is applied to each of the pairs of closed sets

$$\begin{array}{rll} A & \text{and } U(1/2^m)^{\mathrm{c}}, & \text{for } k=1 \\ \overline{U((k-1)/2^{m+1})} & \text{and } U((k+1)/2^{m+1})^{\mathrm{c}} & \text{for } k=3,\dots,2^{m+1}-3, \\ U(1-1/2^m)^{\mathrm{c}} & \text{and } B & \text{for } k=2^{m+1}-1. \end{array}$$

In this way the open sets $U(k/2^{m+1})$ are defined for all $k=0,\dots,2^{m+1}$. Recursively one obtains $U(r)$ for all dyadic rationals $r\in[0,1]$.

We now can define the desired function by

$$f(x) := \inf\bigl\{r : r\in[0,1] \text{ dyadic rational}, x\in U(r)\bigr\}.$$

If $x\in A$, then $x\in U(r)$ for all dyadic rationals, so $f(x)=0$. In turn, if $x\in B$, then $x\notin U(r)$ for any $r\in(0,1)$, so $f(x)=1$. It only remains to prove that f is continuous.

Let $r\in[0,1]$ be given and let $x\in[f<r]$. Then there is a dyadic rational $r'<r$ with $x\in U(r')$. Since for all $y\in U(r')$ one has $f(y)\le r'<r$, we see that $[f<r]$ is open. On the other hand let $x\in[f>r]$. Then there is a dyadic rational $r''>r$ with $f(x)>r''$, so $x\notin U(r'')$, but then $x\notin\overline{U(r')}$ for every $r'<r''$. If $r<r'<r''$ and $y\in\overline{U(r')}^{\mathrm{c}}$, then $f(y)\ge r'>r$. So $\overline{U(r')}^{\mathrm{c}}$ is an open neighbourhood of x belonging to $[f>r]$, hence this latter set is open. Altogether we obtain that f is continuous. □

Tietze's Theorem

Our second classical result is essentially an application of Urysohn's Lemma to the level sets of a continuous function defined on a closed subset of K.

Theorem 4.3 (Tietze). *Let K be a compact space, let $A\subseteq K$ be closed and let $f\in\mathrm{C}(A)$. Then there is $h\in\mathrm{C}(K)$ such that $h|_A=f$.*

Proof. Since the real and imaginary parts of a continuous function are continuous, it suffices to prove the assertion for real-valued functions. Let $f\in\mathrm{C}(A;\mathbb{R})$.

Since A is compact we may suppose that $f(A)\subseteq[0,1]$ and $0,1\in f(A)$. By Urysohn's Lemma 4.2 we obtain a continuous function $g_1 : K\to[0,1/3]$ with $\bigl[f\le\frac{1}{3}\bigr]\subseteq[g_1=0]$ and $\bigl[f\ge\frac{2}{3}\bigr]\subseteq\bigl[g_1=\frac{1}{3}\bigr]$. For $f_1 := f-g_1|_A$ we obtain $f_1\in\mathrm{C}(A;\mathbb{R})$ and $0\le f_1\le\frac{2}{3}$.

Suppose we have defined the continuous functions

$$f_n : A\to[0,(\tfrac{2}{3})^n] \quad\text{and}\quad g_n : K\to[0,\tfrac{1}{3}(\tfrac{2}{3})^n]$$

for some $n\in\mathbb{N}$. Then again by Urysohn's Lemma we obtain a continuous function

$$g_{n+1} : K\to[0,\tfrac{1}{3}(\tfrac{2}{3})^n]$$

with $\bigl[f_n\le\frac{1}{3}(\frac{2}{3})^n\bigr]\subseteq[g_{n+1}=0]$ and $\bigl[f_n\ge\frac{2}{3}(\frac{2}{3})^n\bigr]\subseteq\bigl[g_{n+1}=\frac{2}{3}(\frac{2}{3})^n\bigr]$.

Set $f_{n+1} = f_n - g_{n+1}|_A$. Since $\|g_n\|_\infty \leq (\frac{2}{3})^n$, the function

$$g := \sum_{n=1}^{\infty} g_n$$

is continuous, i.e., $g \in \mathrm{C}(K;\mathbb{R})$. For $x \in A$ we have

$$\begin{aligned} f(x) - (g_1(x) + \cdots + g_n(x)) &= f_1(x) - (g_2(x) + \cdots + g_n(x)) \\ &= f_2(x) - (g_3(x) + \cdots + g_n(x)) = \cdots = f_n(x). \end{aligned}$$

Since $f_n(x) \to 0$ as $n \to \infty$, we obtain $f(x) = g(x)$. □

The Stone–Weierstraß Theorem

Our third classical theorem about continuous functions is the Stone–Weierstrass theorem. A linear subspace $A \subseteq \mathrm{C}(K)$ is called a **subalgebra** if $f, g \in A$ implies $fg \in A$. It is called **conjugation invariant** if $f \in A$ implies $\overline{f} \in A$. We say that A **separates the points** of K if for any pair $x, y \in K$ of distinct points of K there is $f \in A$ such that $f(x) \neq g(x)$. Urysohn's Lemma tells that $\mathrm{C}(K)$ separates the points of K, since singleton sets $\{x\}$ are closed. We can now formulate another central result.

Theorem 4.4 (Stone–Weierstraß). *Let A be a complex conjugation invariant subalgebra of $\mathrm{C}(K)$ containing the constant functions and separating the points of K. Then A is dense in $\mathrm{C}(K)$.*

For the proof we note that $\overline{A}$ also satisfies the hypotheses of the theorem, hence we may suppose without of loss of generality that A is closed. The next result is the key to the proof.

Proposition 4.5. *Let A be a closed conjugation invariant subalgebra of $\mathrm{C}(K)$ containing the constant function $\mathbf{1}$. Then any positive function f has a unique real square root $g \in A$, i.e. $f = g^2 = \overline{g} \cdot g$.*

Proof. We prove that the square root of f, defined pointwise, belongs to A. By normalising first we can assume $\|f\|_\infty \leq 1$. Recall that the binomial series

$$(1+x)^{1/2} = \sum_{n=0}^{\infty} \binom{1/2}{n} x^n$$

is absolutely convergent for $-1 \leq x \leq 1$. Since $0 \leq f(x) \leq 1$ for all $x \in K$ implies $\|f - \mathbf{1}\|_\infty \leq 1$, we can plug $f - \mathbf{1}$ into this series to obtain g. Since A is closed, we have $g \in A$. □

Proof of Theorem 4.4. As noted above, we may suppose that A is closed. Notice first that it suffices to prove that $\mathrm{Re}A$ is dense in $\mathrm{C}(K;\mathbb{R})$, because, by conjugation

invariance, for each $f \in A$ also $\operatorname{Re} f$ and $\operatorname{Im} f$ belong to A. So one can approximate real and imaginary parts of functions $h \in \mathrm{C}(K)$ by elements of $\operatorname{Re} A$ separately.

Since A is a conjugation invariant algebra, for every $f \in A$ also $|f|^2 = f\overline{f} \in A$, thus Proposition 4.5 implies that $|f| \in A$. From this we obtain for every $f, g \in \operatorname{Re} A$ that

$$\max(f,g) = \frac{1}{2}(|f-g| + (f-g)) \in A$$

and

$$\min(f,g) = -\frac{1}{2}(|f-g| - (f-g)) \in A.$$

These facts will be crucial in the following.

Let $a, b \in \mathbb{R}$ and $x, y \in K$, $x \neq y$ be given. We construct a function $f_{x,y} \in A$ with

$$f_{x,y}(x) = a \quad \text{and} \quad f_{x,y}(y) = b.$$

Take $f \in A$ a function which separates the points x and y. By addition of certain multiples of $\mathbf{1}$ to f we may suppose $f(x) < 0 < f(y)$. By defining

$$f_{x,y} := a\frac{\min(f,0)}{f(x)} + b\frac{\max(f,0)}{f(y)},$$

we obtain a function $f_{x,y} \in A$ with the desired properties.

Let $x \in K$, $\varepsilon > 0$ and let $h \in \mathrm{C}(K)$. We now construct a function $f_x \in A$ such that

$$f_x(x) = h(x) \quad \text{and} \quad f_x(y) > h(y) - \varepsilon$$

for all $y \in K$. For every $y \in K$, $y \neq x$, take a function $f_{x,y} \in A$ with

$$f_{x,y}(x) = h(x) \quad \text{and} \quad f_{x,y}(y) = h(y),$$

and set $f_{x,x} = h(x)\mathbf{1}$. For the open sets

$$U_{x,y} := [f_{x,y} > h - \varepsilon\mathbf{1}]$$

we have $y \in U_{x,y}$, whence they cover the whole of K. By compactness we find $y_1, \ldots, y_n \in K$ such that

$$K = U_{x,y_1} \cup \cdots \cup U_{x,y_n}.$$

The function $f_x := \max(f_{y_1}, \ldots, f_{y_n}) \in A$ has the desired property. Indeed, for $y \in K$ there is some $k \in \{1, \ldots, n\}$ with $y \in U_{x,y_k}$, so

$$f_x(y) \geq f_{x,y_k}(y) > h(y) - \varepsilon.$$

Now let $h \in \mathrm{C}(K;\mathbb{R})$ and $\varepsilon > 0$ be arbitrary. For every $x \in K$ consider the function f_x from the above. The open sets $U_x := [f_x < h + \varepsilon\mathbf{1}]$, $x \in K$, cover K. So again by compactness we have a finite subcover $K \subseteq U_{x_1} \cup \cdots \cup U_{x_m}$. We set $f := \min(f_{x_1}, \ldots, f_{x_m}) \in A$ and obtain

$$h(y) - \varepsilon < f(y) \leq f_{x_k}(y) < h(y) + \varepsilon,$$

whenever $y \in U_{x_k}$. Whence $\|h - f\|_\infty \leq \varepsilon$, whence A is dense in $\mathrm{C}(K;\mathbb{R})$. □

Metrisability and Separability

The Stone–Weierstrass theorem can be employed to characterise the metrisability of the compact space K in terms of a topological property of $\mathrm{C}(K)$. Recall that a topological space Ω is called **separable** if there exists a countable and dense subset A of Ω. A Banach space E is separable if and only if there is a countable set $D \subseteq E$ such that $\overline{\mathrm{lin}}(D) = E$.

Lemma 4.6. *Each compact metric space is separable.*

Proof. Fix $m \in \mathbb{N}$. The balls $\mathrm{B}(x, 1/m)$, $x \in K$, cover K and so there is a finite set $F_m \subseteq K$ such that

$$K \subseteq \bigcup_{x \in F_m} B(x, 1/m).$$

Then the set $F := \bigcup_{m\in\mathbb{N}} F_m$ is countable and dense in K. □

Theorem 4.7. *Let K be a compact topological space. Then K is metrisable if and only if $\mathrm{C}(K)$ is separable.*

Proof. Suppose that $\mathrm{C}(K)$ is separable, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathrm{C}(K)$ such that $\{f_n : n \in \mathbb{N}\}$ is dense in $\mathrm{C}(K)$. Define

$$\Phi : K \to \Omega := \prod_{n\in\mathbb{N}} \mathbb{C}, \qquad \Phi(x) := (f_n(x))_{n\in\mathbb{N}},$$

where Ω carries the usual product topology. Then Φ is continuous and injective, by Urysohn's lemma and the density assumption. The topology on Ω is metrisable (see Appendix A.5), hence Φ is a homeomorphism from K onto $\Phi(K)$. Consequently, K is metrisable.

For the converse suppose that $d : K \times K \to \mathbb{R}_+$ is a metric that induces the topology of K. By Lemma 4.6 there is a countable set $A \subseteq K$ with $\overline{A} = K$. Consider the countable(!) set

$$D := \{f \in \mathrm{C}(K) : f \text{ is a finite product of functions } d(\cdot, y),\ y \in A\} \cup \{\mathbf{1}\}.$$

Then $\mathrm{lin}(D)$ is a conjugation invariant subalgebra of $\mathrm{C}(K)$ containing the constants and separating the points of K. By the Stone–Weierstrass theorem, $\overline{\mathrm{lin}}(D) = \mathrm{C}(K)$ and hence $\mathrm{C}(K)$ is separable. □

4.2 The Space $\mathrm{C}(K)$ as a Commutative C^*-Algebra

In this section we show how the compact space K can be recovered if only the space $\mathrm{C}(K)$ is known (see Theorem 4.11 below).

The main idea is readily formulated. Let K be any compact topological space. To $x \in K$ we associate the functional

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

called the **Dirac** or **evaluation functional** at $x \in K$. Then $\delta_x \in \mathrm{C}(K)'$ with $\|\delta_x\| = 1$. By Urysohn's lemma, $\mathrm{C}(K)$ separates the points of K, which means that the map

$$\delta : K \to \mathrm{C}(K)', \qquad x \mapsto \delta_x$$

is injective. Let us endow $\mathrm{C}(K)'$ with the weak*-topology (see Appendix C.5). Then δ is continuous, and since the weak*-topology is Hausdorff, δ is a *homeomorphism onto its image* (Proposition A.4). Consequently,

$$K \cong \{\delta_x : x \in K\} \subseteq \mathrm{C}(K)'. \tag{4.1}$$

In order to recover K from $\mathrm{C}(K)$ we need to distinguish the Dirac functionals within $\mathrm{C}(K)'$. For this we view $\mathrm{C}(K)$ not merely as a Banach space, but rather as a *Banach algebra*.

Recall, e.g. from Appendix C.2, the notion of a *commutative Banach algebra*. In particular, note that for us a Banach algebra is always *unital*, i.e., contains a unit element. Clearly, if K is a compact topological space, then we have

$$\|fg\|_\infty \le \|f\|_\infty \|g\|_\infty \qquad (f, g \in \mathrm{C}(K)),$$

and hence $\mathrm{C}(K)$ is a Banach algebra with respect to pointwise multiplication, unit element $\mathbf{1}$ and the sup-norm. The Banach algebra $\mathrm{C}(K)$ is degenerate if and only if $K = \emptyset$. Since $\mathrm{C}(K)$ is also invariant under complex conjugation and

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

holds, we conclude that $\mathrm{C}(K)$ is even a commutative *C^*-algebra*.

The central notion in the study of $\mathrm{C}(K)$ as a Banach algebra is that of an *(algebra) ideal*, see again Appendix C.2. For a closed subset $F \subseteq K$ we define

$$I_F := \big\{ f \in \mathrm{C}(K) : f \equiv 0 \quad \text{on } F \big\}.$$

Evidently, I_F is closed; and it is proper if and only if $F \neq \emptyset$. The next theorem tells that *each* closed ideal of $\mathrm{C}(K)$ is of this type.

Theorem 4.8. *Let $I \subseteq \mathrm{C}(K)$ be a closed algebra ideal. Then there is a closed subset $F \subseteq K$ such that $I = I_F$.*

Proof. Define

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

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

$$F_\varepsilon \subseteq [\,f_1 > 1\,] \cup \cdots \cup [\,f_k > 1\,]\,.$$

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

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

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

Indeed, one has $g \geq 1$ on F_ε and $|f| < \varepsilon$ on $K \setminus F_\varepsilon$.

Hence for n large enough we have $\|g_n - f\|_\infty \leq \varepsilon$. Since ε was arbitrary and I is closed, we obtain $f \in I$ as desired. □

Recall from Appendix C.2 that an ideal I in a Banach algebra A is called a **maximal** if $I \neq A$, and for every ideal J satisfying $I \subseteq J \subseteq A$ either $J = I$ or $J = A$. The maximal ideals in $\mathrm{C}(K)$ are easy to spot.

Lemma 4.9. *An ideal I of $\mathrm{C}(K)$ is maximal if and only if $I = I_{\{x\}}$ for some $x \in K$.*

Proof. It is straightforward to see that each $I_{\{x\}}$, $x \in K$, is a maximal ideal. Suppose converse that I is a maximal ideal. By Theorem 4.8 it suffices to show that I is closed. Since $\overline{I}$ is again an ideal and I is maximal, $\overline{I} = I$ or $\overline{I} = \mathrm{C}(K)$. In the latter case we can find $f \in I$ such that $\|\mathbf{1} - f\|_\infty < 1$. Then $f = \mathbf{1} - (\mathbf{1} - f)$ has no zeroes and hence $1/f \in \mathrm{C}(K)$. But then $\mathbf{1} = (1/f)f \in I$, whence $I = A$ contradicting the assumption that I is maximal. □

To proceed we observe that the maximal ideal $I_{\{x\}}$ can also be written as

$$I_{\{x\}} = \{f \in \mathrm{C}(K) : f(x) = 0\} = \ker \delta_x,$$

where δ_x is the Dirac functional at x as above. Clearly,

$$\delta_x : \mathrm{C}(K) \to \mathbb{C}$$

is a nonzero *multiplicative* linear functional satisfying $\delta_x(\mathbf{1}) = 1$, i.e., an *algebra homomorphism* from $\mathrm{C}(K)$ into $\mathbb{C}$. (See again Appendix C.2 for the terminology.) As before, also the converse is true.

Lemma 4.10. *A nonzero linear functional $\psi : \mathrm{C}(K) \to \mathbb{C}$ is multiplicative if and only if $\psi = \delta_x$ for some $x \in K$.*

Proof. Let $\gamma : \mathrm{C}(K) \to \mathbb{C}$ be nonzero and multiplicative. Then there is $f \in \mathrm{C}(K)$ such that $\gamma(f) = 1$. Hence

$$1 = \gamma(f) = \gamma(\mathbf{1}f) = \gamma(\mathbf{1})\gamma(f) = \gamma(\mathbf{1}),$$

which means that γ is an algebra homomorphism. Hence $I := \ker(\gamma)$ is a proper ideal, and maximal since it has codimension one. By Lemma 4.9 there is $x \in K$ such that $\ker(\gamma) = I_{\{x\}} = \ker(\delta_x)$. Hence $f - \gamma(f)\mathbf{1} \in \ker \delta_x$ and therefore

$$0 = (f - \gamma(f)\mathbf{1})(x) = f(x) - \gamma(f)$$

for every $f \in \mathrm{C}(K)$. □

Lemma 4.10 yields a purely algebraic distinction of the Dirac functionals among all linear functionals on $\mathrm{C}(K)$. Let us summarise our results in the following theorem.

Theorem 4.11. *Let K be a compact space, and let*

$$\Gamma(\mathrm{C}(K)) := \{\gamma \in \mathrm{C}(K)' : \gamma \textit{ algebra homomorphism}\}.$$

Then the map

$$\delta : K \to \Gamma(\mathrm{C}(K)) \qquad x \mapsto \delta_x$$

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

4.3 The Koopman Operator

We now return to our original setting of a TDS $(K;\varphi)$ with its Koopman operator $T = T_\varphi$ defined at the beginning of this chapter. Actually, we shall first be a little more general, i.e., we shall consider possibly different compact spaces K, L and a mapping $\varphi : L \to K$. Again we can define the **Koopman operator** T_φ mapping functions on K to functions on L. The following lemma says that φ is continuous if and only if $T_\varphi(\mathrm{C}(K)) \subseteq \mathrm{C}(L)$.

Lemma 4.12. *Let K, L be compact spaces, and let $\varphi : L \to K$ be a mapping. Then φ is continuous if and only if $f \circ \varphi$ is continuous for all $f \in \mathrm{C}(K)$.*

Proof. Clearly, if φ is continuous, then also $f \circ \varphi$ is continuous for every $f \in \mathrm{C}(K)$. Conversely, if this condition holds, then $\varphi^{-1}[\,|f| > 0\,] = [\,|f \circ \varphi| > 0\,]$ is open in L for every $f \in \mathrm{C}(K)$. To conclude that φ is continuous, it suffices to show that these sets *form a base of the topology of K* (see Appendix A.2). Note first that these sets

are open. On the other hand, if $U \subseteq K$ is open and $x \in U$ is a point, then by Urysohn's lemma one can find a function $f \in \mathrm{C}(K)$ such that $0 \le f \le 1$, $f(x) = 1$ and $f \equiv 0$ on $K \setminus U$. But then $x \in [|f| > 0] \subseteq U$. □

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

Theorem 4.13. *Let K, L be (nonempty) compact spaces and let $T : \mathrm{C}(K) \to \mathrm{C}(L)$ be linear. Then the following assertions are equivalent.*

(i) *T is an algebra homomorphism.*

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

In this case, φ in (ii) *is uniquely determined and the operator has norm $\|T\| = 1$.*

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

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

is an algebra homomorphism. By Theorem 4.11 there is a unique $x =: \varphi(y)$ such that $T'\delta_y = \delta_{\varphi(y)}$. This means that $(Tf)(y) = f(\varphi(y))$ for all $y \in L$, i.e., $Tf = f \circ \varphi$ for all $f \in \mathrm{C}(K)$. By Lemma 4.12, φ is continuous, whence (ii) is established. □

Theorem 4.13 with $K = L$ shows that *no information is lost* when looking at T_φ in place of φ itself. On the other hand, one has all the tools from linear analysis—in particular spectral theory—to study the linear operator T_φ. Our aim is to show how properties of the TDS $(K;\varphi)$ are reflected by properties of the operator T_φ. Here is a first result in this direction.

Lemma 4.14. *Let K, L be compact spaces, and let $\varphi : L \to K$ be continuous, with Koopman operator $T = T_\varphi : \mathrm{C}(K) \to \mathrm{C}(L)$. Then the following assertion holds.*

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

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

Proof. This is Exercise 1. □

An important consequence is the following.

Corollary 4.15. *The compact spaces K, L are homeomorphic if and only if the algebras $\mathrm{C}(K)$ and $\mathrm{C}(L)$ are isomorphic.*

We now return to TDSs and their invariant sets.

Lemma 4.16. *Let $(K;\varphi)$ be a TDS with Koopman operator $T = T_\varphi$ and let $A \subseteq K$ be a closed subset. Then A is φ-invariant if and only if the ideal I_A is T-invariant.*

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

Using Theorem 4.11, we obtain the following characterisation of minimality.

Corollary 4.17. *Let $(K;\varphi)$ be a TDS and $T = T_\varphi$ its Koopman operator on $\mathrm{C}(K)$. Then the TDS $(K;\varphi)$ is minimal if and only if no nontrivial closed algebra ideal of $\mathrm{C}(K)$ is invariant under T.*

An important object in the study of T_φ is its **fixed space**

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

It is the eigenspace corresponding to the eigenvalue 1, hence a spectral notion. Since $T_\varphi \mathbf{1} = \mathbf{1}$, the fixed space $\mathrm{fix}(T_\varphi)$ is at least one-dimensional (unless $K = \emptyset$ and the TDS is degenerate). If $f \in \mathrm{fix}(T_\varphi)$, then f is constant on each forward orbit $\mathrm{orb}_+(x)$, and, since f is continuous, even on the closure. This observation leads to a spectral property of topological transitivity.

Lemma 4.18. *Let $(K;\varphi)$ be a TDS with Koopman operator $T = T_\varphi$ on $\mathrm{C}(K)$. If $(K;\varphi)$ is topologically transitive, then $\mathrm{fix}(T)$ is one-dimensional.*

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

Exercise 3 shows that the converse statement fails.

An eigenvalue λ of $T = T_\varphi$ of modulus 1 is called a **peripheral eigenvalue**. The set of peripheral eigenvalues

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

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

Theorem 4.19. *Let $(K;\varphi)$ be a TDS with Koopman operator $T = T_\varphi$. If $\mathrm{fix}(T)$ is one-dimensional, then the peripheral point spectrum of T is a subgroup of $\mathbb{T}$, each peripheral eigenvalue is simple, and each corresponding eigenvector is unimodular (up to a multiplicative constant).*

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

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

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

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

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

Example 4.20 (Minimal Rotations). For a rotation system $(G;a)$ the associated Koopman operator is denoted by

$$L_a : \mathrm{C}(G) \to \mathrm{C}(G), \qquad (L_a f)(x) := f(ax) \qquad (f \in \mathrm{C}(G), x \in G)$$

and called the **left rotation (operator)**. Recall from Theorem 3.4 that $(G;a)$ is minimal if and only if $\{a^n : n \in \mathbb{N}_0\}$ is dense in G. In this case, G is Abelian, $\mathrm{fix}(L_a) = \mathbb{C}\mathbf{1}$, and by Theorem 4.19 the peripheral point spectrum $\sigma_p(L_a) \cap \mathbb{T}$ is a subgroup of $\mathbb{T}$.

In order to determine this subgroup, suppose that $\lambda \in \mathbb{T}$ and $\chi \in \mathrm{C}(G)$ such that $|\chi| = \mathbf{1}$ and $L_a\chi = \lambda\chi$, i.e., $\chi(ax) = \lambda\chi(x)$ for all $x \in G$. Without loss of generality we may suppose that $f(1) = 1$. It follows that

$$\chi(a^n a^m) = \chi(a^{n+m}) = \lambda^{n+m} = \lambda^n\lambda^m = \chi(a^n)\chi(a^m)$$

for all $n, m \in \mathbb{N}_0$. By continuity of χ and since the powers of a are dense in G, this implies

$$\chi : G \to \mathbb{T} \text{ continuous}, \qquad \chi(xy) = \chi(x)\chi(y) \qquad (x, y \in G)$$

i.e., χ is a continuous homomorphism into $\mathbb{T}$, a so-called **character** of G. Conversely, each such character χ of G is an unimodular eigenvalue of L_a, with eigenvalue $\chi(a)$. It follows that

$$\sigma_p(L_a) \cap \mathbb{T} = \big\{\chi(a) : \chi : G \to \mathbb{T} \text{ is a character}\big\}$$

is the group of peripheral eigenvalues of L_a. See Section 14.3 for more about characters.

4.4 The Theorem of Gelfand–Naimark

We now come to one of the great theorems of functional analysis. While we have seen above that $\mathrm{C}(K)$ is a commutative C^*-algebra, the following theorem tells that the converse also holds.

Theorem 4.21 (Gelfand–Naimark). *Let A be a commutative C^*-algebra. Then there is a compact space K and an isometric $*$-isomorphism $\Phi : A \to \mathrm{C}(K)$. The space K is unique up to homeomorphism.*

The Gelfand-Naimark theorem is central to our operator theoretic approach to ergodic theory, see Chapter 12 and the Halmos–von Neumann Theorem 17.13.

Let say some words about the proof of Theorem 4.21. By Corollary 4.15 the space K is unique up to homeomorphism. Moreover, Theorem 4.11 leads us the way to find K in the first place. Namely, it must be given as the set of all scalar algebra homomorphisms

$$\Gamma(A) := \big\{ \psi : A \to \mathbb{C} : \ \psi \text{ is an algebra homomorphism} \big\}.$$

The set $\Gamma(A)$, endowed with the restriction of the weak* topology $\sigma(A',A)$, is called the **Gelfand space** of A. Each element of A can be viewed in a natural way as a continuous function on $\Gamma(A)$, and the main problem is to show that in this manner one obtains an isomorphism of C^*-algebras $\mathrm{C}(\Gamma(A)) \cong A$. We postpone detailed proof to the Supplement of this chapter. At this point, let us rather familiarise with the result by looking at some examples.

Examples 4.22. 1) For a compact space K consider the algebra $A = \mathrm{C}(K)$. Then $\Gamma(A) = K$ as we proved in Section 4.2.

2) Consider the C^*-algebra $\mathrm{C}_0(\mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{C}$ vanishing at infinity. Define $A := \mathrm{C}_0(\mathbb{R}) \oplus \langle \mathbf{1} \rangle$. Then A is a C^*-algebra and $\Gamma(A)$ is homeomorphic to the one-point compactification of $\mathbb{R}$, i.e., to the unit circle $\mathbb{T}$. Of course, one can replace $\mathbb{R}$ by any locally compact Hausdorff space and thereby arrive at the **one-point compactification** of such spaces.

3) Let $\mathrm{X} = (X, \Sigma, \mu)$ be a finite measure space. For the C^*-algebra $\mathrm{L}^\infty(\mathrm{X})$, the Gelfand space is the Stone representation space of the **measure algebra** $\Sigma(\mathrm{X})$. See Chapter 5 and Chapter 12, in particular Section 12.4.

4) The Gelfand space of the C^*-algebra $\ell^\infty(\mathbb{N})$ is the **Čech–Stone compactification** of $\mathbb{N}$ and is denoted by $\beta\mathbb{N}$. This space will be studied in some more detail in Chapter 15.

5) Consider the Banach space $E := \ell^\infty(\mathbb{Z})$ and the left shift $T : E \to E$ thereon defined by $T(x_n) = (x_{n+1})$. We call a sequence $x = (x_n) \in \ell^\infty$ **almost periodic** if the set

$$\big\{ T^k x : k \in \mathbb{Z} \big\} \subseteq E$$

is relatively compact in E. The set $\mathrm{ap}(\mathbb{Z})$ of almost periodic sequences is a closed subspace of $\ell^\infty(\mathbb{Z})$ and actually a C^*-subalgebra. The Gelfand representation space of $A = \mathrm{ap}(\mathbb{Z})$ is $\mathrm{b}\mathbb{Z}$ the so-called **Bohr compactification** of $\mathbb{Z}$, see Exercise 13 and Chapter 14. Similar construction can be carried out for all locally compact topological groups.

Supplement: Proof of the Gelfand–Naimark Theorem

We need some preparation for the proof. But as has been already said, the candidate for the sought after compact space is the set of scalar algebra homomorphisms.

Definition 4.23. Let A be a commutative complex Banach algebra. The set

$$\Gamma(A) := \big\{ \psi : A \to \mathbb{C} : \psi \text{ is an algebra homomorphism} \big\}$$

is called the **Gelfand (representation) space** of A.

The first proposition shows that the elements of the Gelfand space are continuous functionals.

Proposition 4.24. *Let A be a commutative complex Banach algebra. If $\psi \in \Gamma(A)$, then ψ is continuous with $\|\psi\| \leq 1$.*

Proof. Suppose by contradiction that there is $a \in A$ with $\|a\| < 1$ and $\psi(a) = 1$. The series

$$b := \sum_{n=1}^{\infty} a^n$$

is absolutely convergent in A with $ab + a = b$. Since ψ is a homomorphism, we obtain

$$\psi(b) = \psi(ab) + \psi(a) = \psi(a)\psi(b) + \psi(a) = \psi(b) + 1,$$

a contradiction. This means that $|\psi(a)| \leq 1$ holds for all $a \in \mathrm{B}(0,1)$. □

As in the case of $A = \mathrm{C}(K)$, we now consider the dual Banach space A' of A endowed with its weak*-topology (see Appendix C.5). Then $\Gamma(A)$ is a weakly* closed subset of A' since

$$\Gamma(A) = \big\{ \psi \in A' : \psi(\mathrm{e}) = 1 \big\} \cap \bigcap_{x,y \in A} \big\{ \psi \in A' : \psi(xy) = \psi(x)\psi(y) \big\}.$$

As a consequence of Proposition 4.24 and the Banach–Alaoglu Theorem C.4 the set $\Gamma(A)$ is compact. Notice, however, that at this point we do not yet know whether $\Gamma(A)$ is nonempty. Later we shall show that $\Gamma(A)$ has sufficiently many elements, but accept this fact for the moment. For $x \in A$ consider the function

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

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

$$\Phi := \Phi_A : A \to \mathrm{C}(\Gamma(A)), \qquad x \mapsto \hat{x}$$

is well-defined and clearly an algebra homomorphism. It is called the **Gelfand representation** or **Gelfand map**.

To conclude the proof of the Gelfand–Naimark theorem it remains to show that

1) Φ commutes with the involutions, i.e., a $*$-*homomorphism*, see Appendix C.2;
2) Φ is isometric;
3) Φ is surjective.

The proof of 1) is simple as soon as one knows that an algebra homomorphism $\psi : A \to \mathbb{C}$ of a C^*-algebra is automatically a $*$-homomorphism. 2) entails that $\Gamma(A)$ is nonempty if A is nondegenerate; the analysis of the proof of Theorem 4.11 shows that in order to prove 2) and 3) we have to study maximal ideals in A and invertibility of elements of the form $\lambda \mathrm{e} - a \in A$.

To prove the remaining parts of the theorem needs an excursion into *Gelfand theory*. Extensive treatment of this topic can be found in [Rudin (1987), Chapter 18], [Rudin (1991), Part III] and in most books on functional analysis.

The Spectrum in Unital Algebras

Let A be a complex algebra with unit e and let $a \in A$. The **spectrum** of a is the set

$$\mathrm{Sp}(a) := \big\{\lambda \in \mathbb{C} : \lambda \mathrm{e} - a \text{ is not invertible}\big\}.$$

The number

$$r(a) := \inf_{n \in \mathbb{N}} \|a^n\|^{1/n}$$

is called the **spectral radius** of a. Here are some first properties.

Lemma 4.25. a) *For every $a \in A$ we have $r(a) \le \|a\|$.*

b) *If $A \neq \{0\}$ then $r(\mathrm{e}) = 1$.*

Proof. a) This follows from the submultiplicativity of the norm. To see b) notice that $\|\mathrm{e}\| = \|\mathrm{ee}\| \le \|\mathrm{e}\|^2$. Since by assumption $\|\mathrm{e}\| \neq 0$, we obtain $\|\mathrm{e}\| \ge 1$. This implies $\|\mathrm{e}\| = \|\mathrm{e}^n\| \ge 1$, and therefore $r(\mathrm{e}) = 1$. □

Just as in the case of bounded linear operators one can replace "inf" by "lim" in the definition of the spectral radius, i.e.

$$r(a) = \lim_{n \to \infty} \|a^n\|^{1/n},$$

see Exercise 6. The spectral radius is connected to the spectrum via the next result.

Lemma 4.26. *Let A be a Banach algebra with unit element* e. *If $a \in A$ is such that $r(a) < 1$, then $\mathrm{e} - a$ is invertible and its inverse is given by the* **Neumann**[1] **series**

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

Moreover, for $\lambda \in \mathbb{C}$ with $|\lambda| > r(a)$ one has $\quad (\lambda \mathrm{e} - a)^{-1} = \lambda^{-1} \sum_{n=0}^{\infty} \lambda^{-n} a^n$.

We leave the proof of this lemma as Exercise 5.

As for bounded linear operators on Banach spaces the spectrum has the following properties.

Proposition 4.27. *Let A be a complex Banach algebra and let $a \in A$. Then the following are true.*

a) *The spectrum $\mathrm{Sp}(a)$ is a compact subset of $\mathbb{C}$ and is contained in the closed ball $\overline{\mathrm{B}}(0, r(a))$.*

b) *The mapping*
$$\mathbb{C} \setminus \mathrm{Sp}(a) \to A, \qquad \lambda \mapsto (\lambda \mathrm{e} - a)^{-1}$$
is holomorphic and vanishes at infinity.

c) *If $A \neq \{0\}$, then there is $\lambda \in \mathrm{Sp}(a)$ with $|\lambda| = r(a)$.*

Proof. a) If $\lambda \in \mathbb{C}$ is such that $|\lambda| > r(a)$, then Lemma 4.26 implies that $\lambda \mathrm{e} - a$ is invertible, hence $\mathrm{Sp}(a) \subseteq \overline{\mathrm{B}}(0, r(a))$. We now show that the complement of $\mathrm{Sp}(a)$ is open, hence compactness of $\mathrm{Sp}(a)$ follows. Indeed, let $\lambda \in \mathbb{C} \setminus \mathrm{Sp}(a)$ be fixed and let $\mu \in \mathbb{C}$ be such that $|\lambda - \mu| < \|(\lambda \mathrm{e} - a)^{-1}\|^{-1}$. Then by Lemma 4.26 we conclude that $(\mu \mathrm{e} - a)^{-1}$ exists and is given by the absolutely convergent series

$$\begin{aligned} (\mu \mathrm{e} - a)^{-1} &= (\lambda \mathrm{e} - a)^{-1} ((\lambda - \mu)(\lambda \mathrm{e} - a)^{-1} - \mathrm{e})^{-1} \\ &= (\lambda \mathrm{e} - a)^{-1} \sum_{n=0}^{\infty} (\lambda - \mu)^n (\lambda \mathrm{e} - a)^{-n}, \end{aligned} \tag{4.2}$$

i.e., $\mu \in \mathbb{C} \setminus \mathrm{Sp}(a)$.

b) By (4.2) the function $\mu \mapsto (\mu \mathrm{e} - a)^{-1}$ is given by a power series in a small neighbourhood of each $\lambda \in \mathbb{C} \setminus \mathrm{Sp}(a)$, so it is holomorphic. From Lemma 4.26 we obtain

$$\|(\lambda \mathrm{e} - a)^{-1}\| \leq |\lambda^{-1}| \sum_{n=0}^{\infty} \|\lambda^{-n} a^n\| \leq |\lambda|^{-1} \sum_{n=0}^{\infty} |\lambda|^{-n} \|a\|^n = \frac{1}{|\lambda| - \|a\|}$$

for $|\lambda| > \|a\|$. This shows the second part of b).

c) Suppose that $\mathrm{Sp}(a) \subseteq \mathrm{B}(0, r_0)$ for some $r_0 > 0$. This means that

$$\mathbb{C} \setminus \overline{\mathrm{B}}(0, r_0) \ni \lambda \mapsto (\lambda \mathrm{e} - a)^{-1}$$

[1] Named after Carl Neumann (1832 – 1925).

is holomorphic and hence given by the series

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

which is uniformly convergent for $|\lambda| \geq r'$ with $r' > r_0$. This implies that $\|a^n\| \leq Mr'^n$ holds for all $n \in \mathbb{N}_0$ and some some $M \geq 0$. From this we conclude $r(a) \leq r'$, hence $r(a) \leq r_0$, and $\mathrm{Sp}(a)$ cannot be contained in any ball smaller than $\overline{\mathrm{B}}(0, r(a))$. It remains to prove that $\mathrm{Sp}(a)$ is nonempty provided $\dim A \geq 1$. If $\mathrm{Sp}(a)$ is empty then, by part b), the mapping

$$\mathbb{C} \to A, \qquad \lambda \mapsto (\lambda \mathrm{e} - a)^{-1}$$

is a bounded holomorphic function. For every $\varphi \in A'$ we can define the holomorphic function

$$f_\varphi : \mathbb{C} \mapsto \mathbb{C}, \quad \lambda \mapsto \varphi((\lambda \mathrm{e} - a)^{-1}),$$

which is again bounded and holomorphic. By Liouville's Theorem from complex analysis each f_φ is constant, but then by the Hahn–Banach Theorem C.3 also the function $\lambda \mapsto (\lambda \mathrm{e} - a)^{-1} \in A$ is constant. By part b), this constant must be zero, which is impossible if $A \neq \{0\}$. □

Theorem 4.28 (Gelfand–Mazur). *Let $A \neq \{0\}$ be a complex Banach algebra such that every nonzero element in A is invertible. Then A is isomorphic to $\mathbb{C}$.*

Proof. Let $a \in A$. Then by Proposition 4.27 there is $\lambda_a \in \mathrm{Sp}(a)$ with $|\lambda_a| = r(a)$. By assumption, $\lambda_a \mathrm{e} - a = 0$, so $a = \lambda_a \mathrm{e}$, hence A is one-dimensional. This proves the assertion. □

Maximal Ideals

If I is a closed ideal in the Banach algebra A, then the quotient vector space A/I becomes a Banach algebra if endowed with the norm

$$\|a + I\| = \inf\big\{ \|a + x\| : x \in I \big\}.$$

For the details we refer to Exercise 7.

Proposition 4.29. *Let A be a commutative complex Banach algebra. If $\psi \in \Gamma(A)$, then $\ker \psi$ is a maximal ideal. Conversely, if $I \subseteq A$ is a maximal ideal, then I is closed and there is a unique $\psi \in \Gamma(A)$ such that $I = \ker \psi$.*

Proof. Clearly, $\ker \psi$ is a closed ideal. Since $\ker \psi$ is of codimension one, it must be maximal.

For the second assertion let I be a maximal ideal. Consider its closure $\overline{I}$, still an ideal. Since $\mathrm{B}(\mathrm{e}, 1)$ consists of invertible elements by Lemma 4.25, we have $\mathrm{B}(\mathrm{e}, 1) \cap I =$

$\emptyset$, hence $\mathrm{B}(\mathrm{e},1)\cap\bar{I}=\emptyset$. In particular, $\bar{I}$ is a proper ideal, hence is equal to I by maximality. Therefore, each maximal ideal is closed.

Now consider the quotient algebra A/I. Since I is maximal, A/I does not contain any proper ideals other than 0. If a is a noninvertible element, then aA is an ideal, so actually it must be the zero-ideal. This yields $a=0$. Hence all nonzero elements in A/I are invertible. By the Gelfand–Mazur Theorem 4.28 we obtain that A/I is isomorphic to $\mathbb{C}$ under some Ψ. The required homomorphism $\psi:A\to\mathbb{C}$ is given by $\psi=\Psi\circ q$, where $q:A\to A/I$ is the quotient map. Uniqueness of ψ can be proved as follows. For every $a\in A$ we have $\psi(a)\mathrm{e}-a\in\ker\psi$. So if $\ker\psi=\ker\psi'$, then $\ker\psi'(\psi(a)\mathrm{e}-a)=0$, hence $\psi(a)=\psi'(a)$. □

As an important consequence of this proposition we obtain that the Gelfand space $\Gamma(A)$ is not empty if the algebra A is not degenerate. Indeed, by an application of Zorn's lemma, every proper ideal I of A is contained in a maximal one, so in a unital Banach algebra there are maximal ideals.

On our way to the proof of the Gelfand–Naimark theorem we need to connect the spectrum and the Gelfand space. This is content of the next theorem.

Theorem 4.30. *Let A be a commutative unital Banach algebra and let $a\in A$. Then*

$$\mathrm{Sp}(a)=\big\{\psi(a)\,:\,\psi\in\Gamma(A)\big\}=\hat{a}(\Gamma(A)).$$

Proof. Since $\psi(\mathrm{e})=1$ for $\psi\in\Gamma(A)$, one has $\psi(\psi(a)\mathrm{e}-a)=0$, so $\psi(a)\mathrm{e}-a$ cannot be invertible, i.e., $\psi(a)\in\mathrm{Sp}(a)$.

On the other hand if $\lambda\in\mathrm{Sp}(a)$, then the principal ideal $(\lambda\mathrm{e}-a)A$ is not the whole algebra A. So by a standard application of Zorn's lemma we obtain a maximal ideal I containing it. By Proposition 4.27 there is $\psi\in\Gamma(A)$ with $\ker\psi=I$. We conclude $\psi(\lambda\mathrm{e}-a)=0$, i.e., $\psi(a)=\lambda$. □

An immediate consequence is the **spectral radius formula**.

Corollary 4.31. *For a commutative unital Banach algebra A and an element $a\in A$ we have*

$$r(a)=\sup\big\{|\psi(a)|\,:\,\psi\in\Gamma(A)\big\}=\|\hat{a}\|_\infty=\|\Phi_A(a)\|_\infty.$$

The Proof of the Gelfand–Naimark Theorem

To complete the proof of Gelfand–Naimark theorem the C^*-algebra structure has to enter the picture.

Lemma 4.32. *Let A be a commutative C^*-algebra. For $a\in A$ with $a=a^*$ the following assertions are true.*

a) $r(a)=\|a\|$.

b) $\mathrm{Sp}(a)\subseteq\mathbb{R}$*; equivalently,* $\psi(a)\in\mathbb{R}$ *for every* $\psi\in\Gamma(A)$.

Proof. a) We have $aa^* = a^2$, so $\|a^2\| = \|a\|^2$ holds. By induction one can prove that $\|a^{2^n}\| = \|a\|^{2^n}$ is true for all $n \in \mathbb{N}_0$. From this we obtain

$$r(a) = \lim_{n\to\infty} \|a^{2^n}\|^{\frac{1}{2^n}} = \|a\|.$$

b) The equivalence of the two formulations is clear by Theorem 4.30. For a fixed $\psi \in \Gamma(A)$ set $\alpha = \operatorname{Re}\psi(a)$ and $\beta = \operatorname{Im}\psi(a)$. Since $\psi(\mathrm{e}) = 1$, we obtain

$$\psi(a + it\mathrm{e}) = \alpha + i(\beta + t)$$

for all $t \in \mathbb{R}$. The inequality

$$\|(a+it\mathrm{e})(a+it\mathrm{e})^*\| = \|a^2 + t^2\mathrm{e}\| \le \|a\|^2 + t^2$$

is easy to see. On the other hand, by using that $\|\psi\| \le 1$ and the C^*-property of the norm we obtain

$$\alpha^2 + (\beta+t)^2 = |\psi(a+it\mathrm{e})|^2 \le \|a+it\mathrm{e}\|^2 = \|(a+it\mathrm{e})(a+it\mathrm{e})^*\|.$$

These two inequalities imply that

$$\alpha^2 + \beta^2 + 2\beta t + t^2 \le \|a\|^2 + t^2 \quad \text{for all } t \in \mathbb{R}.$$

This yields $\beta = 0$. □

Lemma 4.33. *Let A be a commutative C^*-algebra and let $\psi \in \Gamma(A)$. Then*

$$\psi(a^*) = \overline{\psi(a)} \qquad \textit{for all } a \in A,$$

i.e., ψ is a $*$-homomorphism.

Proof. For $a \in A$ fixed define

$$x := \frac{a+a^*}{2} \quad \text{and} \quad y := \frac{b-b^*}{2i}.$$

Then we have $x = x^*$ and $y = y^*$ and $a = x + iy$. By Lemma 4.32.b) $\psi(x), \psi(y) \in \mathbb{R}$, so we conclude

$$\psi(a^*) = \psi(x^*) - i\psi(y^*) = \overline{\psi(x) + i\psi(y)} = \overline{\psi(x+iy)} = \overline{\psi(a)}.$$ □

Proof the Gelfand–Naimark Theorem 4.21. Let $a \in A$ and let $\psi \in \Gamma(A)$. Then by Lemma 4.32.b) we obtain

$$\Phi(a^*)(\psi) = a^*(\psi) = \psi(a^*) = \overline{\psi(a)} = \overline{\widehat{a}(\psi)} = \overline{\Phi(a)}(\psi),$$

hence Φ is a $*$-homomorphism, and therefore a homomorphism between the C^*-algebras A and $\mathrm{C}(\Gamma(A))$.

Next, we prove that the Gelfand map Φ is an isometry, i.e., that $\|a\| = \|\hat{a}\|_\infty$. For $a \in A$ we have $(aa^*)^* = aa^*$, so by Lemma 4.32.a) $\|aa^*\| = r(aa^*)$. From this and from Corollary 4.31 we conclude

$$\|a\|^2 = \|aa^*\| = r(aa^*) = \|\Phi(aa^*)\|_\infty = \|\Phi(a)\overline{\Phi(a)}\|_\infty = \|\Phi(a)\|_\infty^2.$$

Finally, we prove that $\Phi : A \to \mathrm{C}(\Gamma(A))$ is surjective. Since Φ is a homomorphism, $\Phi(A)$ is $*$-subalgebra, i.e., a conjugation invariant subalgebra of $\mathrm{C}(\Gamma(A))$. It trivially separates the points of $\Gamma(A)$. By the Stone–Weierstraß Theorem 4.4 $\Phi(A)$ is dense, but since Φ is an isometry its image is closed. Hence Φ is surjective. □

Exercises

1. Prove Lemma 4.14.

2. Show that $\mathrm{C}(K)$ is separable if and only if K is metrisable.

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

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

5. Prove Lemma 4.26.

6 (Fekete's Lemma). Let $(x_n) \in \mathbb{R}$ be a nonnegative sequence with the property

$$x_{n+m}^{n+m} \leq x_n^n x_m^m \qquad \text{for all } n,m \in \mathbb{N}.$$

Prove that $(x_n)_n$ is convergent and that

$$\inf_{n\in\mathbb{N}} x_n = \lim_{n\to\infty} x_n.$$

Apply this to prove the equality

$$r(a) = \lim_{n\to\infty} \|a^n\|^{1/n}$$

for $a \in A$, A a Banach algebra.

7. Let A be a commutative Banach algebra and let I be a closed ideal in A. Prove that

$$\|a+I\| = \inf\big\{\|a+x\| : x \in I\big\}$$

defines a norm on the quotient vector space A/I, and that A/I becomes a Banach algebra with the multiplication

$$(a+I)(b+I) := ab+I.$$

8. Let A be a Banach algebra and $a,b \in A$ such that $ab = ba$. Prove that $r(ab) \leq r(a)r(b)$.

9. Let A be a complex Banach algebra. Prove that the Gelfand map is an isometry if and only if one has $\|a\|^2 = \|a^2\|$ for all $a \in A$.

10 (Disc Algebra). We let $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ be the open unit disc. Show that

$$\mathscr{D} := \left\{ f \in \mathrm{C}(\overline{\mathbb{D}}) : f\big|_{\mathbb{D}} \text{ holomorphic} \right\}$$

is a commutative complex Banach algebra if endowed with pointwise operations and the sup-norm. Define

$$f^*(z) := \overline{f(\overline{z})}.$$

Show that $f \mapsto f^*$ is an involution on $\mathscr{D}$ and $\|f\|_\infty = \|f^*\|_\infty$. Prove that the Gelfand map $\Phi_{\mathscr{D}}$ is an algebra homomorphism but not a $*$-homomorphism.

11. Let A be a C^*-algebra. Prove that $r(a) = \|a\|$ for each $a \in A$ with $aa^* = a^*a$.

12. Let A be a complex algebra *without* unit element. Show that the vector space $A_+ := A \oplus \mathbb{C}$ becomes a unital algebra with respect to the multiplication

$$(a,x)\cdot(b,y) := (ab+xb+ya, xy) \qquad (a,b \in A, x,y \in \mathbb{C})$$

and with unit element $\mathrm{e} = (0,1)$. Suppose that A is a Banach space, and that $\|ab\| \leq \|a\|\,\|b\|$ for all $a,b \in A$. Show that A_+ becomes Banach algebra with respect to the norm $\|(a,x)\| := \|a\| + |x|$.

13. Show that $\mathrm{ap}(\mathbb{Z})$, the set of almost periodic sequences is a C^*-subalgebra of $\ell^\infty(\mathbb{Z})$.

14. Show that c, the space of convergent scalar sequences, endowed with the obvious operations and the sup-norm is a C^*-algebra. Describe its Gelfand space.

Chapter 5
Measure-Preserving Systems

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

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

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

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

$$\mu[f > \alpha] = \mu[f \circ \varphi > \alpha] \qquad \text{for all } \alpha \in \mathbb{R}.$$

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

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

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

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

A **measurable space** is a pair (X,Σ), where X is a set and Σ is a σ-algebra of subsets of X. Given measurable spaces (X,Σ) and (Y,Σ'), a mapping $\varphi : X \to Y$ is called measurable if

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

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

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

It is convenient to abbreviate a measure space (X,Σ,μ) simply with the single letter X, i.e., $\mathrm{X} = (X,\Sigma,\mu)$. However, if different measure spaces $\mathrm{X},\mathrm{Y},\mathrm{Z},\ldots$ are involved we shall occasionally use $\Sigma_\mathrm{X},\Sigma_\mathrm{Y},\Sigma_\mathrm{Z}\ldots$ for the corresponding σ-algebras, and $\mu_\mathrm{X},\mu_\mathrm{Y},\mu_\mathrm{Z}\ldots$ for the corresponding measures. For integration with respect to the measure of X we introduce the notation

$$\int_\mathrm{X} f\cdot g = \langle f,g\rangle_\mathrm{X} := \int_X f\cdot g \,\mathrm{d}\mu$$

whenever $f\cdot g \in \mathrm{L}^1(\mathrm{X})$. When no confusion can arise we shall drop the subscript X, and write, e.g.,

$$\int fg = \langle f,g\rangle = \langle fg,\mathbf{1}\rangle = \langle g,f\rangle$$

whenever $f\cdot g \in \mathrm{L}^1(\mathrm{X})$.

Definition 5.1. Let X and Y be measure spaces. A measurable mapping $\varphi : X \to Y$ is called **measure-preserving** if $\varphi_*\mu_\mathrm{X} = \mu_\mathrm{Y}$, i.e.,

$$\mu[\,\varphi \in A\,] = \mu_\mathrm{Y}(A) \qquad (A \in \Sigma_\mathrm{Y}).$$

In the case that $\mathrm{X} = \mathrm{Y}$ and φ is measure-preserving, μ is called an **invariant measure** for φ (or φ-invariant, invariant under φ).

Standard arguments from measure theory show that for $\varphi_*\mu_\mathrm{X} = \mu_\mathrm{Y}$ it is sufficient to have $\mu_\mathrm{X}[\,\varphi \in A\,] = \mu_\mathrm{Y}(A)$ for all A belonging to a *generator* $\mathcal{E}$ of the σ-algebra Σ_Y, see Appendix B.1. Furthermore, it is also equivalent to

$$\int_\mathrm{X} (f\circ\varphi) = \int_\mathrm{Y} f$$

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

We can now introduce our main object of interest.

Definition 5.2. A pair $(\mathrm{X};\varphi)$ is called a **measure-preserving dynamical system** (MDS) if $\mathrm{X} = (X,\Sigma,\mu)$ is a probability space, $\varphi : X \to X$ is measurable and μ is φ-invariant.

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

5.1 Examples

1. The Baker's Transformation

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

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

Lebesgue measure is invariant under this transformation since

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

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

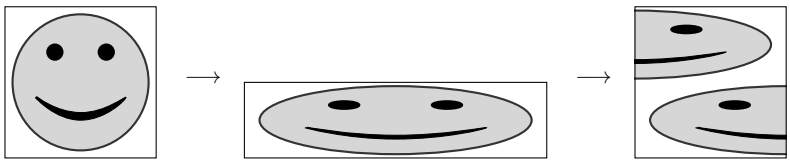


Fig. 5.1 The baker's transformation deforms a square like a baker does with puff pastry when kneading.

2. The Doubling Map

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

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

(cf. also Exercise 2.10). The Lebesgue measure is invariant under φ since

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

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

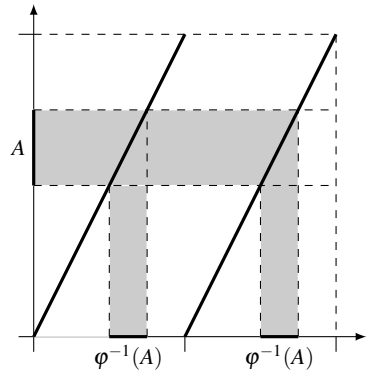


Fig. 5.2 A set A and its inverse image under the doubling map.

3. The Tent Map

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

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

(cf. Exercise 2.11). It is Exercise 1 to show that φ preserves the Lebesgue measure.

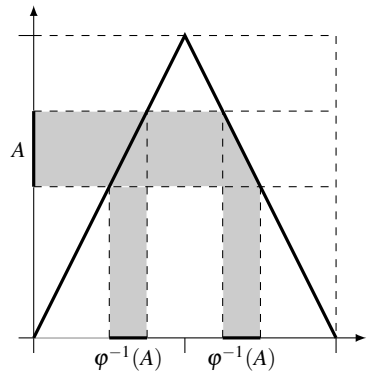


Fig. 5.3 A set A and its inverse image under the tent map.

4. The Gauss Map

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

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

It is easy to see that

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

and

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

for every $y \in [0,1)$. By Exercise 2, the measure

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

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

5. Bernoulli Shifts

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

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

(cf. Example 2.4) with the product σ-algebra $\Sigma := \bigotimes_{n\geq 0} \mathcal{P}(L)$, see Appendix B.6. On (X,Σ) we consider the left shift τ. To see that τ is a measurable mapping, note that the *cylinder sets*, i.e., sets of the form

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

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

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

is again contained in Σ.

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

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

(By Theorem B.9 there is a unique measure μ on Σ satisfying this requirement.) The product measure μ is shift invariant, because for cylinder sets A as in (5.1) we have

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

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

The Bernoulli shift is a special case of a more general construction. Namely, fix a probability space Y and consider the infinite product $X := \prod_{n\geq 0} Y$ with the product measure $\mu_X := \bigotimes_{n\geq 0} \mu_Y$ on the product σ-algebra $\Sigma_{\mathrm{X}} := \bigotimes_{n\geq 0} \Sigma_{\mathrm{Y}}$. Then the shift τ is measurable and μ_{X} is τ-invariant. Obviously, all this can also be done with two-sided shifts.

6. Markov Shifts

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

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

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

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

one has

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

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

$$\mu[\tau \in A] = \mu\left(L \times \{j_0\} \times \{j_1\} \cdots \times \{j_n\} \times \prod_{m>n} L\right) = \sum_{j=0}^{k-1} p_j s_{jj_0} \ldots s_{j_{n-1}j_n}.$$

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

$$p_{j_0} s_{j_0 j_1} \ldots s_{j_{n-1} j_n} = \sum_{j=0}^{k-1} p_j s_{jj_0} \ldots s_{j_{n-1} j_n}$$

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

$$p_l = \sum_{j=0}^{k-1} p_j s_{jl}.$$

This means that $p^t S = p^t$, i.e., p^t is a left fixed vector of S. Such a left fixed probability vector indeed always exists, by Perron's theorem. (See Exercise 9 and Theorem 8.12 for proofs of Perron's theorem.)

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

As in the case of Bernoulli shifts, Markov shifts can be generalised to arbitrary probability spaces. One needs the notion of a probability kernel and the Ionescu–Tulcea theorem (Theorem B.10).

7. Products and Skew-Products

Let $(\mathrm{X}; \varphi)$ be an MDS and let Y be a probability space. Furthermore, let

$$\Phi : X \times Y \to Y$$

be a measurable map such that $\Phi(x,\cdot)$ preserves μ_Y for all $x \in X$. Define

$$\psi : X \times Y \to X \times Y, \qquad \psi(x,y) := (\varphi(x), \Phi(x,y)).$$

Then the product measure $\mu_\mathrm{X} \otimes \mu_\mathrm{Y}$ is ψ-invariant. Hence $(X \times Y, \Sigma_\mathrm{X} \otimes \Sigma_\mathrm{X}, \mu_\mathrm{X} \otimes \mu_\mathrm{X}; \psi)$ is an MDS, called the **skew product** of $(\mathrm{X}; \varphi)$ along Φ.

In the special case that $\Phi(x,\cdot) = \varphi'$ does not depend on $x \in X$, the skew product is called the **product** of the MDSs $(\mathrm{X}; \varphi)$ and $(\mathrm{Y}; \varphi')$.

5.2 Measures on Compact Spaces

This section connects our discussion about topological dynamics with invariant measures. Since this is not a course on measure theory we do not provide all details here. Interested readers may consult [Bogachev (2007)], [Bauer (1981)] or any other standard textbook on measure theory.

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

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

Clearly $\mathrm{Ba}(K) \subseteq \mathrm{Bo}(K)$, and by the proof of Lemma 4.12 the open Baire sets form a base for the topology of K. Exercise 8 shows that $\mathrm{Ba}(K)$ is generated by the compact G_δ-subsets of K. If K is metrisable, then $\mathrm{Ba}(K)$ and $\mathrm{Bo}(K)$ coincide, but in general this is false, see Exercise 10 and [Bogachev (2007), II, 7.1.3].

A (complex) measure μ on $\mathrm{Ba}(K)$ (resp. $\mathrm{Bo}(K)$) is called a **Baire measure** (resp. **Borel measure**).

Lemma 5.3. *Every finite positive Baire measure is* **regular** *in the sense that for every $B \in \mathrm{Ba}(K)$ one has*

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

Proof. This is a standard argument involving Dynkin systems (Theorem B.1); for a different proof see [Bogachev (2007), II, 7.1.8]. □

Combining the regularity with Urysohn's lemma one can show that for a finite positive Baire measure μ, $\mathrm{C}(K)$ is dense in $\mathrm{L}^p(K, \mathrm{Ba}(K), \mu)$, $1 \le p < \infty$ (see also Lemma D.3).

A finite positive *Borel* measure is called **regular** if

$$\begin{aligned}\mu(B) &= \sup\big\{\mu(A) : A \text{ compact}, A \subseteq B\big\}\\ &= \inf\big\{\mu(O) : O \text{ open}, B \subseteq O\big\}.\end{aligned} \tag{5.3}$$

To clarify the connection between Baire and regular Borel measures we state the following result.

Proposition 5.4. *If μ, ν are regular finite positive Borel measures on K, then to each Borel set $A \in \mathrm{Bo}(K)$ there is a Baire set $B \in \mathrm{Ba}(K)$ such that $\mu(A \triangle B) = 0 = \nu(A \triangle B)$.*

Proof. Let $A \subseteq K$ be a Borel set. By regularity, there are open sets O_n, O_n', closed sets L_n, L_n' such that $L_n, L_n' \subseteq A \subseteq O_n, O_n'$ and $\mu(O_n \setminus L_n), \nu(O_n' \setminus L_n') \to 0$. By passing to $O_n \cap O_n'$ and $L_n \cap L_n'$ we may suppose that $O_n = O_n'$ and $L_n = L_n'$. Moreover, by

passing to finite unions of the L_n and finite intersections of the O_n we may suppose that

$$L_n \subseteq L_{n+1} \subseteq A \subseteq O_{n+1} \subseteq O_n$$

for all $n \in \mathbb{N}$. By Urysohn's lemma we can find continuous functions $f_n \in \mathrm{C}(K)$ with $\mathbf{1}_{L_n} \leq f_n \leq \mathbf{1}_{O_n}$. Now define

$$B := \bigcap_{n\in\mathbb{N}} [f_n > 0],$$

which is clearly a Baire set. By construction, $L_n \subseteq B \subseteq O_n$. This yields $A \triangle B \subseteq O_n \setminus L_n$ for all $n \in \mathbb{N}$, whence $\mu(A\triangle B) = 0 = \nu(A\triangle B)$. □

Proposition 5.4 shows that a regular Borel measure is completely determined by its restriction to the Baire algebra. Moreover, as soon one disregards null sets, the two concepts become exchangeable, see also Remark 5.8 below.

Let us denote the linear space of finite complex Baire measures on K by $\mathrm{M}(K)$. It is a Banach space with respect to the total variation norm $\|\mu\|_{\mathrm{M}} := |\mu|(K)$ (see Appendix B.9). Every $\mu \in \mathrm{M}(K)$ defines a linear functional on $\mathrm{C}(K)$ via

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

Since $|\langle f,\mu\rangle| \leq \int_K |f| \,\mathrm{d}\,|\mu| \leq \|f\|_\infty \|\mu\|_{\mathrm{M}}$, we have $\langle\cdot,\mu\rangle \in \mathrm{C}(K)'$, the Banach space dual of $\mathrm{C}(K)$. The following lemma shows in particular that the association $\mu \mapsto \langle\cdot,\mu\rangle$ is injective.

Lemma 5.5. *If $\mu,\nu \in \mathrm{M}(K)$ with $\int_K f\,\mathrm{d}\mu = \int_K g\,\mathrm{d}\nu$ for all $f \in \mathrm{C}(K)$, then $\mu = \nu$.*

Proof. By passing to $\mu - \nu$ we may suppose that $\nu = 0$. By Exercise 8.a) and standard measure theory it suffices to prove that $\mu(A) = 0$ for each compact G_δ-subset A of K. Given such a set, one can find open subsets O_n of K such that $A = \bigcap_{n\in\mathbb{N}} O_n$. By Urysohn's lemma there are continuous functions $f_n \in \mathrm{C}(K)$ such that $\mathbf{1}_A \leq f_n \leq \mathbf{1}_{O_n}$. Then $f_n \to \mathbf{1}_A$ pointwise and $|f_n| \leq 1$. It follows that $\mu(A) = \lim_{n\to\infty}\langle f_n,\mu\rangle = 0$, by the dominated convergence theorem. □

Note that, using the same approximation technique as in the proof, one can show that for a Baire measure $\mu \in \mathrm{M}(K)$

$$\mu \geq 0 \quad \Leftrightarrow \quad \langle f,\mu\rangle \geq 0 \quad \text{for all } 0 \leq f \in \mathrm{C}(K).$$

Lemma 5.5 has an important consequence.

Proposition 5.6. *Let K, L be compact spaces, let $\varphi : K \to L$ be Baire measurable, and let μ,ν be finite positive Baire measures on K, L, respectively. If*

$$\int_K (f\circ\varphi)\,\mathrm{d}\mu = \int_K f\,\mathrm{d}\nu \qquad \textit{for all } f \in \mathrm{C}(L),$$

then $\nu = \varphi_*\mu$.

By Proposition 5.6, if $\varphi : K \to K$ is Baire measurable, then a probability measure μ on K is φ-invariant if and only if $\langle f \circ \varphi, \mu \rangle = \langle f, \mu \rangle$ for all $f \in \mathrm{C}(K)$.

We have seen that every Baire measure on K gives rise to a bounded linear functional on $\mathrm{C}(K)$ by integration. The following important theorem states the converse.

Theorem 5.7 (Riesz' Representation Theorem). *Let K be a compact space. Then the mapping*

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

is an isometric isomorphism.

For the convenience of the reader, we have included a proof in Appendix D; see also [Rudin (1987), 2.14] or [Lang (1993), IX.2]. Justified by the Riesz theorem, we shall identify $\mathrm{M}(K) = \mathrm{C}(K)'$ and often write μ in place of $\langle \cdot, \mu \rangle$.

Remark 5.8. The common form of Riesz' theorem, for instance in [Rudin (1987)], involves *regular Borel* measures instead of Baire measures. This implies in particular that each finite positive Baire measure has a unique extension to a regular Borel measure. It is sometimes convenient to use this extension, and we shall do it without further reference. However, it is advantageous to work with Baire measures in general since regularity is automatic and the Baire algebra behaves well when one forms infinite products (see Exercise 8). From the functional analytic point of view there is anyway no difference between a Baire measure and its regular Borel extension, since the associated L^p-spaces coincide by Proposition 5.4. For a positive Baire (regular Borel) measure μ on K we therefore abbreviate

$$\mathrm{L}^p(K, \mu) := \mathrm{L}^p(K, \mathrm{Ba}(K), \mu) \qquad (1 \le p \le \infty)$$

and note that in place of $\mathrm{Ba}(K)$ we may write $\mathrm{Bo}(K)$ in this definition. Moreover, as already noted, $\mathrm{C}(K)$ is dense in $\mathrm{L}^p(K, \mu)$ for $1 \le p < \infty$.

For each $0 \le \mu \in \mathrm{M}(K)$ it is easily seen that

$$I_\mu := \Big\{ f \in \mathrm{C}(K) : \|f\|_{\mathrm{L}^1(\mu)} = \int_K |f| \, \mathrm{d}\mu = 0 \Big\}$$

is a closed algebra ideal of $\mathrm{C}(K)$. It is the kernel of the canonical mapping

$$\mathrm{C}(K) \to \mathrm{L}^\infty(K, \mu)$$

that sends $f \in \mathrm{C}(K)$ to its equivalence class modulo μ-almost everywhere equality. By Theorem 4.8 there is a closed set $M \subseteq K$ such that $I = I_M$. The set M is called the **(topological) support** of μ and is denoted by $M =: \mathrm{supp}(\mu)$. Here is a measure theoretic description. (See also Exercise 12.)

Proposition 5.9. *Let $0 \le \mu \in \mathrm{M}(K)$. Then*

$$\mathrm{supp}(\mu) = \big\{ x \in K : \mu(U) > 0 \text{ for each open neighbourhood } U \text{ of } x \big\}. \tag{5.4}$$

Proof. Let $M := \operatorname{supp}(\mu)$ and let L denote the right-hand side of (5.4). Let $x \in M$ and U be an open neighbourhood of x. By Urysohn's lemma there is $f \in \mathrm{C}(K)$ such that $x \in [f \neq 0] \subseteq U$. Since $x \in M$, $f \notin I_M$ and hence $\int_K |f| \, \mathrm{d}\mu \neq 0$. It follows that $\mu(U) > 0$ and, consequently, that $x \in L$.

Conversely, suppose that $x \notin M$. Then by Urysohn's lemma there is $f \in \mathrm{C}(K)$ such that $x \in [f \neq 0]$ and $f = 0$ on M. Hence $f \in I_M$ and therefore $\int_K |f| \, \mathrm{d}\mu = 0$. It follows that $\mu[f \neq 0] = 0$, and hence $x \notin L$. □

A positive measure $\mu \in \mathrm{M}(K)$ has **full support** or is **strictly positive** if $\operatorname{supp}(\mu) = K$. Equivalently, the canonical map $\mathrm{C}(K) \to \mathrm{L}^\infty(K, \mu)$ is isometric (Exercise 11).

For a TDS $(K; \varphi)$, each φ-invariant probability Baire measure $\mu \in \mathrm{M}(K)$ gives rise to an MDS $(K, \mathrm{Ba}(K), \mu; \varphi)$, which for convenience we abbreviate by

$$(K, \mu; \varphi).$$

The following classical theorem says that such measures can always be found.

Theorem 5.10 (Krylov–Bogoljubov). *Let $(K; \varphi)$ be a TDS. Then there is at least one φ-invariant Baire probability measure on K.*

Proof. We postpone the proof of this theorem to Chapter 10, see Theorem 10.2. Note however that under the identification $\mathrm{M}(K) = \mathrm{C}(K)'$ from above, the φ-invariance of μ just means that $T'_\varphi(\mu) = \mu$, i.e., μ is a fixed point of the adjoint of the Koopman operator T_φ on $\mathrm{C}(K)$. Hence, fixed point theorems or similar techniques can be applied. □

The following example shows that there may be many different invariant Baire probability measures, and hence many different MDSs $(K, \mu; \varphi)$ associated with a TDS $(K; \varphi)$.

Example 5.11. Consider the rotation TDS $(\mathbb{T}; a)$ for some $a \in \mathbb{T}$. Obviously, the normalised arc-length measure is invariant. If a is an n^{th} root of unity, then the convex combination of point measures

$$\mu := \frac{1}{n} \sum_{j=1}^{n} \delta_{a^j}$$

is another invariant probability measure, see also Exercise 5.

Example 5.12. Fix $k \in \mathbb{N}$, consider the shift $(\mathscr{W}_k^+; \tau)$. Then each of the Markov measures in Example 5.1.6 is an invariant measure for τ on $\mathrm{Ba}(\mathscr{W}_k^+)$ (by Exercise 8 the product σ-algebra is $\Sigma = \mathrm{Ba}(\mathscr{W}_k^+)$).

TDSs with unique invariant probability measures will be studied in Chapter 10.2.

5.3 Haar Measure and Rotations

Let G be a compact topological group. A nontrivial finite Baire measure on G that is invariant under all left rotations $g \mapsto a \cdot g$, $a \in G$, is called a **Haar measure**. It can be shown that such a Haar measure is essentially unique, i.e., for any two Haar measures μ, ν on G there is a positive constant α such that $\mu = \alpha\nu$. In case of compact groups one sometimes reserves the name Haar measure for the *unique invariant probability* measure on G. It is a fundamental fact that on each compact group there is such a Haar measure, usually denoted by m.

The Haar measure is automatically **right-invariant**, i.e., invariant under right rotations, and **inversion invariant**, i.e., invariant under the inversion mapping $g \mapsto g^{-1}$ of the group. Moreover, m has full support $\operatorname{supp}(\mathrm{m}) = G$, i.e., if $\emptyset \neq U \subseteq G$ is open, then $\mu(U) > 0$. We accept these facts for the moment without proof and return to them in Chapter 14, see Theorem 14.13.

In many concrete cases the Haar measure can be described explicitly.

Example 5.13. 1) The Haar measure $\mathrm{d}z$ on $\mathbb{T}$ is given by

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

for measurable $f : \mathbb{T} \to [0,\infty]$.

2) The Haar measure on a finite discrete group is the counting measure. In particular, the Haar probability measure on the cyclic group $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z} = \{0,1,\ldots,n-1\}$ is given by

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

for every $f : \mathbb{Z}_n \to \mathbb{C}$.

3) For the Haar measure on the dyadic integers $\mathbb{A}_2$, see Exercise 6.

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

With the Haar measure at hand, we can prolong our list of examples from the previous section.

Example 5.14 (Rotation Systems). If G is a compact group then the Haar measure m is (by definition) invariant under the left rotation with $a \in G$. Hence the **rotation system** $(G, \mathrm{m}; a)$ is an MDS with respect to the Haar measure. And, since the Haar measure is invariant also under right rotations, each **right rotation system** $(G, \mathrm{m}; \rho_a)$ is an MDS with respect to the Haar measure.

Example 5.15 (Homogeneous Systems). Let G be a compact group with Haar measure m, and let $H \leq G$ be a closed subgroup of G. Then the factor map $q : G \to G/H$ maps m to a measure $\mu := q_*\mathrm{m}$ on H that is invariant under all left rotations on G/H. In particular, for $a \in G$ the **homogeneous system** $(G/H, \mu; a)$ is an MDS with respect to this measure.

Exercises

1. Show that the Lebesgue measure is invariant under the tent map (Example 5.1.3). Consider the continuous mapping $\varphi : [0,1] \to [0,1]$, $\varphi(x) := 4x(1-x)$. Show that the (finite!) measure $\mu := \mathrm{d}x/2\sqrt{x(1-x)}$ on $[0,1]$ is invariant under φ.

2. Consider the Gauss transformation (Example 5.1.4).

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

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

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

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

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

4. Consider the locally compact Abelian group $G := (0,\infty)$ with respect to multiplication. Show that $\mu := \mathrm{d}x/x$ is invariant with respect to all mappings $x \mapsto a \cdot x$, $a > 0$. (Hence μ is "the" Haar measure on G.) Find an isomorphism $\Phi : \mathbb{R} \to G$ that is also a homeomorphism (on $\mathbb{R}$ we consider the additive structure).

5. Determine all invariant measures for $(\mathbb{T}; a)$ where $a \in \mathbb{T}$ is a root of unity.

6. Consider the compact Abelian group $\mathbb{A}_2$ of dyadic integers (Example 2.11) together with the map

$$\Psi : \{0,1\}^{\mathbb{N}_0} \to \mathbb{A}_2$$

described in Exercise 2.8. Furthermore, let μ be the Bernoulli $(1/2, 1/2)$-measure on $K := \{0,1\}^{\mathbb{N}_0}$. Show that the push-forward measure $\mathrm{m} := \Psi_*\mu$ is the Haar probability measure on $\mathbb{A}_2$. (Hint: For $n \in \mathbb{N}$ consider the projection

$$\pi_n : \mathbb{A}_2 \to \mathbb{Z}_{2^n}$$

onto the n^{th} coordinate. This is a homomorphism, and let $H_n := \ker(\pi_n)$ be its kernel. Show that $\mathrm{m}(a + H_n) = 1/2^n$ for any $a \in \mathbb{A}_2$.)

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

8. Let K be a compact space with its Baire algebra $\mathrm{Ba}(K)$.

a) Show that $\mathrm{Ba}(K)$ is generated by the compact G_δ-sets.

b) Show that $\mathrm{Ba}(K) = \mathrm{Bo}(K)$ if K is metrisable.

c) Show that if $A \subseteq O$ with $A \subseteq K$ closed and $O \subseteq K$ open, there exists $A \subseteq A' \subseteq O' \subseteq O$ with $A', O' \in \mathrm{Ba}(K)$, A' closed and O' open.

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

9 (Perron's Theorem). Let A be a column-stochastic $d \times d$-matrix, i.e., all $a_{ij} \geq 0$ and $\sum_{i=1}^d a_{ij} = 1$ for all $j = 1, \ldots, d$. In this exercise we show: *There is a nonzero vector $p \geq 0$ such that $Ap = p$.*

a) Let $\mathbf{1} = (1, 1, \cdots, 1)^t$ and show that $\mathbf{1}^t A = \mathbf{1}$. Conclude that there is a vector $0 \neq v \in \mathbb{R}^d$ such that $Av = v$.

b) Show that $\|Ax\|_1 \leq \|x\|_1$, where $\|x\|_1 := \sum_{j=1}^d |x_j|$ for $x \in \mathbb{R}^d$.

c) Show that for $\lambda > 1$ the matrix $\lambda \mathrm{I} - A$ is invertible and

$$(\lambda - A)^{-1} x = \sum_{n=0}^{\infty} \frac{A^n x}{\lambda^{n+1}} \qquad (x \in \mathbb{R}^d).$$

d) Let $y := |v| = (|v_1|, \ldots, |v_d|)$, where v is the vector from a). Employing c), show that

$$\frac{\|v\|_1}{\lambda - 1} \leq \left\|(\lambda - A)^{-1} y\right\|_1 \qquad (\lambda > 1).$$

e) Fix a sequence $\lambda_n \searrow 1$ and consider

$$y_n := \left\|(\lambda_n - A)^{-1} y\right\|_1^{-1} (\lambda_n - A)^{-1} y.$$

By compactness one may suppose that $y_n \to p$ as $n \to \infty$. Show that $p \geq 0$, $\|p\|_1 = 1$ and $Ap = p$.

10. Let X be an uncountable set, take $p \notin X$ and let $K := X \cup \{p\}$ be the *one-point compactification* of X. To be precise, a set $O \subseteq K$ is open if $O \subseteq X$, or if $p \in O$ and $X \setminus O$ is finite.

a) Show that this indeed defines a compact topology on K.

b) Show that if $f \in \mathrm{C}(K)$, then $f(x) = f(p)$ for all but countably many $x \in X$.

c) Prove that the singleton $\{p\}$ is a Borel, but not a Baire set in K.

11. Let K be a compact space and $0 \leq \mu \in \mathrm{M}(K)$. Show that the following assertions are equivalent.

(i) $\operatorname{supp}(\mu) = K$.

(ii) $\int_K |f| \, \mathrm{d}\mu = 0$ implies that $f = 0$ for every $f \in \mathrm{C}(K)$.

(iii) $\|f\|_{\mathrm{L}^\infty(K,\mu)} = \|f\|_\infty$ for every $f \in \mathrm{C}(K)$.

12. Let K be a compact space, and let $\mu \in \mathrm{M}(K)$ be a regular Borel probability measure on K. Show that

$$\operatorname{supp}(\mu) = \bigcap \{A : A \subseteq K \text{ closed}, \quad \mu(A) = 1\}$$

and that $\mu(\operatorname{supp}(\mu)) = 1$. So $\operatorname{supp}(\mu)$ is the smallest closed subset of K with full measure.

Chapter 6
Recurrence and Ergodicity

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

6.1 The Measure Algebra and Invertible MDS

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

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

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

$$[A] \subseteq [B] \overset{\text{Def.}}{\Longleftrightarrow} \mu_{\mathrm{X}}(B \setminus A) = 0, \quad [A] \cap [B] := [A \cap B], \quad [A] \cup [B] := [A \cup B]$$

and so on. Moreover, the measure μ induces a (σ-additive) map

$$\Sigma(\mathrm{X}) \to [0, \infty], \qquad [A] \mapsto \mu(A).$$

As for equivalence classes of functions, we normally do not distinguish notationally between a set A in Σ and its equivalence class $[A]$ in $\Sigma(\mathrm{X})$.

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

$$d_{\mathrm{X}}(A,B) := \mu(A \triangle B) = \|\mathbf{1}_A - \mathbf{1}_B\|_1 \qquad (A,B \in \Sigma(\mathrm{X})).$$

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

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

Proof. This is just Lemma B.18 from Appendix B. Its proof is standard measure theory using Dynkin systems. □

Suppose that X and Y are measure spaces and $\varphi : X \to Y$ is a measurable mapping, and consider the push-forward measure $\varphi_*\mu_{\mathrm{X}}$ on Y. If φ preserves null-sets, i.e., if

$$\mu_{\mathrm{Y}}(A) = 0 \quad \Rightarrow \quad \mu_{\mathrm{X}}(\varphi^{-1}A) = (\varphi_*\mu_{\mathrm{X}})(A) = 0 \qquad \text{for all } A \in \Sigma_{\mathrm{Y}},$$

the mapping φ induces a mapping φ^* between the measure algebras defined by

$$\varphi^* : \Sigma(\mathrm{Y}) \to \Sigma(\mathrm{X}), \quad [A] \mapsto \left[\varphi^{-1}A\right].$$

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

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

All this is in particular applicable when φ is measure-preserving. In this case and when both μ_{X} and μ_{Y} are probability measures, the induced mapping $\varphi^* : \Sigma(\mathrm{Y}) \to \Sigma(\mathrm{X})$ is an isometry, since

$$d_{\mathrm{X}}(\varphi^*A, \varphi^*B) = \mu_{\mathrm{X}}([\varphi \in A] \triangle [\varphi \in B]) = \mu_{\mathrm{X}}[\varphi \in (A \triangle B)] = \mu_{\mathrm{Y}}(A \triangle B) = d_{\mathrm{Y}}(A,B)$$

for $A,B \in \Sigma_{\mathrm{Y}}$.

We are now able to define invertibility of an MDS.

Definition 6.2. An MDS $(\mathrm{X};\varphi)$ is called **invertible** if the induced map φ^* on $\Sigma(\mathrm{X})$ is bijective.

We now relate this notion of invertibility to a possibly more intuitive one, based on the following definition.

Definition 6.3. Let X and Y be measure spaces, and let $\varphi : X \to Y$ be measure-preserving. A measurable map $\psi : Y \to X$ is called an **essential inverse** to φ if

$$\varphi \circ \psi = \mathrm{id} \quad \mu_{\mathrm{Y}}\text{-a.e.} \quad \text{and} \quad \psi \circ \varphi = \mathrm{id} \quad \mu_{\mathrm{X}}\text{-a.e.}$$

The mapping φ is called **essentially invertible** if it has an essential inverse.

Exercise 2 helps with computations with essential inverses. In particular, it follows that if $\mathrm{X} = \mathrm{Y}$ and ψ is an essential inverse of φ, then ψ^n is an essential inverse to φ^n for every $n \in \mathbb{N}$. See also Exercise 3 for an equivalent characterisation of essential invertibility.

Lemma 6.4. *An essential inverse is unique up to equality almost everywhere. If $\psi : Y \to X$ is an essential inverse to a measure-preserving map $\varphi : X \to Y$, then ψ is also measure-preserving. Moreover, the induced map $\varphi^* : \Sigma(\mathrm{Y}) \to \Sigma(\mathrm{X})$ is bijective with inverse $(\varphi^*)^{-1} = \psi^*$.*

Proof. Let ψ_1, ψ_2 be essential inverses of φ. Observe that

$$\varphi^{-1}[\,\psi_1 \neq \psi_2\,] \subseteq [\,\psi_1 \circ \varphi \neq \mathrm{id}\,] \cup [\,\psi_2 \circ \varphi \neq \mathrm{id}\,].$$

By assumption and since φ is measure-preserving, it follows that $[\,\Psi_1 \neq \Psi_2\,]$ is a null-set. Now suppose that ψ is an essential inverse of φ. Then, since $\mu_{\mathrm{Y}} = \varphi_* \mu$,

$$\mu_{\mathrm{Y}}(\psi^{-1}(A)) = \mu(\varphi^{-1}\psi^{-1}(A)) = \mu\left((\psi \circ \varphi)^{-1} A\right) = \mu(A)$$

since $(\psi \circ \varphi)^{-1} A = A$ modulo a μ-null set. The last assertion follows analogously. □

A direct consequence of Lemma 6.4 is the following.

Corollary 6.5. *An MDS $(\mathrm{X}; \varphi)$ is invertible if φ is essentially invertible.*

Remark 6.6. The converse of Corollary 6.5 does not hold in general: Take $X = \{0,1\}$, $\Sigma = \{\emptyset, X\}$ and $\varphi(x) = 1$ for $x = 0, 1$. However, if the probability space is a so-called *standard Borel space*, then invertibility of the MDS (i.e., invertibility of φ^*) is equivalent to the essential invertibility of φ, see Chapter 12.

Examples 6.7. The baker's transformation and every group rotation is invertible (use Corollary 6.5). The tent map and the doubling map systems are not invertible (Exercise 4). A two-sided Bernoulli system is invertible but a one-sided is not (Exercise 5).

6.2 Recurrence

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

Definition 6.8. Let $(\mathrm{X};\varphi)$ be an MDS.

a) A set $A \in \Sigma_{\mathrm{X}}$ is called **recurrent** if almost every point of A returns to A after some time, or equivalently,

$$A \subseteq \bigcup_{n\geq 1} \varphi^{*n} A \tag{6.1}$$

in the measure algebra $\Sigma(\mathrm{X})$.

b) A set $A \in \Sigma_{\mathrm{X}}$ is **infinitely recurrent** if almost every point of A returns to A infinitely often, or equivalently,

$$A \subseteq \bigcap_{n\geq 1} \bigcup_{k\geq n} \varphi^{*k} A \tag{6.2}$$

in the measure algebra $\Sigma(\mathrm{X})$.

Here is an important characterisation.

Lemma 6.9. *Let $(\mathrm{X};\varphi)$ be an MDS. Then the following statements are equivalent.*

(i) *Every $A \in \Sigma_{\mathrm{X}}$ is recurrent.*

(ii) *Every $A \in \Sigma_{\mathrm{X}}$ is infinitely recurrent.*

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

Proof. The implication (ii)$\Rightarrow$(i) is clear. For the converse take $A \in \Sigma_{\mathrm{X}}$ and apply φ^* to (6.1) to obtain $\varphi^* A \subseteq \bigcup_{n\geq 2} \varphi^{*n} A$. Putting this back into (6.1) we obtain

$$A \subseteq \bigcup_{n\geq 1} \varphi^{*n} A = \varphi^* A \cup \bigcup_{n\geq 2} \varphi^{*n} A \subseteq \bigcup_{n\geq 2} \varphi^{*n} A.$$

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

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

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

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

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

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

Theorem 6.10 (Poincaré). *Every MDS $(\mathrm{X};\varphi)$ is (infinitely) recurrent, i.e., every set $A \in \Sigma_{\mathrm{X}}$ is infinitely recurrent.*

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

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

This means that the sets $(\varphi^{*n}A)_{n\in\mathbb{N}_0}$ are (essentially) disjoint. (A set A with this property is called *wandering*.) On the other hand, all sets $\varphi^{*n}A$ have the same measure, and since μ_X is finite, this measure must be zero. Hence $A = \emptyset$, and the MDS $(\mathrm{X};\varphi)$ is infinitely recurrent by the implication (iii)$\Rightarrow$(i) of Lemma 6.9. □

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

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

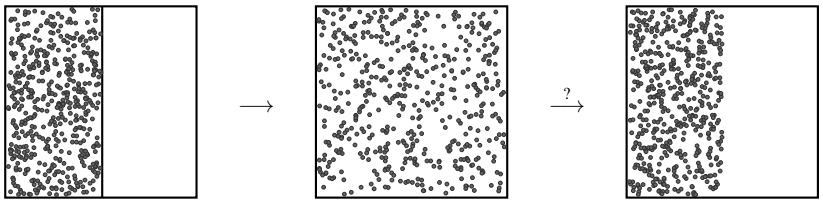


Fig. 6.1 What happens after removing the wall?

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

The mathematical model of this experiment is that of a Markov shift over $L := \{0,\dots,n\}$. The "state sequence" $(j_0, j_1, j_2 \dots)$ records the number $j_m \in L$ of balls in urn 1 after the m^{th} step. To determine the transition matrix P note that if there are $i \geq 1$ balls in urn 1 then the probability to decrease its number to $i-1$ is i/n; and if $i < n$ then the probability to increase the number from i to $i+1$ is $1-i/n$. Hence we have $P = (p_{ij})_{0\leq i,j\leq n}$ with

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

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

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

and it is easy to show that $p = (p_0, \ldots, p_n)$ is indeed a fixed probability row vector, i.e., that $pP = p$ (Exercise 6). (Since P is irreducible, Perron's theorem implies that p is actually the *unique* fixed probability vector.)

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

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

Let $(\mathrm{X}; \varphi)$ be an MDS, fix $A \in \Sigma_{\mathrm{X}}$ and define

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

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

$$A_n := B_n \cap A \qquad (n \in \mathbb{N}), \tag{6.3}$$

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

Lemma 6.11. *Let $(\mathrm{X}; \varphi)$ be an MDS, $\mathrm{X} = (X, \Sigma, \mu)$, and let $A, B \in \Sigma$. Then*

$$\mu(B) = \sum_{k=1}^{n} \mu\left(A \cap \bigcap_{j=1}^{k-1} \varphi^{*j} A^{\mathrm{c}} \cap \varphi^{*k} B\right) \quad + \quad \mu\left(\bigcap_{k=0}^{n-1} \varphi^{*k} A^{\mathrm{c}} \cap \varphi^{*n} B\right) \tag{6.4}$$

for $n \geq 1$.

Proof. Using the φ-invariance of μ we write

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

and this is (6.4) when $n = 1$ and with $\cap_{j=1}^{0} \varphi^{*j} A^{\mathrm{c}} =: X$. Doing the same again with the second summand yields

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

and this is (6.4) for $n = 2$. An induction argument concludes the proof. □

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

$$\Sigma_A := \{B \subseteq A : B \in \Sigma\} \subseteq \mathcal{P}(A)$$

and thereon the induced probability measure μ_A defined by

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

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

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

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

Theorem 6.12. *Let* $(\mathrm{X}; \varphi)$ *be an MDS,* $\mathrm{X} = (X, \Sigma, \mu)$*, and let* $A \in \Sigma$ *with* $\mu(A) > 0$*. Then*

$$\int_A n_A \, \mathrm{d}\mu_A = \frac{\mu\left(\bigcup_{n \geq 0} \varphi^{*n} A\right)}{\mu(A)}. \tag{6.5}$$

Proof. We specialise $B = X$ in Lemma 6.11. Note that

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

since A is recurrent, by Poincaré's theorem. Hence by (6.4) in Lemma 6.11 we obtain

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

Letting $n \to \infty$ yields

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

and interchanging the order of summation we arrive at

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

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

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

$$\bigcup_{n=0}^{\infty} \tau^{*n} A = X, \tag{6.6}$$

then we could conclude from (6.5) in Theorem 6.12 that the expected waiting time for returning to A is $1/\mu(A) = 2^{100}$ steps. Clearly, our lifetimes and that of the universe would not suffice for the waiting, even if we do one step every millisecond.

However, this reasoning builds on (6.6), a property not yet established. This will be done in the next section.

6.3 Ergodicity

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

Lemma 6.13. *Let $(\mathrm{X};\varphi)$ be an MDS and $A \in \Sigma_{\mathrm{X}}$. Then the following assertions are equivalent.*

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

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

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

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

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

Definition 6.14. An MDS $(\mathrm{X};\varphi)$ is called **ergodic** if every invariant set is essentially equal either to $\emptyset$ or to X.

Since invariant sets are the fixed points of φ^* in the measure algebra, a system is ergodic if and only if φ^* on $\Sigma(\mathrm{X})$ has only the trivial fixed points $\emptyset$ and X.

In contrast to minimality in the topological case (Theorem 3.5), an MDS need not have ergodic subsystems: just consider $X = [0,1]$ with the Lebesgue measure and $\varphi = \mathrm{id}_X$, the identity. Another difference to the topological situation is that the presence of a nontrivial invariant set A does not only lead to a *restriction* of the system (to A), but to a decomposition $X = A \cup A^{\mathrm{c}}$. So ergodicity could also be termed *indecomposability*.

The following result characterises ergodicity.

Lemma 6.15. *Let* $(\mathrm{X};\varphi)$ *be an MDS. Then the following statements are equivalent.*

(i) *The MDS* $(\mathrm{X};\varphi)$ *is ergodic.*

(ii) *For every* $\emptyset \neq A \in \Sigma(\mathrm{X})$ *one has*

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

(iii) *For every* $\emptyset \neq A \in \Sigma(\mathrm{X})$ *one has*

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

(iv) *For every* $\emptyset \neq A,B \in \Sigma(\mathrm{X})$ *there is* $n \geq 1$ *such that*

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

Proof. (i)$\Rightarrow$(ii): For a set $A \in \Sigma_{\mathrm{X}}$ and $n \geq 0$ the set

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

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

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

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

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

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

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

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

The implication (i)$\Rightarrow$(ii) of Lemma 6.15 says that in an ergodic system each set of positive measure is visited infinitely often by almost every point. Note also that (iv) is an analogue of condition b) in Proposition 2.29 for MDSs.

Let us now look at some examples. The implication (iv)$\Rightarrow$(i) of Lemma 6.15 combined with the following result shows that a Bernoulli shift is ergodic.

Proposition 6.16. *Let* $(\mathscr{W}_k^+, \Sigma, \mu; \tau) = B(p_0, \dots, p_{k-1})$ *be a Bernoulli shift. Then*

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

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

Proof. We use the notation of Example 5.1.5. Let $\mathcal{E}$ denote the algebra of cylinder sets on $\mathscr{W}_k^+ = L^{\mathbb{N}_0}$. If $B \in \mathcal{E}$ then $B = B_0 \times \prod_{k\geq n_0} L$ for some $n_0 \in \mathbb{N}$ and $B_0 \subseteq L^{n_0}$. Then $\tau^{*n} A = L^n \times A$ and

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

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

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

(Exercise 1), one has $\mu(\tau^{*n} A \cap B_m) \to \mu(\tau^{*n} A \cap B)$ as $m \to \infty$ uniformly in $n \in \mathbb{N}$. Hence one can interchange limits to obtain

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

as claimed. □

Proposition 6.16 actually says that the Bernoulli shift is *strongly mixing*, a property that will be studied in Chapter 9. Here are other examples of ergodic systems.

Examples 6.17. a) *A Markov shift with an irreducible transition matrix is ergodic.* This is Theorem 8.14 below.

b) *Every minimal rotation on a compact group is ergodic.* For the special case of the torus, this is Proposition 7.15, the general case is treated in in Theorem 10.13.

c) *The Gauss map (see page 63) is ergodic*. A proof can be found in [Einsiedler and Ward (2011), Sect. 3.2].

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

Corollary 6.18 (Kac). *Let* $(\mathrm{X}; \varphi)$ *be an ergodic MDS,* $\mathrm{X} = (X, \Sigma, \mu)$*, and* $A \in \Sigma$ *with* $\mu(A) > 0$*. Then for the* **expected return time** *to A one has*

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

Proof. Since the system is ergodic, the implication (i) ⇒ (iii) of Lemma 6.15 shows that $X = \bigcup_{n\geq 0} \varphi^{*n} A$. Hence the claim follows from Theorem 6.12. □

Let us return to our ball experiment. The transition matrix P described on page 80 is indeed irreducible, whence by Example 6.17.a) the corresponding Markov shift is ergodic. Therefore we can apply Kac's theorem concluding that indeed the expected return time to our initial state A is 2^{100} for our choice $n = 100$.

Supplement: Induced Transformation and Kakutani–Rokhlin Lemma

We conclude this chapter with two results of some importance in ergodic theory. However, we shall not immediately use them and one may skip this section on a first reading.

The Induced Transformation

Let $(\mathrm{X};\varphi)$ be an MDS, $\mathrm{X} = (X,\Sigma,\mu)$, and let $A \in \Sigma$ with $\mu(A) > 0$. Define the **induced transformation** φ_A by

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

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

Theorem 6.19. *Let $(\mathrm{X};\varphi)$ be an MDS, $X = (X,\Sigma,\mu)$, and let $A \in \Sigma$ be a set of positive measure. Then the induced transformation φ_A is measurable with respect to Σ_A and preserves the induced measure μ_A.*

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

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

which shows that φ_A is indeed Σ_A-measurable. To see that φ_A preserves μ, we use Lemma 6.11. Note that since $B \subseteq A$,

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

and

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

Since μ is finite, $\sum_n \mu(B_n) < \infty$ and hence $\mu(A^{\mathrm{c}} \cap B_n) \to 0$ as $n \to \infty$. Letting $n \to \infty$ in (6.4) we obtain

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

and that was to be proved. □

By Exercise 8, if the original MDS is ergodic, then $(A, \Sigma_A, \mu_A; \varphi_A)$ is ergodic, too.

The Kakutani–Rokhlin Lemma

The following theorem is due to Kakutani [Kakutani (1943)] and Rokhlin [Rohlin (1948)]. Our presentation is based on [Rosenthal (1988)].

Theorem 6.20 (Kakutani–Rokhlin). *Let $(\mathrm{X}; \varphi)$ be an ergodic MDS, let $A \in \Sigma$ with $\mu(A) > 0$ and $n \in \mathbb{N}$. Then there is a set $B \in \Sigma$ such that*

$$B, \varphi^* B, \ldots, \varphi^{*(n-1)} B \quad \textit{are pairw. disjoint} \quad \textit{and} \quad \mu\Big(\bigcup_{j=0}^{n-1} \varphi^{*j} B\Big) \geq 1 - n\mu(A).$$

Before we give the proof, let us think about the meaning of Theorem 6.20. A **Rokhlin tower** of height $n \in \mathbb{N}$ is a finite sequence of *pairwise disjoint* sets of the form

$$B,\ \varphi^* B,\ \varphi^{*2} B,\ \ldots,\ \varphi^{*(n-1)} B,$$

and B is called the **ceiling** of the tower. Since

$$\varphi^{*k} B \cap \varphi^{*j} B = \varphi^{j*}(\varphi^{*(k-j)} B \cap B) \qquad (1 \leq j < k \leq n-1),$$

in order to have a Rokhlin tower it suffices to show that $\varphi^{*j} B \cap B = \emptyset$ for $j = 1, \ldots, n-1$.

Since the sets in a Rokhlin tower all have the same measure and μ is a probability measure, one has the upper bound

$$\mu(B) \leq \frac{1}{n}$$

for the ceiling B of a Rokhlin tower of height n. On the other hand, the second condition in the Kakutani–Rokhlin result is equivalent to

$$\mu(B) \geq \frac{1}{n} - \mu(A).$$

So if $\mu(A)$ is small, then $\mu(B)$ is close to its maximal value $1/n$.

Corollary 6.21. *Let $(\mathrm{X}; \varphi)$ be an ergodic MDS such that there are sets with arbitrarily small positive measure. Then for each $\varepsilon > 0$ and $n \in \mathbb{N}$ there is a Rokhlin tower $B, \varphi^* B, \varphi^{*2} B, \ldots, \varphi^{*(n-1)} B$ such that*

$$\mu\Big(\bigcup_{j=0}^{n-1}\varphi^{*j}B\Big) > 1-\varepsilon.$$

Let us turn to the proof of Theorem 6.20. Define

$$A_0 := A, \qquad A_{k+1} := \varphi^* A_k \cap A^{\mathrm{c}} \qquad (k \geq 0)$$

Then $A_n = \varphi^{*n}A \cap \bigcap_{j=0}^{n-1}\varphi^{*j}A^{\mathrm{c}}$ is the set of points in A^{c} that hit A for the first time in the n^{th} step. Then $(A_j)_{j\geq 0}$ is a pairwise disjoint sequence of sets. Moreover, we note that $\bigcup_{n\geq 1} A_n = A^{\mathrm{c}}$. This follows from

$$A_0 \cup \bigcup_{n\geq 1} A_n = \bigcup_{j\geq 0}\varphi^{*j}A = X$$

in $\Sigma(\mathrm{X})$ by ergodicity. Now, define the set

$$B_j := \bigcup_{k\geq 1} A_{nk+j} \qquad (j = 0,\ldots,n-1).$$

Then the B_j are pairwise disjoint, and $B_1 \cup \cdots \cup B_{n-1} = A^{\mathrm{c}}$. We claim that

$$\varphi^{*j}B_0 = B_j \cup C_j \quad \text{for some} \quad C_j \subseteq A \cup \varphi^* A \cup \cdots \cup \varphi^{*(j-1)}A \tag{6.7}$$

for $j = 0,\ldots,n-1$. This is an easy induction argument, the induction step being

$$\begin{aligned}\varphi^{*(j+1)}B_0 &= \varphi^*(B_j \cup C_j) = \bigcup_{k\geq 1}\varphi^* A_{nk+j} \cup \varphi^* C_j \\ &= \bigcup_{k\geq 1} A_{nk+j+1} \cup \bigcup_{k\geq 1}(\varphi^* A_{nk+j} \cap A) \cup \varphi^* C_j = B_{j+1} \cup C_{j+1}\end{aligned}$$

with

$$C_{j+1} = \varphi^* C_j \cup \bigcup_{k\geq 1}(\varphi^* A_{nk+j} \cap A) \subseteq \varphi^*\Big(\bigcup_{k=0}^{j-1}\varphi^{*k}A\Big) \cup A = \bigcup_{k=0}^{j}\varphi^{*k}A.$$

From (6.7) it follows immediately that

$$B_0 \cap \varphi^{*j}B_0 = \emptyset \qquad (j = 1,\ldots,n-1)$$

and hence B_0 is the ceiling of a Rokhlin tower of height n. Finally, (6.7) implies that

$$\bigcup_{j=0}^{n-1}\varphi^{*j}B_0 \supseteq \bigcup_{j=0}^{n-1} B_j = \bigcup_{k\geq n} A_k,$$

whence

$$\left(\bigcup_{j=0}^{n-1} \varphi^{*j} B_0\right)^{\mathrm{c}} \subseteq \bigcup_{k=0}^{n-1} A_k \subseteq \bigcup_{k=0}^{n-1} \varphi^{*k} A.$$

Hence we obtain

$$\mu\left(\bigcup_{j=0}^{n-1} \varphi^{*j} B_0\right) \geq 1 - \mu\Big(\bigcup_{k=0}^{n-1} \varphi^{*k} A\Big) \geq 1 - n\mu(A).$$

Exercises

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

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

defines a complete metric on the measure algebra $\Sigma(\mathrm{X})$. Show further that

a) $d_{\mathrm{X}}(A \cap B, C \cap D) \leq d_{\mathrm{X}}(A,C) + d_{\mathrm{X}}(B,D)$;

b) $d_{\mathrm{X}}(A^{\mathrm{c}}, B^{\mathrm{c}}) = d_{\mathrm{X}}(A,B)$;

c) $d_{\mathrm{X}}(A \setminus B, C \setminus D) \leq d_{\mathrm{X}}(A,C) + d_{\mathrm{X}}(B,D)$;

d) $d_{\mathrm{X}}(A \cup B, C \cup D) \leq d_{\mathrm{X}}(A,C) + d_{\mathrm{X}}(B,D)$;

e) $|\mu(A) - \mu(B)| \leq d_{\mathrm{X}}(A,B)$.

In particular, the mappings

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

are continuous with respect to d_{X}. Show also that

$$(A_n)_n \subseteq \Sigma(\mathrm{X}), \quad A_n \nearrow A \quad \Rightarrow \quad d_{\mathrm{X}}(A_n, A) \searrow 0.$$

2. Let $\mathrm{X}_0, \mathrm{X}_1, \mathrm{X}_2$ be probability measure spaces. Let

$$f_0, g_0 : X_0 \to X_1 \quad \text{and} \quad f_1, g_1 : X_1 \to X_2$$

be measurable such that $f_0 = g_0$ and $f_1 = g_1$ almost everywhere and f_0 preserves null sets. Prove the following assertions:

a) g_0 preserves null sets, too;

b) $f_1 \circ f_0 = g_1 \circ g_0$ almost everywhere;

c) $f_0^* = g_0^*$ as mappings $\Sigma(\mathrm{X}_1) \to \Sigma(\mathrm{X}_0)$.

3. Let X and Y be probability spaces, and let $\varphi : X \to Y$ be measure-preserving. Show that the following assertions are equivalent.

(i) φ is essentially invertible;

(ii) There are sets $A \in \Sigma_X$, $A' \in \Sigma_Y$ such that $\mu_X(A) = 1 = \mu_Y(A')$, and φ maps A bijectively onto A' with measurable inverse.

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

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

6. Show that $pP = p$, where P, p are the transition matrix and the probability vector defined above in connection with the ball experiment (see page 80).

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

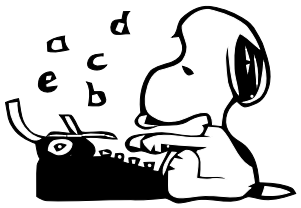


Fig. 6.2 Philosopher at work.

Show also that if Snoopy had been typing since eternity, he almost surely already would *have typed* the book infinitely often.

8. Let $(X;\varphi)$ be an ergodic MDS, $X = (X, \Sigma, \mu)$, and let $A \in \Sigma$ with $\mu(A) > 0$. Show that the induced transformation φ_A is ergodic, too. (Hint: Let the sets $(A_n)_{n\geq 1}$ be defined as in (6.3) and let $B \subseteq A$ with $\varphi_A^* B = B$. Show by induction over $n \in \mathbb{N}$ that $\varphi^{*n} B \cap A \subseteq B$; e.g., use the identity $\varphi^{*n} B \cap A_j = \varphi^{*j}(\varphi^{*n-j} B \cap A) \cap A_j$ for $j < n$.)

Chapter 7
The Banach Lattice L^p and the Koopman Operator

Let X and Y be two measure spaces and suppose that we are given a measurable map $\varphi : \mathrm{Y} \to X$. The **Koopman operator** $T = T_\varphi$, defined by

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

then maps a measurable function f on X to a measurable functions $T_\varphi f$ on Y. If, in addition, φ respects null sets, i.e., we have

$$\mu_{\mathrm{X}}(A) = 0 \quad \Rightarrow \quad \mu_{\mathrm{Y}}[\varphi \in A] = 0 \qquad (A \in \Sigma_{\mathrm{X}}),$$

then

$$f = g \quad \mu_{\mathrm{X}}\text{-almost everywhere} \quad \Rightarrow \quad T_\varphi f = T_\varphi g \quad \mu_{\mathrm{Y}}\text{-almost everywhere}$$

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

$$T_\varphi[f] := \left[T_\varphi f\right] = [f \circ \varphi].$$

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

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

Note that

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

So T_φ acts on equivalence classes of characteristic functions as φ^* acts on the measure algebra $\Sigma(\mathrm{X})$ (cf. Chapter 6.1).

Suppose now in addition that φ is measure-preserving, i.e., $\varphi_* \mu_{\mathrm{Y}} = \mu_{\mathrm{X}}$. Then for any measurable function f on X and $1 \leq p < \infty$ we obtain (by Appendix B.5)

$$\|T_\varphi f\|_p^p = \int_Y |f\circ\varphi|^p \,\mathrm{d}\mu_Y = \int_X |f|^p \,\mathrm{d}(\varphi_*\mu_Y) = \int_X |f|^p \,\mathrm{d}\mu_X = \|f\|_p^p .$$

This shows that

$$T_\varphi : L^p(X) \to L^p(Y) \qquad (1 \le p < \infty)$$

is a linear isometry. The same is true for $p = \infty$ (Exercise 1). Moreover, by a similar computation

$$\int_Y T_\varphi f = \int_X f \qquad (f \in L^1(X)). \tag{7.1}$$

In Chapter 4 we associated with a TDS $(K;\varphi)$ the commutative C^*-algebra $C(K)$ and the Koopman operator $T = T_\varphi$ acting on it. In the case of an MDS $(X;\varphi)$ it is natural to investigate the associated Koopman operator T_φ on each of the spaces $L^p(X)$, $1 \le p \le \infty$.

For the case $p = \infty$ we note that $L^\infty(X)$ is, like $C(K)$ in Chapter 4, a commutative C^*-algebra, and Koopman operators are algebra homomorphisms. (This fact will play an important role in Chapter 12.) For $p \ne \infty$, however, the space L^p is in general not closed under multiplication of functions, so we have to look for a new structural element preserved by the operator T_φ; this will be the *lattice structure*.

Since we do not expect order theoretic notions to be common knowledge, we shall introduce the main abstract concepts in the next section and then proceed with some more specific facts for L^p-spaces. (The reader familiar with Banach lattices may skip these parts and proceed directly to Section 7.3.) Our intention here, however, is mostly terminological. This means that Banach lattice theory provides a convenient framework to address certain features common to L^p- as well as $C(K)$-spaces, but no deep results from the theory are actually needed. The unexperienced reader may without harm stick to these examples whenever we use abstract terminology. For a more detailed account we refer to the monograph [Schaefer (1974)].

7.1 Banach Lattices and Positive Operators

A **lattice** is a partially ordered set $(V,\le)$ such that

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

exist for all $x,y \in V$. (Here $\sup A$ denotes the least upper bound or **supremum** of the set $A \subseteq V$, if it exists. Likewise, $\inf A$ is the greatest lower bound, or **infimum** of A.) A lattice is called **complete** if every nonempty subset has a supremum and an infimum.

A subset $D \subseteq V$ of a lattice $(V,\le)$ is called $\vee$**-stable** ($\wedge$**-stable**) if $a,b \in D$ implies $a \vee b \in D$ ($a \wedge b \in D$). A subset that is $\vee$-stable and $\wedge$-stable is a **sublattice**. If V is a lattice and $A \subseteq V$ is a nonempty set, then A has a supremum if and only if the $\vee$-stable set

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

has a supremum, and in this case these suprema coincide. If $(V, \leq)$ and $(W, \leq)$ are lattices, then a map $\Theta : V \to W$ such that

$$\Theta(x \vee y) = (\Theta x) \vee (\Theta y) \quad \text{and} \quad \Theta(x \wedge y) = (\Theta x) \wedge (\Theta y)$$

is called a **homomorphism** of lattices. If Θ is bijective, then also Θ^{-1} is a homomorphism, and Θ is called a **lattice isomorphism**.

Examples 7.1. 1) The prototype of a complete lattice is the extended real line $\overline{\mathbb{R}} := [-\infty, \infty]$ with the usual order. It is lattice isomorphic with the closed unit interval $[0,1]$ via the isomorphism $x \mapsto \frac{1}{2} + \frac{1}{\pi}\arctan(x)$.

2) For a measure space $\mathrm{X} = (X, \Sigma, \mu)$ we denote by

$$\mathrm{L}^0 := \mathrm{L}^0(\mathrm{X}; \overline{\mathbb{R}})$$

the set of all equivalence classes of measurable functions $f : \mathrm{X} \to [-\infty, \infty]$, and define a partial order by

$$[f] \leq [g] \quad \overset{\text{Def.}}{\Leftrightarrow} \quad f \leq g \quad \mu_{\mathrm{X}}\text{-almost everywhere.}$$

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

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

(To be very precise, this means that the supremum $f \vee g$ is represented by the pointwise supremum of representatives of the equivalence classes f and g.) Analogously we see that in L^0 the supremum (infimum) of every *sequence* exists and is represented by the pointwise supremum of representatives. However, it is even true that the lattice $\mathrm{L}^0(\mathrm{X}; \overline{\mathbb{R}})$ is *complete* (Exercise 10).

3) The subsets $\mathrm{L}^0(\mathrm{X}; \mathbb{R})$ and $\mathrm{L}^p(\mathrm{X}; \mathbb{R})$ for $1 \leq p \leq \infty$ are sublattices of $\mathrm{L}^0(\mathrm{X}; \overline{\mathbb{R}})$.

4) If X is a finite measure space, then the measure algebra $\Sigma(\mathrm{X})$ is a lattice (with respect to the obvious order) and

$$\Sigma(\mathrm{X}) \to \mathrm{L}^\infty(\mathrm{X}), \qquad [A] \mapsto [\mathbf{1}_A] \qquad (A \in \Sigma_{\mathrm{X}})$$

is an injective lattice homomorphism. By Exercise 13, the lattice $\Sigma(\mathrm{X})$ is complete.
The name "measure algebra" derives from the fact that $\Sigma(\mathrm{X})$ is not just a (complete) lattice, but a *Boolean algebra*, cf. Section 12.2 below.

The lattice $\mathrm{L}^0(\mathrm{X}; \overline{\mathbb{R}})$ does not show any reasonable algebraic structure since the presence of the values $\pm\infty$ prevents a global definition of a sum. This problem vanishes when one restricts to the sublattice $\mathrm{L}^p(\mathrm{X}; \mathbb{R})$ for $p = 0$ or $1 \leq p \leq \infty$, which is

also a real vector space. Moreover, the lattice and vector space structures are connected by the rules

$$f \le g \quad \Rightarrow \quad f+h \le g+h, \quad c\cdot f \le c\cdot g, \quad -g \le -f \tag{7.2}$$

for all $f,g,h \in \mathrm{L}^p(\mathrm{X};\mathbb{R})$ and $c \ge 0$. A lattice V which is also a real vector space such that (7.2) holds is called a **(real) vector lattice**. In any vector lattice one can define

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

which—in our space $\mathrm{L}^p(\mathrm{X};\mathbb{R})$—all coincide with the respective pointwise operations. In any real vector lattice X the following formulae hold:

$$\big||f|\big| = |f|, \tag{7.3a}$$
$$|f+g| \le |f|+|g|, \quad |cf| = |c|\cdot|f|, \tag{7.3b}$$
$$\big||f|-|g|\big| \le |f-g|, \tag{7.3c}$$
$$f = f^+ - f^-, \quad |f| = f^+ + f^-, \tag{7.3d}$$
$$f \vee g = \tfrac{1}{2}(f+g+|f-g|), \quad f \wedge g = \tfrac{1}{2}(f+g-|f-g|), \tag{7.3e}$$
$$(f \vee g)+h = (f+h) \vee (g+h), \quad (f \wedge g)+h = (f+h) \wedge (g+h), \tag{7.3f}$$
$$|f \vee g - f_1 \vee g_1| \le |f-f_1|+|g-g_1|, \tag{7.3e}$$
$$|f \wedge g - f_1 \wedge g_1| \le |f-f_1|+|g-g_1|, \tag{7.3g}$$

for all $f,f_1,g,g_1,h \in V$ and $c \in \mathbb{R}$. (Note that these are easy to establish for the special case $V = \mathbb{R}$ and hence carry over immediately to the spaces $\mathrm{L}^p(\mathrm{X};\mathbb{R})$.)

In a real vector lattice V an element $f \in V$ is called **positive** if $f \ge 0$. The **positive cone** is the set $V_+ := \{f \in V : f \ge 0\}$ of positive elements. If $f \in V$ then f^+ and f^- are positive, and hence (7.3d) yields that $V = V_+ - V_+$.

So far, we dealt with real vector spaces. However, for the purpose of spectral theory (which plays an important role in the study of Koopman operators) it is essential to work with the *complex* Banach spaces $\mathrm{L}^p(\mathrm{X}) = \mathrm{L}^p(\mathrm{X};\mathbb{C})$. Any function $f \in \mathrm{L}^p(\mathrm{X};\mathbb{C})$ can be uniquely written as

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

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

$$\mathrm{L}^p(\mathrm{X}) = \mathrm{L}^p(\mathrm{X};\mathbb{R}) \oplus \mathrm{i}\,\mathrm{L}^p(\mathrm{X};\mathbb{R}).$$

Furthermore, the absolute value mapping $|\cdot|$ has an extension to $\mathrm{L}^p(\mathrm{X})$ that satisfies

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

cf. Exercise 2. Finally, the *norm* on $\mathrm{L}^p(\mathrm{X})$ satisfies

$$|f| \le |g| \quad \Rightarrow \quad \|f\|_p \le \|g\|_p .$$

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

Definition 7.2. A **complex Banach lattice** is a complex Banach space E such that there is a real-linear subspace $E_{\mathbb{R}}$ together with an ordering $\le$ of it, and a mapping $|\cdot| : E \to E$, called **absolute value** or **modulus**, such that the following holds.

1) $E = E_{\mathbb{R}} \oplus \mathrm{i}E_{\mathbb{R}}$ as real vector spaces.

The projection onto the first component is denoted by $\mathrm{Re} : E \to E_{\mathbb{R}}$ and called the **real part**; and $\mathrm{Im}\, f := -\mathrm{Re}(\mathrm{i}f)$ is called the **imaginary part**; hence $f = \mathrm{Re}\, f + \mathrm{i}\,\mathrm{Im}\, f$ is the canonical decomposition of f.

2) $(E_{\mathbb{R}}, \le)$ is a real vector lattice.

3) $|f| = \sup\{\mathrm{Re}\, \mathrm{e}^{\pi \mathrm{i} t} f : t \in \mathbb{Q}\} = \sup_{t\in\mathbb{Q}} (\cos \pi t)\, \mathrm{Re}\, f - (\sin \pi t)\, \mathrm{Im}\, f.$

4) $|f| \le |g| \quad \Rightarrow \quad \|f\| \le \|g\|.$

The set $E_+ = \{f \in E_{\mathbb{R}} : f \ge 0\}$ is called the **positive cone** of E. If we write $f \le g$ with $f, g \in E$, we tacitly suppose that $f, g \in E_{\mathbb{R}}$. For $f \in E$ we call $\overline{f} := \mathrm{Re}\, f - \mathrm{i}\,\mathrm{Im}\, f$ its **conjugate**.

Examples 7.3. 1) If X is a measure space, then $\mathrm{L}^p(\mathrm{X})$ is a complex Banach lattice for each $1 \le p \le \infty$.

2) If K is a compact space, then $\mathrm{C}(K)$ is a complex Banach lattice (Exercise 3).

The following lemma lists some properties of Banach lattices, immediate for our standard examples. The proof for general Banach lattices is left as Exercise 4.

Lemma 7.4. *Let E be a complex Banach lattice. Then the following assertions hold for all $f, g \in E$ and $\alpha \in \mathbb{C}$.*

a) $\overline{f+g} = \overline{f} + \overline{g}$ *and* $\overline{\alpha f} = \overline{\alpha}\overline{f}$.

b) $\left|\overline{f}\right| = |f| \ge |\mathrm{Re}\, f|, |\mathrm{Im}\, f| \ge 0.$

c) $f \in E_{\mathbb{R}} \quad \Rightarrow \quad |f| = f \vee (-f).$

d) $f \ge 0 \quad \Leftrightarrow \quad |f| = f.$

e) $|f+g| \le |f| + |g|$ *and* $||f| - |g|| \le |f-g|$ *and* $|\alpha f| = |\alpha|\,|f|.$

f) $\||f|\| = \|f\|$ *and* $\||f| - |g|\| \le \|f - g\|.$

Moreover, (7.3d)–(7.3g) *hold for $f, f_1, g, g_1, h \in E_{\mathbb{R}}$.*

Part c) shows that the modulus mapping defined on E is compatible with the modulus on $E_{\mathbb{R}}$ coming from the lattice structure there. Part f) implies that the modulus mapping $f \mapsto |f|$ is continuous on E. Hence by (7.3e) the mappings

$$E_{\mathbb{R}} \times E_{\mathbb{R}} \to E_{\mathbb{R}}, \qquad (f, g) \mapsto f \vee g,\ f \wedge g$$

are continuous. Furthermore, from b) it follows that $\|\mathrm{Re}\, f\|, \|\mathrm{Im}\, f\| \le \|f\|$, whence the projection onto $E_{\mathbb{R}}$ is bounded, and $E_{\mathbb{R}}$ is closed. Finally, by d) also the positive cone E_+ is closed.

Positive Operators

Let E, F be Banach lattices. A (linear) operator $S : E \to F$ is called **positive** if

$$f \in E,\ f \ge 0 \quad \Rightarrow \quad Sf \ge 0.$$

We write $S \ge 0$ to indicate that S is positive. The following lemma collects the basic properties of positive operators.

Lemma 7.5. *Let E, F be Banach lattices and let $S : E \to F$ be a positive operator. Then the following assertions hold.*

a) $f \le g \quad \Rightarrow \quad Sf \le Sg$ *for all* $f, g \in E_{\mathbb{R}}$;
b) $f \in E_{\mathbb{R}} \quad \Rightarrow \quad Sf \in F_{\mathbb{R}}$;
c) $S(\mathrm{Re}\, f) = \mathrm{Re}\, Sf$ *and* $S(\mathrm{Im}\, f) = \mathrm{Im}\, Sf$ *for all* $f \in E$;
d) $\overline{Sf} = S\overline{f}$ *for all* $f \in E$;
e) $|Sf| \le S|f|$ *for all* $f \in E$;
f) S *is bounded.*

Proof. a) follows from linearity of S. b) holds since S is linear, $S(E_+) \subseteq F_+$ and $E_{\mathbb{R}} = E_+ - E_+$. c) and d) are immediate consequences of b).
e) For every $t \in \mathbb{Q}$ we have $\mathrm{Re}(e^{\pi i t}) f \le |f|$ and applying S yields

$$\mathrm{Re}\big(e^{\pi i t}\big) Sf \le S|f|.$$

Taking the supremum with respect to $t \in \mathbb{Q}$ we obtain $|Sf| \le S|f|$ as claimed.
f) Suppose that S is not bounded. Then there is a sequence $(f_n)_n$ such that $f_n \to 0$ in norm but $\|Sf_n\| \to \infty$. Since $|Sf_n| \le S|f_n|$ one has also $\|S|f_n|\| \to \infty$. Hence we may suppose that $f_n \ge 0$ for all $n \in \mathbb{N}$. By passing to a subsequence we may also suppose that $\sum_n \|f_n\| < \infty$. But since E is a Banach space, $f := \sum_n f_n$ converges. This yields $0 \le f_n \le f$ and hence $0 \le Sf_n \le Sf$, implying that $\|Sf_n\| \le \|Sf\|$. This is a contradiction. □

A linear operator $S : E \to F$ between two complex Banach lattices E, F is called a **lattice homomorphism** if $|Sf| = S|f|$ for all $f \in E$. Every lattice homomorphism is positive and bounded and satisfies

$$S(f \vee g) = Sf \vee Sg \quad \text{and} \quad S(f \wedge g) = Sf \wedge Sg$$

for $f, g \in E_{\mathbb{R}}$. We note that the Koopman operator T_φ associated with a TDS $(K; \varphi)$ or an MDS $(X; \varphi)$ is a lattice homomorphism.

7.2 The Space $L^p(X)$ as a Banach Lattice

Having introduced the abstract concept of a Banach lattice, we turn to some specific properties of the concrete Banach lattices $L^p(X)$ for $1 \leq p < \infty$. To begin with, we note that these spaces have **order continuous norm**, which means that for each decreasing sequence $(f_n)_{n\in\mathbb{N}} \subseteq L^p_+(X)$, $f_n \geq f_{n+1}$, one has

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

This is a direct consequence of the Monotone Convergence Theorem (see Appendix B.5) by considering the sequence $(f_1^p - f_n^p)_{n\in\mathbb{N}}$. Actually, the Monotone Convergence Theorem accounts also for the following statement.

Theorem 7.6. *Let* X *be a measure space and let* $1 \leq p < \infty$. *Let* $\mathscr{F} \subseteq L^p_+(X)$ *be a* $\vee$*-stable set such that*

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

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

Proof. Take a sequence $(f_n)_{n\in\mathbb{N}} \subseteq \mathscr{F}$ with $\|f_n\|_p \to s$. By passing to the sequence $(f_1 \vee f_2 \vee \ldots \vee f_n)_{n\in\mathbb{N}}$ we may suppose that $(f_n)_n$ is increasing. Define $f := \lim_n f_n$ pointwise almost everywhere. By the Monotone Convergence Theorem we obtain $\|f\|_p = s$ hence $f \in L^p$. Since $\inf_n(f - f_n) = 0$, by order continuity of the norm it follows that $\|f - f_n\|_p \to 0$. If h is any upper bound for $\mathscr{F}$, then $f_n \leq h$ for each n, and then $f \leq h$. Hence it remains to show that f is an upper bound of $\mathscr{F}$. Take an arbitrary $g \in \mathscr{F}$. Then $f_n \vee g \nearrow f \vee g$. Since $f_n \vee g \in \mathscr{F}$, $\|f_n \vee g\|_p \leq s$ for each $n \in \mathbb{N}$. The Monotone Convergence Theorem implies that $\|f \vee g\|_p = \lim_n \|f_n \vee g\|_p \leq s$. But $f \vee g \geq f$ and hence

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

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

Remark 7.7. In the previous proof we have used that $|f| \leq |g|$, $|f| \neq |g|$ implies $\|f\|_1 < \|g\|_1$, i.e., that L^1 has **strictly monotone norm**. Each L^p-space ($1 \leq p < \infty$) has strictly monotone norm, cf. Exercise 6.

A Banach lattice E is called **order complete** if any nonempty subset which has an upper bound has even a least upper bound. Equivalently (Exercise 5), for all $f, g \in E_{\mathbb{R}}$ the **order interval**

$$[f,g] := \{h \in E_{\mathbb{R}} : f \leq h \leq g\}$$

is a complete lattice (as defined in Section 7.1.)

Corollary 7.8. *Let* X *be a measure space and let* $1 \le p < \infty$*. Then the Banach lattice* $\mathrm{L}^p(\mathrm{X};\mathbb{R})$ *is order complete.*

Proof. Let $\mathscr{F}' \subseteq \mathrm{L}^p(\mathrm{X};\mathbb{R})$ and suppose that there exists $F \in \mathrm{L}^p(\mathrm{X};\mathbb{R})$ with $f \le F$ for all $f \in \mathscr{F}'$. We may suppose without loss of generality that $\mathscr{F}'$ is $\vee$-stable. Pick $g \in \mathscr{F}'$ and consider the set $g \vee \mathscr{F}' := \{g \vee f : f \in \mathscr{F}'\}$, which is again $\vee$-stable and has the same set of upper bounds as $\mathscr{F}'$. Now $\mathscr{F} := (g \vee \mathscr{F}') - g$ is $\vee$-stable by (7.3f), consists of positive elements and is dominated by $F - g \ge 0$. In particular it satisfies the conditions of Theorem 7.6. Hence it has a supremum h. It is then obvious that $h+g$ is the supremum of $\mathscr{F}'$. By passing to $-\mathscr{F}'$ one can prove that a nonempty family in L^p that is *bounded from below* has an infimum in L^p. □

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

Definition 7.9. Let E be a Banach lattice. A linear subspace $I \subseteq E$ is called a (vector) **lattice ideal** if

$$f, g \in E,\ |f| \le |g|, \quad g \in I \quad \Rightarrow \quad f \in I.$$

If I is a lattice ideal, then $f \in I$ if and only if $|f| \in I$ if and only if $\operatorname{Re} f, \operatorname{Im} f \in I$. It follows from (7.3e) that the real part $I_{\mathbb{R}} := I \cap E_{\mathbb{R}}$ of a lattice ideal I is a vector sublattice of $E_{\mathbb{R}}$.

Immediate examples of closed lattice ideals in $\mathrm{L}^p(\mathrm{X})$ are obtained from measurable sets $A \in \Sigma_{\mathrm{X}}$ by a construction similar to the TDS case (cf. page 43):

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

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

Theorem 7.10. *Let* X *be a finite measure space and* $1 \le p < \infty$*. Then each closed lattice ideal* $I \subseteq \mathrm{L}^p(\mathrm{X})$ *has the form* I_A *for some* $A \in \Sigma_{\mathrm{X}}$*.*

Proof. Let $I \subseteq \mathrm{L}^p(\mathrm{X})$ be a closed lattice ideal. The set

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

is nonempty, closed, $\vee$-stable and has upper bound $\mathbf{1} \in \mathrm{L}^p(\mathrm{X})$ (since μ_{X} is finite). Therefore, by Corollary 7.8 it has even a least upper bound $g \in J$. It follows that $0 \le g \le \mathbf{1}$ and thus $h := g \wedge (\mathbf{1} - g) \ge 0$. Since $0 \le h \le g$ and $g \in I$, the ideal property yields $h \in I$. Since I is a subspace, $g + h \in I$. But $h \le \mathbf{1} - g$, so $g + h \le \mathbf{1}$, and this yields $h + g \in J$. Thus $h + g \le g$, i.e., $h \le 0$. All in all we obtain $g \wedge (\mathbf{1} - g) = h = 0$, hence g must be a characteristic function $\mathbf{1}_{A^{\mathrm{c}}}$ for some $A \in \Sigma$.

We claim that $I = I_A$. To prove the inclusion $\subseteq$ take $f \in I$. Then $|f| \wedge \mathbf{1} \in J$ and therefore $|f| \wedge \mathbf{1} \le g = \mathbf{1}_{A^{\mathrm{c}}}$. This means that $f \in I_A$. To prove the converse inclusion take $f \in I_A$. It suffices to show that $|f| \in I$, hence we may suppose that $f \ge 0$.

Then $f_n := f \wedge (n\mathbf{1}) = n(n^{-1}f \wedge \mathbf{1}) \leq n\mathbf{1}_{A^c} = ng$. Since $g \in J \subseteq I$, we have $ng \in I$, whence $f_n \in I$. Now $(f_n)_n$ is increasing and converges pointwise, hence in norm to its supremum f. Since I is closed, $f \in I$ and this concludes the proof. □

Remarks 7.11. a) In most of the results of this section we required $p < \infty$, for good reasons. The space $\mathrm{L}^\infty(\mathrm{X})$ is a Banach lattice, but its norm is not order continuous (Exercise 6). If the measure is finite, $\mathrm{L}^\infty(\mathrm{X})$ is still order complete (Exercise 10), but this is not true for general measure spaces. Moreover, if L^∞ is not finite dimensional, then there are always closed lattice ideals not of the form I_A.

b) For a finite measure space X we have

$$\mathrm{L}^\infty(\mathrm{X}) \subseteq \mathrm{L}^p(\mathrm{X}) \subseteq \mathrm{L}^1(\mathrm{X}) \quad (1 \leq p \leq \infty).$$

Particularly important in this scale will be the *Hilbert space* $\mathrm{L}^2(\mathrm{X})$.

c) Let K be a compact space. Then the closed lattice ideals in the Banach lattice $\mathrm{C}(K)$ coincide with the closed algebra ideals, i.e., with the sets

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

see Exercise 7.

7.3 The Koopman Operator and Ergodicity

We now study MDSs $(\mathrm{X}; \varphi)$ and the Koopman operator $T = T_\varphi$ on $\mathrm{L}^p(\mathrm{X})$. We know that, for every $1 \leq p \leq \infty$,

1) T is an *isometry* on $\mathrm{L}^p(\mathrm{X})$;
2) T is a *Banach lattice homomorphism* on $\mathrm{L}^p(\mathrm{X})$ (see page 93);
3) $T(fg) = Tf \cdot Tg$ for all $f \in \mathrm{L}^p(\mathrm{X})$, $g \in \mathrm{L}^\infty(\mathrm{X})$;
4) T is a C^*-algebra homomorphism on $\mathrm{L}^\infty(\mathrm{X})$.

We shall see in this section that *ergodicity* of $(\mathrm{X}; \varphi)$ can be characterised by a lattice theoretic property of the associated Koopman operator, namely *irreducibility*.

Definition 7.12. A positive operator $T \in \mathscr{L}(E)$ on a Banach lattice E is called **irreducible** if the only T-invariant closed lattice ideals of E are the trivial ones, i.e.,

$$I \subseteq E \text{ closed lattice ideal, } T(I) \subseteq I \quad \Rightarrow \quad I = \{0\} \text{ or } I = E.$$

If T is not irreducible, it is called **reducible**.

Let us discuss this notion in the finite-dimensional setting.

Example 7.13. Consider the space $\mathrm{L}^1(\{0, \ldots, n-1\}) = \mathbb{R}^n$ and a positive operator T on it, identified with its $n \times n$-matrix. Then the irreducibility of T according to

Definition 7.12 coincides with that notion introduced in Section 2.4 on page 21. Namely, if T is reducible, then there exists a nontrivial T-invariant ideal I_A in $\mathbb{R}^n$, for some $A \subsetneq \{0,1,\dots,n-1\}$. After a permutation of the points we may suppose that $A = \{k,\dots,n-1\}$ for $0 \le k < n$, and this means that the representing matrix (with respect to the canonical basis) has the form:

$$\begin{array}{c} \\ \\ \\ k \end{array} \overset{\overset{k}{\downarrow}}{\left(\begin{array}{cc|c} \star & \star & \star \\ \star & \star & \star \\ \hline 0 & & \star \end{array}\right)}$$

Let us return to the situation of an MDS $(X;\varphi)$. We consider the associated Koopman operator $T = T_\varphi$ on the space $L^1(X)$, but note that T leaves each space L^p invariant. The following result shows that the ergodicity of an MDS is characterised by the irreducibility of the Koopman operator T and the one-dimensionality of its **fixed space**

$$\operatorname{fix}(T) := \{f \in L^1(X) : Tf = f\} = \ker(I - T).$$

(Note that always $\mathbf{1} \in \operatorname{fix}(T)$ whence $\dim \operatorname{fix}(T) \ge 1$.) This is reminiscent (cf. Corollary 4.17) but also contrary to the TDS case, where minimality could *not* be characterised by the one-dimensionality of the fixed space of the Koopman operator.

Proposition 7.14. *Let $(X;\varphi)$ be an MDS with Koopman operator $T := T_\varphi$ on $L^1(X)$. Then for $1 \le p < \infty$ the following statements are equivalent.*

(i) *$(X;\varphi)$ is ergodic.*

(ii) *$\dim \operatorname{fix}(T) = 1$, i.e., 1 is a simple eigenvalue of T.*

(iii) *$\dim(\operatorname{fix}(T) \cap L^p) = 1$.*

(iv) *$\dim(\operatorname{fix}(T) \cap L^\infty) = 1$.*

(v) *T as an operator on $L^p(X)$ is irreducible.*

Proof. Notice that by Lemma 6.13 A is φ-invariant if and only if $\varphi^*[A] = [A]$ if and only if $\mathbf{1}_A \in \operatorname{fix}(T)$. In particular, this establishes the implication (iv) $\Rightarrow$ (i).

(i) $\Rightarrow$ (ii): For any $f \in \operatorname{fix}(T)$ and any $c \in \mathbb{R}$ the set $[f > c]$ is φ-invariant, and hence trivial. Let

$$c_0 := \sup\{c \in \mathbb{R} : \mu[f > c] = 1\}.$$

Then for $c < c_0$ we have $\mu[f \le c] = 0$, and therefore $\mu[f < c_0] = 0$. For $c > c_0$ we have $\mu[f > c] \ne 1$, hence $\mu[f > c] = 0$, and therefore $\mu[f > c_0] = 0$, too. This implies $f = c_0$ almost everywhere.

(ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) is trivial.

(i) $\Leftrightarrow$ (v): By Theorem 7.10 it suffices to show that $A \in \Sigma$ is essentially φ-invariant if and only if $I_A \cap L^p$ is invariant under T. In order to prove this, note that for $f \in L^1$

and $A \in \Sigma$

$$T(|f| \wedge \mathbf{1}_A) = T|f| \wedge T\mathbf{1}_A = |Tf| \wedge T\mathbf{1}_A.$$

So if A is essentially φ-invariant, we have $T\mathbf{1}_A = \mathbf{1}_A$, and $f \in I_A$ implies that $Tf \in I_A$ as well. For the converse suppose that $T(I_A) \subseteq I_A$. Then, since $\mathbf{1}_{A^c} \in I_A$, we must have $\mathbf{1}_{\varphi^* A^c} \in I_A$, which amounts to $\varphi^* A^c \cap A = \emptyset$ in $\Sigma(\mathrm{X})$. Consequently $A \subseteq \varphi^* A$, which means that A is φ-invariant. □

We apply Proposition 7.14 to the rotations on the torus (see Chapter 5.3).

Proposition 7.15. *For $a \in \mathbb{T}$ the rotation MDS $(\mathbb{T}, \mathrm{m}; a)$ is ergodic if and only if a is not a root of unity.*

Proof. Let $T := L_a$ be the Koopman operator on $\mathrm{L}^2(\mathbb{T})$ (cf. Example 4.20) and suppose that $f \in \mathrm{fix}(T)$. The functions $\chi_n : x \mapsto x^n$, $n \in \mathbb{Z}$, form a complete orthonormal system in $\mathrm{L}^2(\mathbb{T})$. Note that $T\chi_n = a^n \chi_n$, i.e., χ_n is an eigenvector of T with corresponding eigenvalue a^n. With appropriate Fourier-coefficients b_n we have

$$\sum_{n\in\mathbb{Z}} b_n \chi_n = f = Tf = \sum_{n\in\mathbb{Z}} b_n T\chi_n = \sum_{n\in\mathbb{Z}} a^n b_n \chi_n.$$

By the uniqueness of the Fourier-coefficients, $b_n(a^n - 1) = 0$ for all $n \in \mathbb{Z}$. This implies that *either* $b_n = 0$ for all $n \in \mathbb{Z} \setminus \{0\}$ (whence f is constant), *or* there is $n \in \mathbb{Z} \setminus \{0\}$ with $a^n = 1$ (whence a is a root of unity). □

Remark 7.16. If $(\mathbb{T}; a)$ is a rational rotation , i.e., a is a root of unity, then a non-constant fixed point of L_a is easy to find: Suppose that $a^n = 1$ for some $n \in \mathbb{N}$, and divide $\mathbb{T}$ into n arcs: $I_j := \{z \in \mathbb{T} : \arg z \in [2\pi(j-1), 2\pi j/n)\}$, $j = 1, \ldots, n$. Take any integrable function on I_1 and "copy it over" to the other segments. The so arising function is a fixed point of L_a.

Proposition 7.15 together with Example 2.33 says that a rotation on the torus is ergodic if and only if it is minimal. Exercise 8 generalises this to rotations on $\mathbb{T}^n$. Actually, the result is true for any rotation on a compact group as we shall prove in Chapter 14.

Similar to the case of minimal TDSs, the peripheral point spectrum of the Koopman operator has a nice structure.

Proposition 7.17. *Let $(\mathrm{X}; \varphi)$ be an ergodic MDS, and consider the Koopman operator $T := T_\varphi$ on $\mathrm{L}^p(\mathrm{X})$ ($1 \le p \le \infty$). Then each eigenvalue of T is unimodular and simple, and each corresponding eigenfunction is unimodular (up to a multiplicative constant). Moreover the (peripheral) point spectrum $\sigma_p(T) = \sigma_p(T) \cap \mathbb{T}$ is a subgroup of $\mathbb{T}$.*

The proof is similar to the proof of Theorem 4.19, see Exercise 9. For an ergodic group rotation $(G, \mathrm{m}; a)$, the analogue to the TDS-case (Example 4.20) holds and is treated in Theorem 14.30.

Supplement: Interplay between Lattice and Algebra Structure

The space $\mathrm{C}(K)$ for a compact space K and $\mathrm{L}^\infty(\mathrm{X})$ for some probability space $\mathrm{X} = (X, \Sigma, \mu)$ are simultaneously commutative C^*-algebras and complex Banach lattices. In this section we shall show that both structures, the $*$-algebraic and the Banach lattice structure, essentially determine each other, in the sense that either one can be reconstructed from the other, and that mappings that preserve one of them also preserve the other.

For the following we suppose that $E = \mathrm{C}(K)$ or $E = \mathrm{L}^\infty(\mathrm{X})$. To begin with, note that

$$|f| = [1 - (1 - |f|^2)]^{1/2} = \sum_{k=0}^{\infty} \binom{1/2}{k} (-1)^k (1 - f\overline{f})^k$$

which is valid for $f \in E$ with $\|f\| \leq 1$; for general f we use a rescaling. This shows that the modulus mapping can be recovered from the C^*-algebraic structure (algebra + conjugation + unit element + norm approximation).

Conversely, the multiplication can be recovered from the lattice structure in the following way. Since multiplication is bilinear and conjugation is present in the lattice structure, it suffices to know the products $f \cdot g$ where f, g are real valued. By the polarisation identity

$$2 \cdot f \cdot g = (f+g)^2 - f^2 - g^2, \tag{7.4}$$

the product of the real elements is determined by taking squares; and since for a real element $h^2 = |h|^2$, it suffices to cover squares of positive elements.

Now, let $0 < p < \infty$. Writing the (continuous and convex/concave) mapping $x \mapsto x^p$ on $\mathbb{R}_+$ as the supremum/infimum over tangent lines, we obtain for $x \geq 0$

$$x^p = \begin{cases} \sup_{t>0} pt^{p-1}x - (p-1)t^p, & \text{if } 1 \leq p < \infty, \\ \inf_{t>0} pt^{p-1}x - (p-1)t^p, & \text{if } 0 < p \leq 1. \end{cases} \tag{7.5}$$

In the case $p \geq 1$ define

$$h_n(x) := \bigvee_{k=1}^{n} pt_k^{p-1}x - (p-1)t_k^p \qquad (x \geq 0,\ n \in \mathbb{N}),$$

where $(t_k)_{k\in\mathbb{N}}$ is any enumeration of $\mathbb{Q}_+$. Then, by (7.5) and continuity, $h_n(x) \nearrow x^p$ pointwise, and hence uniformly on bounded subsets of $\mathbb{R}_+$, by Dini's theorem (Exercise 14). So if $0 \leq f \in E$ then $h_n \circ f \to f^p$ in norm, and $h_n \circ f$ is contained in any vector sublattice of E that contains f and $\mathbf{1}$. In particular, for $p = 2$, it follows that the multiplication can be recovered from the Banach lattice structure and the constant function $\mathbf{1}$. More precisely, we have the following result.

Theorem 7.18. *Let $E = \mathrm{C}(K)$ or $\mathrm{L}^\infty(\mathrm{X})$, and let $A \subseteq E$ be closed, conjugation invariant with $\mathbf{1} \in A$. Then the following assertions are equivalent.*

(i) *A is a subalgebra of E.*

(ii) *A is a vector sublattice of E.*

(iii) *If $f \in A$ then $|f|^p \in A$ for all $0 < p < \infty$.*

Now suppose that A satisfies (i)–(iii)*, and let $\Phi : A \to \mathrm{L}^\infty(\mathrm{Y})$ be a conjugation-preserving linear operator such that $\Phi\mathbf{1} = \mathbf{1}$. Then the following statements are equivalent for $1 \leq p < \infty$.*

(iv) *Φ is a homomorphism of C^*-algebras.*

(v) *Φ is a lattice homomorphism.*

(vi) *$\Phi|f|^p = |\Phi f|^p$ for every $f \in A$.*

Proof. The proof of the first part is left as Exercise 15. In the second part, suppose that (iv) holds. Then it follows from

$$[\Phi|f|]^2 = \Phi|f|^2 = \Phi(f\overline{f}) = (\Phi f)(\Phi\overline{f}) = |\Phi f|^2$$

that Φ is a lattice homomorphism, i.e., (v). Now suppose that (v) is true. Then Φ is positive and hence bounded. Let $f \geq 0$ and $1 < p < \infty$. We have $f^p = \lim_n h_n \circ f$, where h_n is as above. Then

$$\Phi f^p = \lim_n \Phi(h_n \circ f) = \lim_n h_n \circ (\Phi f) = (\Phi f)^p$$

because Φ is a lattice homomorphism. Hence (vi) is established.

Finally, suppose that (vi) holds. Then Φ is positive; indeed, if $0 \leq f \in A$, then by (iii), $g := f^{1/p} \in A$ and thus $\Phi f = \Phi g^p = |\Phi g|^p \geq 0$. Now let $f \in A$ and suppose that $p > 1$. Then by the Hölder inequality (Theorem 7.19) below,

$$\Phi|f| \leq (\Phi|f|^p)^{1/p}(\Phi\mathbf{1}^q)^{1/q} = (\Phi|f|^p)^{1/p} = |\Phi f| \leq \Phi|f|$$

by (vi) and the positivity of Φ. Hence (ii) follows, and this implies (vi) for $p = 2$ as already shown. The polarisation identity (7.4) yields $\Phi(fg) = \Phi(f)\Phi(g)$ for real elements $f, g \in A$. Since Φ preserves conjugation, Φ is multiplicative on the whole of A, and (i) follows. □

The following fact was used in the proof, but is also of independent interest.

Theorem 7.19 (Hölder's Inequality for Positive Operators). *Let $E = \mathrm{C}(K)$ or $\mathrm{L}^\infty(\mathrm{X})$, and let $A \subseteq E$ be a C^*-subalgebra of E. Furthermore, let Y be any measure space, and let $T : A \to \mathrm{L}^0(\mathrm{Y})$ be a positive linear operator. Then*

$$|T(f \cdot g)| \leq (T|f|^p)^{1/p} \cdot (T|g|^q)^{1/q} \tag{7.6}$$

whenever $f, g \in A$ with $f, g \geq 0$.

Proof. We start with the representation

$$x^{1/p} = \inf_{t>0} (1/p)t^{-1/q}x + (1/q)t^{1/p}$$

which follows from (7.5) with p replaced by $1/p$. For fixed $y > 0$, multiply this identity with $y^{1/q}$ and arrive at

$$x^{1/p}y^{1/q} = \inf_{t>0}(1/p)(y/t)^{1/q}x + (1/q)t^{1/p}y^{1/q} = \inf_{s>0}(1/p)s^{-1/q}x + (1/q)s^{1/p}y$$

by a change of parameter $s = ty$. Replace now x by x^p and y by y^q to obtain

$$xy = \inf_{s>0}(1/p)s^{-1/q}x^p + (1/q)s^{1/p}y^q. \tag{7.7}$$

Note that this holds true even for $y = 0$, and by a continuity argument we may restrict the range of the parameter s to positive *rational* numbers. Then, for a fixed $0 < s \in \mathbb{Q}$ (7.7) yields

$$|f \cdot g| \leq (1/p)s^{-1/q}|f|^p + (1/q)s^{1/p}|g|^q.$$

Applying T we obtain

$$|T(f \cdot g)| \leq T|f \cdot g| \leq (1/p)s^{-1/q}T|f|^p + (1/q)s^{1/p}T|g|^q$$

and, taking the infimum over $0 < s \in \mathbb{Q}$, by (7.7) again we arrive at (7.6). □

Remark 7.20. a) Theorem 7.18 remains true if one replaces $L^\infty(Y)$ by a space $C(K')$. (Fix an arbitrary $x \in K'$ and apply Theorem 7.18 with $(Y, \Sigma', \nu) = (K, \mathrm{Bo}(K), \delta_x)$.)

b) Theorem 7.18 can be generalised by means of approximation arguments, see Exercise 16.

Exercises

1. Show that if $\varphi : Y \to X$ is measurable with $\varphi_*\mu_Y = \mu_X$, then the associated Koopman operator $T_\varphi : L^\infty(X) \to L^\infty(Y)$ is an isometry.

2. Let X be a measure space and let $1 \leq p \leq \infty$. Show that

$$|f| = \sup\{\mathrm{Re}(cf) : c \in \mathbb{T}\}$$

with the supremum being taken in the order of $L^p(X;\mathbb{R})$.

3. Prove that $C(K)$, K compact, is a complex Banach lattice.

4. Prove Lemma 7.4.

5. Show that a Banach lattice E is order complete if and only if every order interval $[f,g]$, $f,g \in E_\mathbb{R}$, is a complete lattice.

6. Show that in general a space $L^\infty(X,\Sigma,\mu)$ does neither have strictly monotone nor order continuous norm.

7. Prove that for a space $\mathrm{C}(K)$ the closed lattice ideals coincide with the closed algebra ideals. (Hint: Adapt the proof of Theorem 4.8.)

8. Show that the rotation on the torus $\mathbb{T}^n$ is ergodic if and only if it is minimal. (Hint: Copy the proof of Proposition 7.15 using n-dimensional Fourier coefficients.)

9. Prove Proposition 7.17.

10. Let X be a finite measure space. Show that the lattice

$$V := \mathrm{L}^0(\mathrm{X};[0,1]) = \{f \in \mathrm{L}^1(\mathrm{X}) : 0 \leq f \leq 1\}$$

is complete. (Hint: Use Corollary 7.8.) Show that the lattices

$$\mathrm{L}^0(\mathrm{X};\overline{\mathbb{R}}) \quad \text{and} \quad \mathrm{L}^0(\mathrm{X};[0,\infty])$$

are lattice isomorphic to V and conclude that they are complete. Finally, prove that $\mathrm{L}^\infty(\mathrm{X};\mathbb{R})$ is an order complete Banach lattice, but its norm is not order continuous in general.

11. Let $\mathrm{X} = (X,\Sigma,\mu)$ be a probability space, and let and $\mathscr{F} \subseteq \mathrm{L}^1(\mathrm{X})_+$ be $\wedge$-stable such that $\inf \mathscr{F} = 0$. Show (e.g., by using Theprem 7.6) that $\inf_{f\in\mathscr{F}} \|f\|_1 = 0$.

12. Let X be a finite measure space and $0 \leq f \in \mathrm{L}^0(\mathrm{X};\overline{\mathbb{R}})$. Show that $\sup_{n\in\mathbb{N}}(nf \wedge \mathbf{1}) = \mathbf{1}_{[f>0]}$. Show that $f = \mathbf{1}_A$ for some $A \in \Sigma$ if and only if $cf \wedge \mathbf{1} = f$ for every $c > 1$.

13. Let $\mathrm{X} = (X,\Sigma,\mu)$ be a finite measure space. Show that $\{\mathbf{1}_A : A \in \Sigma\}$ is a complete sublattice of $\mathrm{L}^1(\mathrm{X})$. Conclude that the measure algebra $\Sigma(\mathrm{X})$ is a complete lattice.

14 (Dini). Let K be a compact space and suppose that $(f_n)_n$ is a sequence in $\mathrm{C}(K)$ with $0 \leq f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$. Suppose further that there is $f \in \mathrm{C}(K)$ such that $f_n \to f$ *pointwise*. Then the convergence is uniform.

15. Prove the first part of Theorem 7.18.

16 (Hölder's Inequality for Positive Operators). Let X and Y be measure spaces, $T : \mathrm{L}^1(\mathrm{X}) \to \mathrm{L}^0(\mathrm{Y})$ a positive linear operator. Show that

$$|T(f\cdot g)| \leq (T\,|f|^p)^{1/p} \cdot (T\,|g|^q)^{1/q}$$

whenever $1 < p < \infty$, $1/p + 1/q = 1$ and $f \in \mathrm{L}^p(\mathrm{X})$, $g \in \mathrm{L}^q(\mathrm{X})$.

Chapter 8
The Mean Ergodic Theorem

As was said at the beginning of Chapter 5, the reason for introducing measure-preserving dynamical systems is the intuition of a statistical equilibrium emergent from (very rapid) deterministic interactions of a multitude of particles. A measurement of the system can be considered as a random experiment, where repeated measurements appear to be independent since the time scale of measurements is far larger than the time scale of the internal dynamics.

In such a perspective, one expects that the (arithmetic) averages over the outcomes of these measurements ("time averages") should converge—in some sense or another—to a sort of "expected value". In mathematical terms, given an MDS $(\mathrm{X};\varphi)$ and an "observable" $f : X \to \mathbb{R}$, the time averages take the form

$$\mathrm{A}_n f(x) := \frac{1}{n}\left(f(x) + f(\varphi(x)) + \cdots + f(\varphi^{n-1}(x)\right) \tag{8.1}$$

if $x \in X$ is the initial state of the system; and the expected value is the "space mean"

$$\int_{\mathrm{X}} f$$

of f. In his original approach, Boltzmann assumed the so-called *"Ergodenhypothese"*, which allowed him to prove the convergence $\mathrm{A}_n f(x) \to \int_{\mathrm{X}} f$ ("time mean equals space mean", cf. Chapter 1). However, the Ehrenfests in [Ehrenfest and Ehrenfest (1912)] doubted that this "Ergodenhypothese" is ever satisfied, a conjecture that was confirmed independently by Rosenthal and Plancherel only a few years later (see [Brush (1971)]). After some 20 more years, John von Neumann and George D. Birkhoff made a major step forward by separating the question of convergence of the averages $\mathrm{A}_n f$ from the question whether the limit is the space mean of f or not. Their results—the *"Mean Ergodic Theorem"* and the *"Individual Ergodic Theorem"*—roughly state that under reasonable conditions on f the time averages always converge in some sense, while the limit is the expected "space mean" if and only if the system is ergodic (in our terminology). These theorems gave birth to Ergodic Theory as a mathematical discipline.

The present and the following two chapters are devoted to these fundamental results, starting with von Neumann's theorem. The linear operators induced by dynamical systems and studied in the previous chapters will be the main protagonists now.

8.1 Von Neumann's Mean Ergodic Theorem

Let $(\mathrm{X};\varphi)$ be an MDS and let $T = T_\varphi$ be the Koopman operator. Note that the time mean of a function f under the first $n \in \mathbb{N}$ iterates of T (8.1) can be written as

$$\mathrm{A}_n f = \frac{1}{n}\left(f + f\circ\varphi + \cdots + f\circ\varphi^{n-1}\right) = \frac{1}{n}\sum_{j=0}^{n-1} T^j f.$$

Von Neumann's theorem deals with these averages $\mathrm{A}_n f$ for f from the Hilbert space $\mathrm{L}^2(\mathrm{X})$.

Theorem 8.1 (Von Neumann). *Let $(\mathrm{X};\varphi)$ be an MDS and consider the Koopman operator $T := T_\varphi$. For each $f \in \mathrm{L}^2(\mathrm{X})$ the limit*

$$\lim_{n\to\infty} \mathrm{A}_n f = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} T^j f$$

exists in the L^2-sense and is a fixed point of T.

We shall not give von Neumann's original proof, but take a route that is more suitable for generalisations. For a linear operator T on a vector space E we let

$$\mathrm{A}_n[T] := \frac{1}{n}\sum_{j=0}^{n-1} T^j \qquad (n \in \mathbb{N})$$

be the **Cesàro averages** (or: Cesàro means) of the first n iterates of T. If T is understood, we omit explicit reference and simply write A_n in place of $\mathrm{A}_n[T]$. Further we denote by

$$\mathrm{fix}(T) := \{f \in E : Tf = f\} = \ker(\mathrm{I} - T)$$

the **fixed space** of T.

Lemma 8.2. *Let E be a Banach space and let $T : E \to E$ be a bounded linear operator on E. Then, with $\mathrm{A}_n := \mathrm{A}_n[T]$, the following assertions hold.*

a) *If $f \in \mathrm{fix}(T)$, then $\mathrm{A}_n f = f$ for all $n \in \mathbb{N}$, and hence $\mathrm{A}_n f \to f$.*

b) *One has*

$$\mathrm{A}_n T = T\mathrm{A}_n = \frac{n+1}{n}\mathrm{A}_{n+1} - \frac{1}{n}\mathrm{I} \qquad (n \in \mathbb{N}). \tag{8.2}$$

If $\mathrm{A}_n f \to g$, then $Tg = g$ and $\mathrm{A}_n T f \to g$.

c) *One has*

$$(\mathrm{I}-T)\mathrm{A}_n = \mathrm{A}_n(\mathrm{I}-T) = \frac{1}{n}(\mathrm{I}-T^n) \qquad (n \in \mathbb{N}). \tag{8.3}$$

If $(1/n)T^n f \to 0$ *for all* $f \in E$, *then* $\mathrm{A}_n f \to 0$ *for all* $f \in \mathrm{ran}(\mathrm{I}-T)$.

d) *One has*

$$\mathrm{I}-\mathrm{A}_n = (\mathrm{I}-T)\frac{1}{n}\sum_{k=0}^{n-1}\sum_{j=0}^{k-1} T^j \qquad (n \in \mathbb{N}). \tag{8.4}$$

If $\mathrm{A}_n f \to g$, *then* $f-g \in \overline{\mathrm{ran}}(\mathrm{I}-T)$.

Proof. a) is trivial, and the formulae (8.2)–(8.4) are established by simple algebra. The remaining statements then follow from these formulae. □

From a) and b) of Lemma 8.2 we can conclude the following.

Lemma 8.3. *Let* T *be a bounded linear operator on a Banach space* E. *Then*

$$F := \big\{ f \in E : P_T f := \lim_{n\to\infty} \mathrm{A}_n f \text{ exists} \big\}$$

is a T*-invariant subspace of* E *containing* $\mathrm{fix}(T)$. *Moreover* $P_T : F \to F$ *is a projection onto* $\mathrm{fix}(T)$ *satisfying* $TP_T = P_T T = P_T$ *on* F.

Proof. It is clear that F is a subspace of E and $P_T : F \to E$ is linear. By Lemma 8.2.b, F is T-invariant, $\mathrm{ran}(P_T) \subseteq \mathrm{fix}(T)$ and $P_T T f = T P_T f = P_T f$ for all $f \in F$. Finally it follows from Lemma 8.2.a that $\mathrm{fix}(T) \subseteq F$, and $P_T f = f$ if $f \in \mathrm{fix}(T)$. In particular $P_T^2 = P_T$, i.e., P_T is a projection. □

Definition 8.4. Let T be a bounded linear operator on a Banach space E. Then the operator P_T, defined by

$$P_T f := \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} T^j f \tag{8.5}$$

on the space F of all $f \in E$ where this limit exists, is called the **mean ergodic projection** associated with T. The operator T is called **mean ergodic** if $F = E$, i.e., if the limit in (8.5) exists for every $f \in E$.

Using this terminology we can rephrase von Neumann's result: *The Koopman operator associated with an MDS* $(\mathrm{X};\varphi)$ *is mean ergodic when considered as an operator on* $E = \mathrm{L}^2(\mathrm{X})$. Our proof of this statement consists of two more steps.

Theorem 8.5. *Let* $T \in \mathscr{L}(E)$, E *a Banach space. Suppose that* $\sup_{n\in\mathbb{N}_0} \|\mathrm{A}_n\| < \infty$ *and that* $(1/n)T^n f \to 0$ *for all* $f \in E$. *Then the subspace*

$$F := \{ f : \lim_{n\to\infty} \mathrm{A}_n f \text{ exists} \}$$

is closed, T*-invariant, and decomposes into a direct sum of closed subspaces*

$$F = \mathrm{fix}(T) \oplus \overline{\mathrm{ran}}(\mathrm{I}-T).$$

The operator $T\big|_F \in \mathscr{L}(F)$ *is mean ergodic. Furthermore, the operator*

$$P_T : F \to \operatorname{fix}(T) \qquad P_T f := \lim_{n\to\infty} \mathrm{A}_n f$$

is a bounded projection with kernel $\ker(P_T) = \overline{\operatorname{ran}}(\mathrm{I}-T)$ *and* $P_T T = P_T = T P_T$.

Proof. By Lemma 8.3, all that remains to show is that F is closed, P_T is bounded, and $\ker(P_T) = \overline{\operatorname{ran}}(\mathrm{I}-T)$. The closedness of F and the boundedness of P_T are solely due to the uniform boundedness of the operator sequence $(\mathrm{A}_n)_{n\in\mathbb{N}}$ and the strong convergence lemma (Exercise 1).

To prove the remaining statement, note that $\operatorname{ran}(\mathrm{I}-T) \subseteq \ker(P_T)$ by Lemma 8.2.c; hence $\overline{\operatorname{ran}}(\mathrm{I}-T) \subseteq \ker(P_T)$ since P_T is bounded. Since P_T is a projection, one has $\ker(P_T) = \operatorname{ran}(\mathrm{I}-P_T)$, and by Lemma 8.2.d we finally obtain

$$\ker(P_T) = \operatorname{ran}(\mathrm{I}-P_T) \subseteq \overline{\operatorname{ran}}(\mathrm{I}-T) \subseteq \ker(P_T). \qquad \square$$

Mean ergodic operators will be studied in greater detail in Section 8.4 below. For the moment we note the following important result, which entails von Neumann's theorem as a corollary.

Theorem 8.6 (Mean Ergodic Theorem on Hilbert Spaces). *Let H be a Hilbert space and let $T \in \mathscr{L}(H)$ be a contraction, i.e., such that $\|T\| \le 1$. Then*

$$P_T f := \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} T^j f \qquad \textit{exists for every } f \in H.$$

Moreover, $H = \operatorname{fix}(T) \oplus \overline{\operatorname{ran}}(\mathrm{I}-T)$ is an orthogonal decomposition, and the mean ergodic projection P_T is the orthogonal projection onto $\operatorname{fix}(T)$.

Proof. If T is a contraction, then the powers T^n and hence the Cesàro averages $\mathrm{A}_n[T]$ are contractions, too, and $(1/n)T^n \to 0$. Therefore, Theorem 8.5 can be applied and so the subspace F is closed and $P_T : F \to F$ is a projection onto $\operatorname{fix}(T)$ with kernel $\overline{\operatorname{ran}}(\mathrm{I}-T)$.

Take now $f \in H$ with $f \perp \operatorname{ran}(\mathrm{I}-T)$. Then $(f \,|\, f - Tf) = 0$ and hence $(f \,|\, Tf) = (f \,|\, f) = \|f\|^2$. This implies that

$$\|Tf - f\|^2 = \|Tf\|^2 - 2\operatorname{Re}(f \,|\, Tf) + \|f\|^2 = \|Tf\|^2 - \|f\|^2 \le 0$$

since T is a contraction. Consequently, $f = Tf$, i.e., $f \in \operatorname{fix}(T)$. Hence we have proved that

$$\operatorname{ran}(\mathrm{I}-T)^\perp \subseteq \operatorname{fix}(T).$$

However, from $\overline{\operatorname{ran}}(\mathrm{I}-T) \cap \operatorname{fix}(T) = \{0\}$ we obtain $\operatorname{ran}(\mathrm{I}-T)^\perp = \operatorname{fix}(T)$ as claimed. (Alternatively, note that P_T must be a contraction and use Exercise 2.) $\square$

Let us note the following interesting consequence.

Corollary 8.7. *Let T be a contraction on a Hilbert space H. Then* $\operatorname{fix}(T) = \operatorname{fix}(T^*)$ *and* $P_T = P_{T^*}$.

Proof. Since $\mathrm{A}_n[T] \to P_T$ strongly we conclude that

$$\mathrm{A}_n[T^*] = \mathrm{A}_n[T]^* \to P_T^* = P_T \quad \text{weakly.}$$

But T^* is a contraction as well, whence $\mathrm{A}_n[T^*] \to P_{T^*}$ strongly. It follows that $P_T = P_{T^*}$ and hence $\mathrm{fix}(T) = \mathrm{fix}(T^*)$. □

8.2 The Fixed Space and Ergodicity

Let $(\mathrm{X};\varphi)$ be an MDS. In von Neumann's theorem the Koopman operator was considered as an operator on $\mathrm{L}^2(\mathrm{X})$. However, it is natural to view it also as an operator on L^1 and in fact on any L^p-space with $1 \le p \le \infty$.

Theorem 8.8. *Let $(\mathrm{X};\varphi)$ be an MDS, and consider the Koopman operator $T = T_\varphi$ on the space $\mathrm{L}^1 = \mathrm{L}^1(\mathrm{X})$. Then*

$$P_T f := \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} T^j f \qquad \textit{exists in } \mathrm{L}^p \textit{ for every } f \in \mathrm{L}^p,\ 1 \le p < \infty.$$

Moreover, for $1 \le p \le \infty$, the following assertions hold.

a) *The space $\mathrm{fix}(T) \cap \mathrm{L}^p$ is a Banach sublattice of L^p containing $\mathbf{1}$.*

b) *$P_T : \mathrm{L}^p \to \mathrm{fix}\, T \cap \mathrm{L}^p$ is a positive contractive projection satisfying*

$$\int_{\mathrm{X}} P_T f = \int_{\mathrm{X}} f \qquad (f \in \mathrm{L}^p) \tag{8.6}$$

$$\textit{and} \qquad P_T(f \cdot P_T g) = (P_T f) \cdot (P_T g) \qquad (f \in \mathrm{L}^p, g \in \mathrm{L}^q,\ \tfrac{1}{p} + \tfrac{1}{q} = 1). \tag{8.7}$$

Proof. Let $f \in \mathrm{L}^\infty$. Then, by von Neumann's theorem, the limit $P_T f = \lim_{n\to\infty} \mathrm{A}_n f$ exists in L^2, thus a fortiori in L^1. Moreover, since $|T^n f| \le \|f\|_\infty \mathbf{1}$ for each $n \in \mathbb{N}_0$, we have $|\mathrm{A}_n f| \le \|f\|_\infty \mathbf{1}$ and finally $|P_T f| \le \|f\|_\infty \mathbf{1}$. Then Hölder's inequality yields

$$\|\mathrm{A}_n f - P_T f\|_p^p \le (2\|f\|_\infty)^{p-1} \|\mathrm{A}_n f - P_T f\|_1 \to 0$$

for any $1 \le p < \infty$. Hence we have proved that for each such p the space

$$F_p := \left\{ f \in \mathrm{L}^p : \|\cdot\|_p - \lim_{n\to\infty} \mathrm{A}_n f \ \text{ exists} \right\}$$

contains the dense subspace L^∞; but it is closed in L^p by Theorem 8.5, and so it must be all of L^p.

a) Take $f \in \mathrm{fix}(T)$. Then $T|f| = |Tf| = |f|$ and $T\overline{f} = \overline{Tf} = \overline{f}$. Hence $\mathrm{fix}(T) \cap \mathrm{L}^p$ is a vector sublattice of L^p. Since it is obviously closed in L^p, it is a Banach sublattice of L^p.

b) Equation (8.6) is immediate from the definition of P_T and equation (7.1). For the proof of the remaining part we may suppose by symmetry and density that $g = P_T g \in \mathrm{L}^\infty$. Then $Tg = g$ and hence

$$\begin{aligned} \mathrm{A}_n(fg) &= \frac{1}{n}\sum_{j=0}^{n-1} T^j(fg) = \frac{1}{n}\sum_{j=0}^{n-1}(T^j f)(T^j g) = \frac{1}{n}\sum_{j=0}^{n-1}(T^j f)g \\ &= (\mathrm{A}_n f)\cdot g \to (P_T f)\cdot g \qquad \text{as } n\to\infty. \end{aligned}$$

□

Remark 8.9. We give a probabilistic view on the mean ergodic projection P_T. Define

$$\Sigma_\varphi := \mathrm{fix}(\varphi^*) := \big\{A \in \Sigma_\mathrm{X} : \varphi^* A = A\big\} = \big\{A \in \Sigma_\mathrm{X} : \mathbf{1}_A \in \mathrm{fix}(T)\big\}.$$

This is obviously a sub-σ-algebra of Σ_X, called the **φ-invariant σ-algebra**. We claim that

$$\mathrm{fix}(T) = \mathrm{L}^1(X, \Sigma_\varphi, \mu_\mathrm{X}).$$

Indeed, since step functions are dense in L^1, every $f \in \mathrm{L}^1(X,\Sigma_\varphi,\mu_\mathrm{X})$ is contained in $\mathrm{fix}(T)$. On the other hand, if $f \in \mathrm{fix}(T)$, then for every Borel set $B \subseteq \mathbb{C}$ one has

$$\varphi^*[f \in B] = [\varphi \in [f \in B]] = [f \circ \varphi \in B] = \big[T_\varphi f \in B\big] = [f \in B],$$

which shows that $[f \in B] \in \Sigma_\varphi$. Hence f is Σ_φ-measurable. Now, take $A \in \Sigma_\varphi$ and $f := \mathbf{1}_A$ in (8.7) and use (8.6) to obtain

$$\int_\mathrm{X} \mathbf{1}_A P_T f = \int_\mathrm{X} P_T(\mathbf{1}_A \cdot f) = \int_\mathrm{X} \mathbf{1}_A \cdot f.$$

This shows that $P_T f = \mathbb{E}(f\,|\,\Sigma_\varphi)$ is the **conditional expectation** of f with respect to the φ-invariant σ-algebra Σ_φ.

We can now extend the characterisation of ergodicity by means of the Koopman operator obtained in Proposition 7.14.

Theorem 8.10. *Let $(\mathrm{X};\varphi)$ be an MDS, $\mathrm{X} = (X,\Sigma,\mu)$, with associated Koopman operator $T = T_\varphi$ on $\mathrm{L}^1(\mathrm{X})$, and let $1 \le p < \infty$. The following statements are equivalent.*

(i) *$(\mathrm{X};\varphi)$ is ergodic.*

(ii) *$\dim \mathrm{fix}(T) = 1$, i.e., 1 is a simple eigenvalue of T.*

(iii) *$\dim(\mathrm{fix}(T) \cap \mathrm{L}^\infty) = 1$.*

(iv) *T as an operator on $\mathrm{L}^p(\mathrm{X})$ is irreducible.*

(v) *$P_T f = \lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{j=0}^{n-1} T^j f = \int_\mathrm{X} f \cdot \mathbf{1}$ for every $f \in \mathrm{L}^1(\mathrm{X})$.*

(vi) $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \int_X (f \circ \varphi^j) \cdot g = \left(\int_X f \right) \left(\int_X g \right)$

for all $f \in L^p$, $g \in L^q$ *where* $\frac{1}{p} + \frac{1}{q} = 1$.

(vii) $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(A \cap \varphi^{*j}(B)) = \mu(A)\mu(B)$ *for all* $A, B \in \Sigma$.

(viii) $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(A \cap \varphi^{*j}(A)) = \mu(A)^2$ *for all* $A \in \Sigma$.

Proof. The equivalence of (i)–(iv) has been proved in Proposition 7.14.

(ii) $\Rightarrow$ (v): Let $f \in L^1$. Then $P_T f \in \mathrm{fix}(T)$, so $P_T f = c \cdot \mathbf{1}$ by (ii). Integrating yields $c = \int_X f$.

(v) $\Rightarrow$ (vi) follows by multiplying with g and integrating; (vi) $\Rightarrow$ (vii) follows by specialising $f = \mathbf{1}_A$ and $g = \mathbf{1}_B$, and (vii) $\Rightarrow$ (viii) is proved by specialising $A = B$.

(viii) $\Rightarrow$ (i): Let A be φ-invariant, i.e., $\varphi^*(A) = A$ in the measure algebra. Then

$$\frac{1}{n} \sum_{j=0}^{n-1} \mu(A \cap \varphi^{*j}(A)) = \frac{1}{n} \sum_{j=0}^{n-1} \mu(A) = \mu(A),$$

and hence $\mu(A)^2 = \mu(A)$ by (viii). Therefore $\mu(A) \in \{0, 1\}$, and (i) is proved. □

Remarks 8.11. a) Note that the convergence in (v) is even in the L^p-norm whenever $f \in L^p$, with $1 \le p < \infty$. This follows from Theorem 8.8.

b) Assertion (v) is, in probabilistic language, a kind of "weak law of large numbers". Namely , fix $f \in L^p(X)$ and let $X_j := f \circ \varphi^j$, $j \in \mathbb{N}_0$. Then the X_j are identically distributed random variables on the probability space (X) with common expectation $\mathbb{E}(X_j) = \int_X f =: c$. Property (iii) just says that

$$\frac{X_0 + \cdots + X_{n-1}}{n} \to c \quad \text{as} \quad n \to \infty$$

in L^1-norm. Note that this is slightly stronger than the "classical" weak law of large numbers that asserts convergence in measure (probability) only (see [Billingsley (1979)]). We shall come back to this in Section 11.3.

c) Suppose that $\mu(A) > 0$. Then dividing by $\mu(A)$ in (vii) yields

$$\frac{1}{n} \sum_{j=0}^{n-1} \mu_A\left[\varphi^j \in B\right] \to \mu(B) \qquad (n \to \infty)$$

for every $B \in \Sigma$, where μ_A is the conditional probability given A (cf. Section 6.2). This shows that in an ergodic system the original measure μ is completely determined by μ_A.

d) By a standard density argument one can replace "for all $f \in L^1(X)$" in assertion (v) by "for all f in a dense subset of $L^1(X)$". The same holds for f and

g in assertion (vi). Similarly, it suffices to take in (vii) sets from a $\cap$-stable generator of Σ (Exercise 3).

8.3 Perron's Theorem and the Ergodicity of Markov Shifts

Let $L = \{0, \dots, k-1\}$ and let $S = (s_{ij})_{0 \le i,j < k}$ be a **row-stochastic** matrix, i.e., $S \ge 0$ and $S\mathbf{1} = \mathbf{1}$, where $\mathbf{1} = (1, 1, \dots, 1)^t$. It was remarked in Example 2.3 that S is a contraction on the finite-dimensional Banach space $E := (\mathbb{C}^d, \|\cdot\|_\infty)$. In particular, it satisfies the hypotheses of Theorem 8.5, whence the limit

$$Qx := \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} S^j x$$

exists for every $y \in F := \mathrm{fix}(S) \oplus \mathrm{ran}(\mathrm{I} - S)$. (Note that in finite dimensions all linear subspaces are closed.) The dimension formula from elementary linear algebra shows that $F = E$, i.e., S is mean ergodic, with Q being the mean ergodic projection. Clearly $Q \ge 0$ and $Q\mathbf{1} = \mathbf{1}$ as well, so the rows of Q are probability vectors. Moreover, since $QS = Q$, each row of Q is a left fixed vector of S. Hence we have proved the following version of Perron's theorem, already claimed in Section 5.1 on page 64.

Theorem 8.12 (Perron). *Let S be a row-stochastic $k \times k$-matrix. Then there is at least one probability (column) vector p such that $p^t S = p^t$.*

Recall (from Section 2.4, cf. Section 7.3) that a row-stochastic matrix S is **irreducible** if for every pair of indices $(i, j) \in L \times L$ there is an $r \in \mathbb{N}_0$ such that $s_{ij}^r := [S^r]_{ij} > 0$. Furthermore, let us call a matrix $A = (a_{ij})_{i,j}$ **strictly positive** if $a_{ij} > 0$ for all indices i, j. Then we have the following characterisation.

Lemma 8.13. *For a row-stochastic $k \times k$-matrix the following assertions are equivalent.*

(i) *S is irreducible;*

(ii) *There is $m \in \mathbb{N}$ such that $(\mathrm{I} + S)^m$ is strictly positive;*

(iii) *$Q = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} S^j$ is strictly positive.*

If (i)–(iii) *hold, then* $\mathrm{fix}(S) = \mathbb{C} \cdot \mathbf{1}$*, there is a* unique *probability vector p with $p^t S = p^t$, and p is strictly positive.*

Proof. (i)$\Rightarrow$(ii): Simply expand $(\mathrm{I} + S)^m = \sum_{j=0}^m \binom{m}{j} S^j$. If S is irreducible, then for high enough $m \in \mathbb{N}$, the resulting matrix must have each entry strictly positive.

(ii)$\Rightarrow$(iii): Since $QS = S$, it follows that $Q(\mathrm{I} + S)^m = 2^m Q$. But if $(\mathrm{I} + S)^m$ is strictly positive, so is $Q(\mathrm{I} + S)^m$, since Q has no zero row.

(iii)$\Rightarrow$(i): If Q is strictly positive, then for large n the Cesàro mean $\mathrm{A}_n[S]$ must be strictly positive, and hence S is irreducible.

Suppose that (i)–(iii) hold. We claim that Q has rank 1. If y is a *column* of Q, then $Qy = y$, i.e., Q is a projection onto its column-space. Let $\varepsilon := \min_j y_j$ and $x := y - \varepsilon\mathbf{1}$. Then $x \geq 0$ and $Qx = x$, but since Q is strictly positive and x is positive, either $x = 0$ or x is also strictly positive. The latter is impossible by construction of x, and hence $x = 0$. This means that all entries of y are equal. Since y was an arbitrary column of Q, the claim is proved.

It follows that each column of Q is a strictly positive multiple of $\mathbf{1}$, and so all the rows of Q are identical. That means, there is a strictly positive p such that $Q = \mathbf{1} \cdot p^t$. We have seen above that $p^t S = p^t$. And if q is any probability vector with $q^t S = q^t$, then $q^t = q^t Q = q^t \mathbf{1} p^t = p^t$. □

We remark that in general $\dim \mathrm{fix}(S) = 1$ does *not* imply irreducibility of S, see Exercise 4.

By the lemma, an irreducible row-stochastic matrix S has a *unique* fixed probability vector p, being strictly positive. We are now in the position to prove the following characterisation.

Theorem 8.14. *Let S be a row-stochastic $k \times k$-matrix with fixed probability vector p. Then p is strictly positive and the Markov shift $(\mathscr{W}_k^+, \Sigma, \mu(S,p); \tau)$ is ergodic if and only if S is irreducible.*

Proof. Let $i_0, \ldots, i_l \in L$ and $j_0, \ldots, j_r \in L$. Then for $n \geq 1$ we have

$$\begin{aligned}\mu\big(\{i_0\} \times \cdots \times \{i_l\} \times L^{n-1} \times \{j_0\} \times \cdots \times \{j_r\} \times \prod L\big) \\ = p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} s_{i_l j_0}^n s_{j_0 j_1} \ldots s_{j_{r-1} j_r}\end{aligned}$$

as a short calculation using (5.2) and involving the relevant indices reveals. If we consider the cylinder sets

$$E := \{i_0\} \times \cdots \times \{i_l\} \times \prod L, \qquad F := \{j_0\} \times \cdots \times \{j_r\} \times \prod L,$$

then if $n > l$

$$[\,\tau^n \in F\,] \cap E = \{i_0\} \times \cdots \times \{i_l\} \times L^{n-l-1} \times \{j_0\} \times \cdots \times \{j_r\} \times \prod L,$$

and hence

$$\mu([\,\tau^n \in F\,] \cap E) = p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} s_{i_l j_0}^{n-l} s_{j_0 j_1} \ldots s_{j_{r-1} j_r}.$$

By taking Cesàro averages of this expression and by splitting off the first summands we see that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu([\,\tau^n \in F\,] \cap E) = p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} q_{i_l j_0} s_{j_0 j_1} \ldots s_{j_{r-1} j_r}.$$

If S is irreducible, we know from Lemma 8.13 that $q_{i_l j_0} = p_{j_0}$, and hence

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mu([\,\tau^n\in F\,]\cap E)=\mu(E)\mu(F).$$

Since cylinder sets form a dense subalgebra of Σ, by Remark 8.11.d we conclude that (vii) of Theorem 8.10 holds, i.e., the MDS is ergodic.

Conversely, suppose that p is strictly positive and the MDS is ergodic and fix $i,j\in L$. Specialising $E=\{i\}\times\prod L$ and $F=\{j\}\times\prod L$ above, by Theorem 8.10 we obtain

$$p_ip_j=\mu(E)\mu(F)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mu([\,\tau^n\in F\,]\cap E)=p_iq_{ij}.$$

Since $p_i>0$ by assumption, $q_{ij}=p_j>0$. Hence Q is strictly positive, and this implies that S is irreducible by Lemma 8.13. □

8.4 Mean Ergodic Operators

As a matter of fact, the concept of *mean ergodicity* has applications far beyond Koopman operators coming from MDSs. In Section 8.1 we introduced that concept for bounded linear operators on general Banach spaces, and a short inspection shows that one can study mean ergodicity under more general hypotheses.

Remark 8.15. The assertions of Lemma 8.2 and Lemma 8.3 remain valid if T is merely a continuous linear operator on a Hausdorff topological vector space E. Some of these statements, e.g. the fundamental identity $\mathrm{A}_n-T\mathrm{A}_n=(1/n)(\mathrm{I}-T^n)$ even hold if E is merely a *convex subset* of a Hausdorff topological vector space and $T:E\to E$ is *affine* (meaning $Tf=Lf+g$, with $g\in E$ and $L:E\to E$ linear and continuous).

Accordingly, and in coherence with Definition 8.4, a continuous linear operator $T:E\to E$ on a Hausdorff topological vector space E is called **mean ergodic** if

$$P_Tf:=\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}T^jf$$

exists for every $f\in E$. Up to now we have seen three instances of mean ergodic operators:

1) Contractions on Hilbert spaces (Theorem 8.6);
2) Koopman operators on L^p, $1\le p<\infty$, associated with MDSs (Theorem 8.8);
3) Row-stochastic matrices on $\mathbb{C}^d$ (Section 8.3).

And we shall see more in this and later chapters:

3) Power-bounded operators on reflexive spaces (Theorem 8.22);
4) Dunford–Schwartz operators on finite measure spaces (Theorem 8.24);

5) The Koopman operator on $\mathrm{C}(\mathbb{T})$ of a rotation on $\mathbb{T}$ (Proposition 10.10 and on $\mathrm{R}[0,1]$ of a mod 1 translation on $[0,1)$ (Proposition 10.19);

6) The Koopman operator on $\mathrm{C}(G)$ of a rotation on a compact group G (Corollary 10.12).

In this section we shall give a quite surprising characterisation for mean ergodicity of bounded linear operators on general Banach spaces.

We begin by showing that the hypotheses of Theorem 8.5 are natural. Recall the notation

$$\mathrm{A}_n = \mathrm{A}_n[T] = \frac{1}{n}\sum_{j=0}^{n-1} T^j \qquad (n \in \mathbb{N})$$

for the Cesàro-averages. A bounded operator T on a Banach space E is called **Cesàro-bounded** if $\sup_{n\in\mathbb{N}} \|\mathrm{A}_n\| < \infty$.

Lemma 8.16. *If E is a Banach space and $T \in \mathscr{L}(E)$ is mean ergodic, then T is Cesàro-bounded and $T^n f/n \to 0$ for every $f \in E$.*

Proof. As $\lim_{n\to\infty} \mathrm{A}_n f$ exists for every $f \in E$ and E is a Banach space, it follows from the uniform boundedness principle (Theorem C.1) that $\sup_{n\in\mathbb{N}} \|\mathrm{A}_n\| < \infty$. From Lemma 8.2.b we have $TP_T = P_T$, i.e., $(\mathrm{I}-T)\mathrm{A}_n f \to 0$. The identity $(\mathrm{I}-T)\mathrm{A}_n = (1/n)\,\mathrm{id} - (1/n)T^n$ implies that $T^n f/n \to 0$ for every $f \in E$. □

The next lemma exhibits in a very general fashion the connection between fixed points of an operator T and certain cluster points of the sequence $(\mathrm{A}_n f)_{n\in\mathbb{N}}$. We shall need it also in our proof of the Markov–Kakutani fixed point Theorem 10.1 below.

Lemma 8.17. *Let C be a convex subset of a Hausdorff topological vector space, $T : C \to C$ be a continuous affine mapping and $g \in C$. Then $Tg = g$ if and only if there is $f \in C$ and a subsequence $(n_j)_{j\in\mathbb{N}}$ such that*

$$(1) \quad T^{n_j} f/n_j \to 0 \qquad \textit{and} \qquad (2) \quad g \in \bigcap\nolimits_{k\in\mathbb{N}} \overline{\{\mathrm{A}_{n_j} f : j \geq k\}}.$$

Proof. If $Tg = g$, then (1) and (2) hold for $f = g$ and every subsequence $(n_j)_{j\in\mathbb{N}}$. Conversely, suppose that (1) and (2) hold for some $f \in E$ and a subsequence $(n_j)_{j\in\mathbb{N}}$. Then

$$g - Tg = (\mathrm{I}-T)g \in \overline{\{(\mathrm{I}-T)\mathrm{A}_{n_j} f : j \geq k\}} = \overline{\{(f - T^{n_j} f)/n_j : j \geq k\}}$$

for every $k \in \mathbb{N}$, cf. Remark 8.15. By (1), $(f - T^{n_j} f)/n_j \to 0$ as $j \to \infty$, and since C is a Hausdorff space, 0 is the only cluster point of the convergent sequence $((f - T^{n_j} f)/n_j)_{j\in\mathbb{N}}$. This implies that $g - Tg = 0$, i.e., $Tg = g$. □

The next auxiliary result is essentially a consequence of Lemma 8.17 and Mazur's Theorem C.7.

Proposition 8.18. *Let T be a Cesàro-bounded operator on some Banach space E such that $T^n h/n \to 0$ for each $h \in E$. Then for $f, g \in E$ the following statements are equivalent.*

(i) $\mathrm{A}_n f \to g$ *in the norm of* E *as* $n \to \infty$.

(ii) $\mathrm{A}_n f \to g$ *weakly as* $n \to \infty$.

(iii) g *is a weak cluster point of a subsequence of* $(\mathrm{A}_n f)_n$.

(iv) $g \in \mathrm{fix}(T) \cap \overline{\mathrm{conv}}\{T^n : n \geq 0\}$.

(v) $g \in \mathrm{fix}(T)$ *and* $f - g \in \overline{\mathrm{ran}}(\mathrm{I} - T)$.

Proof. The implication (v)$\Rightarrow$(i) follows from Theorem 8.5 and the implications (i)$\Rightarrow$(ii)$\Rightarrow$(iii) are trivial. If (iii) holds, then $g \in \mathrm{fix}(T)$ by Lemma 8.17. Moreover,

$$g \in \mathrm{cl}_\sigma\{T^n f : n \geq 0\} \subseteq \mathrm{cl}_\sigma \mathrm{conv}\{T^n f : n \geq 0\} = \overline{\mathrm{conv}}\{T^n f : n \geq 0\}$$

by Mazur's theorem. Finally, suppose that (iv) holds and $\sum_n t_n T^n f$ is any convex combination of the vectors $T^n f$. Then

$$f - \sum_n t_n T^n f = \sum_n t_n (f - T^n f) = (\mathrm{I} - T) \sum_n n t_n \mathrm{A}_n f \in \mathrm{ran}(\mathrm{I} - T)$$

by Lemma 8.2.c). It follows that $f - g \in \overline{\mathrm{ran}}(\mathrm{I} - T)$, as was to be proved. □

In order to formulate our main result, we use the following terminology.

Definition 8.19. Let E be a Banach space, and let $F \subseteq E$ and $G \subseteq E'$ be linear subspaces. Then G **separates the points** of F if for any $0 \neq f \in F$ there is $g \in G$ such that $\langle f, g \rangle \neq 0$. The property that F separates the points of G is defined analogously.

Note that by the Hahn–Banach theorem C.3, E' separates the points of E. Our main result is now a characterisation of mean ergodic operators on Banach spaces.

Theorem 8.20 (Mean Ergodic Operators). *Let T be a Cesàro-bounded operator on some Banach space E such that $T^n f/n \to 0$ weakly for each $f \in E$. Further, let $D \subseteq E$ be a dense subset of E. Then the following assertions are equivalent.*

(i) T *is mean ergodic.*

(ii) T *is weakly mean ergodic, i.e.,* weak-$\lim_{n\to\infty} \mathrm{A}_n f$ *exists for each* $f \in E$.

(iii) *For each* $f \in D$ *the sequence* $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ *has a subsequence with a weak cluster point.*

(iv) $\overline{\mathrm{conv}}\{T^j f : j \in \mathbb{N}_0\} \cap \mathrm{fix}(T) \neq \emptyset$ *for each* $f \in D$.

(v) $\mathrm{fix}(T)$ *separates the points of* $\mathrm{fix}(T')$.

(vi) $E = \mathrm{fix}(T) \oplus \overline{\mathrm{ran}}(I - T)$.

Proof. The equivalence of (i)–(iv) and (vi) is immediate from Proposition 8.18 and Theorem 8.5. Suppose that (vi) holds and $0 \neq f' \in \mathrm{fix}(T')$. Then there is $f \in E$ such

that $\langle f, f' \rangle \neq 0$. By (vi) we can write $f = g + h$ with $h \in \overline{\operatorname{ran}}(\mathrm{I} - T)$. Since $T'f' = f'$, f' vanishes on h by Exercise 5, and hence we must have $\langle g, f' \rangle \neq 0$.

Conversely, suppose that (v) holds and consider the space F of vectors where the Cesàro averages converge. By Theorem 8.5 $F = \operatorname{fix}(T) \oplus \overline{\operatorname{ran}}(I - T)$ is closed in E. We employ the Hahn–Banach theorem C.3 to show that F is dense in E. Suppose that $f' \in E'$ vanishes on F. Then f' vanishes in particular on $\operatorname{ran}(\mathrm{I} - T)$ and this just means that $f' \in \operatorname{fix}(T')$ (see Exercise 5). Moreover, f' vanishes also on $\operatorname{fix}(T)$, which by (v) separates $\operatorname{fix}(T')$. This forces $f' = 0$. □

Remarks 8.21. 1) Condition (v) in Theorem 8.20 can also be expressed as $\operatorname{fix}(T') \cap \operatorname{fix}(T)^{\perp} = \{0\}$, cf. Exercise 5.

2) Under the assumptions of Theorem 8.20, $\operatorname{fix}(T')$ always separates the points of $\operatorname{fix}(T)$, see Exercise 6.

3) If $T \in \mathscr{L}(E)$ is mean ergodic, its adjoint T' need not be mean ergodic with respect to the norm topology, see Exercise 7. However, it is mean ergodic with respect to the weak* topology, i.e., the Cesàro means $\mathrm{A}_n[T']$ converge pointwise in the weak*-topology to $P_{T'} = P'_T$. The proof of this is Exercise 8.

In the following we discuss two classes of examples.

Power-bounded Operators on Reflexive Spaces

An operator $T \in \mathscr{L}(E)$ is called **power-bounded** if $\sup_{n \in \mathbb{N}} \|T^n\| < \infty$. Contractions are obviously power-bounded. Moreover, a power-bounded operator T is Cesàro-bounded and satisfies $T^n/n \to 0$. Hence Theorem 8.20 applies in particular to power-bounded operators. We arrive at a famous result due Yosida [Yosida (1938)], Kakutani [Kakutani (1938)] and Lorch [Lorch (1939)].

Theorem 8.22. *Every power-bounded linear operator on a reflexive Banach space is mean ergodic.*

Proof. Let $f \in E$. Then, by power-boundedness of T the set $\{\mathrm{A}_n f : n \in \mathbb{N}\}$ is norm-bounded. Since E is reflexive, it is even relatively weakly compact, hence the sequence $(\mathrm{A}_n f)_n$ has a weak cluster point. □

Dunford–Schwartz Operators

Let X and Y be measure spaces. An bounded operator

$$T : \mathrm{L}^1(\mathrm{X}) \to \mathrm{L}^1(\mathrm{Y})$$

is called a **Dunford–Schwartz operator** or a **complete contraction**, if

$$\|Tf\|_1 \leq \|f\|_1 \quad \text{and} \quad \|Tf\|_\infty \leq \|f\|_\infty \qquad (f \in \mathrm{L}^1 \cap \mathrm{L}^\infty).$$

By denseness, a Dunford–Schwartz operator is an L^1-contraction. The following result states that T "interpolates" to a contraction on each L^p-space.

Theorem 8.23. *Let $T : \mathrm{L}^1(\mathrm{X}) \to \mathrm{L}^1(\mathrm{Y})$ be a Dunford–Schwartz operator. Then*

$$\|Tf\|_p \le \|f\|_p \qquad \text{for all } f \in \mathrm{L}^p \cap \mathrm{L}^1,\ 1 \le p \le \infty.$$

Proof. The claim is a direct consequence of the Riesz–Thorin interpolation theorem [Folland (1999), Thm. 6.27]. If T is *positive*, there is a more elementary proof, which we give for convenience. For $p = 1, \infty$ there is nothing to show, so let $1 < p < \infty$. Let $f \in \mathrm{L}^p \cap \mathrm{L}^1$ such that $A := [f \neq 0]$ has finite measure. Then, by Theorem 7.19 and Exercise 7.16 and since $T\mathbf{1}_A \le \mathbf{1}$,

$$|Tf|^p = |T(f\mathbf{1}_A)|^p \le (T|f|^p) \cdot (T\mathbf{1}_A)^{p/q} \le (T|f|^p) \cdot \mathbf{1}^{p/q} = T|f|^p.$$

Integrating yields $\|Tf\|_p^p \le \|T|f|^p\|_1 \le \|f\|_p^p$. Finally, a standard density argument completes the proof. □

Using interpolation we can prove the following result about mean ergodicity.

Theorem 8.24. *Any Dunford–Schwartz operator T on $\mathrm{L}^1(\mathrm{X})$ over a finite measure space X is mean ergodic.*

Proof. By Theorem 8.23, T restricts to a contraction on L^2, which is a Hilbert space. By Theorem 8.6, T is mean ergodic on L^2, which means that for $f \in \mathrm{L}^2$ the limit $\lim_{n\to\infty} \mathrm{A}_n[T]f$ exists in $\|\cdot\|_2$. As the measure is finite, this limit exists in $\|\cdot\|_1$. Since L^2 is dense in L^1, the claim follows. □

Employing Theorem 8.20 we can give a second proof.

Alternative proof of Theorem 8.24: Let $\mathrm{B} := \{f \in \mathrm{L}^\infty : \|f\|_\infty \le 1\}$. View L^∞ as the dual of L^1 and equip it with the $\sigma(\mathrm{L}^\infty, \mathrm{L}^1)$-topology. By the Banach–Alaoglu theorem, B is weak*-compact. Since the embedding $(\mathrm{L}^\infty, \sigma(\mathrm{L}^\infty, \mathrm{L}^1)) \subseteq (\mathrm{L}^1, \sigma(\mathrm{L}^1, \mathrm{L}^\infty))$ is (obviously) continuous, B is weakly compact in L^1. In addition, B is invariant under the Cesàro averages A_n of T, and hence the sequence $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ has a weak cluster point for each $f \in \mathrm{B}$. Thus (iii) from Theorem 8.20 is satisfied with $D := \mathrm{L}^\infty = \bigcup_{c>0} c\mathrm{B}$. □

Example 8.25. The assumption of a *finite* measure space is crucial in Theorem 8.24. Consider $E = \mathrm{L}^1(\mathbb{R})$ where $\mathbb{R}$ is endowed with the Lebesgue measure, and $T \in \mathscr{L}(E)$, the shift by 1, i.e., $Tf(x) = f(x+1)$. Then $\dim \operatorname{fix} T = 0$ and $\dim \operatorname{fix} T' = \infty$, so T is not mean ergodic, see also Exercise 7.

Supplement: Operations with Mean Ergodic Operators

Let us illustrate how the various conditions in Theorem 8.20 can be used to check mean ergodicity, and thus allow us to carry out certain constructions for mean ergodic operators.

Powers of Mean Ergodic Operators

Theorem 8.26. *Let E be a Banach space and let $S \in \mathscr{L}(E)$ be a power-bounded mean ergodic operator. Let T be a k^{th}-root of S, i.e., $T^k = S$ for some $k \in \mathbb{N}$. Then T is also mean ergodic.*

Proof. Denote by P_S the mean ergodic projection of S. Define $P := \big(\frac{1}{k}\sum_{j=0}^{k-1} T^j\big)P_S$ and observe that $Pf \in \overline{\operatorname{conv}}\{T^j f : j \in \mathbb{N}_0\}$ for all $f \in E$. Since T commutes with S, it commutes also with P_S. We now obtain

$$PT = TP = \Big(\frac{1}{k}\sum_{j=0}^{k-1} T^{j+1}\Big)P_S = P,$$

since $T^k P_S = SP_S = P_S$, and $Pf \in \overline{\operatorname{conv}}\{T^j f : j \in \mathbb{N}_0\} \cap \operatorname{fix}(T)$ by the above. So Theorem 8.20 (iv) implies that T is mean ergodic. It follows also that P is the corresponding mean ergodic projection. □

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

Example 8.27. Take $E = \mathrm{c}$, the space of convergent sequences, and the multiplication operator

$$M : \mathrm{c} \to \mathrm{c}, \qquad (x_n)_n \mapsto (a_n x_n)_n$$

for some sequence $(a_n)_{n=1}^{\infty}$ with $1 \neq a_n \to 1$. Now $\operatorname{fix}(M) = \{0\}$, whereas $\operatorname{fix}(M')$ contains f' defined by $\langle f', (x_n)_{n=1}^{\infty}\rangle := \lim_{n\to\infty} x_n$. By Theorem 8.20 (v) we conclude that M is not mean ergodic.

Consider now a k^{th} root of unity $1 \neq b \in \mathbb{T}$ and define $T_k := bM$. Then it is easy to see that $\operatorname{fix}(T_k') = \{0\}$, and hence, again by Theorem 8.20 (v), T_k is mean ergodic. It follows from Theorem 8.26 that T_k^k is not mean ergodic. Now one can employ a direct sum construction to obtain a Banach space E and a mean ergodic operator $T \in \mathscr{L}(E)$ with no power T^k, $k \geq 2$, being mean ergodic (Exercise 11).

Convex Combinations of Mean Ergodic Operators

Other examples of "new" mean ergodic operators can be obtained by convex combinations of mean ergodic operators. Our first lemma, a nice application of the Kreĭn–Milman Theorem, is due to Kakutani.

Lemma 8.28. *Let E be a Banach space. Then the identity operator* I *is an extreme point of the closed unit ball in $\mathscr{L}(E)$.*

Proof. Take $T \in \mathscr{L}(E)$ such that $\|\mathrm{I}+T\| \leq 1$ and $\|\mathrm{I}-T\| \leq 1$. Then the same is true for the adjoints: $\|\mathrm{I}'+T'\| \leq 1$ and $\|\mathrm{I}'-T'\| \leq 1$. For $f' \in E'$ define $f_1' := (\mathrm{I}'+T')f'$, $f_2' := (\mathrm{I}'-T')f'$ and conclude that $f' = \frac{1}{2}(f_1'+f_2')$ and $\|f_1'\|, \|f_2'\| \leq \|f'\|$. Suppose now that f' is an extreme point of the unit ball in E'. Then we obtain $f' = f_1' = f_2'$ and

hence $T'f' = 0$. The Banach–Alaoglu Theorem C.4 and the Kreĭn–Milman Theorem C.13 imply that $T' = 0$, and hence $T = 0$.

Now suppose that $\mathrm{I} = \frac{1}{2}(R+S)$ for contractions $R,S \in \mathscr{L}(E)$, and define $T := \mathrm{I} - R$. This implies $\mathrm{I} - T = R$ and $\mathrm{I} + T = 2\mathrm{I} - R = S$, so $\|\mathrm{I} - T\| \leq 1$, $\|\mathrm{I} + T\| \leq 1$. By the above considerations it follows that $T = 0$, i.e., $\mathrm{I} = R = S$. □

Lemma 8.29. *Let R,S be two power-bounded, commuting operators on a Banach space E. Consider*

$$T := tR + (1-t)S$$

for some $0 < t < 1$. Then for the fixed spaces $\mathrm{fix}(T)$, $\mathrm{fix}(R)$ and $\mathrm{fix}(S)$ we have

$$\mathrm{fix}(T) = \mathrm{fix}(R) \cap \mathrm{fix}(S).$$

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

$$\mathrm{I}_{\mathrm{fix}(T)} = T|_{\mathrm{fix}(T)} = tR|_{\mathrm{fix}(T)} + (1-t)S|_{\mathrm{fix}(T)}$$

and $R|_{\mathrm{fix}(T)}, S|_{\mathrm{fix}(T)} \in \mathscr{L}(\mathrm{fix}(T))$, since R and S commute. Lemma 8.28 implies that $R|_{\mathrm{fix}(T)} = S|_{\mathrm{fix}(T)} = \mathrm{I}_{\mathrm{fix}(T)}$, i.e., $\mathrm{fix}(T) \subseteq \mathrm{fix}(R) \cap \mathrm{fix}(S)$. □

Now we can prove the main result of this section.

Theorem 8.30. *Let E be a Banach space and R,S two power-bounded, mean ergodic, commuting operators on E. Then for each $0 \leq t \leq 1$ the convex combination $T := tR + (1-t)S$ is again mean ergodic.*

Proof. Let $0 < t < 1$. By Lemma 8.29 we have $\mathrm{fix}(T) = \mathrm{fix}(R) \cap \mathrm{fix}(S)$ and $\mathrm{fix}(T') = \mathrm{fix}(R') \cap \mathrm{fix}(S')$. By Theorem 8.20 (v) it suffices to show that $\mathrm{fix}(R) \cap \mathrm{fix}(S)$ separates $\mathrm{fix}(R') \cap \mathrm{fix}(S')$. To this end pick $0 \neq f' \in \mathrm{fix}(R') \cap \mathrm{fix}(S')$. Then there is $f \in \mathrm{fix}(R)$ with $\langle f, f' \rangle \neq 0$. Since $S\,\mathrm{fix}(R) \subseteq \mathrm{fix}(R)$, we have $P_S f \in \mathrm{fix}(R) \cap \mathrm{fix}(S)$ where P_S denotes the mean ergodic projection corresponding to S. Consequently, we have (see also Remark 8.21.3))

$$\langle P_S f, f' \rangle = \langle f, P_S' f' \rangle = \langle f, P_{S'} f' \rangle = \langle f, f' \rangle \neq 0.$$ □

The following are immediate consequences of the above.

Corollary 8.31. *Let T be a convex combination of two power-bounded, mean ergodic, commuting operators R and $S \in \mathscr{L}(E)$. Denote by P_R, resp. P_S the mean ergodic projections. Then the projection P_T corresponding to T is obtained as*

$$P_T = P_R P_S = P_S P_R = \lim_{n\to\infty} (R_n S_n).$$

Corollary 8.32. *Let $\{R_j\}_{j=1}^m$ be a family of power-bounded, mean ergodic, commuting operators. Then every convex combination $T := \sum_{j=1}^m t_j R_j$ is mean ergodic.*

Exercises

1 (Strong Convergence Lemma). Let E, F be Banach spaces and let $(T_n)_{n\in\mathbb{N}} \subseteq \mathscr{L}(E;F)$. Suppose that there is $M \geq 0$ such that $\|T_n\| \leq M$ for all $n \in \mathbb{N}$. Show that

$$G := \{f \in E : \lim_{n\to\infty} T_n f \text{ exists}\}$$

is a closed subspace of E, and

$$T : G \to F, \qquad Tf := \lim_{n\to\infty} T_n f$$

is a bounded linear operator with $\|T\| \leq \liminf_{n\to\infty} \|T_n\|$.

2. Let H be a Hilbert space and let $P \in \mathscr{L}(H)$ with $P^2 = P$. Show that P is an orthogonal projection if and only if P is self-adjoint if and only if $\|P\| \leq 1$. (See also the supplement to Chapter 13.)

3. Let $(\mathrm{X};\varphi)$ be an MDS, $\mathrm{X} = (X, \Sigma, \mu)$, with $\mathcal{E}$ a $\cap$-stable generator of Σ. Suppose that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n} \mu(\varphi^{*n} A \cap A) = \mu(A)^2$$

for all $A \in \mathcal{E}$. Show that $(\mathrm{X};\varphi)$ is ergodic. (Hint: Use Lemma B.16.)

4. Show that the matrix

$$S = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

is not irreducible, but $\mathrm{fix}(S)$ is one-dimensional. Show that there is a unique probability vector p such that $p^t A = p^t$. Then show that the Markov shift associated with S is ergodic.

5. Let E be a Banach space. For $F \subseteq E$ and $G \subseteq E'$ one defines

$$F^\perp := \left\{f' \in E' : \langle f, f' \rangle = 0 \text{ for all } f \in F\right\} \quad \text{and} \quad G^\top := G^\perp \cap E.$$

For $T \in \mathscr{L}(E)$ show that

$$\mathrm{fix}(T') = \mathrm{ran}(\mathrm{I} - T)^\perp \quad \text{and} \quad \mathrm{fix}(T) = \mathrm{ran}(\mathrm{I} - T')^\top.$$

6. Let T be a Cesàro-bounded linear operator on a Banach space E, and suppose that $T^n f/n \to 0$ as $n \to \infty$ for every $f \in E$. Show that $\mathrm{fix}(T')$ always separates the points of $\mathrm{fix}(T)$.

(Hint: Take $0 \neq f \in \mathrm{fix}(T)$ and consider the set $K := \{f' \in E' : \|f'\| \leq 1, \langle f, f' \rangle = \|f\|\}$. Then K is nonempty by the Hahn–Banach theorem, $\sigma(E', E)$-closed and norm-bounded. Then use Lemma 8.17 to show that $\mathrm{fix}(T') \cap K \neq \emptyset$).

7 (Shift). For a scalar sequence $x = (x_n)_{n\in\mathbb{N}_0}$ the **left shift** L and the **right shift** R are defined by

$$L(x_0,x_1,\dots) := (x_1,x_2,\dots) \quad \text{and} \quad R(x_0,x_1,\dots) := (0,x_0,x_1,x_2,\dots).$$

Clearly, a fixed vector of L must be constant, while 0 is the only fixed vector of R. Prove the following assertions:

a) *The left and the right shift are mean ergodic on each* $E = \ell^p$, $1 < p < \infty$.

b) *The left shift on* $E = \ell^1$ *is mean ergodic, while the right shift is not.*

c) *The left and right shift are mean ergodic on* $E = \mathrm{c}_0$, *the space of null sequences.*

d) *The left shift* is *while the right shift* is not *mean ergodic on* $E = \mathrm{c}$, *the space of convergent sequences.*

e) *Neither the left nor the right shift is mean ergodic on* $E = \ell^\infty$.

8. Let T be a mean ergodic operator on a Banach space E with associated mean ergodic projection $P : E \to \operatorname{fix}(T)$. Show that $\mathrm{A}_n' \to P'$ in the weak*-topology and that P' is a projection with $\operatorname{ran}(P') = \operatorname{fix}(T')$.

9. Let X be a finite measure space and let T be a positive operator on $\mathrm{L}^1(\mathrm{X})$. We identify $\mathrm{L}^\infty(\mathrm{X})$ with $\mathrm{L}^1(\mathrm{X})'$, the dual space of $\mathrm{L}^1(\mathrm{X})$. Show that T is a Dunford–Schwartz operator if and only if $T'\mathbf{1} \le \mathbf{1}$ and $T\mathbf{1} \le \mathbf{1}$.

10. Show that the following operators are *not* mean ergodic:

a) $E = \mathrm{C}[0,1]$ and $(Tf)(x) = f(x^2)$, $x \in [0,1]$. (Hint: Determine $\operatorname{fix}(T)$ and $\operatorname{fix}(T')$, cf. Chapter 3.)

b) $E = \mathrm{C}[0,1]$ and $(Tf)(x) = xf(x)$, $x \in [0,1]$. (Hint: Look at $\mathrm{A}_n\mathbf{1}$, $n \in \mathbb{N}$.)

11. Work out the details of Example 8.27.

Chapter 9
Mixing Dynamical Systems

In the present chapter we consider two mathematical formalisations of the intuitive concept that iterating the dynamics φ provides a "just mixing" of the space (X,Σ,μ). Whereas the main theme of the previous chapter, the mean ergodicity of the associated Koopman operator, involved just the norm topology, now the weak topology of the associated L^p-spaces becomes important.

The dual space E' of a Banach space E and the associated weak topology on a Banach space E have already been touched upon briefly in Section 8.4. For general background about these notions see Appendix C.4 and C.6. In particular, keep in mind that we write $\langle x,x'\rangle$ for the action of $x'\in E'$ on $x\in E$.

In the case that $E=\mathrm{L}^p(\mathrm{X})$ for some probability space $\mathrm{X}=(X,\Sigma,\mu)$ and $p\in[1,\infty)$, a standard result from functional analysis states that the dual space E' can be identified with $\mathrm{L}^q(\mathrm{X})$ via the canonical duality

$$\langle f,g\rangle := \int_{\mathrm{X}} f\,g = \int_X f\,g\,\mathrm{d}\mu \qquad (f\in\mathrm{L}^p(\mathrm{X}), g\in\mathrm{L}^q(\mathrm{X})),$$

see, e.g. [Rudin (1987), Theorem 6.16]. Here $q\in(1,\infty]$ denotes the **conjugate exponent** to p, i.e., it satisfies $1/p+1/q=1$. (We use this as a standing terminology. Also, if the measure space is understood, we often write $\mathrm{L}^p,\mathrm{L}^q$ in order to increase readability.)

In the symbolism of the L^p-L^q-duality, the integral of $f\in\mathrm{L}^1(\mathrm{X})$ reads

$$\int_{\mathrm{X}} f = \int_X f\,\mathrm{d}\mu = \langle f,\mathbf{1}\rangle = \langle \mathbf{1},f\rangle\,.$$

In particular, $\mu(A)=\langle \mathbf{1}_A,\mathbf{1}\rangle=\langle\mathbf{1},\mathbf{1}_A\rangle$ for $A\in\Sigma$. The connection with the standard inner product on the Hilbert space $\mathrm{L}^2(\mathrm{X})$ is

$$(f\,|\,g) = \int_{\mathrm{X}} f\,\overline{g} = \langle f,\overline{g}\rangle \qquad (f,g\in\mathrm{L}^2).$$

After these prelimiaries we can now go *medias in res*.

9.1 Strong Mixing

Following Halmos [Halmos (1956), p. 36] let us consider an MDS $(\mathrm{X};\varphi)$, $\mathrm{X} = (X,\Sigma,\mu)$, that models the action of a particular way of stirring the contents of a glass (of total volume $\mu(X) = 1$) filled with two different liquids, say wine and water. If $A \subseteq X$ is the region originally occupied by the wine and B is any other part of the glass, then, after n repetitions of the stirring operation, the relative amount of wine in B is

$$a_n(B) := \mu(\varphi^{*n}B \cap A)/\mu(B).$$

For a thorough "mixing" of the two liquids, one would require that eventually the wine is "equally distributed" within the glass. More precisely, for large $n \in \mathbb{N}$ the relative amount $a_n(B)$ of wine in B should be close to the relative amount $\mu(A)$ of wine in the whole glass. These considerations lead to the following mathematical concept.

Definition 9.1. An MDS $(\mathrm{X};\varphi)$, $\mathrm{X} = (X,\Sigma,\mu)$, is called **strongly mixing** (or just **mixing**) if for every $A,B \in \Sigma$ one has

$$\mu(\varphi^{*n}A \cap B) \to \mu(A)\mu(B) \qquad \text{as } n \to \infty.$$

Example 9.2 (Bernoulli Shifts). It was shown in Proposition 6.16 that Bernoulli shifts are strongly mixing.

Each mixing system is ergodic. Indeed, by Theorem 8.10 the MDS $(\mathrm{X};\varphi)$ is ergodic if and only if

$$\frac{1}{n}\sum_{j=0}^{n-1} \mu(\varphi^{*j}B \cap A) \to \mu(A)\mu(B) \qquad \text{for all } A,B \in \Sigma.$$

Since ordinary convergence of a sequence $(a_n)_n$ implies the convergence of the aritmetic averages $(a_1 + \cdots + a_n)/n$ to the same limit, ergodicity is a consequence of mixing.

However, mere ergodicity of the system is not strong enough to guarantee mixing. To stay in the picture from above, ergodicity means that we have $a_n(B) \to \mu(A)$ only *on the average*, but it may well happen that again and again $a_n(B)$ is quite far away from $\mu(A)$. For example, if $B = A^{\mathrm{c}}$, then we may find arbitrarily large n with $a_n(B)$ close to 0, which means that for these n the largest part of the wine is situated again in A, the region where it was located in the beginning.

Example 9.3 (Nonmixing Markov shifts). Consider the Markov shift associated with the transition matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

on two points. The Markov measure μ on $\mathscr{W}_2^+$ puts equal weight of $1/2$ on the two points

$$x_1 := (0,1,0,1,\dots) \quad \text{and} \quad x_2 := (1,0,1,0,\dots)$$

and the system oscillates between those two under the shift τ, so this system is ergodic (cf. also Section 8.3). On the other hand, if $A = \{x_1\}$, then $\mu(\varphi^{*n}A \cap A)$ is either 0 or $1/2$, depending whether n is odd or even. Hence this Markov shift is not strongly mixing. We shall characterise mixing Markov shifts in Proposition 9.8 below.

Let us denote by $\mathbf{1} \otimes \mathbf{1}$ the projection

$$\mathbf{1} \otimes \mathbf{1} : \mathrm{L}^1(\mathrm{X}) \to \mathrm{L}^1(\mathrm{X}), \qquad f \mapsto \langle f, \mathbf{1} \rangle \cdot \mathbf{1} = \Big(\int_{\mathrm{X}} f \Big) \cdot \mathbf{1}$$

onto the one-dimensional space of constant functions. Then, by Theorem 8.10, ergodicity of an MDS $(\mathrm{X}; \varphi)$ is characterised by $P_T = \mathbf{1} \otimes \mathbf{1}$, where P_T is the mean ergodic projection associated with the Koopman operator $T = T_\varphi$ on $\mathrm{L}^1(\mathrm{X})$. In other words

$$P_T = \lim_{n \to \infty} \mathrm{A}_n[T] f = \mathbf{1} \otimes \mathbf{1}$$

in the strong (equivalently, in the weak) operator topology. In an analogous fashion one can characterise strongly mixing MDSs.

Theorem 9.4. *For an MDS $(\mathrm{X}; \varphi)$, $\mathrm{X} = (X, \Sigma, \mu)$, its associated Koopman operator $T = T_\varphi$ and $p \in [1, \infty)$ the following assertions are equivalent.*

(i) *$(\mathrm{X}; \varphi)$ is strongly mixing.*

(ii) *$\mu(\varphi^{*n}A \cap A) \to \mu(A)^2$ for every $A \in \Sigma$.*

(iii) *$T^n \to \mathbf{1} \otimes \mathbf{1}$ as $n \to \infty$ in the weak operator topology on $\mathrm{L}^p(\mathrm{X})$, i.e.,*

$$\int_{\mathrm{X}} (T^n f)\, g = \langle T^n f, g \rangle \to \langle f, \mathbf{1} \rangle \cdot \langle \mathbf{1}, g \rangle = \Big(\int_{\mathrm{X}} f \Big) \cdot \Big(\int_{\mathrm{X}} g \Big)$$

for all $f \in \mathrm{L}^p(\mathrm{X})$, $g \in \mathrm{L}^q(\mathrm{X})$.

Proof. The implications (i) $\Rightarrow$ (ii) and (iii) $\Rightarrow$ (i) are trivial, so suppose that (ii) holds. Fix $A \in \Sigma$ and $k, l \in \mathbb{N}_0$, and let $f := T^k \mathbf{1}_A$, $g := T^l \mathbf{1}_A$. Then for $n \geq l - k$ we have

$$\langle T^n f, g \rangle = \Big\langle T^{n+k} \mathbf{1}_A, T^l \mathbf{1}_A \Big\rangle = \Big\langle T^{n+k-l} \mathbf{1}_A, \mathbf{1}_A \Big\rangle = \mu(\varphi^{*(n+k-l)} A \cap A).$$

By (ii) it follows that

$$\lim_{n \to \infty} \langle T^n f, g \rangle = \mu(A)^2 = \Big\langle T^k \mathbf{1}_A, \mathbf{1} \Big\rangle \Big\langle \mathbf{1}, T^l \mathbf{1}_A \Big\rangle = \langle f, \mathbf{1} \rangle \cdot \langle \mathbf{1}, g \rangle$$

Let $E_A = \overline{\mathrm{lin}}\{\mathbf{1}, T^k \mathbf{1}_A : k \geq 0\}$ in $\mathrm{L}^2(\mathrm{X})$. Then by the power-boundedness of T an approximation argument yields that the convergence

$$\langle T^n f, g \rangle \to \langle f, \mathbf{1} \rangle \cdot \langle \mathbf{1}, g \rangle \tag{9.1}$$

holds for all $f, g \in E_A$. But if $g \in E_A^{\perp}$ then, since E_A is T-invariant, $\langle T^n f, g\rangle = 0 = \langle \mathbf{1}, g\rangle$ for $f \in E_A$. Therefore (9.1) holds true for all $f \in E_A$ and $g \in \mathrm{L}^2(\mathrm{X})$. If we take $f = \mathbf{1}_A$ and $g = \mathbf{1}_B$ we arrive at (i). Moreover, another approximation argument yields (9.1) for arbitrary $f \in \mathrm{L}^p(\mathrm{X})$ and $g \in \mathrm{L}^q(\mathrm{X})$, i.e., (iii). □

Remark 9.5. In the situation of Theorem 9.4 let $D \subseteq \mathrm{L}^p$ and $F \subseteq \mathrm{L}^q$ such that $\mathrm{lin}(D)$ is norm-dense in L^p and $\mathrm{lin}(F)$ is norm-dense in L^1. Then by standard approximation arguments we can add to Theorem 9.4 the additional equivalence

(iii') $\lim_{n\to\infty} \int_{\mathrm{X}} (T^n f) \cdot g \;=\; \Big(\int_{\mathrm{X}} f\Big) \cdot \Big(\int_{\mathrm{X}} g\Big)$ *for every* $f \in D$, $g \in F$.

Based on (iii'), one can add another equivalence:

(i') $\lim_{n\to\infty} \mu(\varphi^{*n} A \cap B) = \mu(A)\mu(B)$ for all $A \in \mathcal{D}$, $B \in \mathcal{F}$',

where $\mathcal{D}, \mathcal{F}$ are given $\cap$-stable generators of Σ (use Lemma B.16).

Let T be a power-bounded operator on a Banach space E. A vector $f \in E$ is called **weakly stable** with respect to T if $T^n f \to 0$ weakly. It is easy to see that the set of weakly stable vectors

$$E_{\mathrm{ws}}(T) := \big\{ f \in E \,:\, T^n f \to 0 \text{ weakly} \big\}$$

is a closed T-invariant subspace of E, and that $\mathrm{fix}(T) \cap E_{\mathrm{ws}}(T) = \{0\}$.

Theorem 9.6. *For an MDS* $(\mathrm{X}; \varphi)$ *with Koopman operator* $T = T_\varphi$ *on* $E = \mathrm{L}^p(\mathrm{X})$, $1 \le p < \infty$, *the following assertions are equivalent.*

(i) $(\mathrm{X}; \varphi)$ *is strongly mixing.*

(iv) *For all* $f \in E$, *if* $\int_{\mathrm{X}} f = 0$ *then* $T^n f \to 0$ *weakly.*

(v) $\mathrm{fix}(T) = \mathrm{lin}\{\mathbf{1}\}$ *and* $T^{n+1} - T^n \to 0$ *in the weak operator topology.*

(vi) $E = \mathrm{lin}\{\mathbf{1}\} \oplus E_{\mathrm{ws}}(T)$.

Proof. (i)$\Leftrightarrow$(vi) and (i)$\Rightarrow$(v) is clear from (iii) of Theorem 9.4.

(v)$\Rightarrow$(vi): By mean ergodicity of T, we have $E = \mathrm{fix}(T) \oplus \overline{\mathrm{ran}}(\mathrm{I} - T)$. By (v), $\mathrm{fix}(T) = \mathrm{lin}\{\mathbf{1}\}$ and $\mathrm{ran}(\mathrm{I} - T) \subseteq E_{\mathrm{ws}}(T)$. Since $E_{\mathrm{ws}}(T)$ is closed, (iv) follows.

(vi)$\Rightarrow$(i): Write $f \in E$ as $f = c\mathbf{1} + g$ with $g \in E_{\mathrm{ws}}(T)$. Then $T^n f = c\mathbf{1} + T^n g \to c\mathbf{1}$ weakly. It follows that $\langle f, \mathbf{1}\rangle = \langle T^n f, \mathbf{1}\rangle \to c$, and this is (i), by (iii) of Theorem 9.4. □

Given an MDS $(\mathrm{X}; \varphi)$ and a fixed natural number $k \in \mathbb{N}$, then φ^k is measure-preserving, too. We thus obtain a new MDS $(\mathrm{X}; \varphi^k)$, called the k^{th} **iterate** of the original system. The next result shows that the class of (strongly) mixing systems is stable under forming iterates. In fact, we can prolong our list characterisations once more.

Corollary 9.7. *For an MDS* $(\mathrm{X}; \varphi)$ *and* $k \in \mathbb{N}$ *the following assertions are equivalent.*

(i) *The MDS* $(\mathrm{X};\varphi)$ *is strongly mixing.*

(vii) *Its* k^{th} *iterate* $(\mathrm{X};\varphi^k)$ *is strongly mixing.*

Proof. By (iv) of Theorem 9.6 it suffices to show that $E_{\mathrm{ws}}(T) = E_{\mathrm{ws}}(T^k)$. The inclusion $\subseteq$ is trivial. For the inclusion $\supseteq$ take $f \in E_{\mathrm{ws}}(T^k)$. Then for each $0 \leq j < k$ the sequence $(T^{nk+j} f)_{n\in\mathbb{N}}$ tends to 0 weakly, and hence $T^n f \to 0$ weakly as well. □

Let us turn to the characterisation of mixing Markov shifts.

Mixing of Markov Shifts

Let $(\mathscr{W}_k^+, \Sigma, \mu(S,p);\tau)$ be the Markov shift on $L := \{0,\dots,k-1\}$ associated with a row-stochastic matrix $S = (s_{ij})_{0\leq i,j<k}$ and a fixed probability vector $p = (p_j)_{0\leq j<k}$. Since ergodicity is necessary for strong mixing, we suppose in addition that S is irreducible (cf. Theorem 8.14). As in Section 8.3 we introduce

$$Q := \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} S^j.$$

Then Q is strictly positive and each row of Q is equal to p.

Proposition 9.8. *In the situation described above, the following assertions are equivalent.*

(i) *The Markov shift* $(\mathscr{W}_k^+, \Sigma, \mu(S,p);\tau)$ *is strongly mixing.*

(ii) $S^n \to Q$ *as* $n \to \infty$.

(iii) *There is* $n \geq 0$ *such that* S^n *is strictly positive.*

(iv) $\sigma(S) \cap \mathbb{T} = \{1\}$.

Proof. Since Q is strictly positive, (ii) implies (iii). That (iv) implies (ii) can be seen by using the ergodic decomposition $\mathbb{C}^k = \mathrm{fix}(S) \oplus \mathrm{ran}(\mathrm{I} - S)$ and noting that S has spectral radius *strictly* less than 1 when restricted to the second direct summand. For the implication (iii)$\Rightarrow$(iv) we may suppose without loss of generality that S itself is strictly positive. Indeed, suppose we know that $\sigma(A) \cap \mathbb{T} = \{1\}$ for every strictly positive row-stochastic matrix A. Now, if we take $A = S^n$, we obtain

$$\sigma(S)^n \cap \mathbb{T} = \sigma(S^n) \cap \mathbb{T} = \{1\}$$

for all large $n \in \mathbb{N}$. (We used here that the product of a strictly positive matrix with a positive one is again strictly positive.) This implies that eventually $\lambda^n = \lambda^{n+1} = 1$ for every peripheral eigenvalue λ of S, hence (iv) holds.

So suppose that A is a strictly positive row-stochastic matrix. Then there is $\varepsilon \in (0,1)$ and a row-stochastic matrix T such that $A = (1-\varepsilon)T + \varepsilon E$, where $E = (1/k)\mathbf{1}\cdot\mathbf{1}^\top$ is the matrix having each entry equal to $1/k$. Since row-stochastic matrices act on row vectors as contractions for the 1-norm, we have

$$\left\|x^tA - y^tA\right\|_1 = (1-\varepsilon)\left\|(x-y)^tT\right\|_1 \le (1-\varepsilon)\left\|x-y\right\|_1$$

for all probability vectors $x, y \in \mathbb{C}^k$. Banach's fixed-point theorem yields that $\lim_{n\to\infty} x^tA^n$ must exist for every probability vector $x \in \mathbb{C}^k$, hence for *every* $x \in \mathbb{C}^k$. Therefore 1 is the only peripheral eigenvalue of A. This was the missing piece of puzzle in the above.

Suppose now that (ii) holds and let

$$E = \{i_0\} \times \cdots \times \{i_l\} \times \prod L, \quad F = \{j_0\} \times \cdots \times \{j_r\} \times \prod L$$

for certain $i_0, \ldots, i_l, j_0, \ldots, j_r \in L := \{0, \ldots, k-1\}$. As in Section 8.3,

$$\mu([\,\tau^n \in F\,] \cap E) = p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} [S^{n-l}]_{i_l j_0} s_{j_0 j_1} \ldots s_{j_{r-1} j_r}$$

for $n > l$. By (ii) this converges to

$$p_{i_0} s_{i_0 i_1} \ldots s_{i_{l-1} i_l} p_{j_0} s_{j_0 j_1} \ldots s_{j_{r-1} j_r} = \mu(F)\mu(E).$$

Since sets E, F as above generate Σ, (i) follows. Conversely, if (i) holds, then by taking $r = l = 0$ in the above we obtain

$$p_{i_0} [S^n]_{i_0 j_0} = \mu([\,\tau^n \in F\,] \cap E) \to \mu(E)\mu(F) = p_{i_0} p_{j_0}.$$

Since p is strictly positive, (ii) follows. □

If the matrix S satisfies the equivalent statements from above, it is called **primitive**. Another term often used is **aperiodic**; this is due to the fact that a row-stochastic matrix is primitive if and only if the greatest common divisor of the lengths of cycles in the associated transition graph over L is equal to 1 (see [Billingsley (1979), p.106] for more information).

The Blum–Hanson Theorem

Let us return to the more theoretical aspects. The following striking result was proved in [Blum and Hanson (1960)]. We state it without proof.

Theorem 9.9 (Blum–Hanson, 1960). *Let* $(\mathrm{X}; \varphi)$ *be an MDS with Koopman operator* $T = T_\varphi$*, and let* $1 \le p < \infty$*. Then* $(\mathrm{X}; \varphi)$ *is strongly mixing if and only if for every subsequence* $(n_k)_{k\in\mathbb{N}}$ *of* $\mathbb{N}$ *and every* $f \in \mathrm{L}^p(\mathrm{X})$ *the strong convergence*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} T^{n_j} f = \Big(\int_{\mathrm{X}} f \Big) \cdot \mathbf{1}$$

holds.

Remarks 9.10. 1) The equivalence stated in this theorem is rather elementary if strong convergence is replaced by weak convergence. Namely, it follows from

the fact that a sequence $(a_n)_n$ in $\mathbb{C}$ is convergent if and only if each of its subsequences is Cesàro convergent (see Exercise 4).

2) As a matter of fact, Theorem 9.9 is purely of operator theoretic content. In [Jones and Kuftinec (1971)] and [Akcoglu and Sucheston (1972)] the following was proved.

Theorem. *Let T be a contraction on a Hilbert space H and denote by P the corresponding mean ergodic projection. Then the following assertions are equivalent.*

(i) *$T^n \to P$ in the weak operator topology.*

(ii) *$\frac{1}{n}\sum_{j=1}^{\infty} T^{n_j} \to P$ in the strong operator topology for every subsequence $(n_j)_j$ of $\mathbb{N}$.*

4) Müller and Tomilov have shown that the statement may fail for power-bounded operators on Hilbert spaces [Müller and Tomilov (2007)].

Strong mixing is generally considered to be a "difficult" property. For instance, Katok and Hasselblatt write:

> "... It [strong mixing] is, however, one of those notions, that is easy and natural to define but very difficult to study..."

[Katok and Hasselblatt (1995), p.748]. One reason may be that the characterisations of mixing from above require knowledge about all the powers T^n of T, and hence are usually not easy to verify. Another reason is that up to now *no purely spectral characterisation of strong mixing is known*. This is in contrast with ergodicity, which is characterised by the purely spectral condition that 1 is a simple eigenvalue of T.

In the rest of the present chapter, we shall deal with a related mixing concept, which proves to be much more well-behaved, so-called *weak mixing*.

9.2 Weakly Mixing Systems

Let us begin with some terminological remarks. Any *strictly increasing* sequence $(j_n)_n$ of natural numbers defines an *infinite* subset $J = \{j_n : n \in \mathbb{N}\}$ of $\mathbb{N}$. Conversely, any infinite subset $J \subseteq \mathbb{N}$ can be written in this way, where the sequence $(j_n)_n$ is uniquely determined and called the **enumeration** of A. As a consequence we often do not distinguish between the set J and its enumeration sequence $(j_n)_n$ and simply call it a **subsequence of** $\mathbb{N}$. If $J \subseteq \mathbb{N}$ is a subsequence and $(x_n)_n$ is a sequence in some topological space E and $x \in E$, then we say that $x_n \to x$ **along** J if $\lim_{n\to\infty} x_{j_n} = x$, where $(j_n)_j$ is the enumeration of J. We write

$$\lim_{n\in J, n\to\infty} x_n = x \qquad \text{or} \qquad x_n \to x \quad \text{as } n \in J, n \to \infty \qquad \text{or} \qquad \lim_{n\in J} x_n = x$$

to denote that $x_n \to x$ along J.

For any subset $J \subseteq \mathbb{N}_0$ we define its **(asymptotic) density** by

$$\mathrm{d}(J) := \lim_{n\to\infty} \frac{\mathrm{card}(J \cap [1,n])}{n}$$

provided this limit exists. The density yields a measure of "largeness" of a set of natural numbers. (See Exercises 1 and 2 for some properties.) Subsequences of density 1 are particularly important. If $(x_n)_n$ is sequence in a topological space and $x \in E$, we say that $x_n \to x$ **in density**, written

$$\text{D-}\lim_n x_n = x,$$

if there is a subsequence $J \subseteq \mathbb{N}$ with density $\mathrm{d}(J) = 1$ such that $\lim_{n\in J} x_n = x$.

With this terminology at hand we can introduce the weaker type of mixing that was announced above.

Definition 9.11. An MDS $(\mathrm{X};\varphi)$ is called **weakly mixing** if

$$\text{D-}\lim_{n\to\infty} \mu(\varphi^{*n}A \cap B) = \mu(A)\mu(B) \qquad \text{for every } A,B \in \Sigma. \tag{9.2}$$

Obviously, a strongly mixing system is weakly mixing, and every weakly mixing system is ergodic (specialise $A = B$ in (9.2) with A being φ-invariant). It turns out that no nontrivial group rotation system is weakly mixing, so ergodic group rotations provide examples of ergodic systems that are not weakly mixing, see Example 9.19 and the remarks thereafter.

In contrast to this, it is much harder to construct weakly mixing systems that are not strongly mixing. This had been an open problem for some while, until finally Chacon in [Chacon (1967)] succeeded by using his so-called *stacking method* (see also [Chacon (1969)] and Petersen [Petersen (1989), Section 4.5]). Curiously enough, the mere *existence* of such systems had been established much earlier, based on Baire category arguments. Namely, Halmos [Halmos (1944)] showed that the set of weakly mixing transformations is residual (i.e., its complement is of first category) in the complete metric space G of all measure preserving transformations with respect to the metric coming from the strong operator topology. And four years later, Rohlin [Rohlin (1948)] proved that the set of all strongly mixing transformations is of first category in G. Thus, in this sense, the generic MDS is weakly but not strongly mixing, cf. [Eisner (2010)].

A large part of the theory of weakly mixing systems can be reduced to pure operator theory.

Theorem 9.12. *Let T be a power-bounded operator T on a Banach space E, let $x \in E$, and let $M \subseteq E'$ such that* $\mathrm{lin}(M)$ *is norm-dense in E'. Then the following assertions are equivalent:*

(i) $\text{D-}\lim_{n\to\infty} \langle T^n x, x' \rangle = 0$ *for all $x' \in M$.*

(ii) $\text{D-}\lim_{n\to\infty} \langle T^n x, x' \rangle = 0$ *for all $x' \in E'$.*

(iii) $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \left|\langle T^j x, x'\rangle\right| = 0$ *for all* $x' \in M$.

(iv) $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \left|\langle T^j x, x'\rangle\right| = 0$ *for all* $x' \in E'$.

The proof of Theorem 9.12 rests on the following general fact, first observed in [Koopman and von Neumann (1932)].

Lemma 9.13 (Koopman–von Neumann). *For a bounded sequence* $(x_n)_n$ *in some Banach space* E *and* $x \in E$ *the following conditions are equivalent.*

(i) $\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \|x_j - x\| = 0$.

(ii) $\operatorname{D-lim}_{n\to\infty} x_n = x$.

Proof. We may suppose that $x = 0$ and, by passing to the norm of the elements, that $E = \mathbb{R}$ and $0 \le x_n \in \mathbb{R}$ for all $n \in \mathbb{N}$.

(i)$\Rightarrow$(ii): Let $L_k := \{n \in \mathbb{N} : x_n \ge 1/k\}$. Then $L_1 \subseteq L_2 \subseteq \ldots$, and $\mathrm{d}(L_k) = 0$ since

$$\frac{\operatorname{card}(L_k \cap [1,n])}{n} \le \frac{\sum_{j=1}^{n} k x_j}{n} = k \cdot \frac{1}{n}\sum_{j=1}^{n} x_j \to 0 \qquad \text{as } n \to \infty.$$

Therefore, we can choose integers $1 \le n_1 < n_2 < \ldots$ such that

$$\frac{\operatorname{card}(L_k \cap [1,n])}{n} < \frac{1}{k} \qquad (n \ge n_k).$$

Now we define

$$L := \bigcup_{k\ge 1} L_k \cap [n_k, \infty)$$

and claim that $\mathrm{d}(L) = 0$. Indeed, let $m \in \mathbb{N}$ such that $n_k \le m < n_{k+1}$. Then $L \cap [1,m] \subseteq L_k \cap [1,m]$ (because the L_k increase with k) and so

$$\frac{\operatorname{card}(L \cap [1,m])}{m} \le \frac{\operatorname{card}(L_k \cap [1,m])}{m} \le \frac{1}{k}.$$

Hence $\mathrm{d}(L) = 0$ as claimed. For $J := \mathbb{N} \setminus L$ we have $\mathrm{d}(J) = 1$ and $\lim_{n\in J} x_n = 0$.

(ii)$\Rightarrow$(i): Let $\varepsilon > 0$ and $c := \sup\{x_n : n \in \mathbb{N}\}$. Choose $n_\varepsilon \in \mathbb{N}$ such that $n > n_\varepsilon$ and $n \in J$ imply that $x_n < \varepsilon$. If $n > n_\varepsilon$, we conclude that

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} x_j &\le \frac{c n_\varepsilon}{n} + \frac{1}{n}\sum_{j=n_\varepsilon+1}^{n} x_j \le \frac{c n_\varepsilon}{n} + \varepsilon\frac{n-n_\varepsilon}{n} + \frac{1}{n}\sum_{j\in(n_\varepsilon,n]\setminus J} x_j \\ &\le \frac{c n_\varepsilon}{n} + \varepsilon + c\frac{\operatorname{card}([1,n]\setminus J)}{n}.\end{aligned}$$

Since $\mathrm{d}(\mathbb{N}\setminus J) = 0$, this implies

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} x_j \le \varepsilon.$$

Since ε was arbitrary, the proof is complete. □

Proof of Theorem 9.12. The equivalences (i)⇔(iii) and (ii)⇔(iv) follow from the Koopman–von Neumann lemma. It is clear that (iv) implies (iii) and the converse holds by approximation. (Note that the sequence $(T^n x)_n$ is bounded.)

Proposition 9.14 (Jones–Lin). *One can add the assertion*

$$\text{(v)} \quad \sup_{x'\in E', \|x'\|\le 1} \frac{1}{n} \sum_{j=0}^{n-1} \left|\langle T^j x, x'\rangle\right| \underset{n\to\infty}{\longrightarrow} 0.$$

to the equivalences in Theorem 9.12. That is, the convergence in (iv) *there is even uniform in x' from the dual unit ball.*

Proof. [Jones and Lin (1976)] First, we suppose that T is a contraction. The dual unit ball $\mathrm{B}' := \{x' \in E' : \|x'\| \le 1\}$ is then compact with respect to the weak*-topology (Banach–Alaoglu theorem). Since T is a contraction, T' leaves B' invariant, and thus gives rise to a TDS $(\mathrm{B}'; T')$. We consider its Koopman operator S on $\mathrm{C}(\mathrm{B}')$, i.e.,

$$(Sf)(x') := f(T'x') \qquad (x' \in \mathrm{B}').$$

Consider the function $f \in \mathrm{C}(\mathrm{B}')$ defined by $f(x') := |\langle x, x'\rangle|$. Hypothesis (iv) of Theorem 9.12 simply says that $\mathrm{A}_n[S]f \to 0$ pointwise. As a consequence of the Riesz representation theorem (see Theorem 5.7) and the dominated convergence theorem, $\mathrm{A}_n[S]f \to 0$ weakly, hence in norm, by Proposition 8.18.

In the case that T is just power-bounded we pass to the equivalent norm $|||y||| := \sup_{n\ge 0} \|T^n y\|$ for $y \in E$. With respect to this new norm T is a contraction, and the induced norm on the dual space is equivalent to the original dual norm, see Exercise 3. Consequently, the contraction case can be applied to conclude the proof. □

Let E be a Banach space and $T \in \mathscr{L}(E)$. A vector $x \in E$ is called **almost weakly stable** (with respect to T) if it satisfies the equivalent conditions (ii) and (iv) of Theorem 9.12 and (v) of Proposition 9.14. The set of almost weakly stable vectors is denoted by

$$E_{\mathrm{s}} = E_{\mathrm{s}}(T) := \left\{x \in E : x \text{ is almost weakly stable w.r.t. } T\right\} \tag{9.3}$$

and called the **almost stable subspace**.

Corollary 9.15. *Let T be a power-bounded operator on the Banach space E. Then the set $E_{\mathrm{s}}(T)$ of weakly almost stable vectors is a closed T-invariant subspace of E contained in $\overline{\mathrm{ran}}(\mathrm{I}-T)$. Moreover, $E_{\mathrm{s}}(T) = E_{\mathrm{s}}(T^k)$ holds for all $k \in \mathbb{N}$.*

Proof. It is trivial from (ii) in Theorem 9.12 that E_{s} is T-invariant. That it is a closed subspace follows from characterisation (iv) (see Exercise 6). If $x \in E_{\mathrm{s}}$ then by (iv)

$\mathrm{A}_n[T]x \to 0$ weakly, hence $x \in \overline{\mathrm{ran}}(\mathrm{I}-T)$ by Proposition 8.18. Finally, suppose that $x \in E_{\mathrm{s}}(T)$ and let $k \in \mathbb{N}$. Then for $n \in \mathbb{N}$ and $x' \in E'$ we have

$$\frac{1}{n}\sum_{j=0}^{n-1}\left|\left\langle T^{kj}x,x'\right\rangle\right| \le k\cdot\frac{1}{kn}\sum_{j=0}^{nk-1}\left|\left\langle T^{j}x,x'\right\rangle\right| \to 0 \quad \text{as } n\to\infty,$$

so $x \in E_{\mathrm{s}}(T^k)$. The converse inclusion $E_{\mathrm{s}}(T^k) \subseteq E_{\mathrm{s}}(T)$ is left as Exercise 7. □

After these general operator theoretic considerations, we return to dynamical systems. Recall our standing terminology that for $p \in [1,\infty]$ the letter q denotes the exponent conjugate to p.

Theorem 9.16. *Let $(\mathrm{X};\varphi)$ by an MDS with associated Koopman operator $T = T_\varphi$ on $E := \mathrm{L}^p(\mathrm{X})$, $p \in [1,\infty)$. Then the following assertions are equivalent.*

(i) *The MDS $(\mathrm{X};\varphi)$ is weakly mixing.*

(ii) *For every $f \in \mathrm{L}^p(\mathrm{X})$, $g \in \mathrm{L}^q(\mathrm{X})$ there is a subsequence $J \subseteq \mathbb{N}$ of density $\mathrm{d}(J) = 1$ with*

$$\lim_{n\in J}\int_{\mathrm{X}}(T^n f)\cdot g \;=\; \Big(\int_{\mathrm{X}} f\Big)\cdot\Big(\int_{\mathrm{X}} g\Big).$$

(iii) $\dfrac{1}{n}\displaystyle\sum_{k=0}^{n-1}\left|\left\langle T^k f,g\right\rangle - \langle f,\mathbf{1}\rangle\cdot\langle\mathbf{1},g\rangle\right| \underset{n\to\infty}{\longrightarrow} 0$ *for all $f \in \mathrm{L}^p$, $g \in \mathrm{L}^q$.*

(iv) *For each $f \in E$ with $\int_{\mathrm{X}} f = 0$ one has $f \in E_{\mathrm{s}}(T)$.*

(v) $\mathrm{fix}(T) = \mathrm{lin}\{\mathbf{1}\}$ *and* $\mathrm{ran}(\mathrm{I}-T) \subseteq E_{\mathrm{s}}(T)$.

(vi) $E = \mathrm{lin}\{\mathbf{1}\} \oplus E_{\mathrm{s}}(T)$.

Proof. Note that (ii) simply says that $f - \langle f,\mathbf{1}\rangle\,\mathbf{1} \in E_{\mathrm{s}}(T)$ for all $f \in E$. Hence the equivalence (ii)$\Leftrightarrow$(iii) follows from Theorem 9.12, and the equivalences (ii)$\Leftrightarrow$(iv) and (ii)$\Leftrightarrow$(vi) are straightforward.

(v)$\Leftrightarrow$(vi): By Corollary 9.15 we have $E_{\mathrm{s}} \subseteq \overline{\mathrm{ran}}(\mathrm{I}-T)$. So the asserted equivalence follows since T is mean ergodic.

(ii)$\Rightarrow$(i): Simply specialise $f = \mathbf{1}_A, g = \mathbf{1}_B$ in (ii) for $A,B \in \Sigma_{\mathrm{X}}$.

(i)$\Rightarrow$(ii): Fix $A \in \Sigma_{\mathrm{X}}$ and let $f := \mathbf{1}_A$. Then, by definition of weak mixing,

$$\operatorname*{D\text{-}lim}_{n\to\infty}\langle T^n(f-\langle f,\mathbf{1}\rangle\,\mathbf{1}),\mathbf{1}_B\rangle = \operatorname*{D\text{-}lim}_{n\to\infty}\langle T^n f,\mathbf{1}_B\rangle - \mu(A)\mu(B) = 0$$

for all $B \in \Sigma_{\mathrm{X}}$. Hence Theorem 9.12 with $M = \{\mathbf{1}_B : B \in \Sigma_{\mathrm{X}}\}$ yields that $f - \langle f,\mathbf{1}\rangle\,\mathbf{1} \in E_{\mathrm{s}}(T)$. Since $E_{\mathrm{s}}(T)$ is a closed linear subspace, by approximation it follows that $f - \langle f,\mathbf{1}\rangle\,\mathbf{1} \in E_{\mathrm{s}}$ for all $f \in E$, i.e., (ii). □

Remark 9.17. In the situation of Theorem 9.16 let $D \subseteq \mathrm{L}^p$ and $M \subseteq \mathrm{L}^q$ such that $\mathrm{lin}(D)$ is norm-dense in L^p and $\mathrm{lin}(M)$ is norm-dense in L^1. Then we can add to Theorem 9.16 the following additional equivalences.

(ii') *For every $f \in D, g \in M$ there is a subsequence $J \subseteq \mathbb{N}$ of density $\mathrm{d}(J) = 1$ with*

$$\lim_{n\in J} \int_{\mathrm{X}} (T^n f)\cdot g \;=\; \Big(\int_{\mathrm{X}} f\Big)\cdot\Big(\int_{\mathrm{X}} g\Big).$$

(iii') $\dfrac{1}{n}\displaystyle\sum_{k=0}^{n-1}\Big|\Big\langle T^k f, g\Big\rangle - \langle f,\mathbf{1}\rangle\cdot\langle \mathbf{1}, g\rangle\Big| \underset{n\to\infty}{\longrightarrow} 0$ *for each $f \in D$, $g \in M$.*

(vi') $f - \langle f,\mathbf{1}\rangle\,\mathbf{1} \in E_{\mathrm{s}}(T)$ *for each $f \in D$.*

Moreover, based on (iii') and Lemma B.16 one can add another equivalence:

(i') $\mathrm{D}\text{-}\lim_{n\to\infty} \mu(\varphi^{*n}A \cap B) = \mu(A)\mu(B)$ for all $A \in \mathscr{D}, B \in \mathcal{F}$,

where $\mathcal{D}, \mathcal{F}$ are given $\cap$-stable generators of Σ_{X}. The proofs for these claims are left as Exercise 8.

From these results we obtain as in Corollary 9.7 a characterisation of weakly mixing systems by their iterates.

Corollary 9.18. *For an MDS $(\mathrm{X};\varphi)$ and $k \in \mathbb{N}$ the following assertions are equivalent:*

(i) *The MDS $(\mathrm{X};\varphi)$ is weakly mixing.*

(vii) *Its k^{th} iterate $(\mathrm{X};\varphi^k)$ is weakly mixing.*

Proof. By (iv) from 9.16 it suffices to have $E_{\mathrm{s}}(T) = E_{\mathrm{s}}(T^k)$, which has already been established in Corollary 9.15. □

An MDS $(\mathrm{X};\varphi)$ is called **strongly ergodic** if all its k^{th} iterates are ergodic. It is easy to construct ergodic systems which are not strongly ergodic (Exercise 10). By Corollary 9.18 and Theorem 9.16 every weakly mixing MDS is strongly ergodic. On the other hand, any ergodic (=irrational) rotation $(\mathbb{T};a)$ is strongly ergodic, but not weakly mixing, as the following example shows.

Example 9.19 (Nonmixing Group Rotations). Consider any rotation $(\mathbb{T};a)$ with Koopman operator T and $f(z) := z$. Then $\langle f,\mathbf{1}\rangle = 0$. But $\langle T^n f, \overline{f}\rangle = a^n$ does not converge to $0 = \langle f,\mathbf{1}\rangle$ along *any* subsequence $J \subseteq \mathbb{N}$, so $(\mathbb{T};a)$ is not weakly mixing.

We shall see in Chapter 16 that *no* nontrivial ergodic group rotation is a weakly mixing system. Moreover, we shall give a new description of $E_{\mathrm{s}}(T)$ as a summand of the so-called *Jacobs–de Leeuw–Glicksberg decomposition*

$$E = E_{\mathrm{r}}(T) \oplus E_{\mathrm{s}}(T)$$

of a reflexive Banach space E with respect to a Markov operator $T \in \mathscr{L}(E)$. As a consequence, we shall obtain further characterisations on weakly mixing systems, see Theorem 16.20. In particular, we shall prove in Corollary 16.35 that an ergodic MDS $(\mathrm{X};\varphi)$ is weakly mixing if and only if 1 is the only eigenvalue of T_φ of modulus 1. Hence weakly mixing systems can be characterised like ergodic systems by a purely spectral condition for the Koopman operator.

Product Systems

Weak mixing can also be equivalently described by means of product systems. Recall from Section 5.1 that given two MDSs $(\mathrm{X};\varphi)$ and $(\mathrm{Y};\psi)$ their product is

$$(\mathrm{X}\otimes\mathrm{Y};\varphi\times\psi), \quad \text{where} \quad \mathrm{X}\otimes\mathrm{Y} := (X\times Y, \Sigma_{\mathrm{X}}\otimes\Sigma_{\mathrm{Y}}, \mu_{\mathrm{X}}\otimes\mu_{\mathrm{Y}}).$$

and $(\varphi\times\psi)(x,y) := (\varphi(x),\psi(y))$. It is sometimes convenient to write $\varphi\times\psi$ to refer to that system.

For functions f on X and $g\in Y$ the function $f\otimes g$ on the product is defined by

$$(f\otimes g)(x,y) := f(x)g(y) \qquad (x\in X, y\in Y)$$

see Appendix B.6. Standard product measure theory yields that for $1\le p<\infty$ the linear span of

$$\big\{f\otimes g : f\in \mathrm{L}^p(\mathrm{X}),\ g\in \mathrm{L}^p(\mathrm{Y})\big\}$$

is dense in $\mathrm{L}^p(\mathrm{X}\otimes\mathrm{Y})$, see Corollary B.19. Since we have the product measure on the product space, the canonical L^p-L^q-duality satisfies

$$\langle f\otimes g, u\otimes v\rangle = \langle f,u\rangle\cdot\langle g,v\rangle$$

for L^p-functions f,g and L^q-functions u,v (where q is, of course, the conjugate exponent to p).

Let T and S denote the Koopman operators associated with φ and ψ, respectively. The Koopman operator associated with $\varphi\times\psi$ is denoted by $T\otimes S$, since it satisfies

$$(T\otimes S)(f\otimes g) = Tf\otimes Sg \qquad (f\in\mathrm{L}^p(\mathrm{X}), g\in\mathrm{L}^p(\mathrm{Y})).$$

(By density, $T\otimes S$ is uniquely determined by these identities.) After these preliminary remarks we can state and prove the announced additional characterisation of weak mixing.

Theorem 9.20. *Let $(\mathrm{X};\varphi)$ and $(\mathrm{Y};\psi)$ be MDSs. Then the following assertions are equivalent.*

(i) *Both $(\mathrm{X};\varphi)$ and $(\mathrm{Y};\psi)$ are weakly mixing.*

(ii) *The product system $(\mathrm{X}\otimes\mathrm{Y};\varphi\times\psi)$ is weakly mixing.*

Proof. Suppose that (i) holds and let $f,u\in\mathrm{L}^2(\mathrm{X})$ and $g,v\in\mathrm{L}^2(\mathrm{Y})$. By hypothesis and Theorem 9.16 one can find subsequences $J,J'\subseteq\mathbb{N}$ of density 1 such that

$$\lim_{n\in J}\langle T^n f,u\rangle = \langle f,\mathbf{1}\rangle\langle\mathbf{1},u\rangle \quad \text{and} \quad \lim_{n\in J'}\langle S^n g,v\rangle = \langle g,\mathbf{1}\rangle\langle\mathbf{1},v\rangle$$

Since $\mathrm{d}(J\cap J')=1$ (Exercise 1) we may suppose that $J=J'$. Hence,

$$\lim_{n\in J}\langle (T\otimes S)^n(f\otimes g),u\otimes v\rangle = \lim_{n\in J}\langle T^nf\otimes S^ng,u\otimes v\rangle$$
$$= \lim_{n\in J}\langle T^nf,u\rangle\langle S^ng,v\rangle = \langle f,\mathbf{1}\rangle\langle \mathbf{1},u\rangle\langle g,\mathbf{1}\rangle\langle \mathbf{1},v\rangle = \langle f\otimes g,\mathbf{1}\rangle\langle \mathbf{1},u\otimes v\rangle .$$

By the equivalence (ii') in Remark 9.17 with $D=E=\{h\otimes k : h\in \mathrm{L}^2(\mathrm{X}),k\in \mathrm{L}^2(\mathrm{Y})\}$ we conclude that the product system is weakly mixing.

Conversely, suppose that (ii) holds. Then for $f,u\in \mathrm{L}^2(\mathrm{X})$ we have

$$\langle T^nf,u\rangle = \langle T^nf,u\rangle\cdot\langle \mathbf{1},\mathbf{1}\rangle = \langle T^nf\otimes S^n\mathbf{1},u\otimes\mathbf{1}\rangle = \langle (T\otimes S)^n(f\otimes\mathbf{1}),u\otimes\mathbf{1}\rangle .$$

The latter converges to $\langle f\otimes\mathbf{1},\mathbf{1}\rangle\cdot\langle \mathbf{1},u\otimes\mathbf{1}\rangle = \langle f,\mathbf{1}\rangle\langle \mathbf{1},u\rangle$ in density, whence $(\mathrm{X};\varphi)$ is weakly mixing. By symmetry, we obtain (i). □

Theorem 9.20 allows us to enlarge the list of equivalences of Theorem 9.16.

Corollary 9.21. *For an MDS* $(\mathrm{X};\varphi)$ *the following assertions are equivalent.*

(i) $(\mathrm{X};\varphi)$ *is weakly mixing.*

(viii) $(\mathrm{X}\otimes\mathrm{Y};\varphi\times\psi)$ *is weakly mixing for some/each weakly mixing system* $(\mathrm{Y};\psi)$.

(ix) $(\mathrm{X}\otimes\mathrm{X};\varphi\times\varphi)$ *is weakly mixing.*

We remark that the product of two *ergodic* systems need not be ergodic, see Exercise 9.

9.3 Weak Mixing of All Orders

For our mixing concepts hitherto we considered pairs of sets $A,B\in\Sigma$ or pairs of functions f,g. In this section we introduce a new notion, which intuitively describes the mixing of an arbitrary finite number of sets.

Definition 9.22. An MDS $(\mathrm{X};\varphi)$, $\mathrm{X}=(X,\Sigma,\mu)$, is called **weakly mixing of order** $k\in\mathbb{N}$ if

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\left|\mu(A_0\cap\varphi^{*n}A_1\cap\ldots\cap\varphi^{*(k-1)n}A_{k-1})-\mu(A_0)\mu(A_1)\cdots\mu(A_{k-1})\right| = 0$$

holds for every $A_0,\ldots,A_{k-1}\in\Sigma$. An MDS is **weakly mixing of all orders** if it has the latter property for all $k\in\mathbb{N}$.

Notice first that weak mixing of order 2 is just the same as weak mixing. We begin with some elementary observations.

Proposition 9.23. a) *If an MDS* $(\mathrm{X};\varphi)$ *is weakly mixing of order* $m\in\mathbb{N}$*, then it is weakly mixing of order* k *for all* $k\le m$.

b) *An MDS* $(\mathrm{X};\varphi)$, $\mathrm{X}=(X,\Sigma,\mu)$, *is weakly mixing of order* $k\in\mathbb{N}$ *if and only if*

$$\operatorname*{D-lim}_{n\to\infty}\mu(A_0\cap\varphi^{*n}A_1\cap\ldots\cap\varphi^{*(k-1)n}A_{k-1})=\mu(A_0)\mu(A_1)\cdots\mu(A_{k-1})$$

for all $A_0,A_1,\ldots,A_{k-1}\in\Sigma$.

Proof. a) To check weak mixing of order k specialise $A_k=\cdots=A_{m-1}=X$. For b) Use the Koopman–von Neumann Lemma 9.13. □

As seen above, a system that is weakly mixing of all orders is weakly mixing. Our aim in this section is to show that, surprisingly, the converse is also true: A weakly mixing system is even weakly mixing of *every* order $k\in\mathbb{N}$ (Theorem 9.29 below).

We begin with an auxiliary lemma that plays a central role in this context.

Lemma 9.24 (Van der Corput). *Let H be a Hilbert space and $(u_n)_{n\in\mathbb{N}_0}\subseteq H$ with $\|u_n\|\le 1$. For $j\in\mathbb{N}_0$ define*

$$\gamma_j:=\limsup_{N\to\infty}\Big|\frac{1}{N}\sum_{n=0}^{N-1}\big(u_n\,\big|\,u_{n+j}\big)\Big|.$$

Then the inequality

$$\limsup_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1}u_n\Big\|^2\le\limsup_{J\to\infty}\frac{1}{J}\sum_{j=0}^{J-1}\gamma_j$$

holds. In particular, $\lim_{N\to\infty}\frac{1}{N}\sum_{j=0}^{N-1}\gamma_j=0$ *implies* $\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}u_n=0$.

Proof. We shall use the notation $\mathrm{o}(1)$ to denote terms converging to 0 as $N\to\infty$. First observe that for a fixed $J\in\mathbb{N}$ we have

$$\frac{1}{N}\sum_{n=0}^{N-1}u_n=\frac{1}{N}\sum_{n=0}^{N-1}u_{n+j}+\mathrm{o}(1)$$

for every $0\le j\le J-1$ and hence

$$\frac{1}{N}\sum_{n=0}^{N-1}u_n=\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{J}\sum_{j=0}^{J-1}u_{n+j}+\mathrm{o}(1).$$

By the Cauchy–Schwarz inequality this implies

$$\begin{aligned}\Big\|\frac{1}{N}\sum_{n=0}^{N-1}u_n\Big\|&\le\frac{1}{N}\sum_{n=0}^{N-1}\Big\|\frac{1}{J}\sum_{j=0}^{J-1}u_{n+j}\Big\|+\mathrm{o}(1)\le\left(\frac{1}{N}\sum_{n=0}^{N-1}\Big\|\frac{1}{J}\sum_{j=0}^{J-1}u_{n+j}\Big\|^2\right)^{1/2}+\mathrm{o}(1)\\&\le\left(\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{J^2}\sum_{j_1,j_2=0}^{J-1}\big(u_{n+j_1}\,\big|\,u_{n+j_2}\big)\right)^{1/2}+\mathrm{o}(1).\end{aligned}$$

Letting $N \to \infty$ in the above and using the definition of γ_j, we thus obtain

$$\limsup_{N\to\infty} \Big\| \frac{1}{N} \sum_{n=0}^{N-1} u_n \Big\|^2 \le \frac{1}{J^2} \sum_{j_1,j_2=0}^{J-1} \gamma_{|j_1-j_2|}. \tag{9.4}$$

It thus remains to show that

$$\limsup_{J\to\infty} \frac{1}{J^2} \sum_{j_1,j_2=0}^{J-1} \gamma_{|j_1-j_2|} \le \limsup_{J\to\infty} \frac{1}{J} \sum_{j=0}^{J-1} \gamma_j =: d. \tag{9.5}$$

To this aim, observe first that

$$\sum_{j_1,j_2=0}^{J-1} \gamma_{|j_1-j_2|} = J\gamma_0 + 2\sum_{j=1}^{J-1} (J-j)\gamma_j \le 2\sum_{j=0}^{J-1} (J-j)\gamma_j = 2\sum_{k=1}^{J-1} k\Big(\frac{1}{k}\sum_{j=0}^{k} \gamma_j\Big). \tag{9.6}$$

Take $\varepsilon > 0$ and k_0 such that $\frac{1}{k}\sum_{j=0}^{k} \gamma_j < d+\varepsilon$ for every $k \ge k_0$. By

$$2\sum_{k=k_0}^{J-1} k\Big(\frac{1}{k}\sum_{j=0}^{k} \gamma_j\Big) < 2(d+\varepsilon)\sum_{k=1}^{J-1} k = J(J-1)(d+\varepsilon),$$

the estimate (9.6) implies that

$$\limsup_{J\to\infty} \frac{1}{J^2} \sum_{j_1,j_2=0}^{J-1} \gamma_{|j_1-j_2|} \le d+\varepsilon,$$

which implies (9.5). □

For more on van der Corput's lemma see, e.g., [Tao (2008b)]. Here we shall employ it to derive the following "multiple mean ergodic theorem".

Theorem 9.25. *Let* $(\mathrm{X};\varphi)$ *be a weakly mixing MDS and let* $T := T_\varphi$ *be the Koopman operator on* $E := \mathrm{L}^2(X,\Sigma,\mu)$. *Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (T^n f_1)(T^{2n} f_2)\cdots(T^{(k-1)n} f_{k-1}) = \Big(\int_{\mathrm{X}} f_1 \cdots \int_{\mathrm{X}} f_{k-1}\Big)\cdot \mathbf{1} \tag{9.7}$$

in $\mathrm{L}^2(\mathrm{X})$ *for every* $k \ge 2$ *and every* $f_1,\dots,f_{k-1} \in \mathrm{L}^\infty(\mathrm{X})$.

Note that since we take the functions $f_1,\dots,f_k$ from $\mathrm{L}^\infty(\mathrm{X})$, the above products remain in L^2.

Proof. We prove the theorem by induction on $k \ge 2$. For $k = 2$ the assertion reduces to the mean ergodic theorem for ergodic systems, see Theorem 8.10. So suppose that the assertion holds for some $k \ge 2$ and take $f_1,\dots,f_k \in \mathrm{L}^\infty$. Since the MDS defined by φ is weakly mixing we can decompose $f_k = \langle f_k,\mathbf{1}\rangle\,\mathbf{1} + (f - \langle f_k,\mathbf{1}\rangle\,\mathbf{1})$, where the latter part is contained in E_{s}. Hence without loss of generality we my consider the

cases $f_k \in E_s$ and $f_k \in \mathbb{C}\mathbf{1}$ separately. Since in the latter case the assertions reduces to the induction hypothesis, we are left with the case that $f_k \in E_s$. Then $\langle f_k, \mathbf{1}\rangle = 0$ (since $E_s \subseteq \overline{\mathrm{ran}}(\mathrm{I} - T_k)$, the kernel of the mean ergodic projection), so we have to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (T^n f_1)(T^{2n} f_2) \cdots (T^{kn} f_k) = 0.$$

To achieve this we employ van der Corput's Lemma 9.24. Define $u_n := T_1^n f_1 \cdot T_2^n f_2 \cdots T_k^n f_k$. Then the invariance of μ and the multiplicativity T imply

$$\begin{aligned}
\left(u_n \,\middle|\, u_{n+j}\right) &= \int_X (T^n f_1 \cdot T^{2n} f_2 \cdots T^{kn} f_k) \cdot (T^{n+j}\overline{f_1} \cdot T^{2n+2j}\overline{f_2} \cdots T^{kn+kj}\overline{f_k}) \\
&= \int_X (f_1 \cdot T^n f_2 \cdots T^{(k-1)n} f_k) \cdot (T^j \overline{f_1} \cdot T^{n+2j}\overline{f_2} \cdots T^{(k-1)n+kj}\overline{f_k}) \\
&= \int_X (f_1 T^j \overline{f_1}) \cdot T^n (f_2 T^{2j}\overline{f_2}) \cdots T^{(k-1)n} (f_k T^{kj}\overline{f_k}).
\end{aligned}$$

Thus, by the induction hypothesis, the Cesàro means of $\left(u_n \,\middle|\, u_{n+j}\right)$ converge to

$$\int_X f_1 T^j \overline{f_1} \cdot \int_X f_2 T^{2j}\overline{f_2} \cdots \int_X f_k T^{kj}\overline{f_k} = \prod_{m=1}^k \left(f_m \,\middle|\, T^{jm} f_m\right).$$

So we obtain

$$\gamma_j := \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=0}^{N-1} \left(u_n \,\middle|\, u_{n+j}\right) \right| = \prod_{m=1}^k \left| \left(f_m \,\middle|\, T^{jm} f_m\right) \right|.$$

Now, each T^{jm} is a contraction and $f_k \in E_s(T) = E_s(T^k)$ (Corollary 9.15), hence it follows that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \gamma_j \le \|f_1\|^2 \ldots \|f_{k-1}\|^2 \limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \left(f_k \,\middle|\, T^{kj} f_k\right) \right| = 0,$$

where we used Theorem 9.16 (iii) for the last equality. An application of van der Corput's lemma concludes the proof. □

Remark 9.26. Suitably adapted, the proof works for $T, T^{2n}, \ldots, T^{(k-1)n}$ replaced by $T^{m_1 n}, T^{m_2 n}, \ldots, T^{m_{k-1} n}$ for given $1 \le m_1 \le \cdots \le m_{k-1}$. (The proof is left as Exercise 12.)

By specialising the f_j to characteristic functions in the Theorem 9.25 we obtain the following corollary.

Corollary 9.27. *Let* $(\mathrm{X}; \varphi)$, $\mathrm{X} = (X, \Sigma, \mu)$, *be a weakly mixing MDS. Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\left(A_0 \cap \varphi^{*n} A_1 \cap \ldots \cap \varphi^{*(k-1)n} A_{k-1}\right) = \mu(A_0)\mu(A_1)\cdots\mu(A_{k-1})$$

for every $k \in \mathbb{N}$ *and every* $A_0, \ldots, A_{k-1} \in \Sigma$.

For the final step we need a simple lemma.

Lemma 9.28. *Let H be a Hilbert space, $(f_n)_n \subseteq H$ and $f \in H$. Then the following assertions are equivalent.*

(i) $f_n \to f$ *in density.*

(ii) $\frac{1}{n}\sum_{j=1}^{n} \|f_j - f\| \to 0.$

(iii) $\frac{1}{n}\sum_{j=1}^{n} \|f_j - f\|^2 \to 0.$

(iv) $\frac{1}{n}\sum_{j=1}^{n} f_j \to f$ *weakly* and $\frac{1}{n}\sum_{j=1}^{n} \|f_j\|^2 \to \|f\|^2.$

Proof. The Koopman–von Neumann Lemma accounts for the equivalence of (i)–(iii). The equivalence of (iii) and (iv) follows from the identity

$$\|f_j - f\|^2 = \|f_j\|^2 - 2\,\mathrm{Re}\,(f_j \,|\, f) + \|f\|^2. \qquad \square$$

We can now prove the following result explaining the title of this section.

Theorem 9.29. *Every weakly mixing MDS $(\mathrm{X};\varphi)$, $\mathrm{X} = (X,\Sigma,\mu)$, is* **weakly mixing of all orders***, i.e.,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \mu\left(A_0 \cap \varphi^{*n} A_1 \cap \ldots \cap \varphi^{*(k-1)n} A_{k-1}\right) - \mu(A_0)\mu(A_1)\cdots\mu(A_{k-1}) \right| = 0$$

for every $k \in \mathbb{N}$ and every $A_0,\ldots,A_{k-1} \in \Sigma$.

Proof. Since $(\mathrm{X}\otimes\mathrm{X};\varphi\times\varphi$ is again weakly mixing by Corollary 9.21, we can apply Corollary 9.18 to $\varphi\times\varphi$ and the sets $A_0\times A_0,\ldots,A_{k-1}\times A_{k-1}$ to obtain

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\left(A_0 \cap \ldots \cap T^{(k-1)n} A_{k-1}\right)^2 \\
&\quad = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (\mu\times\mu)\left((A_0\times A_0) \cap \ldots \cap (T^{(k-1)n} A_{k-1} \times T^{(k-1)n} A_{k-1})\right) \\
&\quad = \mu(A_0)^2 \ldots \mu(A_{k-1})^2.
\end{aligned}$$

This combined with Corollary 9.27 implies, by Lemma 9.28, that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \mu\left(A_0 \cap \ldots \cap T^{(k-1)n} A_{k-1}\right) - \mu(A_0)\ldots\mu(A_{k-1}) \right| = 0,$$

and the theorem is proved. $\square$

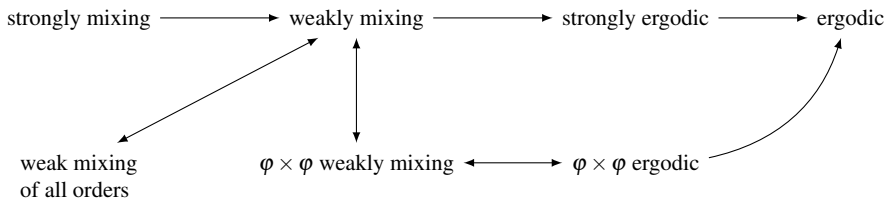
Remark 9.30. It is a long-standing open problem whether the analogue of Theorem 9.29 for *strongly* mixing systems is true. That is, it is not known even for $k = 3$ whether for a general strongly mixing system $(\mathrm{X};\varphi)$ one has

$$\lim_{n\to\infty} \mu\left(A_0 \cap \varphi^{*n} A_1 \cap \ldots \cap \varphi^{*(k-1)n} A_{k-1}\right) = \mu(A_0)\mu(A_1)\cdots\mu(A_{k-1})$$

for every $A_0, \ldots, A_{k-1} \in \Sigma$.

Concluding remarks

Let us summarise the important relations between these many notions in the following diagram:



Exercises

1 (Upper and Lower Density). For a subset $J \subseteq \mathbb{N}_0$ the **upper** and **lower** density are defined as

$$\overline{\mathrm{d}}(J) := \limsup_{n\to\infty} \frac{\mathrm{card}(J \cap [1,n])}{n} \quad \text{and} \quad \underline{\mathrm{d}}(J) := \liminf_{n\to\infty} \frac{\mathrm{card}(J \cap [1,n])}{n}.$$

Show that $\overline{\mathrm{d}}(J)$ and $\underline{\mathrm{d}}(J)$ do not change if we change J by finitely many elements. Then show that

$$\overline{\mathrm{d}}(J) = 1 - \underline{\mathrm{d}}(\mathbb{N} \setminus J) \qquad \text{for every } J \subseteq \mathbb{N}$$

and

$$\overline{\mathrm{d}}(J \cup K) \leq \overline{\mathrm{d}}(J) + \overline{\mathrm{d}}(K) \qquad \text{for every } J, K \subseteq \mathbb{N}.$$

Conclude that if $\mathrm{d}(J) = \mathrm{d}(K) = 1$ then $\mathrm{d}(J \cap K) = 1$.

2. Prove the following assertions.

a) For $\alpha, \beta \in [0,1]$, $\alpha \leq \beta$ there is a set $A \subseteq \mathbb{N}$ with $\underline{\mathrm{d}}(A) = \alpha$ and $\overline{\mathrm{d}}(A) = \beta$.

b) Suppose that $A \subseteq \mathbb{N}$ is **relatively dense**, i.e., has bounded gaps (Definition 3.9). Then $\underline{\mathrm{d}}(A) > 0$. The converse implication, however, is not true.

c) Let $\mathbb{N} = A_1 \cup A_2 \cup \cdots \cup A_k$. Then there is $1 \leq j \leq k$ with $\overline{\mathrm{d}}(A_j) > 0$. The assertion is not true if $\overline{\mathrm{d}}$ is replaced by $\underline{\mathrm{d}}$.

3. Let T be a power-bounded operator on a Banach space E and define

$$|||y||| := \sup_{n\geq 0} \|T^n y\| \quad \text{for} \quad y \in E.$$

Show that $|||\cdot|||$ is a norm on E, equivalent to the original norm. Show that T is a contraction with respect to this norm. Then show that the corresponding dual norm

$$|||x'||| = \sup\{\,|\langle x, x'\rangle| \,:\, |||x||| \le 1\,\}$$

is equivalent to the original dual norm.

4. Prove that a sequence $(a_n)_n$ in $\mathbb{C}$ converges if and only if each of its subsequences is Cesàro convergent.

5. Let $(a_n)_n, (b_n)_{n\in\mathbb{N}} \in \ell^\infty$, and suppose that $a_n \to a$ along a subsequence J of $\mathbb{N}$ with density 1, and $\frac{1}{n}\sum_{j=1}^n b_j \to b$. Show that then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} a_j b_j = ab.$$

Note that the Cesàro limit is not multiplicative in general! (Hint: Use the Koopman–von Neumann lemma and a 3ε-argument.)

6. Let T be a power-bounded operator on a Banach space E. Prove that $E_s(T)$ is a closed subspace of E, cf. Corollary 9.15. (Hint: Use the description (iv) of Theorem 9.12.)

7. Let T be a power-bounded operator on a Banach space E and let $k \in \mathbb{N}$. Prove the inclusion $E_s(T^k) \subseteq E_s(T)$, cf. Corollary 9.15. (Hint: Use the description (iv) of Theorem 9.12 and employ an argument as in the proof of Corollary 9.7.)

8. Prove the characterisations of weakly mixing systems stated in Remark 9.17.

9. Let $(\mathbb{T}, a)$ be an ergodic rotation. Show that the product system $(\mathbb{T}\times\mathbb{T}; (a,a))$ is not ergodic. (Hint: Consider the function $f(z,w) := z\overline{w}$ on $\mathbb{T}\times\mathbb{T}$.)

10. Give an example of an ergodic MDS $(\mathrm{X};\varphi)$ such that the MDS $(X,\Sigma,\mu;\varphi^k)$ is not ergodic for some $k \ge 2$.

11. Give an interpretation of weak mixing (Definition 9.11) and mixing of all orders (Theorem 9.29) for the wine/water-example from the beginning of this chapter.

12. Prove Remark 9.26.

13. Prove that for an MDS $(\mathrm{X};\varphi)$ the following assertions are equivalent.

a) $(\mathrm{X};\varphi)$ is strongly mixing.

b) $(\mathrm{X}\otimes\mathrm{X};\varphi\times\varphi)$ is strongly mixing.

c) $(\mathrm{X}\otimes\mathrm{Y};\varphi\times\psi)$ is strongly mixing for any strongly mixing MDS $(Y,\Sigma',\nu;\psi)$.

14. Let $(\mathrm{X};\varphi)$ be an MDS, with $\mathcal{F}$ a $\cap$-stable generator of Σ. Suppose that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n} \mu(\varphi^{*n}A \cap A) = \mu(A)^2$$

for all $A \in \mathcal{F}$. Show that $(X;\varphi)$ is ergodic. (Hint: Imitate the proof of (ii)$\Rightarrow$(iii) in Theorem 9.4 and use Lemma B.16.)

Chapter 10
Mean Ergodic Operators on $\mathrm{C}(K)$

In this chapter we consider a TDS $(K;\varphi)$ and ask under which conditions its Koopman operator $T = T_\varphi$ is mean ergodic on $\mathrm{C}(K)$. In contrast to the L^p-case, it is much more difficult to establish weak compactness of subsets of $\mathrm{C}(K)$. Consequently, property (iii) of Theorem 8.20 plays a minor role here. Using instead condition (v)

$$\text{“}\mathrm{fix}(T) \quad \text{separates} \quad \mathrm{fix}(T')\text{”}$$

of that theorem, we have already seen simple examples of TDSs where the Koopman operator is not mean ergodic (see Exercise 8.10.a)). Obviously, mean ergodicity of T is connected with $\mathrm{fix}(T')$ being “not too large” (as compared to $\mathrm{fix}(T)$). Now, if we identify $\mathrm{C}(K)' = \mathrm{M}(K)$ by virtue of the Riesz Representation Theorem 5.7, we see that a complex Baire measure $\mu \in \mathrm{M}(K)$ is a fixed point of T' if and only if it is φ-invariant, i.e.,

$$\mathrm{M}_\varphi(K) := \big\{\mu \in \mathrm{M}(K) : \mu \text{ is } \varphi\text{-invariant}\big\} = \mathrm{fix}(T').$$

Mean ergodicity of T therefore becomes more likely if there are only few φ-invariant measures. We first discuss the existence of invariant measures as announced in Section 5.2.

10.1 Invariant Measures

Given a TDS $(K;\varphi)$ with Koopman operator T on $\mathrm{C}(K)$ we want to find a *fixed point* of T' which is a probability measure. For this purpose let us introduce

$$\mathrm{M}^1(K) := \big\{\mu \in \mathrm{M}(K) : \mu \geq 0,\ \langle \mu, \mathbf{1}\rangle = 1\big\}, \qquad \text{and}$$
$$\mathrm{M}^1_\varphi(K) := \mathrm{M}^1(K) \cap \mathrm{M}_\varphi(K) = \mathrm{M}^1(K) \cap \mathrm{fix}(T')$$

the set of all, and of all φ-invariant probability measures, respectively. Recall that

$$\mu \geq 0 \quad \Leftrightarrow \quad \langle f, \mu \rangle \geq 0 \quad \text{for all } 0 \leq \mathrm{C}(K).$$

It follows that both $\mathrm{M}^1(K)$ and $\mathrm{M}^1_\varphi(K)$ are convex and weakly* closed. Since

$$\mathrm{M}^1(K) \subseteq \big\{ \mu \in \mathrm{M}(K) : \|\mu\| \leq 1 \big\}$$

and the latter set is weakly* compact by the Banach–Alaoglu theorem, we conclude that $\mathrm{M}^1(K)$ and $\mathrm{M}^1_\varphi(K)$ are convex and compact (for the weak* topology). Moreover, $\mathrm{M}^1(K)$ is not empty, since it contains all point measures. To see that $\mathrm{M}^1_\varphi(K)$ is not empty either, we employ one of the classical fixed point theorems.

Theorem 10.1 (Markov–Kakutani). *Let C be a nonempty compact, convex subset of a Hausdorff topological vector space, and let Γ be a collection of pairwise commuting affine and continuous mappings $T : C \to C$. Then these mappings have a common fixed point in C.*

Proof. First we suppose that $\Gamma = \{T\}$ is a singleton. Let $f \in C$. Then, by compactness of C, the sequence $\mathrm{A}_n[T]f$ has a cluster point $g \in C$. Since C is compact, $(1/n)(f - T^n f) \to 0$. Hence by Lemma 8.17 we conclude that $Tg = g$.

For the general case it suffices, by compactness, that the sets $\mathrm{fix}(T)$, $T \in \Gamma$, have the finite intersection property. We prove that by induction. Suppose for $T_1, \ldots, T_n \in \Gamma$ we already know that $C_1 := \mathrm{fix}(T_1) \cap \cdots \cap \mathrm{fix}(T_{n-1})$ is nonempty. Now C_1 is compact and convex, and any $T_{n+1} \in \Gamma$ leaves C_1 invariant (because T_{n+1} commutes with each T_j, $1 \leq j \leq n$). Hence by the case already treated, we obtain $\mathrm{fix}(T_{n+1}) \cap C_1 \neq \emptyset$. □

With this result at hand we can prove the existence of invariant measures for any TDS.

Theorem 10.2 (Krylov–Bogoljubov). *For every TDS $(K; \varphi)$ there exists at least one φ-invariant probability measure on K. More precisely, for every $0 \neq f \in \mathrm{fix}(T_\varphi)$ in $\mathrm{C}(K)$ there exists $\mu \in \mathrm{M}^1_\varphi(K)$ such that $\langle f, \mu \rangle \neq 0$.*

Proof. We apply the Markov–Kakutani theorem to $\Gamma = \{T'_\varphi\}$ and to the weakly* compact, convex set

$$C := \mathrm{M}^1(K) \cap \big\{ \mu \in \mathrm{M}(K) : \langle f, \mu \rangle = \|f\|_\infty \big\}.$$

By compactness there is $x_0 \in K$ with $\delta_{x_0} \in C$, so C is nonempty. Since $f \in \mathrm{fix}(T_\varphi)$, T'_φ leaves C invariant, by Theorem 10.1 the assertion is proved. □

Remarks 10.3. 1) Let $T = T_\varphi$. As has been mentioned, $\mu \in \mathrm{M}(K)$ is φ-invariant if and only if $T'\mu = \mu$, i.e., $\mu \in \mathrm{fix}(T')$. Hence Theorem 10.2 implies in particular that

$$\mathrm{fix}(T') \quad \text{separates} \quad \mathrm{fix}(T).$$

Remarkably enough, this holds for any power-bounded operator on a general Banach space (see Exercise 8.6).

2) Theorem 10.2 has a generalisation to so-called **Markov operators**, i.e., operators T on $\mathrm{C}(K)$ such that $T \geq 0$ and $T\mathbf{1} = \mathbf{1}$ (see Exercise 2).

3) The proof of the Krylov–Bogoljubov theorem relies on the case of the Markov–Kakutani theorem when $\Gamma = \{T'_\varphi\}$ is a singleton. Here the use of Lemma 8.17 makes it possible to consider any subsequence $(\mathrm{A}_{n_k}[T'_\varphi])$, a cluster point of which will be a fixed point of T'_φ. This "trick" can be used to produce invariant measures with particular properties, as will be done in Chapter 18.

The following result shows that the extreme points of the convex compact set $\mathrm{M}^1_\varphi(K)$ (see Appendix C.7) are of special interest.

Proposition 10.4. *Let $(K;\varphi)$ be a TDS. Then $\mu \in \mathrm{M}^1_\varphi(K)$ is an extreme point of $\mathrm{M}^1_\varphi(K)$ if and only if the MDS $(K,\mu;\varphi)$ is ergodic.*

A φ-invariant probability measure $\mu \in \mathrm{M}^1_\varphi(K)$ is called **ergodic** if the corresponding MDS $(K,\mu;\varphi)$ is ergodic. By the lemma, μ is ergodic if and only if it is an extreme point of $\mathrm{M}^1_\varphi(K)$. By the Kreĭn–Milman theorem (Theorem C.13) such extreme points do always exist. Furthermore, we have

$$\mathrm{M}^1_\varphi(K) = \overline{\mathrm{conv}}\big\{\mu \in \mathrm{M}^1_\varphi(K) : \mu \text{ is ergodic}\big\}, \tag{10.1}$$

where the closure is to be taken in the weak* topology. Employing the so-called *"Choquet theory"* one can go even further and represent an arbitrary φ-invariant measure as a *"barycenter"* of ergodic measures, see [Phelps (1966)].

Proof. Assume that φ is not ergodic. Then there exists a set A with $0 < \mu(A) < 1$ such that $\varphi^* A = A$. The probability measure μ_A, defined by

$$\mu_A(B) := \mu(A \cap B)/\mu(A) \qquad (B \in \mathrm{Ba}(K)),$$

is clearly φ-invariant, i.e., $\mu_A \in \mathrm{M}^1_\varphi(K)$. Analogously, $\mu_{A^\mathrm{c}} \in \mathrm{M}^1_\varphi(K)$. But

$$\mu = \mu(A)\mu_A + (1-\mu(A))\mu_{A^\mathrm{c}}$$

is a representation of μ as a nontrivial convex combination, and hence μ is not an extreme point of $\mathrm{M}^1_\varphi(K)$.

Conversely, suppose that μ is ergodic and that $\mu = (\mu_1+\mu_2)/2$ for some $\mu_1, \mu_2 \in \mathrm{M}^1_\varphi(K)$. This implies in particular that $\mu_1 \leq 2\mu$, hence

$$|\langle f, \mu_1\rangle| \leq 2\,\langle |f|, \mu\rangle = 2\,\|f\|_{\mathrm{L}^1(\mu)} \qquad (f \in \mathrm{C}(K)).$$

Since $\mathrm{C}(K)$ is dense in $\mathrm{L}^1(K,\mu)$, the measure μ_1 belongs to $\mathrm{L}^1(K,\mu)'$. Let us consider the Koopman operator T_φ on $\mathrm{L}^1(K,\mu)$. Since μ is ergodic, $\mathrm{fix}(T_\varphi)$ is one-dimensional. By Example 8.24 the operator T_φ is mean ergodic, and by Theorem 8.20(v) $\mathrm{fix}(T_\varphi)$ separates $\mathrm{fix}(T'_\varphi)$. Hence the latter is one-dimensional too, and since

$\mu, \mu_1 \in \mathrm{fix}(T'_\varphi)$ it follows that $\mu_1 = \mu$. Consequently, μ is an extreme point of $\mathrm{M}^1_\varphi(K)$. □

By this lemma, if there are two different φ-invariant probability measures, then there is also a nonergodic one. Conversely, all φ-invariant measures are ergodic if and only if there is exactly one φ-invariant probability measure. This leads us to the next section.

10.2 Uniquely and Strictly Ergodic Systems

Topological dynamical systems that have a unique invariant probability measure are called **uniquely ergodic**. To analyse this notion further, we need the following useful information.

Proposition 10.5. *Let $(K;\varphi)$ be a TDS. Then $\mathrm{M}_\varphi(K)$ is a lattice, i.e., if $\nu \in \mathrm{M}_\varphi(K)$, then also $|\nu| \in \mathrm{M}_\varphi(L)$. Consequently, $\mathrm{M}^1_\varphi(K)$ linearly spans $\mathrm{M}_\varphi(K)$.*

Proof. Let T be the associated Koopman operator on $\mathrm{C}(K)$ and take $\nu \in \mathrm{fix}(T') = \mathrm{M}_\varphi(K)$. By Exercise 8 we obtain $|\nu| = |T'\nu| \le T'|\nu|$ and

$$\langle \mathbf{1}, |\nu| \rangle \le \langle \mathbf{1}, T'|\nu| \rangle = \langle \mathbf{1}, |\nu| \rangle.$$

This implies that $\langle \mathbf{1}, T'|\nu| - |\nu| \rangle = 0$, hence $|\nu| \in \mathrm{fix}(T')$. To prove the second statement, note that $\nu \in \mathrm{fix}(T')$ if and only if $\mathrm{Re}\,\nu, \mathrm{Im}\,\nu \in \mathrm{fix}(T')$. But if ν is a real (=signed) measure in $\mathrm{fix}(T')$, then $\nu = \nu^+ - \nu^-$ with $\nu^+ = 1/2(|\nu| + \nu)$ and $\nu^- = 1/2(|\nu| - \nu)$ both in $\mathrm{M}_\varphi(K)_+$ (see Chapter 7.2). □

As a consequence we can characterise unique ergodicity.

Theorem 10.6. *Let $(K;\varphi)$ be a TDS with Koopman operator T on $\mathrm{C}(K)$. The following assertions are equivalent.*

(i) *$(K;\varphi)$ is uniquely ergodic, i.e., $\mathrm{M}^1_\varphi(K)$ is a singleton.*

(ii) $\dim \mathrm{M}_\varphi(K) = 1$.

(iii) *Every invariant probability measure is ergodic.*

(iv) *T is mean ergodic and* $\dim \mathrm{fix}(T) = 1$.

(v) *For each $f \in \mathrm{C}(K)$ there is $c(f) \in \mathbb{C}$ such that*

$$(\mathrm{A}_n[T]f)(x) \to c(f) \quad \text{as } n \to \infty \qquad \text{for all } x \in K.$$

Proof. The equivalence (i) $\Leftrightarrow$ (ii) is clear by Proposition 10.5. The equivalence (i) $\Leftrightarrow$ (iii) follows from (10.1).

If (ii) holds then $\mathrm{fix}(T)$ is one-dimensional, since by Theorem 10.2 $\mathrm{fix}(T') = \mathrm{M}_\varphi(K)$ separates the points of $\mathrm{fix}(T)$. But then also $\mathrm{fix}(T)$ separates the points of $\mathrm{fix}(T')$,

and hence by Theorem 8.20, T is mean ergodic. If (iv) holds then $Pf := \lim_{n\to\infty} \mathrm{A}_n f$ exists uniformly, and is a constant function $Pf = c(f)\mathbf{1}$, hence (v) holds.

Finally, suppose that (v) holds. If $\nu \in \mathrm{M}^1_\varphi(K)$ is arbitrary and $f \in \mathrm{C}(K)$ then by dominated convergence,

$$\langle f, \nu \rangle = \langle \mathrm{A}_n f, \nu \rangle \to \langle c(f)\mathbf{1}, \nu \rangle = c(f).$$

Hence $\mathrm{M}^1_\varphi(K)$ consists of at most one point, and hence (i) holds. □

Let K be compact and let $\mu \in \mathrm{M}^1(K)$ be any probability measure on K. Recall from Section 5.2 that the **support** $\operatorname{supp}(\mu)$ of μ is the unique closed set $M \subseteq K$ such that for $f \in \mathrm{C}(K)$

$$f \in I_M := \big\{ g \in \mathrm{C}(K) : g \equiv 0 \quad \text{on } M \big\} \quad \Leftrightarrow \quad \int_K |f| \, \mathrm{d}\mu = 0. \tag{10.2}$$

Equivalently, we have

$$\begin{aligned} \operatorname{supp}(\mu) &= \big\{ x \in K : \mu(U) > 0 \text{ for each open set } x \in U \subseteq K \big\} \\ &= \bigcap \big\{ A \subseteq K : A \text{ is closed and } \mu(A) = 1 \big\} \end{aligned}$$

(see Lemma 5.9 and Exercise 5.12). The following observation is quite important.

Lemma 10.7. *Let $(K;\varphi)$ be a TDS and $\mu \in \mathrm{M}^1(K)$ a φ-invariant measure. Then the support $\operatorname{supp}(\mu)$ of μ is φ-stable, i.e., it satisfies $\varphi(\operatorname{supp}(\mu)) = \operatorname{supp}(\mu)$.*

Proof. Let $M := \operatorname{supp}(\mu)$. Since T is a lattice homomorphism,

$$\int_K |Tf| \, \mathrm{d}\mu = \int_K T\,|f| \, \mathrm{d}\mu = \int_K |f| \, \mathrm{d}\mu,$$

whence by (10.2) $f \in I_M$ if and only if $Tf \in I_M$. By Lemma 4.14 we conclude that $\varphi(M) = M$. The remaining statement follows readily. □

The measure $\mu \in \mathrm{M}^1(K)$ has **full support** or is **strictly positive** if $\operatorname{supp}(\mu) = K$. With the help of Lemma 10.7 we obtain a new characterisation of minimal systems.

Proposition 10.8. *For a TDS $(K;\varphi)$ the following assertions hold.*

a) *Every minimal set $M \subseteq K$ is the support of an invariant probability measure on K.*

b) *$(K;\varphi)$ is minimal if and only if every invariant probability measure on K is strictly positive.*

Proof. By Lemma 10.7 every invariant probability measure for a minimal TDS has full support. Then a) follows by the Krylov–Bogoljubov theorem applied to the subsystem $(M;\varphi)$, and b) follows from a). □

A TDS $(K;\varphi)$ is called **strictly ergodic** if it is uniquely ergodic and the unique invariant probability measure μ has full support. We obtain the following characterisation.

Corollary 10.9. *Let $(K;\varphi)$ be a TDS with Koopman operator T on $\mathrm{C}(K)$. The following assertions are equivalent.*

(i) *The TDS $(K;\varphi)$ is strictly ergodic.*

(ii) *The TDS $(K;\varphi)$ is minimal and T is mean ergodic.*

Proof. (i)$\Rightarrow$(ii): Suppose that $(K;\varphi)$ is strictly ergodic, and let μ be the unique invariant probability measure on K. By Theorem 10.6, T is mean ergodic. Now let $\emptyset \neq M \subseteq K$ be φ-invariant. Then by Theorem 10.2 there is $\nu \in \mathrm{M}^1_\varphi(M)$, which then can be canonically extended to all K such that $\mathrm{supp}(\nu) \subseteq M$, see also Exercise 11. But then $\mu = \nu$ and hence $M = K$.

(ii)$\Rightarrow$(i): If $(K;\varphi)$ is minimal, it is forward transitive, so $\mathrm{fix}(T) = \mathbb{C}\cdot\mathbf{1}$. Mean ergodicity of T implies that $\mathrm{fix}(T)$ separates $\mathrm{fix}(T') = \mathrm{M}_\varphi(K)$, which is possible only if $\dim \mathrm{M}_\varphi(K) = 1$, too. Hence, $(K;\varphi)$ is uniquely ergodic, with unique invariant probability measure μ, say. By minimality, μ has full support (Proposition 10.8). □

Minimality is *not* characterised by unique ergodicity. An easy example of a uniquely ergodic system that is not minimal (not even transitive) is given in Exercise 4. Minimal systems that are not uniquely ergodic are much harder to obtain, see [Furstenberg (1961)] and [Raimi (1964)].

10.3 Mean Ergodicity of Group Rotations

We turn our attention to various ergodic theoretic properties of rotations on a compact group G. By Theorem 8.8 the associated Koopman operator is mean ergodic on $\mathrm{L}^p(G)$ for $p \in [1,\infty)$. We investigate now its behaviour on $\mathrm{C}(G)$, starting with the special case $G = \mathbb{T}$.

Proposition 10.10. *The Koopman operator L_a associated with a rotation system $(\mathbb{T};a)$ is mean ergodic on $\mathrm{C}(\mathbb{T})$.*

Proof. We write $T = L_a$ for the Koopman operator and $\mathrm{A}_n = \mathrm{A}_n[T]$ its Cesàro averages. The linear hull of the functions $\chi_n : z \mapsto z^n$, $n \in \mathbb{Z}$, is a dense subalgebra of $\mathrm{C}(\mathbb{T})$, by the Stone–Weierstrass Theorem 4.4. Since T is power-bounded and therefore Cesàro-bounded, it suffices to show that $\mathrm{A}_n\chi_m$ converges for every $m \in \mathbb{Z}$. Note that $T\chi_m = a^m\chi_m$ for $m \in \mathbb{N}$, hence if $a^m = 1$, then $\chi_m \in \mathrm{fix}(T)$ and there is nothing to show. So suppose that $a^m \neq 1$. Then

$$\mathrm{A}_n\chi_m = \frac{1}{n}\sum_{j=0}^{n-1} T^j\chi_m = \left(\frac{1}{n}\sum_{j=0}^{n-1} a^{mj}\right)\chi_m = \frac{1}{n}\,\frac{1-a^{mn}}{1-a^m}\,\chi_m \to 0$$

as $n \to \infty$. (Compare this with the final problem in Chapter 1.) □

To establish the analogous result for general group rotations, we need an auxiliary result.

Proposition 10.11. *Let G be a compact group and let $f \in \mathrm{C}(G)$. Then the mapping*

$$G \to \mathrm{C}(G), \quad a \mapsto L_a f$$

is continuous.

Proof. For $g, h \in \mathrm{C}(G)$ define $(g \otimes h)(x,y) := g(x)h(y)$ and consider the linear space

$$Y := \operatorname{lin}\big\{g \otimes h : g, h \in \mathrm{C}(G)\big\} \subseteq \mathrm{C}(G \times G).$$

This is a conjugation invariant subalgebra of $\mathrm{C}(G \times G)$, trivially separating the points of $G \times G$. By the Stone–Weierstraß Theorem 4.4, Y is dense in $\mathrm{C}(G \times G)$. Define $F(x,y) := (L_x f)(y) = f(xy)$, which is, by the continuity of the multiplication on G, a continuous function, i.e., $F \in \mathrm{C}(G \times G)$. By what is said above, we can take a sequence $(F_n)_{n\in\mathbb{N}} \subseteq Y$ converging to F. Let $a \in G$ and $\varepsilon > 0$ be fixed. For every $x \in G$ we have

$$\begin{aligned}\|L_a f - L_x f\|_\infty &= \|F(a,\cdot) - F(x,\cdot)\|_\infty \\ &\le \|F(a,\cdot) - F_n(a,\cdot)\|_\infty + \|F_n(a,\cdot) - F_n(x,\cdot)\|_\infty + \|F_n(x,\cdot) - F(x,\cdot)\|_\infty \\ &\le 2\varepsilon + \|F_n(a,\cdot) - F_n(x,\cdot)\|_\infty,\end{aligned}$$

if we choose $n \in \mathbb{N}$ sufficiently large. Since the function F_n is a linear combination of functions of the form $g \otimes h$ and since $n \in \mathbb{N}$ is fixed, from

$$\|g \otimes h(a,\cdot) - g \otimes h(x,\cdot)\|_\infty = |g(a) - g(x)| \cdot \|h\|_\infty$$

it follows that

$$\|L_a f - L_x f\|_\infty \le 2\varepsilon + \|F_n(a,\cdot) - F_n(x,\cdot)\|_\infty < 3\varepsilon$$

if $x \in G$ lies in a suitably small neighbourhood of a. □

(See also Corollary 14.12 for a proof of Proposition 10.11 that avoids the use of the Stone–Weierstrass theorem.) As a consequence we obtain the following generalisation of Proposition 10.10.

Corollary 10.12. *The Koopman operator associated with a compact group rotation system $(G;a)$ is mean ergodic on $\mathrm{C}(G)$.*

Proof. Let $f \in \mathrm{C}(G)$. Then by Proposition 10.11 the set $\{L_g f : g \in G\} \subseteq \mathrm{C}(G)$ is compact. A fortiori, the set $\{L_a^n f : n \in \mathbb{N}_0\}$ is relatively compact. Consequently, its closed convex hull is compact, whence Theorem 8.20 (iii) concludes the proof. □

Combining the results of this chapter so far, we arrive at the following important characterisation. (We include also the equivalences known from previous chapters.)

Theorem 10.13 (Minimal Group Rotations). *Let G be a compact group with Haar probability measure* m *and consider a group rotation $(G;a)$ with associated Koopman operator L_a on* C(G). *Then the following assertions are equivalent.*

(i) $(G;a)$ *is minimal.*

(ii) $(G;a)$ *is (forward) transitive.*

(iii) $\{a^n : n \geq 0\}$ *is dense in G.*

(iv) $\{a^n : n \in \mathbb{Z}\}$ *is dense in G.*

(v) $\dim \mathrm{fix}(L_a) = 1$.

(vi) $(G;a)$ *is uniquely ergodic.*

(vii) *The Haar measure* m *is the only invariant probability measure for $(G;a)$.*

(viii) $(G;a)$ *is strictly ergodic.*

(ix) $(G,\mathrm{m};a)$ *is ergodic.*

Proof. The equivalence of (i)–(iv) has been shown in Theorems 2.32 and 3.4. The implication (ii)$\Rightarrow$(v) follows from Lemma 4.18. Combining $\dim \mathrm{fix}(L_a) = 1$ with the mean ergodicity of L_a (Corollary 10.12) we obtain the implication (v)$\Rightarrow$(vi) by Theorem 10.6. (vi)$\Rightarrow$(vii) is clear since the Haar measure is obviously invariant, and (vii)$\Rightarrow$(viii) follows since the Haar measure has full support (Theorem 14.13.) By Corollary 10.9, (viii) implies that $(G;a)$ is minimal, hence (i).

By Theorem 10.6, (vii) implies (ix). Conversely, suppose that $(G,\mathrm{m};a)$ is ergodic. Then $\mathrm{fix}(L_a) = \mathbb{C}\mathbf{1}$, where L_a is considered as an operator on $\mathrm{L}^1(G,\mathrm{m})$. Since m has full support, the canonical map $\mathrm{C}(G) \hookrightarrow \mathrm{L}^1(G;\mathrm{m})$ is injective, whence (v) follows. □

10.4 Furstenberg's Theorem on Group Extensions

Let G be a compact group with Haar measure m, and let $(K;\varphi)$ be a TDS. Furthermore, let $\Phi : K \to G$ be continuous and consider the group extension along Φ with dynamics

$$\psi : K \times G \to K \times G, \qquad \psi(x,y) := (\varphi(x), \Phi(x)y),$$

see Section 2.2.2. Let $\pi : K \times G \to K$, $\pi(x,g) := x$ be the corresponding factor map. Of course, for any φ-invariant probability measure μ on K the measure $\mu \otimes \mathrm{m}$ is an invariant measure for the TDS $(K \times G;\psi)$. Even if $(K;\varphi)$ is uniquely ergodic, the product need not be so. As an example consider the product of two commensurable irrational rotations, which is not minimal (see Example 2.34), so by Theorem 10.13 it is not uniquely ergodic.

Therefore, in this section we look for conditions guaranteeing that the group extension is uniquely ergodic. We need the following preparations. The group G acts on G by multiplication from the right by $a \in G$, with associated Koopman operator

$$R_a g(y) = g(ya) \qquad (a, y \in G, g \in \mathrm{C}(G)). \tag{10.3}$$

Also the right rotation ρ_a by an element $a \in G$ in the second coordinate is an automorphism of the group extension $(K \times G; \psi)$, see Section 2.2.2. The associated Koopman operator is $\mathrm{I} \otimes R_a$, which acts on $f \otimes g$ as $f \otimes R_a g$, and commutes with the Koopman operator T_ψ, i.e., $T_\psi(\mathrm{I} \otimes R_a) = (\mathrm{I} \otimes R_a)T_\psi$.

Theorem 10.14. *Let ν be a ψ-invariant probability measure on $K \times G$, and let $\mu := \pi_* \nu$. If $\mu \otimes \mathrm{m}$ is ergodic, then $\nu = \mu \otimes \mathrm{m}$.*

Proof. The product measure $\mu \otimes \mathrm{m}$ is not only ψ-invariant, but also invariant under ρ_a for all $a \in G$. Fix $f \in \mathrm{C}(K)$ and $g \in \mathrm{C}(G)$. Since the MDS $(K \times G, \mu \otimes \mathrm{m}; \psi)$ is ergodic, as a consequence of von Neumann's Theorem 8.1, we find a *subsequence* (n_k) such that for $\mathrm{A}_{n_k} := \mathrm{A}_{n_k}[T_\psi]$ we have

$$\mathrm{A}_{n_k}(f \otimes g) \to \langle f, \mu \rangle \langle g, \mathrm{m} \rangle \cdot \mathbf{1} \qquad \mu \otimes \mathrm{m}\text{-almost everywhere.} \tag{10.4}$$

For $a \in G$ we define

$$E(a) := \big\{(x,y) \in K \times G : \mathrm{A}_{n_k}(f \otimes R_a g)(x,y) \to \langle f, \mu \rangle \langle g, \mathrm{m} \rangle\big\}.$$

Note that $\rho_a(x,y) \in E(1)$ if and only if $(x,y) \in E(a)$ by (10.3). Hence, by (10.4) we have $(\mu \otimes \mathrm{m})(E(a)) = 1$ for all $a \in G$.

Since the mapping

$$G \times \mathrm{C}(G) \to \mathrm{C}(G), \quad (a, g) \mapsto R_a g$$

is continuous (by Proposition 10.11), the orbit

$$R_G g := \big\{R_a g : a \in G\big\}$$

is a compact subset of $\mathrm{C}(G)$. In particular, it is separable, and we find a countable subset $C_g \subseteq G$ such that $\{R_a g : a \in C_g\}$ is dense in R_G. For such a countable set define

$$E := \bigcap\nolimits_{a \in C_g} E(a),$$

which satisfies $(\mu \otimes \mathrm{m})(E) = 1$. We claim that actually $E = \bigcap_{a \in G} E(a)$.

Proof of Claim. The inclusion $\supseteq$ is clear. For the converse inclusion fix $a \in G$ and take $(a_m)_m \subseteq C_g$ such that $\|R_{a_m} g - R_a g\|_\infty \to 0$ as $m \to \infty$. Then $\mathrm{A}_{n_k}(f \otimes R_{a_m} g) \to \mathrm{A}_{n_k}(f \otimes R_a g)$ uniformly in $k \in \mathbb{N}$ as $m \to \infty$. Hence, if $(x,y) \in E(a_m)$ for each $m \in \mathbb{N}$, then by a standard argument, $(x,y) \in E(a)$, too. □

Now, for $x \in K$ and $y, a \in G$ we have the following chain of equivalences:

$$\begin{aligned}(x, ya) \in E &\Leftrightarrow \forall\, b \in G : (x, ya) \in E(b) \Leftrightarrow \forall\, b \in G : (x, yab) \in E(1) \\ &\Leftrightarrow \forall\, b \in G : (x,y) \in E(ab) \Leftrightarrow (x,y) \in E.\end{aligned}$$

Hence, $E = E_1 \times G$, where $E_1 \subseteq K$ is the section $E_1 := \{x \in K : (x,1) \in E\}$. Since E is a Baire set in $K \times G$, and $\mathrm{Ba}(K \times G) = \mathrm{Ba}(K) \otimes \mathrm{Ba}(G)$ (Exercise 5.8), by standard

measure theory E_1 is measurable, whence $\mu(E_1) = 1$. By definition of μ, this means that $\nu(E) = 1$, i.e.,

$$A_{n_k}(f \otimes g) \to \langle f, \mu \rangle \langle g, \mathrm{m} \rangle \qquad \nu\text{-almost everywhere.}$$

From the dominated convergence we obtain

$$\langle f \otimes g, \nu \rangle = \left\langle \mathrm{A}_{n_k}(f \otimes g), \nu \right\rangle \to \langle f, \mu \rangle \langle g, \mathrm{m} \rangle .$$

As we saw in the proof of Proposition 10.11 the linear hull of functions $f \otimes g$ is dense in $\mathrm{C}(K \times G)$, so by the Riesz Representation Theorem 5.7 the proof is complete. □

A direct consequence is the following result of Furstenberg.

Corollary 10.15 (Furstenberg). *If $(K; \varphi)$ is uniquely ergodic with invariant probability measure μ, and if $\mu \otimes \mathrm{m}$ is ergodic, then $(K \times G; \psi)$ is uniquely ergodic.*

The usual proof of Furstenberg's result relies on the *pointwise ergodic theorem* (which is the subject of the next chapter) and on *generic points* (see Exercise 11.5). We now turn to an important example of the situation from above.

Unique Ergodicity of Skew Shifts

Let $\alpha \in [0,1)$ be an irrational number. By Theorem 10.13, the rotation $\varphi_\alpha : x \mapsto x + \alpha \pmod 1$ defines a uniquely ergodic TDS. We consider here its group extension along the function $\Phi : [0,1) \to [0,1)$, $\Phi(x) = x$, i.e.,

$$K := [0,1]^2 \pmod 1 \quad \text{and} \quad \psi_\alpha(x,y) = (\varphi_\alpha(x), \Phi(x) + y) = (x + \alpha, x + y).$$

Then $(K; \psi_\alpha)$ is a TDS, see Chapter 2, Example 2.20.

Our aim is to show that this TDS, called the **skew rotation**, is strictly ergodic. The two-dimensional Lebesgue measure λ_2 on K is ψ-invariant, so it remains to prove that this measure is actually the only ψ-invariant probability measure on K. The next is an auxiliary result, needed for the application of Corollary 10.15.

Proposition 10.16. *Let $\alpha \in [0,1)$ be an irrational number. Then the skew rotation $(K, \lambda_2; \psi_\alpha)$ is ergodic.*

Proof. To prove the assertion we use Proposition 7.14 and show that the fixed space of the Koopman operator T of ψ_α on $\mathrm{L}^2(K, \lambda_2)$ consists of the constant functions only. For $f, g \in \mathrm{L}^2[0,1]$ the function $f \otimes g : (x,y) \mapsto f(x)g(y)$ belongs to $\mathrm{L}^2(K, \lambda_2)$. Now every $F \in \mathrm{L}^2(K, \lambda_2)$ can be written uniquely as an L^2-sum

$$F = \sum_{n \in \mathbb{Z}} f_n \otimes e_n,$$

where $(f_n)_{n \in \mathbb{Z}} \subseteq \mathrm{L}^2[0,1]$ and $(e_n)_{n \in \mathbb{Z}}$ is the orthonormal basis

$$e_n(y) := \mathrm{e}^{2\pi \mathrm{i} n y} \qquad (y \in [0,1],\ n \in \mathbb{Z})$$

of $\mathrm{L}^2[0,1]$. Applying T to F yields

$$TF = \sum_{n\in\mathbb{Z}} (e_n L_\alpha f_n) \otimes e_n,$$

hence the condition $F \in \mathrm{fix}(T)$ is equivalent to

$$L_\alpha f_n = e_{-n} f_n \quad \text{for every } n \in \mathbb{Z}.$$

For $n \in \mathbb{Z}$ with $f_n \neq 0$ we have to show that $n = 0$. Since

$$L_\alpha |f_n| = |L_\alpha f_n| = |e_{-n} f_n| = |f_n|$$

and α is irrational, $|f_n|$ is a constant (nonzero) function. Now take $\theta \in [0,1)$ rationally independent from α and consider the Koopman operator $L_\theta \in \mathscr{L}(\mathrm{L}^2[0,1])$ associated with the translation by θ on $[0,1)$. For $g_n := f_n/(L_\theta f_n)$ we then have

$$L_\alpha g_n = L_\alpha \frac{f_n}{L_\theta f_n} = \frac{L_\alpha f_n}{L_\theta L_\alpha f_n} = \frac{e_{-n} f_n}{e_{-n}(\theta) e_{-n} L_\theta f_n} = e_n(\theta) g_n.$$

By Exercise 10 the eigenvalues of L_α are of the form $\mathrm{e}^{2\pi \mathrm{i} m\alpha}$ for some $\mathbb{Z}$, so $e_n(\theta) = \mathrm{e}^{2\pi \mathrm{i} m\alpha} = e_m(\alpha)$ for some $m \in \mathbb{Z}$. But since θ is rationally independent from α, it follows that $n = 0$ and thus $F = f_0 \otimes e_0$. Since $L_\alpha f_0 = f_0$ and φ_α is ergodic, we have that f_0 and hence F is constant. □

Since λ_2 is strictly positive, the TDS $(K;\psi_\alpha)$ is even strictly ergodic. So by Corollary 10.9 we obtain the next result.

Corollary 10.17. *For irrational $\alpha \in \mathbb{R}$ consider the mod 1 translation system $([0,1);\alpha)$ and the corresponding skew rotation $(K;\psi_\alpha)$ on $K = [0,1)\times[0,1)$. Then $(K;\psi_\alpha)$ is strictly ergodic ergodic and its associated Koopman operator is mean ergodic on $\mathrm{C}(K)$.*

10.5 Application: Equidistribution

A sequence $(\alpha_n)_{n=0}^\infty \subseteq [0,1)$ is called **equidistributed** in [0,1) if

$$\lim_{n\to\infty} \frac{\mathrm{card}\{j : 0 \le j < n,\ \alpha_j \in [a,b]\}}{n} = b - a$$

for all a,b with $0 \le a < b \le 1$. In other words, the relative frequency of the first n elements of the sequence falling into an arbitrary fixed interval $[a,b]$ converges to the length of that interval (independently of its location).

In this section we show how mean ergodicity of the Koopman operator on various spaces can be used to study equidistribution of some sequences.

A classical theorem from [Weyl (1916)] gives an important example of an equidistributed sequence and is a quantitative version of Kronecker's theorem in Example 2.33.

Theorem 10.18 (Weyl). *Let $\alpha \in [0,1) \setminus \mathbb{Q}$. Then the sequence $(n\alpha \pmod 1)_{n=0}^{\infty}$ is equidistributed in $[0,1)$.*

We consider the translation system $([0,1);\alpha)$ as in Example 2.6. Recall from Example 2.13 that this TDS is isomorphic to the rotation system $(\mathbb{T};a)$, where $a = \mathrm{e}^{2\pi i\alpha}$. As we saw in the previous section, the Koopman operator L_a is mean ergodic on $\mathrm{C}(\mathbb{T})$, hence so is the Koopman operator L_α of $([0,1);\alpha)$ on $\mathrm{C}([0,1))$. The latter space is, however, too small for our purposes.

Instead, a better candidate to work with seems to be $\mathrm{R}[0,1]$, the space of bounded, 1-periodic functions on $\mathbb{R}$ that are Riemann integrable over compact intervals, because this space contains characteristic functions of intervals. Equipped with the sup-norm, $\mathrm{R}[0,1]$ becomes a Banach space, and the Koopman operator T associated with the translation $x \mapsto x+\alpha$ leaves $\mathrm{R}[0,1]$ invariant, i.e, restricts to an isometric isomorphism of $\mathrm{R}[0,1]$.

We shall use Riemann's criterion of integrability: a bounded real-valued function $f : [0,1] \to \mathbb{R}$ is Riemann integrable if and only if for every $\varepsilon > 0$ there exist step functions g_ε, h_ε such that

$$g_\varepsilon \le f \le h_\varepsilon \quad \text{and} \quad \int_0^1 (h_\varepsilon - g_\varepsilon)(t)\,\mathrm{d}t \le \varepsilon.$$

Clearly, if f is 1-periodic (i.e., if $f(0) = f(1)$), then g_ε and h_ε can be chosen 1-periodic as well.

Proposition 10.19. *Let $\alpha \in [0,1) \setminus \mathbb{Q}$. Then the Koopman operator associated with the translation system $([0,1);\alpha)$ is mean ergodic on $\mathrm{R}[0,1]$, and the mean ergodic projection is given by*

$$Pf = \int_0^1 f(t)\,\mathrm{d}t \cdot \mathbf{1} \qquad (f \in \mathrm{R}[0,1]).$$

Proof. We denote as usual the Koopman operator by T and its Cesàro averages by A_n, $n \in \mathbb{N}$. We shall identify functions defined on $[0,1)$ with their 1-periodic extension to $\mathbb{R}$.

Let χ be any characteristic function of an interval (open, closed or half-open) contained $[0,1)$, and let $\varepsilon > 0$. A moment's thought reveals that there exist *continuous* 1-periodic functions g_ε, h_ε on $\mathbb{R}$ satisfying

$$g_\varepsilon \le \chi \le h_\varepsilon \quad \text{and} \quad \int_0^1 (h_\varepsilon - g_\varepsilon)(t)\,\mathrm{d}t \le \varepsilon.$$

By Proposition 10.10 and by the isomorphism of the rotation on the torus and the mod 1 translation on $[0,1)$, $\mathrm{A}_n g_\varepsilon$ converges uniformly to $(\int_0^1 g_\varepsilon(t)\,\mathrm{d}t)\cdot\mathbf{1}$, while $\mathrm{A}_n h_\varepsilon$ converges uniformly to $(\int_0^1 h_\varepsilon(t)\,\mathrm{d}t)\cdot\mathbf{1}$. Since $\mathrm{A}_n g_\varepsilon \le \mathrm{A}_n\chi \le \mathrm{A}_n h_\varepsilon$ for all $n\in\mathbb{N}$, we obtain

$$\left(\int_0^1 g_\varepsilon(t)\,\mathrm{d}t-\varepsilon\right)\cdot\mathbf{1}\le \mathrm{A}_n\chi\le\left(\int_0^1 h_\varepsilon(t)\,\mathrm{d}t+\varepsilon\right)\cdot\mathbf{1}$$

for sufficiently large n. So we have for such n that

$$\left\|\mathrm{A}_n\chi-\int_0^1\chi(t)\,\mathrm{d}t\cdot\mathbf{1}\right\|_\infty\le 2\varepsilon$$

by the choice of g_ε and h_ε. Therefore

$$\|\cdot\|_\infty-\lim_{n\to\infty}\mathrm{A}_n\chi=\left(\int_0^1\chi(t)\,\mathrm{d}t\right)\cdot\mathbf{1}.$$

Clearly the convergence on these characteristic functions extends to all 1-periodic step functions. Take now a real-valued $f\in\mathrm{R}[0,1]$. As noted above, for $\varepsilon>0$ we find 1-periodic step functions $g_\varepsilon, h_\varepsilon$ with $g_\varepsilon\le f\le h_\varepsilon$ and $\int_0^1(h_\varepsilon-g_\varepsilon)(t)\,\mathrm{d}t\le\varepsilon$. By a similar argument as above we then obtain

$$\|\cdot\|_\infty-\lim_{n\to\infty}\mathrm{A}_n f=\left(\int_0^1 f(t)\,\mathrm{d}t\right)\cdot\mathbf{1}.\qquad\square$$

By using the mean ergodicity of L_α on $\mathrm{R}[0,1]$ we can now prove Weyl's theorem.

Proof of Weyl's Theorem 10.18. Denote $\alpha_n := n\alpha \pmod 1$ and let

$$f=\begin{cases}\mathbf{1}_{[a,b]} & \text{if } b<1,\\ \mathbf{1}_{\{0\}\cup[a,b]} & \text{if } b=1.\end{cases}$$

Then $f\in\mathrm{R}[0,1]$ and

$$\mathrm{A}_n f(0)=\frac{1}{n}\sum_{j=0}^{n-1}L_\alpha^j f(0)=\frac{1}{n}\sum_{j=0}^{n-1}f(\alpha_j)=\frac{\mathrm{card}\{j:0\le j<n,\ \alpha_j\in[a,b]\}}{n}.$$

By Proposition 10.19, L_α is mean ergodic on $\mathrm{R}[0,1]$ and the Cesàro averages converge uniformly to Pf. In particular

$$\mathrm{A}_n f(0)\to\int_0^1 f(s)\mathrm{d}s=b-a.\qquad\square$$

Note that the last step of the proof consisted essentially in showing that if

$$\frac{1}{n}\sum_{j=0}^{n-1} f(\alpha_j) \to \int_0^1 f(s)\mathrm{d}s \quad (n\to\infty)$$

for all $f \in \mathrm{R}[0,1]$, then the sequence $(\alpha_n)_n$ is equidistributed. The following is a converse to this fact, its proof is left as Exercise 9.

Proposition 10.20. *A sequence $(\alpha_n)_{n=0}^{\infty} \subseteq [0,1)$ is equidistributed if and only if*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(\alpha_j) = \int_0^1 f(s)\,\mathrm{d}s$$

holds for every bounded and Riemann integrable (equivalently, for every continuous) function $f : [0,1] \to \mathbb{C}$.

This result applied to exponential functions $s \mapsto \mathrm{e}^{2\pi \mathrm{i} n s}$ was the basis of Weyl's original proof. More on this circle of ideas can be found in [Hlawka (1979)] or [Kuipers and Niederreiter (1974)].

By using the mean ergodicity of the Koopman operator corresponding to the skew rotation from Corollary 10.17, we obtain the following result on equidistribution.

Proposition 10.21. *Let $\alpha \in [0,1) \setminus \mathbb{Q}$. Then the sequence $(n^2\alpha \pmod 1)_{n=0}^{\infty}$ is equidistributed in $[0,1]$, i.e.,*

$$\lim_{n\to\infty} \frac{\mathrm{card}\{j : 0 \le j < n,\ j^2\alpha \in [a,b] \pmod 1\}}{n} = b - a$$

holds for all a, b with $0 \le a < b \le 1$.

Proof. Consider the TDS $(K;\psi_{2\alpha})$ as in Section 10.4, whose Koopman operator $T := L_{2\alpha}$ is now known to be mean ergodic on $\mathrm{C}(K)$ by Corollary 10.17. In particular, the Cesàro means A_n of T converge to the mean ergodic projection P in the sup-norm, hence pointwise. We saw in Corollary 10.9 that $Pf = \int_K f\,\mathrm{d}\lambda_2$ for every $f \in \mathrm{C}(K)$. Let now $g \in \mathrm{C}([0,1))$ be arbitrary and set $f(x,y) := g(y)$. It is easy to see that one has $\psi_{2\alpha}^n(\alpha,0) = ((2n+1)\alpha, n^2\alpha)$. So the convergence

$$\frac{1}{n}\sum_{j=0}^{n-1} g(n^2\alpha \pmod 1) = \mathrm{A}_n f(\alpha,0) \to \int_K f\,\mathrm{d}\lambda = \int_0^1 g(s)\,\mathrm{d}s$$

and Proposition 10.20 conclude the proof. □

Exercises

1 (Mean Ergodicity on $\mathrm{C}(K)$). Let $(K;\varphi)$ be a TDS with an invariant probability measure μ. Suppose that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f(\varphi^j(x)) = \int_K f\,\mathrm{d}\mu$$

for every $f \in \mathrm{C}(K)$ and every $x \in K$. Show that the Koopman operator T_φ on $\mathrm{C}(K)$ is mean ergodic and that $(K,\mu;\varphi)$ is ergodic. (Hint: Use the Dominated Convergence Theorem to prove *weak mean ergodicity* and then use part (iii) of Theorem 8.20.)

2 (Markov Operators). Let K be a compact topological space. A bounded linear operator $T : \mathrm{C}(K) \to \mathrm{C}(K)$ is called a **Markov operator** if it is *positive* (i.e., $f \geq 0$ implies $Tf \geq 0$) and satisfies $T\mathbf{1} = \mathbf{1}$.

Let T be a Markov operator on $\mathrm{C}(K)$. Show that $\|T\| = 1$ and that there exists a probability measure $\mu \in \mathrm{M}^1(K)$ such that

$$\int_K Tf\,\mathrm{d}\mu = \int_K f\,\mathrm{d}\mu \quad \text{for all} \quad f \in \mathrm{C}(K).$$

(Hint: Imitate the proof of the Krylov–Bogoljubov theorem.)

3. A linear functional ψ on $\ell^\infty = \ell^\infty(\mathbb{N})$ with the following properties is called a **Banach limit**.

(i) ψ is positive, i.e., $(x_n)_{n\in\mathbb{N}_0} \in \ell^\infty$, $x_n \geq 0$ imply $\psi(x_n)_{n\in\mathbb{N}_0} \geq 0$.

(ii) ψ is translation invariant, i.e., $\psi(x_n)_{n\in\mathbb{N}_0} = \psi(x_{n+1})_{n\in\mathbb{N}}$.

(iii) $\psi(\mathbf{1}) = 1$, where $\mathbf{1}$ denotes the constant 1 sequence.

Prove the following assertions.

a) A Banach limit ψ is continuous, i.e., $\psi \in (\ell^\infty)'$.

b) If $(x_n)_{n\in\mathbb{N}_0} \in \ell^\infty$ is periodic with period $k \geq 1$, then for each Banach limit ψ one has

$$\psi(x_n) = \frac{1}{k}\sum_{j=0}^{k-1} x_j.$$

c) If $(x_n) \in \mathrm{c}$ and ψ is a Banach limit, then $\psi(x_n) = \lim_{n\to\infty} x_n$.

d) If $(x_n) \in \ell^\infty$ is Cesàro convergent (i.e., $c_n := \frac{1}{n}\sum_{j=0}^{n-1} x_j$ converges), then for each Banach limit ψ one has

$$\psi(x_n) = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} x_j.$$

e) There exist Banach limits. (Hint: Imitate the proof of the Krylov–Bogoljubov theorem.)
For any $(x_n)_{n\in\mathbb{N}} \in \ell^\infty(\mathbb{R})$ and $\alpha \in [\liminf_{n\to\infty} x_n, \limsup_{n\to\infty} x_n]$ there is a Banach limit ψ with $\psi(x_n) = \alpha$ (see Remark 10.3.3)).

4. Let $K := [0,1]$ and $\varphi(x) = 0$ for all $x \in [0,1]$. Show that $(K;\varphi)$ is uniquely ergodic, but not minimal.

5. Take $K = \{z \in \mathbb{C} : |z| \leq 1\}$ and $\varphi_a(z) = az$ for some $a \in \mathbb{T}$ satisfying $a^n \neq 1$ for all $n \in \mathbb{N}$. Find the ergodic measures for φ_a.

6. For the doubling map on $K = [0,1)$ (cf. Exercise 2.10) and the tent map on $[0,1]$ (cf. Exercise 2.11) answer the following questions:

1) Is the TDS uniquely ergodic?

2) Is the Koopman operator on $\mathrm{C}(K)$ mean ergodic?

7. Let $a \in \mathbb{T}$ with $a^n \neq 1$ for all $n \in \mathbb{N}$. Show that the Koopman operator $T = T_{\varphi_a}$ induced by the rotation φ_a is not mean ergodic on the space $\mathrm{B}(\mathbb{T})$ of bounded, Borel measurable functions endowed with sup-norm. (Hint: Take a 0-1-sequence $\{c_n\}_{n=1}^{\infty}$ which is not Cesàro convergent and consider the characteristic function of the set $\{a^n : c_n = 1\}$.)

8. Let K, L be compact spaces and let $\varphi : K \to L$ be continuous. Let $T := T_\varphi : \mathrm{C}(L) \to \mathrm{C}(K)$ be the associated Koopman operator. Show that $T'\mu = \varphi_*\mu$ is the push-forward measure for every $\mu \in \mathrm{M}(K)$. Then show that $|T'\mu| \leq T'|\mu|$ for every $\mu \in \mathrm{M}(K)$, e.g., by using the definition of the modulus of a measure in Appendix B.9.

9. Prove Proposition 10.20.

10. Let $\alpha \in [0,1)$ be an irrational number and consider the shift $\varphi_\alpha : x \mapsto x + \alpha \pmod 1$ on $[0,1]$. Show that for the point spectrum of the Koopman operator $T = L_\alpha$ on $\mathrm{L}^2[0,1]$ we have $\sigma_p(T) = \{\mathrm{e}^{2\pi i m\alpha} : m \in \mathbb{Z}\}$. (Hint: Use a Fourier expansion as in the proof of Proposition 7.15.)

11. Let $(K;\varphi)$ be a TDS and let $L \subseteq K$ be a compact subset with $\varphi(L) \subseteq L$. Let $R : \mathrm{C}(K) \to \mathrm{C}(M)$ be the restriction operator. Show that the dual operator $R' : \mathrm{M}(L) \to \mathrm{M}(K)$ is an isometric lattice homomorphism satisfying

$$R'(\mathrm{M}^1_\varphi(L)) = \{\mu \in \mathrm{M}^1_\varphi(K) : \mathrm{supp}(\mu) \subseteq L\}.$$

(Hint: First use Tietze's theorem to show that R' is an isometry; then use Exercise 8 and the isometry property of R' to show that $|R'\mu| = |\mu|$ for every $\mu \in \mathrm{M}(L)$; finally use Proposition 5.6 to show that $R'(\mathrm{M}_\varphi(L)) \subseteq \mathrm{M}_\varphi(K)$.)

12. Show the following result:

Proposition. *Let $(K_s;\varphi)$ be the maximal surjective subsystem of a TDS $(K;\varphi)$ (see Corollary 2.24), and let $R : \mathrm{C}(K) \to \mathrm{C}(K_s)$ be the restriction mapping. Then the dual operator $R' : \mathrm{M}(K_s) \to \mathrm{M}(K)$ restricts to an isometric lattice isomorphism*

$$R' : \mathrm{M}_\varphi(K_s) \to \mathrm{M}_\varphi(K).$$

Chapter 11
The Pointwise Ergodic Theorem

While von Neumann's mean ergodic theorem is powerful and far reaching, it does not actually solve the original problem of establishing that "time mean equals space mean" for a given MDS $(\mathrm{X};\varphi)$. For this we need the *pointwise* limit

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

for states $x \in X$ and observables $f : X \to \mathbb{R}$. Of course, by Theorem 8.10 this limit equals the "space mean" $\int_X f\,\mathrm{d}\mu$ for each observable f only if the system is ergodic. Moreover, due to the presence of null-sets we cannot expect the convergence to hold for *all* points $x \in X$. Hence we should ask merely for convergence *almost everywhere*.

Shortly after and inspired by J. von Neumann's mean ergodic theorem, G.D. Birkhoff in [Birkhoff (1931)] proved the desired result.

Theorem 11.1 (Birkhoff). *Let $(X,\Sigma,\mu;\varphi)$ be an MDS and let $f \in \mathrm{L}^1(\mathrm{X})$. Then the limit*

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

exists for μ-almost every $x \in X$.

Birkhoff's theorem is also called the *Pointwise (or: Individual) Ergodic Theorem*. Using the Koopman operator $T = T_\varphi$ and its Cesàro averages $\mathrm{A}_n = \mathrm{A}_n[T]$ we obtain by Birkhoff's theorem that for every $f \in \mathrm{L}^1(\mathrm{X})$, the sequence $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ converges pointwise μ-almost everywhere. Since we already know that T is mean ergodic, the Cesàro averages $\mathrm{A}_n f$ converge in L^1-norm to the projection onto the fixed space of T. Hence Birkhoff's theorem in combination with Theorem 8.10 (v) implies the following characterisation of ergodic MDSs.

Corollary 11.2. *An MDS* $(\mathrm{X};\varphi)$, $\mathrm{X} = (X,\Sigma,\mu)$, *is ergodic if and only if for every ("observable")* $f \in \mathrm{L}^1(\mathrm{X})$ *one has "time mean equals space mean", i.e.,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\varphi^j(x)) = \int_X f \,\mathrm{d}\mu$$

for μ*-almost every ("state")* $x \in X$.

As in the case of von Neumann's theorem, Birkhoff's result is operator theoretic in nature. We therefore proceed as in Chapter 8 and use an abstract approach.

11.1 Pointwise Ergodic Operators

Definition 11.3. Let X be a measure space and let $1 \le p \le \infty$. A bounded linear operator T on $\mathrm{L}^p(\mathrm{X})$ is called **pointwise ergodic** if the limit

$$\lim_{n\to\infty} \mathrm{A}_n f = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} T^j f$$

exists μ-almost everywhere for every $f \in \mathrm{L}^p(\mathrm{X})$.

Birkhoff's theorem says that Koopman operators arising from MDSs are pointwise ergodic. We shall obtain Birkhoff's theorem from a more general result which goes back to [Hopf (1954)] and [Dunford and Schwartz (1958)]. Recall that an operator T on $\mathrm{L}^1(\mathrm{X})$ is called a Dunford–Schwartz operator (or a complete contraction) if one has

$$\|Tf\|_1 \le \|f\|_1 \quad \text{and} \quad \|Tf\|_\infty \le \|f\|_\infty$$

for all $f \in \mathrm{L}^\infty \cap \mathrm{L}^1$. The Hopf–Dunford–Schwartz result then reads as follows.

Theorem 11.4 (Pointwise Ergodic Theorem). *Let* X *be a measure space and let* T *be a positive Dunford–Schwartz operator on* $\mathrm{L}^1(\mathrm{X})$. *Then* T *is pointwise ergodic.*

Before we turn to the proof of Theorem 11.4 let us discuss some further results.

Dunford and Schwartz [Dunford and Schwartz (1958)] have shown that one can omit the condition of positivity from Theorem 11.4. This is due to the fact that for a Dunford– Schwartz operator T there always exists a positive Dunford-Schwartz operator $S \ge 0$ dominating it, in the sense that $|Tf| \le S|f|$ for all $f \in \mathrm{L}^1(\mathrm{X})$. On the other hand, a general positive contraction on $\mathrm{L}^1(\mathrm{X})$ need not be pointwise ergodic. A first example was given in [Chacon (1964)]. Shortly after, A. Ionescu Tulcea proved even that the class of positive isometric isomorphisms on $\mathrm{L}^1[0,1]$ which are not pointwise ergodic is "rich" in the sense of category [IonescuTulcea (1965)].

One can weaken the condition on L^∞-contractivity, though. E.g., Hopf has shown that in place of $T\mathbf{1} \le \mathbf{1}$ (cf. Exercise 8.9) one may suppose that there is a strictly

positive function f such that $Tf \leq f$ [Krengel (1985), Theorem 3.3.5]. For general positive L^1-contractions there is the following result from [Krengel (1985), Theorem 3.4.9].

Theorem 11.5 (Stochastic Ergodic Theorem). *Let* X *be a finite measure space and let T be a positive contraction on* $\mathrm{L}^1(\mathrm{X})$. *Then the Cesàro averages* $(\mathrm{A}_n[T]f)_n$ *converge in measure for every* $f \in \mathrm{L}^1(\mathrm{X})$.

If we pass to L^p spaces with $1 < p < \infty$, the situation improves. Namely, building on [IonescuTulcea (1964)] Akcoglu established in [Akcoglu (1975)] the following celebrated result.

Theorem 11.6 (Akcoglu's Ergodic Theorem). *Let* X *be a measure space and let T be a positive contraction on* $\mathrm{L}^p(\mathrm{X})$ *for some* $1 < p < \infty$. *Then T is pointwise ergodic.*

For $p = 2$ and T self-adjoint, this is due to Stein and has an elementary proof, see [Stein (1961b)]. In the general case the proof is quite involved and beyond the scope of this book, see [Krengel (1985), Section 5.2] or [Kern et al. (1977)] and [Nagel and Palm (1982)]. Burkholder [Burkholder (1962)] has shown that if $p = 2$, the condition of positivity cannot be omitted. (The question whether this is true also for $p \neq 2$ seems still to be open.)

Let us return to Theorem 11.4 and its proof. For simplicity, suppose that $\mathrm{X} = (X, \Sigma, \mu)$ is a *finite* measure space, and T a Dunford–Schwartz operator on X. By Theorem 8.24 we already know that T is mean ergodic, whence

$$\mathrm{L}^1(\mathrm{X}) = \mathrm{fix}(T) \oplus \overline{\mathrm{ran}}(\mathrm{I} - T).$$

Since $\mathrm{L}^\infty(\mathrm{X})$ is dense in $\mathrm{L}^1(\mathrm{X})$, the space

$$F := \mathrm{fix}(T) \oplus (\mathrm{I} - T)\mathrm{L}^\infty(\mathrm{X})$$

is still dense in $\mathrm{L}^1(\mathrm{X})$. As before, we write $\mathrm{A}_n := \mathrm{A}_n[T]$ for $n \in \mathbb{N}$.

Lemma 11.7. *In the situation described above, the sequence* $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ *is* $\|\cdot\|_\infty$*-convergent for every* $f \in F$.

Proof. Write $f = g + (\mathrm{I} - T)h$ with $g \in \mathrm{fix}(T)$ and $h \in \mathrm{L}^\infty(\mathrm{X})$. Then as already seen before, $\mathrm{A}_n f = g + (1/n)(h - T^n h)$, and since $\sup_n \|T^n h\|_\infty \leq \|h\|_\infty$, it follows that $\|\mathrm{A}_n f - g\|_\infty \to 0$ for $n \to \infty$. □

By the lemma we have a.e.-convergence of $(\mathrm{A}_n f)_{n\in\mathbb{N}}$ for every f from the dense subspace F of $\mathrm{L}^1(\mathrm{X})$. What we need is a tool that allows us to pass from F to its closure. This is considerably more difficult than in the case of norm convergence, and will be treated in the next section.

11.2 Banach's Principle and Maximal Inequalities

Let $\mathrm{X} = (X, \Sigma, \mu)$ be a measure space, $1 \le p \le \infty$, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of bounded linear operators on $E := \mathrm{L}^p(\mathrm{X})$. The statement "$\lim_{n\to\infty} T_n f$ exists almost everywhere" can be reformulated as

$$\limsup_{k,l\to\infty} |T_k f - T_l f| := \inf_{n\in\mathbb{N}} \sup_{k,l\ge n} |T_k f - T_l f| = 0, \tag{11.1}$$

where suprema and infima are taken in the complete lattice $\mathrm{L}^0 := \mathrm{L}^0(\mathrm{X};\overline{\mathbb{R}})$ (see Section 7.2).

If (11.1) is already established for f from some dense subspace F of E, one would like to infer that it holds for every $f \in E$. To this purpose we consider the associated **maximal operator** $T^* : E \to \mathrm{L}^0$ defined by

$$T^* f := \sup_{n\in\mathbb{N}} |T_n f| \qquad (f \in E).$$

Note that $T^* f \ge 0$ and $T^*(\alpha f) = |\alpha|\, T^* f$ for every $f \in E$ and $\alpha \in \mathbb{C}$. Moreover, the operator T^* is only *subadditive*, i.e., $T^*(f+g) \le T^* f + T^* g$ for all $f, g \in E$.

Definition 11.8. We say that the sequence of operators $(T_n)_n$ satisfies an (abstract) **maximal inequality** if there is a function $c : (0,\infty) \to [0,\infty)$ with $\lim_{\lambda\to\infty} c(\lambda) = 0$ such that

$$\mu[\, T^* f > \lambda \,] \le c(\lambda) \qquad (\lambda > 0, f \in E, \|f\| \le 1).$$

The following result shows that an abstract maximal inequality is exactly what we need.

Proposition 11.9 (Banach's Principle). *Let $\mathrm{X} = (X, \Sigma, \mu)$ be a measure space, $1 \le p < \infty$, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of bounded linear operators on $E = \mathrm{L}^p(\mathrm{X})$. If the associated maximal operator T^* satisfies a maximal inequality, then the set*

$$F := \big\{ f \in E : (T_n f)_{n\in\mathbb{N}} \text{ is a.e.-convergent} \big\}$$

is a closed subspace of E.

Proof. Since the operators T_n are linear, F is a subspace of E. To see that it is closed, let $f \in E$ and $g \in F$. For any natural numbers k, l we have

$$|T_k f - T_l f| \le |T_k(f-g)| + |T_k g - T_l g| + |T_l(g-f)| \le 2T^*(f-g) + |T_k g - T_l g|\,.$$

Taking the limsup in L^0 with $k, l \to \infty$ (see (11.1)) one obtains

$$h := \limsup_{k,l\to\infty} |T_k f - T_l f| \le 2T^*(f-g) + \limsup_{k,l\to\infty} |T_k g - T_l g| = 2T^*(f-g)$$

since $g \in F$. For $\lambda > 0$ we thus have $[\, h > 2\lambda \,] \subseteq [\, T^*(f-g) > \lambda \,]$ and hence

$$\mu[h > 2\lambda] \le \mu[T^*(f-g) > \lambda] \le c\left(\frac{\lambda}{\|f-g\|}\right).$$

If $f \in \overline{F}$ we can make $\|f-g\|$ arbitrarily small, and since $\lim_{t\to\infty} c(t) = 0$, we obtain $\mu[h > 2\lambda] = 0$. Since $\lambda > 0$ was arbitrary, it follows that $h = 0$. This shows that $f \in F$, whence F is closed. □

Maximal Inequalities for Dunford–Schwartz Operators

In the following $\mathrm{X} = (X, \Sigma, \mu)$ denotes a general measure space, and T denotes a positive Dunford–Schwartz operator. By Theorem 8.23, T is contractive for the p-norm on $\mathrm{L}^1 \cap \mathrm{L}^p$ for each $1 \le p \le \infty$, and by a standard approximation T extends in a coherent way to a positive contraction on each space $\mathrm{L}^p(\mathrm{X})$ for $1 \le p < \infty$.

Since the measure is not necessarily finite, it is unclear, however, whether T extends to L^∞. What we will use are the following simple observations.

Lemma 11.10. *Let $1 \le p < \infty$ and $0 \le f \in \mathrm{L}^p(\mathrm{X};\mathbb{R})$ and $\lambda > 0$. Then the following assertions hold.*

a) $\mu[f > \lambda] \le \lambda^{-p}\|f\|_p^p < \infty$.

b) $(f-\lambda)^+ \in \mathrm{L}^1 \cap \mathrm{L}^p$.

c) $Tf - \lambda \le T(f-\lambda)^+$.

Proof. a) Let $A := [f > \lambda]$. Then $\lambda^p \mathbf{1}_A \le f^p \mathbf{1}_A$, and integrating proves the claim. For b) use the same set A to write

$$(f-\lambda)^+ = (f-\lambda)\mathbf{1}_A = f\mathbf{1}_A - \lambda\mathbf{1}_A$$

Since $\mu(A) < \infty$, the claim follows. Finally note that $|f - (f-\lambda)^+| \le \lambda$, which is easily checked by distinguishing what happens on the sets A and A^{c}. Since T is a Dunford–Schwartz operator we obtain

$$Tf - T(f-\lambda)^+ \le \left|T(f-(f-\lambda)^+)\right| \le \lambda,$$

and c) is proved. □

We now turn to the so-called "maximal ergodic theorem". For $0 \le f \in \mathrm{L}^p$ (with $1 \le p < \infty$) and $\lambda > 0$ we write

$$S_k f := \sum_{j=0}^{k-1} T^j f, \qquad \mathrm{A}_n^* f = \max_{1\le k\le n} \mathrm{A}_k f, \qquad M_n^\lambda f := \max_{1\le k\le n} (S_k f - k\lambda).$$

Then

$$[\mathrm{A}_n^* f > \lambda] = \left[M_n^\lambda f > 0\right] \subseteq \bigcup_{k=1}^n [S_k f > k\lambda]$$

has finite measure and $(M_n^\lambda f)^+ \in \mathrm{L}^1 \cap \mathrm{L}^p$, by Lemma 11.10.a) and b) above.

Theorem 11.11 (Maximal Ergodic Theorem). *Let T be a positive Dunford–Schwartz operator on some space $\mathrm{L}^1(\mathrm{X})$, and let $p \in [1,\infty)$ and $0 \le f \in \mathrm{L}^p(\mathrm{X})$. Then for each $\lambda > 0$*

$$\mu[\,\mathrm{A}_n^* f > \lambda\,] \le \frac{1}{\lambda} \int_{[\mathrm{A}_n^* f > \lambda]} f \,\mathrm{d}\mu.$$

Proof. Fix $2 \le k \ge n$. Then by Lemma 11.10.c)

$$\begin{aligned} S_k f - k\lambda &= f - \lambda + T S_{k-1} f - (k-1)\lambda \le f - \lambda + T(S_{k-1} f - (k-1)\lambda)^+ \\ &\le f - \lambda + T(M_n^\lambda f)^+. \end{aligned}$$

Taking the maximum we obtain $M_n^\lambda f \le f - \lambda + T(M_n^\lambda f)^+$. Now we integrate and estimate

$$\begin{aligned} \int_X (M_n^\lambda f)^+ \,\mathrm{d}\mu &= \int_{[M_n^\lambda f > 0]} M_n^\lambda f \,\mathrm{d}\mu \le \int_{[M_n^\lambda f > 0]} f - \lambda \,\mathrm{d}\mu + \int_X T(M_n^\lambda f)^+ \,\mathrm{d}\mu \\ &\le \int_{[M_n^\lambda f > 0]} f - \lambda \,\mathrm{d}\mu + \int_X (M_n^\lambda f)^+ \,\mathrm{d}\mu \end{aligned}$$

by the L^1-contractivity of T. It follows that

$$\lambda\,\mu[\,\mathrm{A}_n^* f > \lambda\,] = \lambda\,\mu[M_n^\lambda f > 0] \le \int_{[M_n^\lambda f > 0]} f \,\mathrm{d}\mu = \int_{[\mathrm{A}_n^* f > \lambda]} f \,\mathrm{d}\mu,$$

which concludes the proof. □

Corollary 11.12 (Maximal Inequality). *Let T be a positive Dunford–Schwartz operator on some $\mathrm{L}^1(\mathrm{X})$, and let $p \in [1,\infty)$ and $f \in \mathrm{L}^p(\mathrm{X})$. Then*

$$\mu[\,\mathrm{A}^* f > \lambda\,] \le \lambda^{-p} \|f\|_p^p \qquad (\lambda > 0).$$

Proof. By the Maximal Ergodic Theorem 11.11 and Hölder's inequality

$$\mu[\,\mathrm{A}_n^* |f| > \lambda\,] \le \frac{1}{\lambda} \int_{[\mathrm{A}_n^* |f| > \lambda]} f \,\mathrm{d}\mu \le \frac{1}{\lambda} \|f\|_p \,\mu[\,\mathrm{A}_n^* |f| > \lambda\,]^{1/p'},$$

which leads to

$$\mu[\,\mathrm{A}_n^* |f| > \lambda\,]^{1/p} \le \frac{\|f\|_p}{\lambda}.$$

Now, since T is positive, each A_n is positive and hence $\max_{1\le k\le n} |\mathrm{A}_k f| \le \mathrm{A}_n^* |f|$. It follows that

$$\mu\left[\max_{1\le k\le n} |\mathrm{A}_k f| > \lambda\right] \le \mu[\,\mathrm{A}_n^* |f| > \lambda\,] \le \lambda^{-p} \|f\|_p^p.$$

We let $n \to \infty$ and the claim is proved. □

We remark that there is a better estimate in the case $1 < p < \infty$, see Exercise 7. However, we shall not need that in the following, where we turn to the proof of Theorem 11.4.

Proof of the Pointwise Ergodic Theorem

Let, as before, T be a positive Dunford–Schwartz operator on some $\mathrm{L}^1(\mathrm{X})$. If the measure space is finite we saw in Lemma 11.7 that $\mathrm{A}_n f$ converges almost everywhere for all f from the dense subspace $F = \mathrm{fix}(T) \oplus (\mathrm{I} - T)\mathrm{L}^\infty$, so by virtue of Banach's principle and the Maximal Ergodic Theorem, also for each $f \in \mathrm{L}^1(\mathrm{X})$.

In the general case, consider the assertion

$$\lim_{n\to\infty} \mathrm{A}_n f \quad \textit{exists pointwise almost everywhere.} \tag{11.2}$$

We shall establish (11.2) for larger and larger classes of functions f by virtue of Banach's principle. Note that T is coherently defined simultaneously on all spaces L^p with $1 \le p < \infty$. If we want to explicitly consider its version on L^p we shall write T_p; note again that T_p is a contraction.

First of all, it is clear that (11.2) holds if $Tf = f$. It also holds if $f \in (\mathrm{I} - T)(\mathrm{L}^1 \cap \mathrm{L}^\infty)$, because then $\mathrm{A}_n f \to 0$ uniformly by the L^∞-contractivity of T (cf. Lemma 11.7). Next, let $1 < p < \infty$. We know by the Mean Ergodic Theorem that

$$\mathrm{L}^p = \mathrm{fix}\, T_p + \overline{\mathrm{ran}}(\mathrm{I} - T_p).$$

Since a maximal inequality holds for L^p (Corollary 11.12) and $\mathrm{L}^1 \cap \mathrm{L}^\infty$ is dense in L^p, (11.2) holds for all $f \in \mathrm{L}^p$. Finally, since a maximal inequality holds for L^1 (Corollary 11.12 again) and $\mathrm{L}^2 \cap \mathrm{L}^1$ is dense in L^1, (11.2) holds for all $f \in \mathrm{L}^1$. □

Remark 11.13. The Maximal Ergodic Theorem and a related result, called "Hopf's Lemma" (Exercise 8) are from [Hopf (1954)] generalising results from [Yosida and Kakutani (1939)]. A slightly weaker form was already obtained by [Wiener (1939)]. Our proof is due to [Garsia (1965)].

The role of maximal inequalities for almost everywhere convergence results is known at least since [Kolmogoroff (1925)] and is demonstrated impressively in [Stein (1993)]. Employing a Baire category argument one can show that an abstract maximal inequality is indeed *necessary* for pointwise convergence results of quite a general type [Krengel (1985), Ch. 1, Theorem 7.2]; a thorough study of this connection has been carried out in [Stein (1961a)].

Finally, we recommend [Garsia (1970)] for more results on almost everywhere convergence.

11.3 Applications

Weyl's Theorem Revisited

Let $\alpha \in [0,1] \setminus \mathbb{Q}$. In Chapter 10 we proved *Weyl's theorem* stating that the sequence $(n\alpha \pmod 1)_{n\in\mathbb{N}}$ is equidistributed in $[0,1)$. The mean ergodicity of the involved Koopman operator with respect to the sup-norm (see Proposition 10.19) implies the following stronger statement.

Corollary 11.14. *If $\alpha \in [0,1] \setminus \mathbb{Q}$, then for every interval $B = [a,b] \subseteq [0,1]$ we have*

$$\frac{1}{n} \operatorname{card}\big\{ j \in [0,n) \cap \mathbb{N}_0 : x + j\alpha \pmod 1 \in B \big\} \to b - a$$

uniformly for $x \in [0,1]$.

The Pointwise Ergodic Theorem accounts for the analogous statement allowing for general Borel set B in place of just intervals. However, one has to pay the price of loosing convergence *everywhere*. Denoting by λ the Lebesgue measure on $\mathbb{R}$ we obtain the following result.

Corollary 11.15. *If $\alpha \in [0,1] \setminus \mathbb{Q}$, then for every Borel set $B \subseteq [0,1]$ we have*

$$\frac{1}{n} \operatorname{card}\big\{ j \in [0,n) \cap \mathbb{N}_0 : x + j\alpha \pmod 1 \in B \big\} \to \lambda(B)$$

for almost every $x \in [0,1]$.

Proof. This is Exercise 3. □

Borel's Theorem on (Simply) Normal Numbers

A number $x \in [0,1]$ is called **simply normal** (in base 10) if in its decimal expansion

$$x = 0, x_1 x_2 x_3 \ldots \qquad x_j \in \{0,\ldots,9\}, \quad j \in \mathbb{N},$$

each digit appears asymptotically with frequency $1/10$. The following goes back to [Borel (1909)].

Theorem 11.16 (Borel). *Almost every number $x \in [0,1]$ is simply normal.*

Proof. First of all note that the set of numbers with nonunique decimal expansion is countable. Let $\varphi(x) =: 10x \pmod 1$ for $x \in [0,1)$. As in Example 12.3, the MDS $([0,1], \lambda; \varphi)$ is isomorphic to the Bernoulli shift (Section 5.1.5)

$$B(1/10,\ldots,1/10) = (\mathscr{W}_{10}^{+}, \Sigma, \mu; \tau).$$

The isomorphism is induced by the point isomorphism (modulo null sets)

$$\mathscr{W}_{10}^{+} \to [0,1], \qquad (x_1, x_2, \dots) \mapsto 0, x_1 x_2 \dots .$$

(For the more details on isomorphism we refer to Section 12.1 below.) Since the Bernoulli shift is ergodic (Proposition 6.16), the MDS $([0,1], \Lambda, \lambda; \varphi)$, Λ the Lebesgue σ-algebra, is ergodic as well. Fix a digit $k \in \{0, \dots, 9\}$ and consider

$$A := \big\{x \in [0,1) : x_1 = k\big\} = [k/10, (k+1)/10),$$

so $\lambda(A) = 1/10$. Then

$$\frac{\operatorname{card}\{1 \le j \le n : x_j = k\}}{n} = \frac{1}{n} \sum_{j=0}^{n-1} \mathbf{1}_A(\varphi^j(x)) \to \int_{[0,1]} \mathbf{1}_A \, \mathrm{d}\lambda = \lambda(A) = 1/10$$

for almost every $x \in [0,1]$ by Corollary 11.2. □

We refer to Exercise 6 for the definition and analogous property of *normal* numbers.

The Strong Law of Large Numbers

In Chapter 10 we briefly pointed at a connection between the (Mean) Ergodic Theorem and the (weak) law of large numbers. In this section we shall make this more precise and prove the following classical result, going back to [Kolmogoroff (1933)]. We freely use the terminology common in probability theory, cf. [Billingsley (1979)].

Theorem 11.17 (Kolmogorov). *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, and let $(X_n)_{n\in\mathbb{N}} \subseteq \mathrm{L}^1(\Omega, \mathcal{F}, \mathrm{P})$ be a sequence of independent and identically distributed real random variables. Then*

$$\lim_{n\to\infty} \frac{1}{n}(X_1 + \dots + X_n) \;=\; \mathbb{E}(X_1) \qquad \mathrm{P}\text{-}\textit{almost surely.}$$

Proof. Since the X_j are identically distributed, $\nu := \mathrm{P}_{X_j}$ (the distribution of X_j) is a Borel probability measure on $\mathbb{R}$, independent of j, and

$$\mathbb{E}(X_1) = \int_{\mathbb{R}} t \, \mathrm{d}\nu(t)$$

is the common expectation. Define the product space

$$\mathrm{X} := (X, \Sigma, \mu) := \Big(\mathbb{R}^{\mathbb{N}}, \bigotimes\nolimits_{\mathbb{N}} \mathrm{Bo}(\mathbb{R}), \bigotimes\nolimits_{\mathbb{N}} \nu\Big).$$

As mentioned in Chapter 5, the left shift τ is a measurable transformation of (X, Σ) and μ is τ-invariant. The MDS $(\mathrm{X}; \tau)$ is ergodic, what can be shown in exactly the same way as it was done for the Bernoulli shift (Proposition 6.16).

For $n \in \mathbb{N}$ let $Y_n : X \to \mathbb{R}$ be the n^{th} projection and write $g := Y_1$. Then $Y_{j+1} = g \circ \tau^j$ for every $j \ge 0$, hence Corollary 11.2 yields that

$$\lim_{n\to\infty}\frac{1}{n}(Y_1+\cdots+Y_n)=\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}(g\circ\tau^j)=\int_X g\,\mathrm{d}\mu$$

pointwise μ-almost everywhere. Note that $g_*\mu=\nu$ and hence

$$\int_X g\,\mathrm{d}\mu=\int_{\mathbb{R}} t\,\mathrm{d}\nu(t).$$

It remains to show that the $(Y_n)_{n\in\mathbb{N}}$ are in a certain sense "the same" as the originally given $(X_n)_{n\in\mathbb{N}}$. This is done by devising an injective lattice homomorphism

$$\Phi:\mathrm{L}^0(\mathrm{X};\overline{\mathbb{R}})\to\mathrm{L}^0(\Omega,\mathcal{F},\mathrm{P};\overline{\mathbb{R}})$$

which carries Y_n to $\Phi(Y_n)=X_n$ for every $n\in\mathbb{N}$. Define

$$\varphi:\Omega\to X,\quad \varphi(\omega):=(X_n(\omega))_{n\in\mathbb{N}}\qquad(\omega\in\Omega).$$

Since $(X_n)_{n\in\mathbb{N}}$ is an independent sequence, the push-forward measure satisfies $\varphi_*\mathrm{P}=\mu$. Let $\Phi=T_\varphi:f\mapsto f\circ\varphi$ be the Koopman operator induced by φ mapping $\mathrm{L}^0(\mathrm{X};\overline{\mathbb{R}})$ to $\mathrm{L}^0(\Omega,\mathcal{F},\mathrm{P};\overline{\mathbb{R}})$. The operator Φ is well-defined since φ is measure-preserving.

By construction, $\Phi Y_n=Y_n\circ\varphi=X_n$ for each $n\in\mathbb{N}$. Moreover, Φ is clearly a homomorphism of lattices (see Chapter 7) satisfying

$$\sup_{n\in\mathbb{N}}\Phi\big(f_n\big)=\Phi\big(\sup_{n\in\mathbb{N}}f_n\big)\quad\text{and}\quad\inf_{n\in\mathbb{N}}\Phi\big(f_n\big)=\Phi\big(\inf_{n\in\mathbb{N}}f_n\big)$$

for every sequence $(f_n)_{n\in\mathbb{N}}\subseteq\mathrm{L}^0(\mathrm{X};\overline{\mathbb{R}})$. Since the almost everywhere convergence of a sequence can be described in purely lattice theoretic terms involving only countable suprema and infima (cf. also (11.1)), one has

$$\lim_{n\to\infty}f_n=f\quad\mu\text{-a.e.}\quad\Rightarrow\quad\lim_{n\to\infty}\Phi(f_n)=\Phi(f)\quad\text{P-almost surely.}$$

This, for $f_n:=(1/n)(Y_1+\cdots+Y_n)$ and $f:=\mathbb{E}(X_1)\mathbf{1}$, concludes the proof. □

Remark 11.18. By virtue of the same product construction one can show that the Mean Ergodic Theorem implies a general *weak* law of large numbers.

Final Remark: Birkhoff vs. von Neumann

From the point of view of statistical mechanics, Birkhoff's theorem seems to outrun von Neumann's. By virtue of the Dominated Convergence Theorem and the denseness of L^∞ in L^2, the latter is even a corollary of the former (cf. the presentation in [Walters (1982), Corollary 1.14.1]). Reed and Simon take a moderate viewpoint when they write in [Reed and Simon (1972), p.60] (annotation in square brackets by the authors).

> This [i.e., the pointwise ergodic] theorem is closer to what one wants to justify [in?] statistical mechanics than the von Neumann theorem, and it is fashionable to say that the von Neumann theorem is unsuitable for statistical mechanics. We feel that this is an exaggeration. If we had only the von Neumann theorem we could probably live with it quite well. Typically, initial conditions are not precisely measurable anyway, so that one could well associate initial states with measures $f\,\mathrm{d}\mu$ where $\int f\,\mathrm{d}\mu = 1$, in which case the von Neumann theorem suffices. However, the Birkhoff theorem does hold and is clearly a result that we are happier to use in justifying the statement that phase-space averages and time averages are equal.

However, von Neumann's theorem inspired the operator theoretic concept of mean ergodicity and an enormous amount of research in the field of asymptotics of discrete (and continuous) operator semigroups with tantamount applications to various other fields. Certainly it would be too much to say that Birkhoff's theorem is overrated, but von Neumann's theorem should not be underestimated either.

Exercises

1. Let K be a compact topological space, let T be a Markov operator on $\mathrm{C}(K)$ (see Exercise 2), and let $\mu \in \mathrm{M}^1(K)$ be such that $T'\mu \le \mu$. Show that T extends uniquely to a positive Dunford–Schwartz operator on $\mathrm{L}^1(K,\mu)$.

2. Let X be a measure space, $1 \le p < \infty$, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of bounded linear operators on $E = \mathrm{L}^p(\mathrm{X})$. Moreover, let T be a bounded linear operator on E. If the associated maximal operator T^* satisfies a maximal inequality, then the set

$$C := \big\{ f \in E \,:\, T_n f \to T f \text{ almost everywhere} \big\}$$

is a closed subspace of E.

3. Prove Corollaries 11.14 and 11.15.

4. Let $\alpha \in [0,1)$ be an irrational number and consider the shift $\varphi_\alpha : x \mapsto x + \alpha \pmod 1$ on $[0,1]$. Show that for the point spectrum of the Koopman operator $T = T_{\varphi_\alpha}$ on $\mathrm{L}^2[0,1]$ we have $\sigma_p(T) = \{\mathrm{e}^{2\pi \mathrm{i} m\alpha} : m \in \mathbb{Z}\}$. (Hint: Use a Fourier expansion as in the proof of Proposition 7.15.)

5. Let $(K;\varphi)$ be a TDS with Koopman operator $T = T_\varphi$ and associated Cesàro averages $\mathrm{A}_n = \mathrm{A}_n[T]$, $n \in \mathbb{N}$. Let μ be a φ-invariant probability measure on K. A point $x \in K$ is called **generic** for μ if

$$\lim_{n\to\infty} (\mathrm{A}_n f)(x) = \int_K f\,\mathrm{d}\mu \tag{11.3}$$

for *all* $f \in \mathrm{C}(K)$.

a) Show that $x \in K$ is generic for μ if (11.3) holds for each f from a dense subset of $\mathrm{C}(K)$.

b) Show that in the case that $\mathrm{C}(K)$ is separable and μ is ergodic, μ-almost every $x \in K$ is generic for μ. (Hint: Apply Corollary 11.2 to every f from a countable dense set $D \subseteq \mathrm{C}(K)$.)

6. A number $x \in [0,1]$ is called **normal** (in base 10) if every finite combination (of length k) of the digits $\{0,1,\dots,9\}$ appears in the decimal expansion of x with asymptotic frequency 10^{-k}. Prove that almost all numbers in $[0,1]$ are normal.

7 (Dominated Ergodic Theorem). Let $\mathrm{X} = (X,\Sigma,\mu)$ be a measure space and let $f \in \mathrm{L}^0_+(\mathrm{X})$. Show that

$$\int_X f \,\mathrm{d}\mu = \int_0^\infty \mu[f > t] \,\mathrm{d}t$$

and derive from this that $\|f\|_p^p = \int_0^\infty p t^{p-1} \mu[|f| > t] \,\mathrm{d}t$ for all $f \in \mathrm{L}^0(\mathrm{X})$ for $1 \leq p < \infty$.

Now let T be a Dunford–Schwartz operator on $\mathrm{L}^1(\mathrm{X})$ with its Cesàro averages $(\mathrm{A}_n)_{n\in\mathbb{N}}$ and the associated maximal operator A^*. Use the Maximal Ergodic Theorem to show that

$$\|\mathrm{A}^* f\|_p \leq \frac{p}{p-1} \|f\|_p \qquad (1 < p < \infty, \quad f \in \mathrm{L}^p(\mathrm{X})).$$

8 (Hopf's Lemma). For a positive contraction T on $\mathrm{L}^1(\mathrm{X})$ define

$$S_n f := \sum_{j=0}^{n-1} T^j f \quad \text{and} \quad M_n f := \max_{1\leq k\leq n} S_k f \qquad (f \in \mathrm{L}^1(\mathrm{X};\mathbb{R}))$$

for $n \in \mathbb{N}$. Prove that then

$$\int_{[M_n f \geq 0]} f \,\mathrm{d}\mu \geq 0$$

for every $f \in \mathrm{L}^1(\mathrm{X};\mathbb{R})$.

(Hint: Replace $\lambda = 0$ in the proof of the Maximal Ergodic Theorem 11.11 and realise that only the L^1-contractivity is needed.)

Chapter 12
Isomorphisms and Topological Models

In Chapter 10 we showed how a TDS $(K;\varphi)$ gives rise to an MDS by choosing a φ-invariant measure on K. The existence of such a measure is guaranteed by the Krylov–Bogoljubov theorem. In general there are many invariant measures, but

We also investigated how minimality of the TDS is reflected in properties of the constructed MDS. It is our aim now to go in the other direction: To a given MDS construct some TDS (sometimes called **topological model**) and an invariant measure so that the resulting MDS is isomorphic to the original one. By doing this, methods from TDS theory will be available, an we may gain further insight to MDSs. Thus, by switching back and forth between the measure theoretic and the topological situation, we can deepen our understanding of dynamical systems. In particular, this procedure will be carried out in Chapter 17.

Let us first discuss the question when two MDSs can be considered as identical, i.e., isomorphic.

12.1 Point Isomorphisms and Factor Maps

Isomorphisms of topological systems were defined in Chapter 2. The corresponding definition for measure-preserving systems is more subtle and is the subject of the present section.

Definition 12.1. Let $(\mathrm{X};\varphi)$ and $(\mathrm{Y};\psi)$ be two MDSs. A **point factor map** or a **point homomorphism** is a measure-preserving map $\theta : X \to Y$ such that

$$\psi \circ \theta = \theta \circ \varphi \quad \mu\text{-almost everywhere.}$$

We shall indicate this by writing

$$\theta : (\mathrm{X};\varphi) \to (\mathrm{Y};\psi).$$

A **point isomorphism** is a factor map which is essentially invertible (see Definition 6.3). The two MDSs are called **point isomorphic** if there is a point isomorphism between them.

In other words, a measure-preserving map $\theta : X \to Y$ is a factor map if the diagram

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

commutes almost everywhere.

Let $\theta : X \to Y$ be a point isomorphism. Recall that by Lemma 6.4 an essential inverse η of θ is uniquely determined up to equality almost everywhere and it is also measure-preserving. Moreover, since $\theta \circ \varphi = \psi \circ \theta$ almost everywhere and η is measure-preserving, it follows from Exercise 5.2 that

$$\varphi \circ \eta = \eta \circ \theta \circ \varphi \circ \eta = \eta \circ \psi \circ \theta \circ \eta = \eta \circ \psi$$

almost everywhere. To sum up, if θ is a point isomorphism with essential inverse η, then η is a point isomorphism with essential inverse θ.

Remarks 12.2. 1) The composition of point factor maps/isomorphisms of MDSs is again a point isomorphism/factor map, see Exercise 1. In particular, it follows that point isomorphy is an equivalence on the class of MDSs.

2) Sometimes an alternative characterisation of point isomorphic MDSs is used in the literature, e.g., in [Einsiedler and Ward (2011), Definition 2.7]. See Exercise 2.

Now let us give some examples of point isomorphic systems.

Example 12.3 (Doubling Map $\cong$ Bernoulli Shift). *The doubling map*

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

on $[0,1]$ *is point isomorphic to the one-sided Bernoulli shift* $B(1/2, 1/2)$.

Proof. We define

$$\Psi : \{0,1\}^{\mathbb{N}} \to [0,1], \qquad \Psi(x) := \sum_{j=1}^{\infty} \frac{x_j}{2^j}.$$

Then Ψ is a pointwise limit of measurable maps, hence measurable. On the other hand, let

$$\Phi : [0,1] \to \{0,1\}^{\mathbb{N}}, \qquad [\Phi(a)]_n := \lfloor 2^n a \rfloor \pmod 2 \in \{0,1\}.$$

In order to see that Φ is measurable and measure-preserving, we fix $a_1,\dots,a_n \in \{0,1\}$ and note that

$$\Phi^{-1}[x_j = a_j : 1 \le j \le n] = \sum_{j=1}^{n} \frac{a_j}{2^j} + [0, 2^{-n}].$$

The claim follows since the cylinder sets form a generator of the σ-algebra on $\{0,1\}^{\mathbb{N}}$. Next, we note that $\Psi \circ \Phi = \mathrm{id}$ except on the null set $\{1\}$, since the map Φ just produces a dyadic expansion of a number in $[0,1)$. Conversely, $\Phi \circ \Psi = \mathrm{id}$ except on the null set

$$N := \bigcup\nolimits_{n\ge 1} \bigcap\nolimits_{k\ge n} [x_k = 1] \subseteq \{0,1\}^{\mathbb{N}}$$

of all $\{0,1\}$-sequences that are eventually constant 1. Finally, it is an easy computation to show that $\tau \circ \Phi = \Phi \circ \varphi$ everywhere on $[0,1]$. □

Example 12.4 (Tent Map ≅ Bernoulli Shift). *The tent map*

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

on $[0,1]$ *is point isomorphic to the one-sided Bernoulli shift* $B(1/2, 1/2)$.

Proof. We define $\Phi = (\Phi_n)_{n\ge 0} : [0,1] \to \{0,1\}^{\mathbb{N}_0}$ via

$$\Phi_n(x) := \begin{cases} 0, & 0 \le \varphi^n(x) < \frac{1}{2}, \\ 1 & \frac{1}{2} \le \varphi^n(x) \le 1. \end{cases}$$

Then it is clear that Φ is measurable. By induction on $n \in \mathbb{N}_0$ one proves that the graph of φ^n on a dyadic interval of the form $[(j-1)2^{-n}, j2^{-n}]$ is either a line of slope 2^n starting in 0 and rising to 1 (if j is odd), or a line of slope -2^n starting at 1 and falling down to 0 (if j is even). Applying φ once more, we see that

$$\varphi^{n+1}(x) = 0 \quad \text{if and only if} \quad x = \frac{j}{2^n} \quad (j = 0,\dots,2^n).$$

Again by induction on $n \in \mathbb{N}_0$, we can prove that given $a_0,\dots,a_n \in \{0,1\}$ the set

$$\bigcap_{j=0}^{n} [\Phi_j(x) = a_j] = \Phi^{-1}\Big(\{a_0\} \times \cdots \times \{a_n\} \times \prod_{k>n} \{0,1\}\Big)$$

is a dyadic interval (closed, half-open or open) of length $2^{-(n+1)}$. Since the algebra of cylinder sets is generating, it follows that Φ is measure-preserving. Moreover, if τ denotes the shift on $\{0,1\}^{\mathbb{N}_0}$, then clearly $\tau \circ \Phi = \Phi \circ \varphi$ everywhere, whence Φ is a point homomorphism.

An essential inverse of Φ is constructed as follows. Define

$$A_n^0 := \left[0 \le \varphi^n \le \tfrac{1}{2}\right], \quad A_n^1 := \left[\tfrac{1}{2} \le \varphi^n \le 1\right] \qquad (n \in \mathbb{N}_0).$$

Then there is a one-to-one correspondence between dyadic intervals of length $2^{-(n+1)}$ and finite sequences $(a_0, \ldots, a_n) \in \{0,1\}^{n+1}$ given by

$$[(k-1)2^{-(n+1)}, k2^{-(n+1)}] = \bigcap_{j=0}^{n} A_j^{a_j}. \tag{12.1}$$

Let $\Psi : \{0,1\}^{\mathbb{N}_0} \to [0,1]$ be defined by

$$\{\Psi(a)\} = \bigcap_{j=0}^{\infty} A_j^{a_j}.$$

If $x \in [0,1]$, then $x \in A_n^{\Phi_n(x)}$ by definition of Φ, and hence $\Psi(\Phi(x)) = x$. This shows that $\Psi \circ \Phi = \mathrm{id}$.

In order to show that Ψ is measurable, suppose first that a, b are $0-1$-sequences such that $a \neq b$ but $\Psi(a) = x = \Psi(b)$. Then there is a minimal $n \ge 0$ such that $b_n \neq a_n$. Since $x \in A_n^{a_n} \cap A_n^{b_n}$, it follows that $\varphi^n(x) = 1/2$. Hence x is a dyadic rational of the form $j2^{-(n+1)}$ with j odd. Moreover, $\varphi^k(x) = 0$ for all $k \ge n+2$, and this forces $a_k = b_k = 0$ for all $k \ge n+2$. It follows that a and b both come from the countable set N of sequences that are eventually constant 0.

Now, given a dyadic interval with associated $(a_0, \ldots, a_n)$ as in (12.1), we have that

$$\begin{aligned}\{a_0\} \times \ldots \{a_n\} \times \prod_{k>n} \{0,1\} &\subseteq \Psi^{-1}\left(\bigcap_{j=0}^{n} A_j^{a_j}\right) \\ &\subseteq N \cup \Big(\{a_0\} \times \ldots \{a_n\} \times \prod_{k>n} \{0,1\}\Big).\end{aligned}$$

Since N is countable, every subset of N is measurable, and so $\Psi^{-1}\big(\bigcap_{j=0}^{n} A_j^{a_j}\big)$ is measurable. And since dyadic intervals form a $\cap$-stable generator of the Borel σ-algebra of $[0,1]$, Ψ is measurable.

Finally, suppose that $b := \Phi(\Psi(a)) \neq a$ for some $a \in \{0,1\}^{\mathbb{N}_0}$. Then, since $\Psi \circ \Phi = \mathrm{id}$ everywhere, we have $\Psi(b) = \Psi(a)$ and hence $a \in N$ as seen above. This means that $[\Phi \circ \Psi \neq \mathrm{id}] \subseteq N$ is a null set, whence $\Phi \circ \Psi = \mathrm{id}$ almost everywhere. □

Example 12.5. *The baker's transformation*

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

on $X = [0,1] \times [0,1]$ is isomorphic to the two-sided Bernoulli $B(1/2, 1/2)$-shift.
The proof is left as Exercise 3.

We indicated in Chapter 6 that for the study of an MDS $(\mathrm{X};\varphi)$, the relevant feature is the action of φ on the measure algebra and not on the set X itself. This motivates a notion of "isomorphism" that is based on the measure algebra, too.

12.2 Algebra Isomorphisms of Measure Preserving Systems

Recall the notions of a *lattice* and a *lattice homomorphism* from Section 7.1. A lattice $(V,\leq)$ is called a **distributive lattice** ' if the **distributive laws**

$$x\wedge(y\vee z)=(x\wedge y)\vee(x\wedge z),\qquad x\vee(y\wedge z)=(x\vee y)\wedge(x\vee z)$$

hold for all $x,y,z\in V$. If V has a greatest element $\top$ and least element $\bot$, then a **complement** of $x\in V$ is an element $y\in V$ such that

$$x\vee y=\top\quad\text{and}\quad x\wedge y=\bot\,.$$

A **Boolean algebra** is a distributive lattice $(V,\leq)$ with $\top$ and $\bot$ such that every element has a complement. Such a complement is then unique, see [Birkhoff (1948), Thm. X.1], and is usually denoted by x^c. An **(abstract) measure algebra** is a complete Boolean algebra V together with a map $\mu:V\to[0,1]$ such that

1) $\mu(\top)=1$,
2) $\mu(x)=0$ if and only if $x=\bot$,
3) μ is σ-additive in the sense that

$$(x_n)_{n\in\mathbb{N}}\subseteq V,\quad x_n\wedge x_m=\bot\ (n\neq m)\quad\Rightarrow\quad\mu\Big(\bigvee\nolimits_{n\in\mathbb{N}}x_n\Big)=\sum\nolimits_{n\in\mathbb{N}}\mu(x_n).$$

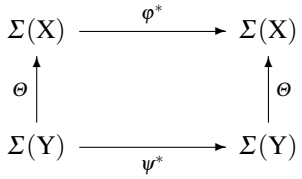
A lattice homomorphism $\Theta:V\to W$ of (abstract) measure algebras $(V,\mu),(W,\nu)$ is called a **(measure algebra) homomorphism** if $\nu(\Theta(x))=\mu(x)$ for all $x\in V$.

By the results of Chapter 6.1, the measure algebra $\Sigma(\mathrm{X})$ associated with a measure space X is an abstract measure algebra (see also Example 7.1.4). Moreover, the map φ^* acts as a measure algebra homomorphism on $\Sigma(\mathrm{X})$. In this way, $(\Sigma(\mathrm{X}),\mu;\varphi^*)$ can be viewed as a "measure algebra dynamical system" and one can form the corresponding notion of a homomorphism of such systems. This leads to the following definition.

Definition 12.6. An (algebra) **homomorphism** of two MDSs $(\mathrm{X};\varphi)$ and $(\mathrm{Y};\psi)$ is a homomorphism

$$\Theta:(\Sigma(\mathrm{Y}),\mu_{\mathrm{Y}})\to(\Sigma(\mathrm{X}),\mu_{\mathrm{X}})$$

of the corresponding measure algebras such that $\Theta\circ\psi^*=\varphi^*\circ\Theta$, i.e., the diagram



is commutative. An (algebra) **isomorphism** is a bijective homomorphism. Two MDSs are (algebra) **isomorphic** if there exists an (algebra) isomorphism between them.

Note that by Exercise 5 each measure algebra homomorphism is an embedding, isometric with respect to the canonical metric coming from the measure(s).

Remark 12.7. Let $\theta : (\mathrm{X};\varphi) \to (\mathrm{Y};\psi)$ be a point factor map. Then the associated map

$$\theta^* : (\Sigma(\mathrm{Y}), \mu_\mathrm{Y}; \psi^*) \to (\Sigma(\mathrm{X}), \mu_\mathrm{X}; \varphi^*)$$

(see Section 6.1) is a measure algebra homomorphism. If θ is a point isomorphism, then θ^* is an algebra isomorphism.

The following example shows that not every algebra isomorphism is induced by a point isomorphism as in Remark 12.7.

Example 12.8. Take $X = \{0\}$, $\Sigma = \{\emptyset, X\}$, $\mu(X) = 1$, $\varphi = \mathrm{id}$ and $(\mathrm{Y};\psi)$ with $Y = \{0,1\}$, $\Sigma' = \{\emptyset, Y\}$, $\mu_\mathrm{Y}(Y) = 1$, $\psi = \mathrm{id}$. The two MDSs are not point isomorphic, but (algebra) isomorphic.

Let us call two probability spaces X and Y **point isomorphic** or **algebra isomorphic** if $(\mathrm{X};\mathrm{id}_X)$ and $(\mathrm{Y};\mathrm{id}_Y)$ are point isomorphic or algebra isomorphic, respectively. Example 12.8 shows that these concepts may differ in general. In certain situations, however, the distinction between "algebra isomorphic" and "point isomorphic" is unnecessary. Indeed, J. von Neumann proved in [von Neumann (1932a)] that two compact metric (Borel) probability spaces are algebra isomorphic if and only if they are point isomorphic. (For a proof and further reading we refer to [Royden (1963), Sec. 14.5] or Bogachev [Bogachev (2007), Ch. 9]). From this one can deduce the following classical result.

Theorem 12.9 (Von Neumann). *Two MDSs on compact metric probability spaces are isomorphic if and only if they are point isomorphic.*

Markov Embeddings

In the remaining part of this section we shall see that (measure) algebra homomorphisms correspond in a natural way to certain operators on the associated L^1-spaces.

Definition 12.10. An operator $S : \mathrm{L}^1(\mathrm{X}) \to \mathrm{L}^1(\mathrm{Y})$ is called a **Markov operator** if

$$S \geq 0, \quad S\mathbf{1} = \mathbf{1}, \quad \text{and} \quad \int_Y Sf = \int_X f \quad \text{for all } f \in \mathrm{L}^1(\mathrm{X}).$$

If, in addition, one has

$$|Sf| = S|f| \qquad \text{for every } f \in \mathrm{L}^1(\mathrm{Y}).$$

then S is called a **Markov embedding**.

Note that S is a Markov operator if and only if $S \geq 0$, $S\mathbf{1} = \mathbf{1}$ and $S'\mathbf{1} = \mathbf{1}$. A Markov operator is a Markov embedding if it is a homomorphism of Banach lattices. It is easy to see that in this case $\|Sf\|_1 = \|f\|_1$ for each $f \in \mathrm{L}^1$, so each Markov embedding is an isometry.

Markov operators form the basic operator class in ergodic theory and shall play a central role below. We shall look at them in more detail in Chapter 13. For now we only remark that of course each Koopman operator associated with an MDS is a Markov embedding.

A Markov embedding S is called a **Markov isomorphism** or an **isometric lattice isomorphism** if it is surjective. In this case, S is bijective and S^{-1} is a Markov embedding, too. The next theorem clarifies the relation between Markov embeddings and measure algebra homomorphisms.

Theorem 12.11. *Each Markov embedding $S : \mathrm{L}^1(\mathrm{Y}) \to \mathrm{L}^1(\mathrm{X})$ induces a map $\Theta : \Sigma(\mathrm{Y}) \to \Sigma(\mathrm{X})$ via*

$$S\mathbf{1}_A = \mathbf{1}_{\Theta A} \qquad \textit{for all} \quad A \in \Sigma(\mathrm{Y}), \tag{12.2}$$

and Θ is a (measure) algebra homomorphism.

Conversely let $\Theta : \Sigma(\mathrm{Y}) \to \Sigma(\mathrm{X})$ be a measure algebra homomorphism. Then there is a unique bounded operator $S : \mathrm{L}^1(\mathrm{Y}) \to \mathrm{L}^1(\mathrm{X})$ satisfying (12.2)*, and S is a Markov embedding.*

Finally, Θ is surjective if and only if S is a Markov isomorphism.

Proof. As S is a positive operator, it maps real-valued functions to real-valued functions. Now, a moment's thought reveals that a real-valued function $f \in \mathrm{L}^1$ is a characteristic function if and only if $f \wedge (\mathbf{1} - f) = 0$. Hence S maps characteristic functions on Y to characteristic functions on X, which gives rise to a map

$$\Theta : \Sigma(\mathrm{X}) \to \Sigma(\mathrm{Y}) \quad \text{with} \quad S\mathbf{1}_A = \mathbf{1}_{\Theta A} \qquad (A \in \Sigma(\mathrm{Y})).$$

It then is easy to see that Θ is a measure algebra homomorphism (Exercise 6).

For the converse, let $\Theta : \Sigma(\mathrm{Y}) \to \Sigma(\mathrm{X})$ be a measure algebra homomorphism. For a step function f on Y we choose a representation

$$f = \sum_{j=1}^n \alpha_j \mathbf{1}_{A_j}$$

with pairwise disjoint $A_j \in \Sigma_\mathrm{Y}$, $\mu_\mathrm{Y}(A_j) > 0$, and $\alpha_j \in \mathbb{C}$. Then we define

$$Sf := \sum_{j=1}^n \alpha_j \mathbf{1}_{\Theta A_j}.$$

From the properties of Θ it follows in a standard way that S is well-defined and satisfies $S\mathbf{1} = \mathbf{1}$, $\int_X Sf = \int_Y f$, $|S| = S|f|$ and hence $\|Sf\|_1 = \|f\|_1$ all step functions f. By approximation, S extends uniquely to an isometry of the L^1-spaces. This extension is clearly a Markov embedding.

Finally, if Θ is surjective, then the set $\mathrm{Step}(X)$ of step functions is contained in the range of S, and since S is an isometry and $\mathrm{Step}(X)$ is dense in $L^1(X)$, S is surjective. Conversely, if S is surjective, it is bijective and S^{-1} is also a Markov embedding. Hence if $B \in \Sigma(X)$ then $S^{-1}\mathbf{1}_B = \mathbf{1}_A$ for some $A \in \Sigma(Y)$, and therefore $\Theta(A) = B$. Consequently, Θ is surjective. □

Remark 12.12. If $(X;\varphi)$ is a measure-preserving system, then its Koopman operator T_φ on $L^1(X)$ is a Markov embedding, and the associated map on the measure algebra is just φ^*. By Definition 6.2, the system is invertible if φ^* is bijective, which by Theorem 12.11 is the case if and only if the Koopman operator T_φ is a Markov isomorphism.

Suppose that $(X;\varphi)$ and $(Y;\psi)$ are measure-preserving systems, and Θ, S is a pair of corresponding maps

$$\Theta : (\Sigma(Y), \mu_Y) \to (\Sigma(X), \mu_X) \qquad S : L^1(Y) \to L^1(X)$$

related by $S\mathbf{1}_A = \mathbf{1}_{\Theta(A)}$ for $A \in \Sigma(Y)$ as before. Then

$$ST_\psi \mathbf{1}_A = S\mathbf{1}_{\psi^* A} = \mathbf{1}_{\Theta(\psi^* A)} \quad \text{and} \quad T_\varphi S\mathbf{1}_A = T_\varphi \mathbf{1}_{\Theta A} = \mathbf{1}_{\varphi^*(\Theta A)}.$$

Therefore, Θ is an algebra homomorphism of the two MDSs if and only if its associated operator $S : L^1(Y) \to L^1(X)$ satisfies $ST_\psi = T_\varphi S$, i.e., if the diagram

$$\begin{array}{ccc} L^1(X) & \xrightarrow{T_\varphi} & L^1(X) \\ {\scriptstyle S}\uparrow & & \uparrow{\scriptstyle S} \\ L^1(Y) & \xrightarrow{T_\psi} & L^1(Y) \end{array} \tag{12.3}$$

commutes. In yet other words: S **intertwines** the Koopman operators of the systems. In this case, we call S a **Markov embedding of the dynamical systems** and write

$$S : (L^1(Y); T_\psi) \to (L^1(X); T_\varphi)$$

The following is an immediate consequence.

Corollary 12.13. *Two measure-preserving systems $(X;\varphi)$ and $(Y;\psi)$ are algebra isomorphic if and only if there is a Markov isomorphism $S : L^1(Y) \to L^1(X)$ that satisfies $ST_\psi = T_\varphi S$.*

A Markov isomorphisms restricts to a lattice isomorphism between the corresponding L^p spaces, whence one can obtain the corresponding characterisation of isomorphism of MDSs in terms of the Koopman operators on the L^p spaces.

Remark 12.14. Corollary 12.13 should be compared with the following consequence of Theorems 4.13 and 7.18:

For two TDSs $(K;\varphi)$ and $(L;\psi)$ the following assertions are equivalent.

(i) *$(K;\varphi)$ and $(L;\psi)$ are isomorphic.*

(ii) *There is an C^*-algebra isomorphism $\Phi : \mathrm{C}(K) \to \mathrm{C}(L)$ with $T_\psi \circ \Phi = \Phi \circ T_\varphi$.*

(iii) *There is an Banach lattice isomorphism $\Phi : \mathrm{C}(K) \to \mathrm{C}(L)$ with $\Phi\mathbf{1} = \mathbf{1}$ and $T_\psi \circ \Phi = \Phi \circ T_\varphi$.*

12.3 Topological Models

A **topological MDS** is any MDS of the form $(K,\mu;\varphi)$, where $(K;\varphi)$ is a TDS and μ is an invariant (Baire) probability measure. The topological MDS $(K,\mu;\varphi)$ is called **metric** if K is endowed with a metric inducing its topology. And it is called **faithful** if $\mathrm{supp}(\mu) = K$, i.e., if the canonical map $\mathrm{C}(K) \to \mathrm{L}^\infty(K,\mu)$ is injective. The following is a consequence of Lemma 10.7.

Lemma 12.15. *If $(K,\mu;\varphi)$ is a faithful topological MDS, then $\varphi(K) = K$, i.e., $(K;\varphi)$ is a surjective TDS.*

Theorem 10.2 of Krylov and Bogoljubov tells that every TDS $(K;\varphi)$ has at least one invariant probability measure, and hence gives rise to at least one topological MDS. (By the lemma above, this topological MDS cannot be faithful if $(K;\varphi)$ is not a surjective system. But even if the TDS is surjective and uniquely ergodic, the arising MDS need not be faithful, as Exercise 9 shows.) Conversely, one may ask:

> Is every measure-preserving system (algebra) isomorphic to a topological one?

Let us define a (faithful) **topological model** of an MDS $(\mathrm{X};\varphi)$ to be any (faithful) topological MDS $(K,\mu;\psi)$ together with a Markov isomorphism

$$\Phi : (\mathrm{L}^1(K,\mu);T_\psi) \to (\mathrm{L}^1(\mathrm{X});T_\varphi)$$

of dynamical systems. We shall abbreviate this by writing

$$\Phi : (K,\mu;\psi) \to (\mathrm{X};\varphi)$$

in the following. Then the question above can be rephrased as: Does every MDS have a topological model? In the following we shall answer this question.

Suppose that $(\mathrm{X};\varphi)$ is a MDS, with Koopman operator $T = T_\varphi$ and let $A \subseteq \mathrm{L}^\infty(\mathrm{X})$ be a C^*-subalgebra. (Recall that this means that A is norm-closed and conjugation invariant, and that $\mathbf{1} \in A$.) By the Gelfand–Naimark Theorem 4.21, there is a compact space K and a (unital) C^*-algebra isomorphism $\Phi : \mathrm{C}(K) \to A$. By the Riesz representation theorem there is a unique probability measure $\mu \in \mathrm{M}^1(K)$ such that

$$\int_K f \,\mathrm{d}\mu = \int_X \Phi f \qquad (f \in \mathrm{C}(K)). \tag{12.4}$$

By construction, the measure μ has full support. By Theorem 7.18 one has in addition

$$|\Phi f| = \Phi\,|f| \qquad (f \in \mathrm{C}(K)), \tag{12.5}$$

and this yields

$$\|\Phi f\|_{\mathrm{L}^1(\mathrm{X})} = \int_X \Phi\,|f| = \int_K |f| \,\mathrm{d}\mu = \|f\|_{\mathrm{L}^1(K,\mu)}$$

for every $f \in \mathrm{C}(K)$, i.e., Φ is an L^1-isometry. Consequently, Φ extends uniquely to an isometric embedding

$$\Phi : \mathrm{L}^1(K,\mu) \to \mathrm{L}^1(\mathrm{X})$$

with range $\operatorname{ran}\Phi = \mathrm{cl}_{\mathrm{L}^1}(A)$, the L^1-closure of A. Moreover, it follows by approximation and (12.4) and (12.5) that Φ is a Markov embedding.

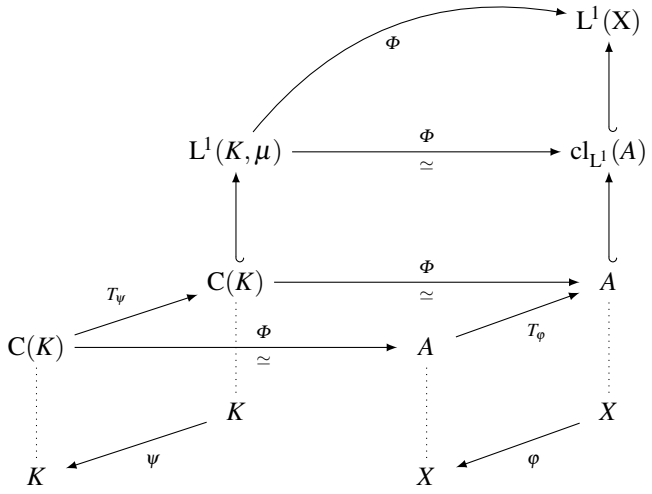
So far, we have not looked at the dynamics yet. So suppose in addition that A is T_φ-invariant, i.e., $T_\varphi(A) \subseteq A$. Then

$$\Phi^{-1} T_\varphi \Phi : \mathrm{C}(K) \to \mathrm{C}(K)$$

is an algebra homomorphism. Hence, by Theorem 4.13 there is a unique continuous map $\psi : K \to K$ such that $\Phi^{-1} T_\varphi \Phi = T_\psi$. Moreover, the measure μ is ψ-invariant, since

$$\int_K f \circ \psi \,\mathrm{d}\mu = \int_K \Phi^{-1} T_\varphi \Phi f \,\mathrm{d}\mu = \int_X T_\varphi \Phi f = \int_X \Phi f = \int_K f \,\mathrm{d}\mu$$

for every $f \in \mathrm{C}(K)$. It follows that $(K, \mu; \psi)$ is a faithful topological MDS, and that $\Phi : \mathrm{L}^1(K,\mu) \to \mathrm{L}^1(\mathrm{X})$ is a Markov embedding that intertwines the Koopman operators, i.e., a Markov embedding of the dynamical systems.



We have proved the nontrivial part of the following theorem. (The remaining part is left as Exercise 10.)

Theorem 12.16. *Let $(\mathrm{X};\varphi)$ be a measure-preserving system with Koopman operator T_φ. Then $A \subseteq \mathrm{L}^\infty(\mathrm{X})$ is a T_φ-invariant C^*-subalgebra if and only if there exists a faithful topological MDS $(K,\mu;\psi)$ and a Markov embedding $\Phi : \mathrm{L}^1(K,\mu) \to \mathrm{L}^1(\mathrm{X})$ with $T_\varphi \Phi = \Phi T_\psi$ and such that $A = \Phi(\mathrm{C}(K))$.*

Let us call an invariant subalgebra A as in Theorem 12.16 **full** if A is dense in $\mathrm{L}^1(\mathrm{X})$. In this case

$$\Phi : (\mathrm{L}^1(K,\mu);T_\psi) \to (\mathrm{L}^1(\mathrm{X});T_\varphi)$$

is a Markov isomorphism of dynamical systems, and we have constructed a (faithful) topological model of $(\mathrm{X};\varphi)$. For the choice $A := \mathrm{L}^\infty(\mathrm{X})$ we hence obtain a distinguished model of $(\mathrm{X};\varphi)$, which will be studied in more detail in Section 12.4 below. For now let us answer the question posed above.

Corollary 12.17. *Every measure-preserving system is (algebra) isomorphic to a faithful topological MDS.*

Note that in the construction above we can choose *each* full subalgebra, uniqueness of a model cannot be expected. Rather, one may construct by a diligent choice of A a model with "nice" properties. Here is a first example for this.

Theorem 12.18 (Metric Models). *An MDS $(\mathrm{X};\varphi)$ has a metric model if and only if $\mathrm{L}^1(\mathrm{X})$ is a separable Banach space.*

Proof. Let $(K,\mu;\psi)$ be a metric model for $(\mathrm{X};\varphi)$. By Theorem 4.7, $\mathrm{C}(K)$ is a separable Banach space, and as any dense subset of $\mathrm{C}(K)$ is also dense in $\mathrm{L}^1(K,\mu)$, the latter space must be separable as well.

Conversely, suppose that $\mathrm{L}^1(\mathrm{X})$ is separable, and let $M \subseteq \mathrm{L}^1(\mathrm{X})$ be a countable dense set. Since L^∞ is dense in L^1 we can approximate each member of M by a sequence in L^∞, and hence we may suppose without loss of generality that $\mathbf{1} \in M \subseteq \mathrm{L}^\infty$. By passing to $\bigcup_{n\geq 0} T_\varphi^n(M)$ we may further suppose that M is T_φ-invariant. Let $A := \mathrm{cl}_{\mathrm{L}^\infty}\mathrm{alg}(M)$ be the smallest C^*-subalgebra of L^∞ containing M. Then A is separable (Exercise 11), T_φ-invariant, and full. Applying Theorem 12.16 yields a topological model $\Phi : (K, \mu; \psi) \to (\mathrm{X}; \varphi)$ with $\Phi(\mathrm{C}(K)) = A$. Since $\Phi : \mathrm{C}(K) \to A$ is an isomorphism of C^*-algebras, $\mathrm{C}(K)$ is a separable Banach space. By Theorem 4.7, K is metrisable. □

12.4 The Stone Representation

Let $(\mathrm{X}; \varphi)$ be any given MDS. Then we can apply the construction preceding Theorem 12.16 to the algebra $A = \mathrm{L}^\infty(\mathrm{X})$, yielding a distinguished topological model $(K, \mu; \psi)$ of the whole system $(\mathrm{X}; \varphi)$. This MDS is called its **Stone representation** or its **Stone model**.

The name derives from an alternative description of the compact space K. Consider the measure algebra $V := \Sigma(\mathrm{X})$, which is a Boolean algebra (see Section 12.2). By the Stone representation theorem (see, e.g., [Birkhoff (1948), Sec. IX.9]) there exists a unique compact and *totally disconnected* space K_V such that V is isomorphic to the Boolean algebra of all clopen subsets of K_V. The compact space K_V is called the **Stone representation space** of the Boolean algebra $\Sigma(\mathrm{X})$, and one can prove that K and K_V are actually homeomorphic. By Example 7.1.4 the Boolean algebra $\Sigma(\mathrm{X})$ is complete, in which case the Stone representation theorem asserts that the space $K \simeq K_V$ is *extremally disconnected*. We shall prove this fact directly using the Banach lattice $\mathrm{C}(K)$.

Proposition 12.19. *The compact space K is extremally disconnected, i.e., the closure of every open set is open.*

Proof. We know that $\mathrm{L}^\infty(\mathrm{X})$ and $\mathrm{C}(K)$ are isomorphic Banach lattices. Since $\mathrm{L}^\infty(\mathrm{X})$ is order complete (Remark 7.11), so is $\mathrm{C}(K)$. Let $G \subseteq K$ be an open set. Consider the function

$$f := \inf\big\{g : g \in \mathrm{C}(K),\, g(x) \geq \mathbf{1}_G(x) \text{ for all } x \in K\big\},$$

where the infimum is to be understood in the order complete lattice $\mathrm{C}(K)$, hence f is continuous. If $x \notin \overline{G}$, then by Urysohn's Lemma 4.2 there is a real-valued continuous function g vanishing on a small neighbourhood of x and greater than 1 on $\overline{G}$, so $f(x) = 0$. If $x \in G$, then each g appearing in the above infimum is 1 on a fixed neighbourhood of x, so $f(x) = 1$. Since f is continuous, we have $f(x) = 1$ for all $x \in \overline{G}$. This implies $f = \mathbf{1}_{\overline{G}}$, hence $\overline{G}$ is open. □

Recall that the measure (positive functional) μ on K is induced via the isomorphism

$$\mathrm{C}(K) \cong \mathrm{L}^\infty(\mathrm{X})$$

and hence is strictly positive, i.e., has full support $\mathrm{supp}(\mu) = K$. Moreover, μ is **order continuous**, i.e., if $\mathscr{F} \subseteq \mathrm{C}(K)_+$ is a $\wedge$-stable set such that $\inf \mathscr{F} = 0$, then $\inf_{f\in\mathscr{F}} \langle f, \mu\rangle = 0$ (see Exercise 11). In the following it will be sometimes convenient to use the regular Borel extension of the (Baire) measure μ.

Lemma 12.20. *The μ-null sets in K are precisely the nowhere dense subsets of K.*

Proof. We shall use that $\mathrm{C}(K)$ is order complete, extremally disconnected and that μ induces a strictly positive order continuous linear functional on $\mathrm{C}(K)$; moreover, we shall use that a characteristic function $\mathbf{1}_U$ is continuous if and only if U is a clopen subset of K.

Let $F \subseteq K$ be a nowhere dense set. Since a set is nowhere dense if and only if its closure is nowhere dense, we may suppose without loss of generality that F is closed and has empty interior. Consider the continuous function f defined as

$$f := \inf\big\{\mathbf{1}_U : F \subseteq U, U \text{ is clopen}\big\}$$

in the (order complete) lattice $\mathrm{C}(K)$. Then trivially $f \geq 0$ and we claim that actually $f = 0$. Indeed, if $f \neq 0$ there is some $x \in K \setminus F$ with $f(x) > 0$ (since $K \setminus F$ is dense). Then the compact sets $\{x\}$ and F can be separated by disjoint open sets and by Proposition 12.19 we find a clopen set $U \supseteq F$ with $x \notin U$. It follows by definition of f that $f \leq \mathbf{1}_U$, which in turn implies that $f(x) = 0$, a contradiction.

We now can use that μ is order continuous and

$$\mu(F) \leq \inf\big\{\mu(U) : F \subseteq U, U \text{ is clopen}\big\} = \inf_{F\subseteq U \text{ clopen}} \langle \mathbf{1}_U, \mu\rangle = \langle f, \mu\rangle = 0,$$

whence F is a μ-null set.

For the converse implicaton, suppose that $A \in \mathrm{Bo}(K)$ is a μ-null set. By regularity we find a decreasing sequence $(U_n)_{n\in\mathbb{N}}$ of open sets containing A such that $\lim_{n\to\infty} \mu(U_n) = \mu(A) = 0$. By what was proved above, the nowhere dense closed sets ∂U_n have μ-measure 0, so $\mu(U_n) = \mu(\overline{U_n})$. The closed set $B := \bigcap_{n\in\mathbb{N}} \overline{U_n}$ contains A and satisfies $\mu(B) = 0$ since

$$\mu(B) \leq \mu(U_n) \to 0 \quad \text{as} \quad n \to \infty.$$

By the strict positivity of μ, B has empty interior, whence A is nowhere dense. □

We now can prove the following result, showing in particular that constructing Stone models repeatedly does not lead to something new.

Proposition 12.21. *The space is $\mathrm{L}^\infty(K,\mu)$ is canonically isomorphic to $\mathrm{C}(K)$, i.e., the canonical map $\mathrm{C}(K) \to \mathrm{L}^\infty(K,\mu)$ is bijective. In other words, every equivalence class $f \in \mathrm{L}^\infty(K,\mu)$ contains a (unique) continuous function.*

Proof. Since μ is strictly positive, the canonical map $J : \mathrm{C}(K) \to \mathrm{L}^\infty(K,\mu)$ (mapping each continuous function onto its equivalence class modulo equality μ-almost everywhere) is an isometry (Exercise 5.11).

Let now $A \in \mathrm{Bo}(K)$ be arbitrary. By regularity there are open sets $G_n \supseteq A$ so that for $H = \bigcap_n G_n$ we have $A \subseteq H$ and $\mu(H \setminus A) = 0$. Since ∂G_n is nowhere dense, by Lemma 12.20 it is of μ-measure 0, and we may suppose that G_n are clopen, whence H is closed. Also by regularity there are compact sets $K_n \subseteq A$ so that for $F = \bigcup_n K_n$ we have $F \subseteq A$ and $\mu(A \setminus F) = 0$. Since ∂K_n is nowhere dense, we obtain the open set $G = \bigcup_n K_n^\circ$ with $\mu(A \setminus G) = 0$. From $G \subseteq A \subseteq H$ and from the fact that H is closed we conclude that $\overline{G}$ is a clopen set with $\nu(\overline{G} \triangle A) = 0$. Therefore $\mathbf{1}_A$ coincides μ-almost everywhere with the continuous function $\mathbf{1}_{\overline{G}}$.

Since step functions are dense in $\mathrm{L}^\infty(K, \mu)$ and J is an isometry, J is an isomorphism. □

The next question is whether ergodicity of $(\mathrm{X}; \varphi)$ reappears in its Stone representation. We note the following result.

Proposition 12.22. *Let $(\mathrm{X}; \varphi)$ be an MDS, let $A \subseteq \mathrm{L}^\infty(\mathrm{X})$ be an invariant C^*-subalgebra, and let $(K, \mu; \psi)$ be a faithful topological MDS and*

$$\Phi : (\mathrm{L}^1(K, \mu), T_\psi) \to (\mathrm{L}^1(\mathrm{X}), T_\varphi)$$

a Markov embedding of dynamical systems with $\Phi(\mathrm{C}(K)) = A$. Then $(K, \mu; \psi)$ is strictly ergodic if and only if $A \cap \mathrm{fix}(T_\varphi) = \mathbb{C}\mathbf{1}$ and T_φ is mean ergodic on A.

Proof. By construction T_ψ is mean ergodic on $\mathrm{C}(K)$ if and only if T_φ is mean ergodic on A. The fixed spaces are related by

$$\mathrm{fix}(T_\psi \cap \mathrm{C}(K)) = \Phi^{-1}(\mathrm{fix}(T_\psi) \cap A).$$

Hence, the assertions follows from Theorem 10.6 and the fact that μ is strictly positive. □

Note that by Corollary 10.9 a strictly ergodic TDS is minimal. If we apply Proposition 12.22 to $A = \mathrm{L}^\infty(\mathrm{X})$, we obtain the following result.

Corollary 12.23. *If an MDS $(\mathrm{X}; \varphi)$ is ergodic and its Koopman operator is mean ergodic on $\mathrm{L}^\infty(\mathrm{X})$, then its Stone representation TDS is strictly ergodic (and hence minimal).*

In the next section we shall see that mean ergodicity on L^∞ is a *very strong* assumption: Essentially no interesting MDS has this property. However, by virtue of Proposition 12.22 one may try to find a strictly ergodic model of a given ergodic system $(\mathrm{X}; \varphi)$ by looking at suitable full subalgebras A of $\mathrm{L}^\infty(\mathrm{X})$. This is the topic of the following section.

12.5 Mean Ergodic Subalgebras

In Chapter 8 we showed that a Koopman operator of an MDS $(\mathrm{X}; \varphi)$ is always mean ergodic on $\mathrm{L}^p(\mathrm{X})$ for $1 \leq p < \infty$. In this section we are interested in the missing

case $p = \infty$. However, mean ergodicity cannot be expected on the whole of L^∞. In fact, the Koopman operator of an ergodic rotation on the torus is not mean ergodic on $\mathrm{L}^\infty(\mathbb{T}, \mathrm{d}z)$, see Exercise 12. The next proposition shows that this is no exception.

Proposition 12.24. *For an ergodic MDS* $(\mathrm{X}; \varphi)$ *the following assertions are equivalent.*

(i) *The Koopman operator* T_φ *is mean ergodic on* $\mathrm{L}^\infty(\mathrm{X})$.

(ii) $\mathrm{L}^\infty(\mathrm{X})$ *is finite dimensional.*

Proof. Since a finite-dimensional Banach space is reflexive, by Theorem 8.22 every power-bounded operator thereon is mean ergodic. Therefore the implication (ii)$\Rightarrow$ (i) follows.

For the converse we use the Stone representation TDS $(K; \psi)$ with its (unique) invariant measure μ. This TDS is minimal by Corollary 12.23. For an arbitrary $x \in K$, the orbit

$$\mathrm{orb}_+(x) = \big\{ \psi^n(x) : n \in \mathbb{N}_0 \big\}$$

is dense in K, so by Lemma 12.20 $\mu(\mathrm{orb}_+(x)) > 0$. Since the orbit is an at most countable set, there is $n \in \mathbb{N}_0$ such that the singleton $\{\psi^n\}$ has positive measure $\alpha := \mu\{\psi^n(x)\} > 0$. But then, by the ψ-invariance of μ,

$$\mu\{\psi^{n+k}(x)\} = \mu\Big[\, \psi^k \in \{\psi^{n+k}(x)\} \,\Big] \geq \mu\{\psi^n(x)\} = \alpha$$

for *each* $k \geq 0$. Since the measure is finite, the orbit $\mathrm{orb}_+(x)$ must be finite. But since it is dense in K, it follows that $K = \mathrm{orb}_+(x)$ is finite. Consequently, $\mathrm{L}^\infty(K, \mu)$ is finite-dimensional. □

Having seen that $T = T_\varphi$ is (in general) not mean ergodic on $\mathrm{L}^\infty(\mathrm{X})$, we look for a smaller (but still full) subalgebra A of $\mathrm{L}^\infty(\mathrm{X})$ on which mean ergodicity is guaranteed. (Let us call such a subalgebra A a **mean ergodic subalgebra**.) In the case of an ergodic rotation on the torus there are natural examples of such subalgebras, e.g. $A = \mathrm{C}(\mathbb{T})$ (Corollary 10.12) or $A = \mathrm{R}(\mathbb{T}) := \{t \mapsto f(\mathrm{e}^{2\pi \mathrm{i} t}) : f \in \mathrm{R}[0,1]\}$, the space of Riemann integrable functions on the torus (Proposition 10.19). But, of course, the question arises, whether such a choice is always possible.

In any case, each mean ergodic subalgebra A has to be contained in the closed space

$$\mathrm{fix}(T_\varphi) \oplus \overline{(\mathrm{I} - T_\varphi)\mathrm{L}^\infty(\mathrm{X})},$$

since by Theorem 8.5 this is the largest subspace of L^∞ on which the Cesàro averages of T_φ converge. Would this be an appropriate choice for an algebra? Here is another bad news.

Proposition 12.25. *For every ergodic MDS* $(\mathrm{X}; \varphi)$ *the following assertions are equivalent.*

(i) $\mathrm{lin}\{\mathbf{1}\} \oplus \overline{(\mathrm{I} - T_\varphi)\mathrm{L}^\infty}$ *is a* C^**-subalgebra of* $\mathrm{L}^\infty(\mathrm{X})$.

(ii) $\mathrm{L}^\infty(\mathrm{X})$ *is finite dimensional.*

(iii) *The Koopman operator T_φ is mean ergodic on* $\mathrm{L}^\infty(\mathrm{X})$.

Proof. By Proposition 12.24 (ii) and (iii) are equivalent. Let $A := \mathrm{lin}\{\mathbf{1}\} \oplus \overline{(\mathrm{I}-T_\varphi)\mathrm{L}^\infty}$, which is a closed *subspace* of L^∞. Since $(\mathrm{X};\varphi)$ is ergodic, one has $\mathrm{fix}\, T_\varphi = \mathrm{lin}\{\mathbf{1}\}$ by Proposition 7.14, and hence (iii) implies (i).

Conversely, suppose that (i) holds. We identify $(\mathrm{X};\varphi)$ with its Stone representation $(K,\mu;\varphi)$. Under this identification A is a closed $*$-subalgebra of $\mathrm{C}(K)$ containing $\mathbf{1}$. If we can prove that A separates the points of K, by the Stone–Weierstraß Theorem 4.4 it follows that $A = \mathrm{C}(K)$, and therefore (iii).

Suppose first that $(K;\varphi)$ has a fixed point $x \in K$. Then $f(x) = 0$ for every $f \in \mathrm{ran}(\mathrm{I}-T_\varphi)$ and hence for every $f \in \overline{\mathrm{ran}}(\mathrm{I}-T_\varphi)$ (closure in sup-norm). For $f \in \mathrm{ran}(\mathrm{I}-T_\varphi) \subseteq A$ we have $|f|^2 = \overline{f}f \in A$ and $|f|^2(x) = 0$. Since x is a fixed point, $\mathrm{A}_n|f|^2(x) = 0$ for every $n \in \mathbb{N}$. But T_φ is mean ergodic on A, so the Cesàro means $\mathrm{A}_n|f|^2$ converge in sup-norm. By Theorem 8.10, the limit is the constant function $\int_K |f|^2 \mathrm{d}\mu \cdot \mathbf{1}$, hence $\int_K |f|^2 \,\mathrm{d}\mu = 0$. From the strict positivity of μ we obtain $|f|^2 = 0$ and hence $f = 0$. This means that $(\mathrm{I}-T_\varphi)\mathrm{L}^\infty = \{0\}$, i.e., $T_\varphi = \mathrm{I}$. By ergodicity, $\dim \mathrm{L}^1 = 1$, i.e., (ii).

Now suppose that $(K;\varphi)$ does not have any fixed points. We shall show that already $\mathrm{ran}(\mathrm{I}-T_\varphi)$ separates the points of K. Take two different points $x, y \in K$. We need to find a continuous function $f \in \mathrm{C}(K)$ such that

$$f(x) - f(y) \neq f(\varphi(x)) - f(\varphi(y)).$$

Since $x \neq \varphi(x)$ and $y \neq \varphi(y)$ such a function is found easily by an application of Urysohn's lemma: If $\varphi(x) = \varphi(y)$, let f be any continuous function separating x and y; and if $\varphi(x) \neq \varphi(y)$, let $f \in \mathrm{C}(K)$ be such that $f(x) = f(\varphi(y)) = 0$ and $f(y) = f(\varphi(x)) = 1$. □

After these rather negative results it becomes clear that our task consists in finding "large" subalgebras contained in $\langle\mathbf{1}\rangle \oplus \overline{(\mathrm{I}-T_\varphi)\mathrm{L}^\infty}$. This was achieved by R. I. Jewett [Jewett (1970)] (in the weak mixing case, cf. Chapter 9) and W. Krieger [Krieger (1972)].

Theorem 12.26 (Jewett–Krieger).

a) *Let* $(\mathrm{X};\varphi)$ *be an ergodic MDS with Koopman operator T_φ. Then there exists a T_φ-invariant C^*-subalgebra of* $\mathrm{L}^\infty(\mathrm{X})$*, dense in* $\mathrm{L}^1(\mathrm{X})$*, on which T_φ is mean ergodic.*

b) *Every ergodic MDS* $(\mathrm{X};\varphi)$ *with Σ_X countably generated is isomorphic to an MDS determined by a strictly ergodic TDS on a totally disconnected compact metric space.*

The proof is, unfortunately, beyond the scope of this book.

Exercises

1. Let $(X_j, \Sigma_j, \mu_j; \varphi_j)$, $j = 1,2,3$ be given MDSs, and let $\theta_1 : X_1 \to X_2$ and $\theta_2 : X_2 \to X_3$ be point factor maps/isomorphisms. Show that $\theta_2 \circ \theta_1$ is a point factor map/isomorphism, too. (Hint: Exercise 6.2.)

2. Prove the following alternative characterisation of point isomorphic MDSs.

Proposition. *For MDSs* $(\mathrm{X}; \varphi)$ *and* $(\mathrm{Y}; \psi)$ *the following assertions are equivalent.*

(i) $(\mathrm{X}; \varphi)$ *and* $(\mathrm{Y}; \psi)$ *are point isomorphic.*

(ii) *There are sets* $A \in \Sigma$ *and* $A' \in \Sigma'$ *and a bijection* $\theta : A \to A'$ *such that the following hold.*

1) $\mu(A) = 1 = \mu'(A')$,

2) θ *is bijective, bi-measurable and measure-preserving,*

3) $A \subseteq \varphi^{-1}(A)$, $A' \subseteq \psi^{-1}(A')$ *and* $\theta \circ \varphi = \psi \circ \theta$ *on* A.

(Hint: For the implication (i)$\Rightarrow$(ii), define

$$A := \bigcap_{n \geq 0} [\, \eta \circ \theta \circ \varphi^n = \varphi^n \,] \cap [\, \psi^n \circ \theta = \theta \circ \varphi^n \,] \ \in \Sigma$$

and $A' \in \Sigma'$ likewise, cf. also Exercise 6.3.)

3. a) Consider the Bernoulli shift $B(1/2, 1/2) = (\mathscr{W}_2^+, \Sigma, \mu; \tau)$ and endow $[0,1]$ with the Lebesgue measure. Show that the measure spaces $(\mathscr{W}_2, \Sigma, \mu)$ and $([0,1], \Lambda, \lambda)$, Λ the Lebesgue σ-algebra, are point isomorphic. Prove that $B(1/2, 1/2)$ and $([0,1], \Lambda, \lambda; \varphi)$, φ the doubling map, are point isomorphic (see Example 5.1.2).

b) Prove the analogous statements for the baker's transformation (Example 5.1.1) and the *two-sided* Bernoulli-shift $B(1/2, 1/2) = (\mathscr{W}_2, \Sigma, \mu; \tau)$.

(Hint: Write the numbers in $[0,1]$ in binary form.)

4. Let φ be the tent map and ψ be the doubling map on $[0,1]$. Show that

$$\varphi : ([0,1], \lambda; \psi) \to ([0,1], \lambda; \varphi)$$

is a point homomorphism of MDSs.

5. a) Let (V, μ) be an abstract measure algebra. Show that

$$d_\mu(x,y) := \mu((x \wedge y^c) \vee (y \wedge x^c) \qquad (x, y \in V)$$

is a metric on V.

b) Let $\Theta : (V, \mu) \to (W, \nu)$ be a homomorphism of measure algebras. Show that $\Theta(\bot) = \bot$, $\Theta(\top) = \top$ and $\Theta(x^c) = \Theta(x)^c$ for all $x \in V$. Then show that Θ is isometric for the metric(s) d_μ, d_ν defined as in a).

c) Let $\Theta : (V, \mu) \to (W, \nu)$ be a *surjective* homomorphism of measure algebras. Show that Θ is bijective, and Θ^{-1} is a measure algebra homomorphism as well.

6. Work out in detail the proof of Theorem 12.11.

7. Show that ergodicity and invertibility of MDSs are (algebra) isomorphism invariants.

8. Let $(\mathrm{X}; \varphi)$ and $(\mathrm{Y}; \psi)$ be two MDSs. Show that if S is a Markov isomorphism of the corresponding L^1-spaces, the respective measure algebra isomorphism Θ is a σ-algebra isomorphism, i.e., preserves countable disjoint unions.

9. Let $K = \mathbb{Z} \cup \{\infty\}$ be the one-point compactification of $\mathbb{Z}$ and let $\varphi : K \to K$ be given by

$$\varphi(x) := \begin{cases} x+1 & \text{if } x \in \mathbb{Z}, \\ \infty & \text{if } x = \infty. \end{cases}$$

Show that δ_∞ is the only φ-invariant probability measure on K, so $(K; \varphi)$ is uniquely ergodic, surjective, but not minimal. (Cf. Example 2.27.)

10. Let $(\mathrm{X}; \varphi)$ be a measure-preserving system with Koopman operator T_φ, let $(K, \mu_K; \varphi_K)$ be a faithful topological MDS, and let $\Phi : \mathrm{L}^1(K, \mu) \to \mathrm{L}^1(\mathrm{X})$ be a Markov embedding of the dynamical systems. Prove that $A := \Phi(\mathrm{C}(K))$ is a T_φ-invariant C^*-subalgebra. (Hint: Use Theorem 7.18.)

11. Let A be a C^*-algebra let $M \subseteq A$ be countable subset of A, and let $B := \mathrm{clalg}(M)$ be the smallest C^*-subalgebra of A containing M. Show that B is separable.

12. Let $a \in \mathbb{T}$ with $a^n \neq 1$ for all $n \in \mathbb{N}$. Show that the Koopman operator L_a induced by the rotation with a is not mean ergodic on $\mathrm{L}^\infty(\mathbb{T}, \mathrm{d}z)$. (Hint: Consider $M := \{a^n : n \in \mathbb{N}\}$ and

$$I := \big\{[f] \in \mathrm{L}^\infty(\mathbb{T}, \mathrm{d}z) \,:\, f \in \mathrm{B}(\mathbb{T}) \text{ vanishes on a neighbourhood of } M\big\}.$$

Show that I is an ideal of $\mathrm{L}^\infty(\mathbb{T}, \mathrm{d}z)$ with $\mathbf{1} \notin J := \bar{I}$. Conclude that there is $\nu \in \mathrm{L}^\infty(\mathbb{T}, \mathrm{d}z)'$ which vanishes on J and satisfies $\langle \mathbf{1}, \nu \rangle = 1$. Use this property for $T'^n \nu$ and $\mathrm{A}'_n \nu$ and then employ a weak*-compactness argument.)

13. Consider the rotation system $(\mathbb{T}, \mathrm{d}z; a)$ for some $a \in \mathbb{T}$.

a) Show that $\mathrm{R}(\mathbb{T})$ is a full C^*-subalgebra of $\mathrm{L}^\infty(\mathbb{T})$.

b) Show that if the system is ergodic, then there is precisely one normalised positive invariant functional on $\mathrm{R}(\mathbb{T})$ (namely, $f \mapsto \int f \, \mathrm{d}z$).

c) Show that on $\mathrm{L}^\infty(\mathbb{T}, \mathrm{d}z)$ functionals as in b) abound.

14. Give an example of an MDS $(\mathrm{X}; \varphi)$ such that the space $\mathrm{L}^\infty(\mathrm{X})$ is infinite dimensional and the Koopman operator T_φ is mean ergodic thereon.

15. Illustrate the Jewett–Krieger theorem on an irrational rotation on the torus $\mathbb{T}$.

Chapter 13
Markov Operators

In this chapter we shall have a closer look at Markov operators introduced in Section 12.2 above. In the whole chapter $\mathrm{X}, \mathrm{Y}, \mathrm{Z}$ are given probability spaces. Recall from Definition 12.10 that an operator $S : \mathrm{L}^1(\mathrm{X}) \to \mathrm{L}^1(\mathrm{Y})$ is a **Markov operator** if it satisfies

$$S \geq 0, \quad S\mathbf{1} = \mathbf{1}, \quad \text{and} \quad \int_{\mathrm{Y}} Sf = \int_{\mathrm{X}} f \quad (f \in \mathrm{L}^1(\mathrm{X})),$$

the latter being equivalently expressed as $S'\mathbf{1} = \mathbf{1}$. We denote the set of all Markov operators by

$$\mathrm{M}(\mathrm{X};\mathrm{Y}) := \big\{S : \mathrm{L}^1(\mathrm{X}) \to \mathrm{L}^1(\mathrm{Y}) \,:\, S \text{ is Markov}\big\}$$

and abbreviate $\mathrm{M}(\mathrm{X}) := \mathrm{M}(\mathrm{X};\mathrm{X})$. The standard *topology* on $\mathrm{M}(\mathrm{X};\mathrm{Y})$ is the weak operator topology (see Appendix C.8).

There is an obvious conflict of terminology with Markov operators between $\mathrm{C}(K)$-spaces defined in Exercise 10.2. Markov operators between L^1-spaces, besides being positive and one-preserving also preserve the integral, or equivalently: T' as an operator between L^∞-spaces is also positive and one-preserving. Hence our "Markov operators" should rather be called **bi-Markov operators**.

This would also be in coherence with the finite dimensional situation. Namely, if $\Omega = \{1, \ldots, d\}$ is a finite set with probability measure $\mu = {}^1\!/_d \sum_{j=1}^d \delta_{\{j\}}$, and operators on $\mathrm{L}^1(\Omega, \mu) = \mathbb{C}^d$ are identified with matrices, then a matrix represents a Markov operator in the sense of definition above if and only if it is *bi*stochastic (or: doubly stochastic).

However, our definition reflects the common practice in the field, see [Glasner (2003), Definition 6.12], so we stick to it. As operator theorists we regret the conflicting meanings of the term "Markov operator", but at least in this book, where Markov operators on $\mathrm{C}(K)$-spaces play a minor role, no problems should arise.

13.1 Examples and Basic Properties

Let us begin with a collection of examples.

Examples 13.1. a) The operator $\mathbf{1}\otimes\mathbf{1}$ defined by

$$(\mathbf{1}\otimes\mathbf{1})f := \langle f,\mathbf{1}\rangle\cdot\mathbf{1} = \left(\int_{\mathrm{X}} f\right)\mathbf{1} \qquad (f\in \mathrm{L}^1(\mathrm{X}))$$

is a Markov operator from $\mathrm{L}^1(\mathrm{X})$ to $\mathrm{L}^1(\mathrm{Y})$. It is the unique Markov operator $S\in \mathrm{M}(\mathrm{X};\mathrm{Y})$ with $\dim \mathrm{ran}(S)=1$, and is called the **trivial** Markov operator.

b) The identity mapping I is a Markov operator, $\mathrm{I}\in \mathrm{M}(\mathrm{X})$.

c) Clearly, every Markov embedding as defined in Section 12.2 is a Markov operator. In particular, every Koopman operator T_φ associated with an MDS $(\mathrm{X};\varphi)$ is a Markov operator.

The following theorem summarises the basic properties of Markov operators.

Theorem 13.2. a) *Composition of Markov operators yields a Markov operator. The set* $\mathrm{M}(\mathrm{X};\mathrm{Y})$ *of all Markov operators is a compact convex subset of* $\mathscr{L}(\mathrm{L}^1(\mathrm{X});\mathrm{L}^1(\mathrm{Y}))$ *with the weak operator topology. In case* $\mathrm{X}=\mathrm{Y}$ *it is a subsemigroup.*

b) *Every Markov operator* $S\in\mathrm{M}(\mathrm{X};\mathrm{Y})$ *is a Dunford–Schwartz operator, i.e., it restricts to a contraction on each space* $\mathrm{L}^p(\mathrm{X})$ *for* $1\le p\le\infty$*:*

$$\|Sf\|_p \le \|f\|_p \qquad (f\in\mathrm{L}^p(\mathrm{X}),\, 1\le p\le\infty).$$

c) *The adjoint* S' *of a Markov operator* $S:\mathrm{L}^1(\mathrm{X})\to\mathrm{L}^1(\mathrm{Y})$ *extends uniquely to a Markov operator* $S':\mathrm{L}^1(\mathrm{Y})\to\mathrm{L}^1(\mathrm{X})$*. The mapping*

$$\mathrm{M}(\mathrm{X};\mathrm{Y})\to\mathrm{M}(\mathrm{Y};\mathrm{X}), \qquad S\mapsto S'$$

is affine and continuous. Moreover, one has $(S')'=S$*. If* $R\in\mathrm{M}(\mathrm{Y};\mathrm{Z})$ *is another Markov operator, then* $(RS)'=S'R'$*.*

Proof. Assertion a) is straightforward. For b) note first that $S\in\mathrm{M}(\mathrm{X};\mathrm{Y})$ is an L^1-contraction. Indeed, by Lemma 7.5 we have $|Sf|\le S|f|$, and integrating yields

$$\|Sf\|_1 = \int_Y |Sf| \le \int_Y S|f| = \int_X |f| = \|f\|_1$$

for every $f\in\mathrm{L}^1$. Furthermore, S is an L^∞-contraction, since $|Sf|\le S|f|\le\|f\|_\infty S\mathbf{1} = \|f\|_\infty\mathbf{1}$ for every $f\in\mathrm{L}^\infty$. The rest is then Theorem 8.23.

For c), take $0\le g\in\mathrm{L}^\infty(\mathrm{Y})$. Then

$$\int_{\mathrm{X}} S'g\cdot f = \int_{\mathrm{Y}} g\cdot Sf \ge 0 \quad \text{whenever } 0\le f\in\mathrm{L}^\infty(\mathrm{X}).$$

Hence S' is a positive operator. It follows that $|S'g| \le S'|g|$ (Lemma 7.5) and hence

$$\|S'g\|_1 = \int_X |S'g| \le \int_X S'|g| = \int_Y |g| \cdot S\mathbf{1} = \int_Y |g| = \|g\|_1$$

for every $g \in L^\infty(Y)$. So S' is an L^1-contraction, and as such has a unique extension to a contraction $S' : L^1(Y) \to L^1(X)$. Since the positive cone of L^∞ is dense in the positive cone of L^1, S' is indeed a positive operator. An easy argument shows that

$$(S')' = S\big|_{L^\infty(X)},$$

in particular $(S')'\mathbf{1} = S\mathbf{1} = \mathbf{1}$, whence S' is a Markov operator. The proofs for the remaining assertions are left as Exercise 1. □

By definition, the adjoint S' of a Markov operator $S \in M(X;Y)$ satisfies

$$\langle Sf, g\rangle_Y = \langle f, S'g\rangle_X$$

for $f \in L^1(X)$ and $g \in L^\infty(Y)$. In order to extend this to other pairs of functions f and g, we need the following lemma.

Lemma 13.3. *Let $1 \le p < \infty$ and let $f, g \in L^p(X)$. Then $f \cdot g \in L^p(X)$ if and only if there are sequences $(f_n)_n, (g_n)_n \subseteq L^\infty(X)$ such that*

$$\|f_n - f\|_p, \|g_n - g\|_p \to 0, \quad \sup_n \|f_n \cdot g_n\|_p < \infty.$$

In this case one can choose $(f_n)_n$ and $(g_n)_n$ with the additional properties

$$|f_n| \le |f|, \quad |g_n| \le |g|, \quad \|f_n g_n - fg\|_p \to 0.$$

Proof. For one implication, pass to subsequences that converge almost everywhere and then apply Fatou's lemma to conclude that the product belongs to L^p. For the converse use the approximation

$$f_n := f \cdot \mathbf{1}_{[|f| \le n]}, \qquad g_n := f \cdot \mathbf{1}_{[|g| \le n]}$$

and the dominated convergence theorem. □

With the help of Lemma 13.3 we can relate the Markov adjoint and the Hilbert space adjoint.

Proposition 13.4. *Let $S \in M(X;Y)$ and $f \in L^1(X)$, $g \in L^1(Y)$. If $g \cdot S|f| \in L^1(Y)$, then $f \cdot S'|g| \in L^1(X)$ and one has*

$$\int_Y Sf \cdot g = \int_X f \cdot S'g.$$

Moreover, $S'\big|_{L^2} = \big(S\big|_{L^2}\big)^*$, the Hilbert space adjoint of $S\big|_{L^2}$.

Proof. Take $g_n \in \mathrm{L}^\infty$ with $|g_n| \le |g|$ and $g_n \to g$ almost everywhere. Hence $g_n Sf \to gSf$ in L^1, by the dominated convergence theorem. On the other hand, we have $S'g_n \to S'g$ and

$$\big\|fS'g_n\big\|_1 \le \int_{\mathrm{X}} |f|\, S'|g_n| = \int_{\mathrm{Y}} \big(S|f|\big)\, |g_n| \le \|g \cdot S|f|\|_1$$

for all $n \in \mathbb{N}$. Hence, by Lemma 13.3, $f \cdot S'|g| \in \mathrm{L}^1(\mathrm{X})$ and $f \cdot S'g_n \to f \cdot S'g$ in $\mathrm{L}^1(\mathrm{X})$. Therefore, $\langle Sf, g\rangle_{\mathrm{Y}} = \langle f, S'g\rangle_{\mathrm{X}}$ as claimed. The remaining assertion follows from the computation

$$\big(S'f \,\big|\, g\big)_{\mathrm{L}^2} = \int_{\mathrm{X}} S'f \cdot \overline{g} = \int_{\mathrm{X}} f \cdot S\overline{g} = \int_{\mathrm{X}} f \cdot \overline{Sg} = (f \,|\, Sg)_{\mathrm{L}^2}$$

for $f \in \mathrm{L}^2(\mathrm{X})$ and $g \in \mathrm{L}^2(\mathrm{Y})$. □

Remark 13.5. Since L^2 is dense in L^1, the mapping

$$\mathrm{M}(\mathrm{X};\mathrm{Y}) \to \mathscr{L}\big(\mathrm{L}^2(\mathrm{X});\mathrm{L}^2(\mathrm{Y})\big), \qquad S \mapsto S\big|_{\mathrm{L}^2}$$

preserves convex combinations, is continuous and injective. Hence we can switch back and forth between the L^1-setting and the L^2-setting.

Finally in this section we describe a convenient method of constructing Markov operators, already employed in Section 12.3 above. The simple proof is left as Exercise 2.

Proposition 13.6. . *Let K be a compact space and $S : \mathrm{C}(K) \to \mathrm{L}^1(\mathrm{X})$ a linear operator satisfying $S \ge 0$ and $S\mathbf{1} = \mathbf{1}$. Then there is a unique $\mu \in \mathrm{M}^1(K)$, namely $\mu = S'\mathbf{1}$, such that S extends to a Markov map $S^\sim : \mathrm{L}^1(K,\mu) \to \mathrm{L}^1(\mathrm{X})$. If S is injective, then μ has full support.*

What we mean by "extension" here is of course that $S = S^\sim \circ J$, where $J : \mathrm{C}(K) \to \mathrm{L}^1(K,\mu)$ is the canonical map assigning to $f \in \mathrm{C}(K)$ its equivalence class Jf modulo equality μ-almost everywhere. We usually suppress reference to this map and write again simply S in place of $S^\sim$ when it is safe to do so.

13.2 Embeddings, Factor Maps and Isomorphisms

Recall from Section 12.2 that a Markov operator is an **embedding** if it is a lattice homomorphism. Let

$$\mathrm{Emb}(\mathrm{X};\mathrm{Y}) := \big\{S : S \in \mathrm{M}(\mathrm{X};\mathrm{Y}) \text{ is a Markov embedding}\big\}$$

denote the set of Markov embeddings. We can now give a comprehensive characterisation.

Theorem 13.7. *For a Markov operator $S \in \mathrm{M}(\mathrm{X};\mathrm{Y})$ the following assertions are equivalent.*

(i) $S(f \cdot g) = Sf \cdot Sg$ *for all* $f, g \in \mathrm{L}^\infty(\mathrm{X})$.

(ii) $S'S = \mathrm{I}_\mathrm{X}$.

(iii) *There is* $T \in \mathrm{M}(\mathrm{Y};\mathrm{X})$ *such that* $TS = \mathrm{I}_\mathrm{X}$.

(iv) $\|Sf\|_p = \|f\|_p$ *for all* $f \in \mathrm{L}^p(\mathrm{X})$ *and some/all* $1 \le p < \infty$.

(v) $|Sf|^p = S|f|^p$ *for all* $f \in \mathrm{L}^\infty(\mathrm{X})$ *and some/all* $1 \le p < \infty$.

(vi) $S \in \mathrm{Emb}(\mathrm{X};\mathrm{Y})$, *i.e.,* $|Sf| = S|f|$ *for all* $f \in \mathrm{L}^1(\mathrm{X})$.

Moreover, (ii) *implies also that* $\|Sf\|_\infty = \|f\|_\infty$ *for all* $f \in \mathrm{L}^\infty(\mathrm{X})$.

Proof. If (i) holds, then $\langle Sf, Sg\rangle = \int (Sf)\cdot(Sg) = \int S(f\cdot g) = \int f\cdot g = \langle f, g\rangle$ for all $f, g \in \mathrm{L}^\infty(\mathrm{X})$. This means that $S'Sf = f$ for all $f \in \mathrm{L}^\infty(\mathrm{X})$, whence (ii) follows by approximation. The implication (ii) $\Rightarrow$ (iii) is trivial. If (iii) holds, then for $f \in \mathrm{L}^p(\mathrm{X})$ and $1 \le p \le \infty$ on has $\|f\|_p = \|TSf\|_p \le \|Sf\|_p \le \|f\|_p$, hence $\|Sf\|_p = \|f\|_p$.

Suppose that (iv) holds for some $p \in [1,\infty)$ and $f \in \mathrm{L}^p(\mathrm{X})$. Then $0 \le |Sf|^p \le S|f|^p$ by Theorem 7.19 with $g = \mathbf{1}$. Integrating yields

$$\|f\|_p^p = \|Sf\|_p^p = \int_\mathrm{Y} |Sf|^p \le \int_\mathrm{Y} S|f|^p = \int_\mathrm{X} |f|^p = \|f\|_p^p .$$

Hence $|Sf|^p = S|f|^p$, i.e., (v) is true. If we start from (v) and use the identity for f and $|f|$, we obtain

$$|Sf|^p = S|f|^p = \big|S|f|\big|^p = \big(S|f|\big)^p .$$

Taking p^{th} roots yields $|Sf| = S|f|$ for every $f \in \mathrm{L}^\infty(\mathrm{X})$, and by density we obtain (vi). Finally, the implication (vi) $\Rightarrow$ (i) follows from the second part of Theorem 7.18. □

Remark 13.8. Property (i) can be strengthened to

(i') $S(f \cdot g) = Sf \cdot Sg$ *for all* $f, g \in \mathrm{L}^1(\mathrm{Y})$ *such that* $f \cdot g \in \mathrm{L}^1(\mathrm{Y})$.

This follows from (i) and Lemma 13.3.

Factor Maps

A Markov operator $P \in \mathrm{M}(\mathrm{X};\mathrm{Y})$ is called a **(Markov) factor map** if P' is an embedding. Here is a characterisation of factor maps.

Theorem 13.9. *For a Markov operator $P \in \mathrm{M}(\mathrm{X};\mathrm{Y})$ the following assertions are equivalent.*

(i) *P is a factor map, i.e., P' is an embedding.*

(ii) $PP' = \mathrm{I}_\mathrm{Y}$.

(iii) *There is* $T \in \mathrm{M}(\mathrm{Y};\mathrm{X})$ *such that* $PT = \mathrm{I}_{\mathrm{Y}}$.

(iv) $P((P'f)\cdot g) = f\cdot(Pg)$ *for all* $f \in \mathrm{L}^\infty(\mathrm{Y})$, $g \in \mathrm{L}^\infty(\mathrm{X})$.

Proof. If (i) holds, then P' is an embedding, whence $\mathrm{I} = (P')'P' = PP'$, i.e., (ii) follows. The implication (ii)$\Rightarrow$(iii) is trivial. If $PT = \mathrm{I}$ then taking adjoints we obtain $T'P' = \mathrm{I}' = \mathrm{I}$, hence by Theorem 13.7 we conclude that P' is an embedding, and that is (i).

Finally, observe that (iv) implies (ii) (just take $g = \mathbf{1}$ and apply a density argument). On the other hand, if P' is an embedding, we have by Theorem 13.7.(i),

$$\int_{\mathrm{Y}} P(P'f\cdot g)\cdot h = \int_{\mathrm{X}} P'f\cdot g\cdot P'h = \int_{\mathrm{X}} P'(fh)\cdot g = \int_{\mathrm{Y}} f\cdot Pg\cdot h.$$

for $f,h \in \mathrm{L}^\infty(\mathrm{Y})$ and $g \in \mathrm{L}^\infty(\mathrm{X})$. This is the weak formulation of (iv). □

Remark 13.10. Assertion (iv) can be strengthened to

(iv') $P((P'f)\cdot g) = f\cdot(Pg)$ $\quad (f \in \mathrm{L}^1(\mathrm{Y}),\ g \in \mathrm{L}^1(\mathrm{X}),\ \ P'f\cdot g \in \mathrm{L}^1(\mathrm{X}))$.

Proof. By denseness, the claim is true if $g \in \mathrm{L}^\infty(\mathrm{X})$. In the general case approximate $g_n \to g$ a.e., with $g_n \in \mathrm{L}^\infty(\mathrm{X})$ and so that $|g_n| \le |g|$. Then $P'f\cdot g_n \to P'f\cdot g$ a.e. and

$$\left|P'f\cdot g_n\right| \le \left|P'f\cdot g\right|$$

hence $P'f\cdot g_n \to P'f\cdot g$ by the dominated convergence theorem. This implies that

$$f\cdot Pg_n = P(P'f\cdot g_n) \to P(P'f\cdot g)$$

in $\mathrm{L}^1(\mathrm{Y})$. But $Pg_n \to Pg$ in $\mathrm{L}^1(\mathrm{Y})$ and by passing to an a.e.-convergent subsequence one concludes (iv'). □

Isomorphisms

Recall from Section 12.2 that a surjective Markov embedding $S \in \mathrm{M}(\mathrm{X};\mathrm{Y})$ is called **Markov isomorphism**. The next is a characterisation of such embeddings.

Corollary 13.11. *For a Markov operator* $S \in \mathrm{M}(\mathrm{X};\mathrm{Y})$ *the following assertions are equivalent.*

(i) *There are* $T_1, T_2 \in \mathrm{M}(\mathrm{Y};\mathrm{X})$ *such that* $T_1S = \mathrm{I}_{\mathrm{X}}$ *and* $ST_2 = \mathrm{I}_{\mathrm{Y}}$.

(ii) *S is an embedding and a factor map.*

(iii) *S and* S' *are both embeddings (both factor maps).*

(iv) *S is a surjective embedding, i.e., a Markov isomorphism.*

(v) *S is an injective factor map.*

(vi) *S is bijective and* S^{-1} *is positive.*

In this case $S^{-1} = S'$.

Proof. By Theorems 13.7 and 13.9 it is clear that (i)–(iii) are equivalent, and any one of them implies (iv) and (v). Suppose that (iv) holds. Then $S'S = \mathrm{I}$ and S is surjective. Hence it is bijective and $S^{-1} = S'$ is a Markov operator, whence (i). The proof of (v)$\Rightarrow$(i) is similar. Clearly (i)–(v) all imply (vi). Conversely, suppose that (vi) holds, i.e., S is bijective and $S^{-1} \geq 0$. Then

$$|Sf| \leq S|f| = S\left|S^{-1}Sf\right| \leq SS^{-1}|Sf| = |Sf|,$$

for each $f \in \mathrm{L}^1(\mathrm{X})$, whence S is an embedding, i.e., (iv). □

A Markov **automorphism** S of X is a self-isomorphism of X. We introduce the following notations

$$\begin{aligned} \mathrm{Iso}(\mathrm{X};\mathrm{Y}) &:= \big\{S : S \in \mathrm{M}(\mathrm{X};\mathrm{Y})\ S \text{ is a Markov isomorphism}\big\}, \\ \mathrm{Aut}(\mathrm{X}) &:= \big\{S : S \in \mathrm{M}(\mathrm{X})\ S \text{ is a Markov automorphism}\big\}. \end{aligned}$$

We note that even if S is a bijective Markov operator, its inverse S^{-1} need not be positive, hence not a Markov operator. As an example consider the bistochastic matrix

$$S := \begin{pmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix} \quad \text{with inverse} \quad S^{-1} = \begin{pmatrix} -1 & 2 \\ 2 & -1 \end{pmatrix},$$

which is not positive. This shows that one cannot drop the requirement of positivity of S^{-1} in (ii) of Corollary 13.11.

13.3 Markov Projections

We turn to our last class of distinguished Markov operators. A Markov operator $Q \in \mathrm{M}(\mathrm{X})$ is called a **Markov projection** if it is a projection, i.e., $Q^2 = Q$. Alternatively, Markov projections are called **conditional expectations**, for reasons that will become clear below (Remark 13.17).

The following lemma characterises Markov projections in terms of the Hilbert space $\mathrm{L}^2(\mathrm{X})$.

Lemma 13.12. *Every Markov projection* $Q \in \mathrm{M}(\mathrm{X})$ *is self-adjoint, i.e., satisfies* $Q' = Q$ *and restricts to an orthogonal projection on the Hilbert space* $\mathrm{L}^2(\mathrm{X})$. *Conversely, if* Q *is an orthogonal projection on* $\mathrm{L}^2(\mathrm{X})$ *such that* $Q \geq 0$ *and* $Q\mathbf{1} = \mathbf{1}$, *then it extends uniquely to a Markov projection on* $\mathrm{L}^1(\mathrm{X})$.

Proof. If Q is a Markov projection, it restricts to a contractive projection on the Hilbert space $H := \mathrm{L}^2(\mathrm{X})$. Hence, by Theorem 13.31 in the supplement below, Q is an orthogonal projection and in particular self-adjoint. Then $Q = Q'$ by Proposition 13.4.

Suppose that Q is an orthogonal projection on H, satisfying $Q \geq 0$ and $Q\mathbf{1} = \mathbf{1}$. Then Q is self-adjoint and hence for $f \in \mathrm{L}^2$ one has

$$\|Qf\|_1 = \int_{\mathrm{X}} |Qf| \leq \int_{\mathrm{X}} Q|f| = \int_{\mathrm{X}} |f| \cdot Q\mathbf{1} = \int_{\mathrm{X}} |f| = \|f\|_1 .$$

By approximation Q extends to a positive contraction on L^1 which is clearly a Markov operator. It follows that $Q^2 = Q$, i.e., Q is a Markov projection. □

Corollary 13.13. *Every Markov projection $Q \in \mathrm{M}(\mathrm{X})$ is uniquely determined by its range* $\mathrm{ran}(Q)$.

Proof. Let P, Q be Markov projections with $\mathrm{ran}(P) = \mathrm{ran}(Q)$. Then

$$\mathrm{ran}(P|_{\mathrm{L}^2}) = \mathrm{ran}(P) \cap \mathrm{L}^2 = \mathrm{ran}(Q) \cap \mathrm{L}^2 = \mathrm{ran}(Q|_{\mathrm{L}^2}).$$

Hence $P = Q$ on L^2 (Corollary 13.32.c)), and thus on L^1 by approximation. □

Before we continue, let us note the following useful fact, based on the L^2-theory of orthogonal projections.

Corollary 13.14. *Let X, Y be probability spaces, and let $S \in \mathrm{M}(\mathrm{X};\mathrm{Y})$ and $Q \in \mathrm{M}(\mathrm{Y})$ such that Q is a projection and QS is an embedding. Then $QS = S$.*

Proof. Let $f \in \mathrm{L}^2(\mathrm{X})$. Since QS is an embedding, it is isometric on L^2, and hence

$$\|f\|_2 = \|QSf\|_2 \leq \|Sf\|_2 \leq \|f\|_2 .$$

Hence $QSf = Sf$ by Corollary 13.32.a) below. Since L^2 is dense in L^1, $QS = S$ follows. □

We now aim at a characterisation of Markov projections. Our main auxiliary result is the following.

Proposition 13.15. *Let $\mathrm{X} = (X, \Sigma, \mu)$ be a probability space. Then the following assertions hold.*

a) *The range $F := \mathrm{ran}(Q)$ of a Markov projection $Q \in \mathrm{M}(\mathrm{X})$ is a Banach sublattice of $\mathrm{L}^1(\mathrm{X})$ containing $\mathbf{1}$.*

b) *If F is a Banach sublattice of $\mathrm{L}^1(\mathrm{X})$ containing $\mathbf{1}$, then*

$$\Sigma_F := \{A \in \Sigma : \mathbf{1}_A \in F\}$$

is a sub-σ-algebra of Σ, and $F = \mathrm{L}^1(X, \Sigma_F, \mu)$.

c) *If Σ' is a sub-σ-algebra of Σ, then the canonical injection*

$$J : \mathrm{L}^1(X, \Sigma', \mu) \to \mathrm{L}^1(X, \Sigma, \mu), \qquad Jf = f$$

is a Markov embedding.

d) *If* $S : \mathrm{L}^1(\mathrm{Y}) \to \mathrm{L}^1(\mathrm{X})$ *is a Markov embedding, then* $Q := SS'$ *is a Markov projection with* $\mathrm{ran}(Q) = \mathrm{ran}(S)$.

Proof. a) Obviously $\mathbf{1} = Q\mathbf{1} \in F$. For $f = Qf \in F$ we have

$$|f| = |Qf| \leq Q|f|, \quad \text{i.e.,} \quad Q|f| - |f| \geq 0.$$

Therefore, since $Q^2 = Q$ and Q is a Markov operator,

$$0 \leq \int_{\mathrm{X}} \big(Q|f| - |f|\big) = \int_{\mathrm{X}} Q\big(Q|f| - |f|\big) = \int_{\mathrm{X}} \big(Q^2 - Q\big)|f| = 0.$$

This implies that $|f| = Q|f| \in F$, whence F is a vector sublattice. Since F is closed (being the range of a bounded projection), it is a Banach sublattice.

b) It is easily seen that Σ_F is a sub-σ-algebra of Σ. Now, the inclusion $\mathrm{L}^1(X, \Sigma_F, \mu) \subseteq F$ is clear, the step functions being dense in L^1. For the converse inclusion, let $0 \leq f \in F$. Then $\mathbf{1}_{[f>0]} = \sup_{n\in\mathbb{N}} nf \wedge \mathbf{1} \in F$. Applying this to $(f - c\mathbf{1})^+ \in F$ yields that $\mathbf{1}_{[f>c]} \in F$ for each $c \in \mathbb{R}$. Hence f is Σ_F-measurable. Since F is a sublattice, $F \subseteq \mathrm{L}^1(X, \Sigma_F, \mu)$, as claimed.

c) is trivial.

d) Since S is an embedding, $S'S = \mathrm{I}$ (Theorem 13.7) and hence $Q^2 = (SS')(SS') = S(S'S)S' = SS' = Q$. By $Q = SS'$ it is clear that $\mathrm{ran}(Q) \subseteq \mathrm{ran}(S)$, but since $QS = SS'S = S$, we also have $\mathrm{ran}(S) \subseteq \mathrm{ran}(Q)$ as claimed. □

Suppose $Q \in \mathrm{M}(\mathrm{X})$ is a Markov projection. Following the steps a)–d) in Proposition 13.15 yields

$$Q \mapsto F = \mathrm{ran}(Q) \mapsto \Sigma_F \mapsto JJ' = P$$

and P is a Markov projection with $\mathrm{ran}(P) = \mathrm{ran}(J) = \mathrm{L}^1(X, \Sigma_F, \mu) = F = \mathrm{ran}(Q)$. By Corollary 13.13, it follows that $P = Q$.

If, however, we start with a Banach sublattice F of $\mathrm{L}^1(\mathrm{X})$ containing $\mathbf{1}$ and follow the steps from above, we obtain

$$F \mapsto \Sigma_F \mapsto Q = JJ' \mapsto \mathrm{ran}(Q),$$

with $\mathrm{ran}(Q) = \mathrm{ran}(J) = \mathrm{L}^1(X, \Sigma_F, \mu) = F$.

Finally, start with a sub-σ-algebra $\Sigma' \subseteq \Sigma$ and follow the above steps to obtain

$$\Sigma' \mapsto Q = JJ' \mapsto F = \mathrm{ran}(Q) = \mathrm{ran}(J) = \mathrm{L}^1(X, \Sigma', \mu) \mapsto \Sigma_F.$$

One has $A \in \Sigma_F$ if and only if $\mathbf{1}_A \in \mathrm{L}^1(X, \Sigma', \mu)$, which holds if, and only if, $\mu(A \triangle B) = 0$ for some $B \in \Sigma'$. Hence $\Sigma' = \Sigma_F$ if and only if Σ' is **relatively complete**, which means that it contains all μ-null sets. (Note that Σ_F as in Proposition 13.15.b) is always relatively complete.)

We have proved the following characterisation result.

Theorem 13.16. *Let* $\mathrm{X} = (X, \Sigma, \mu)$ *be a probability space. The assignments*

$$Q \mapsto P = Q|_{\mathrm{L}^2}, \quad Q \mapsto \mathrm{ran}(Q), \quad Q \mapsto \Sigma' = \left\{A \in \Sigma : Q\mathbf{1}_A = \mathbf{1}_A\right\}$$

constitute one-to-one correspondences between the following four sets of objects.

1) *Markov projections* Q *on* $\mathrm{L}^1(\mathrm{X})$.
2) *Orthogonal projections* P *on* $\mathrm{L}^2(\mathrm{X})$ *satisfying* $P \geq 0$ *and* $P\mathbf{1} = \mathbf{1}$.
3) *Banach sublattices* $F \subseteq \mathrm{L}^1(\mathrm{X})$ *containing* $\mathbf{1}$.
4) *Relatively complete sub-*σ*-algebras* Σ' *of* Σ.

In the following we sketch the connection of all these results with probability theory.

Remark 13.17 (Conditional Expectations). Let Σ' be a sub-σ-algebra of Σ, and let Q be the associated Markov projection, i.e., the unique Markov projection $Q^2 = Q \in \mathrm{M}(\mathrm{X})$ with

$$\mathrm{ran}(Q) = \mathrm{L}^1(X, \Sigma', \mu).$$

Let us abbreviate $\mathrm{Y} := (X, \Sigma', \mu)$. Then, for $A \in \Sigma'$ and $f \in \mathrm{L}^1(\mathrm{X})$, we have

$$\int_A Qf \,\mathrm{d}\mu = \int_{\mathrm{X}} \mathbf{1}_A \cdot JJ'f = \int_{\mathrm{Y}} \mathbf{1}_A \cdot J'f = \int_{\mathrm{X}} (J\mathbf{1}_A) \cdot f = \int_A f \,\mathrm{d}\mu.$$

This means that $Q = \mathbb{E}(\cdot\,|\,\Sigma')$ is the **conditional expectation** operator, common in probability theory [Billingsley (1979), Section 34].

Let us turn to characterisation of Markov projections in the spirit of Theorems 13.7 and 13.9.

Theorem 13.18. *For a Markov operator* $Q \in \mathrm{M}(\mathrm{X})$ *the following assertions are equivalent.*

(i) $Q^2 = Q$, *i.e.,* Q *is a Markov projection.*

(ii) $Q = SS'$ *for some Markov embedding* S.

(iii) $Q = P'P$ *for some Markov factor map* P.

(iv) $Q(Qf \cdot g) = Qf \cdot Qg$ *for all* $f, g \in \mathrm{L}^\infty(\mathrm{X})$.

Proof. By the duality of embeddings and factor maps, (ii) and (iii) are equivalent. The equivalence of (i) and (ii) follows from Proposition 13.15 and the remarks following it. The implication (iii)$\Rightarrow$(iv) follows from

$$\begin{aligned} Q(Qf \cdot g) &= P'P(P'Pf \cdot g) = P'\left[P\left(P'(Pf) \cdot g\right)\right] = P'(Pf \cdot Pg) \\ &= P'Pf \cdot P'Pg = Qf \cdot Qg \end{aligned}$$

by virtue of Theorems 13.7 and 13.9. The implication (iv)$\Rightarrow$(i) is proved by specialising $g = \mathbf{1}$ and by denseness of $\mathrm{L}^\infty(\mathrm{X})$ in $\mathrm{L}^1(\mathrm{X})$. □

Remark 13.19. Assertion (iv) can be strengthened to

(iv') $Q(Qf \cdot g) = Qf \cdot Qg$ *for all* $f, g \in \mathrm{L}^1(\mathrm{X})$ *such that* $Qf \cdot g \in \mathrm{L}^1(\mathrm{X})$.

This follows, as in Remarks 13.10 and 13.8, by approximation.

Example 13.20 (Mean Ergodic Projections). Every Markov operator $T \in \mathrm{M}(\mathrm{X})$, $\mathrm{X} = (X, \Sigma, \mu)$, is a Dunford–Schwartz operator. Hence, it is mean ergodic (Theorem 8.24), i.e., the limit

$$P_T := \lim_{n\to\infty} \mathrm{A}_n[T] = \lim_{n\to\infty} \frac{1}{n} \sum\nolimits_{j=0}^{n-1} T^j$$

exists in the strong operator topology. The operator P_T is a projection with $\mathrm{ran}(P_T) = \mathrm{fix}(T)$. Since T is a Markov operator, so is P_T, i.e., P_T is a Markov projection. By Proposition 13.15, $\mathrm{fix}(T)$ is a Banach sublattice. Note that by Corollary 8.7 we have

$$\mathrm{fix}(T) = \mathrm{fix}(T') \quad \text{and} \quad P_T = P_{T'}.$$

The corresponding σ-algebra is

$$\Sigma_{\mathrm{fix}} = \big\{ A \in \Sigma : T\mathbf{1}_A = \mathbf{1}_A \big\}$$

and the mean ergodic projection P_T coincides with the conditional expectation $P_T = \mathbb{E}(\cdot \,|\, \Sigma_{\mathrm{fix}})$. (Cf. Remark 8.9 for the special case that T is a Koopman operator.)

13.4 Factors and their Models

We now apply the result of the previous section to dynamical systems. We begin with an auxiliary result employing the notation from above.

Lemma 13.21. *Let* X, Y *be probability spaces, and let* $P \in \mathrm{M}(\mathrm{X})$ *and* $Q \in \mathrm{M}(\mathrm{Y})$ *be Markov projections with ranges* $F = \mathrm{ran}(P)$ *and* $G = \mathrm{ran}(Q)$. *Then for a Markov operator* $T \in \mathrm{M}(\mathrm{X};\mathrm{Y})$ *the following equivalences are true.*

a) $T(F) \subseteq G \quad \Leftrightarrow \quad QTP = TP.$

b) $T(F) \subseteq G, \quad T'(G) \subseteq F \quad \Leftrightarrow \quad QT = TP \quad \Leftrightarrow \quad PT' = T'Q.$

Proof. The equivalence a) is trivial. For the equivalences in b) we use a) and note that $TF \subseteq G$ and $T'G \subseteq F$ are equivalent to $QTP = TP$ and $PT'Q = T'Q$. By taking adjoints in the latter equality we obtain $QTP = QT$. This implies $QT = TP$. The last equivalence is trivial. □

Remark 13.22. Under the hypotheses of Lemma 13.21, let Σ_F and Σ_G be the associated sub-σ-algebras of Σ_{X} and Σ_{Y}, respectively. Furthermore, suppose that $T = T_\varphi$ is the Koopman operator of a measure-preserving measurable map $\varphi : Y \to X$.

Then it is easy to see that the equivalent assertions in Lemma 13.21.a) are further equivalent to φ being Σ_G-Σ_F-measurable. If φ is essentially invertible, then

$$T' = T^{-1} = T_{\varphi^{-1}},$$

and hence the statements in b) are equivalent to φ being Σ_G-Σ_F-measurable and φ^{-1} being Σ_F-Σ_G-measurable. See Exercise 4.

We can now introduce one of the main notions in structural ergodic theory. A **factor** of an MDS $(\mathrm{X};\varphi)$ is any T_φ-invariant Banach sublattice of $\mathrm{L}^1(\mathrm{X})$ containing $\mathbf{1}$. A factor F is called **bi-invariant** if F is also T'_φ-invariant.

The following characterises factors in terms of Markov projections. Its proof is straightforward from Lemma 13.21 above.

Theorem 13.23. *For an MDS $(\mathrm{X};\varphi)$ with Koopman operator T the following assertions are true.*

a) *The map*

$$Q \mapsto F = \mathrm{ran}(Q)$$

establishes a one-to-one correspondence between the factors F of $(\mathrm{X};\varphi)$ and the Markov projections $Q \in \mathrm{M}(\mathrm{X})$ satisfying $TQ = QTQ$.

b) *If the MDS $(\mathrm{X};\varphi)$ is invertible, then under the given map in* a) *the bi-invariant factors F correspond to Markov projections satisfying $TQ = QT$.*

Equivalently, by Remark 13.22 above and the results of the previous sections one can describe factors in terms of sub-σ-algebras.

Theorem 13.24. *For an MDS $(\Omega,\Sigma,\mu;\varphi)$, $\mathrm{X} = (X,\Sigma,\mu)$, the following assertions hold.*

a) *There is a one-to-one correspondence between factors F of $(\mathrm{X};\varphi)$ and relatively complete sub-σ-algebras Σ_F of Σ with respect to which φ is measurable, given by*

$$F \mapsto \Sigma_F := \left\{A \in \Sigma : \mathbf{1}_A \in F\right\} \quad \text{and} \quad \Sigma' \mapsto F_{\Sigma'} := \mathrm{L}^1(X,\Sigma',\mu).$$

b) *If φ is essentially invertible, then F is a bi-invariant factor if and only if both φ and φ^{-1} are Σ_F-measurable. Alternatively, F is bi-invariant if it is also a factor of the MDS $(\mathrm{X};\varphi^{-1})$.*

Finally, let us characterise factors in the spirit of Chapter 12, i.e., in the framework of topological MDSs.

Theorem 13.25. *Let $(\mathrm{X};\varphi)$ be an MDS with Koopman-operator $T = T_\varphi$, and let $F \subseteq \mathrm{L}^1(\mathrm{X})$. Then the following assertions are equivalent.*

(i) *F is a factor of $(\mathrm{X};\varphi)$.*

(ii) *$F = \mathrm{cl}_{\mathrm{L}^1}(A)$ for some T_φ-invariant C^*-subalgebra $A \subseteq \mathrm{L}^\infty(\mathrm{X})$.*

(iii) *$F = \mathrm{ran}(S)$ for some Markov embedding of MDSs*

$$S : (\mathrm{L}^1(\mathrm{Y});T_\psi) \to (\mathrm{L}^1(\mathrm{X});T_\varphi)$$

and some MDS $(\mathrm{Y};\psi)$.

(iv) $F = \mathrm{ran}(\Phi)$ *for some Markov embedding of MDSs*

$$\Phi : (\mathrm{L}^1(K,\mu);T_\psi) \to (\mathrm{L}^1(\mathrm{X});T_\varphi)$$

and some faithful topological MDS $(K,\mu;\psi)$.

Proof. The implications (iv)$\Rightarrow$(iii)$\Rightarrow$(i) are trivial, and (ii)$\Rightarrow$(iv) is Theorem 12.16.

For the implication (i)$\Rightarrow$(ii) let F be a T_φ-invariant Banach sublattice of $\mathrm{L}^1(\mathrm{X})$ containing $\mathbf{1}$. We claim that $A := F \cap \mathrm{L}^\infty$ satisfies the requirements. Indeed, A is certainly closed in L^∞ and T_φ-invariant as well as conjugation invariant. Moreover, it contains $\mathbf{1}$ and it is a vector sublattice of L^∞. Hence, by Theorem 7.18, A is also a subalgebra. It remains to show that A is L^1-dense in F. Since F is a sublattice, it suffices to show that each $0 \le f \in F$ can be approximated by elements of A. But this can be done simply by taking $f_n := f \wedge n\mathbf{1}$. So (ii), and hence the whole theorem is proved. □

Let F be a factor of the MDS $(\mathrm{X};\varphi)$. Then each embedding S as in (iii) of Theorem 13.25 is called a **model** of the factor. If the model is given as a faithful topological MDS $(K,\nu;\psi)$ as in (iv), then it is called a **topological model** of the factor. A straightforward way to obtain a model for F is to form the MDS $(\mathrm{X}_F;\varphi)$ on the probability space

$$\mathrm{X}_F := (X,\Sigma_F,\mu_\mathrm{X}), \qquad \Sigma_F = \big\{A \in \Sigma_\mathrm{X} : \mathbf{1}_A \in F\big\}.$$

However, this is in general not a topological model, even if the original system is topological.

Let $S : (\mathrm{L}^1(\mathrm{Y});T_\psi) \to (\mathrm{L}^1(\mathrm{X});T_\varphi)$ be a model of the factor $F = \mathrm{ran}(S)$, $S \in \mathrm{M}(\mathrm{X})$. Then the Markov adjoint $P := S'$ is a factor map satisfying

$$T'_\psi P T_\varphi = P.$$

Indeed, from $ST_\psi = T_\varphi S$ it follows $T'_\psi P = T'_\psi S' = S' T'_\varphi = P T'_\varphi$, and hence $T'_\psi P T_\varphi = P T'_\varphi T_\varphi = P$.

It is convenient to denote this situation in a short-hand way with

$$P : (\mathrm{X};\varphi) \to (\mathrm{Y};\psi), \tag{13.1}$$

understanding that P is actually an *operator* between the L^1-spaces and not a point map on the underlying sets. The factor F can be reconstructed from P via $F = \mathrm{ran}(P') = \mathrm{ran}(S)$ and the associated conditional expectation is $Q = PP'$. So all the essential information is contained in the factor map P, and it is common to abuse language and simply call the whole situation (13.1) a **factor**.

Examples 13.26. 1) If $\theta : (\mathrm{X};\varphi) \to (\mathrm{Y};\psi)$ is a point factor map (Definition 12.1) then its associated Koopman operator $T_\theta : (\mathrm{L}^1(\mathrm{Y}),T_\psi) \to (\mathrm{L}^1(\mathrm{X}),T_\varphi)$ is a

Markov embedding of MDSs, and hence constitutes a factor. The associated factor map is T'_θ, so we can denote this situation also with

$$T'_\theta : (\mathrm{X}; \varphi) \to (\mathrm{Y}; \psi)$$

by our convention above.

2) For each $(\mathrm{X}; \varphi)$, the fixed space $F = \mathrm{fix}(T_\varphi)$ is a factor (Theorem 8.8), and on each faithful topological model for it the dynamics is trivial. If the MDS is ergodic, i.e., $\dim \mathrm{fix}(T_\varphi) = 1$, then F has a one-point space $K = \{\emptyset\}$ as a topological model.

Exercise 5 shows that in Theorem 13.25 one can replace the role of $\mathrm{L}^1(\mathrm{X})$ by $\mathrm{L}^p(\mathrm{X})$ for any $1 \leq p < \infty$.

Remark 13.27. Suppose that the factor F is not only T_φ-invariant, but satisfies $T_\varphi(F) = F$. Then, equivalently, T_φ restricts to an isomorphism on F, and in the construction of a topological model one can assume $T_\varphi(A) = A$. Equivalently, the dynamics ψ on the representation space K is invertible. Hence the factors on which the dynamics is invertible are characterized by admitting invertible (topological) models.

If, actually, the MDS $(\mathrm{X}; \varphi)$ is itself invertible, then it is natural to consider *only* bi-invariant factors, i.e., factors that are also invariant under T_φ^{-1}. As seen above, these are characterised by admitting invertible models. In a context where one exclusively considers invertible MDSs it is hence convenient (and very common) to restrict the name "factor" to include the bi-invariance.

Supplement: Operator Theory on Hilbert Space

In this supplement we review some facts from elementary Hilbert space operator theory. In the following H and K denote Hilbert spaces with inner product denoted by $(\,\cdot\,|\,\cdot\,)$. The Hilbert space adjoint of a linear operator $T : K \to H$ is denoted by $T^* : H \to K$, see Appendix C.10.

We begin a useful lemma concerning the relation of weak and strong convergence in a Hilbert space.

Lemma 13.28. *For a net $(f_\alpha)_\alpha$ in a Hilbert space H and some $f \in H$ the following assertions are equivalent.*

(i) $\|f_\alpha - f\| \to 0$.

(ii) $f_\alpha \to f$ *weakly, and* $\|f_\alpha\| \to \|f\|$.

Proof. The equivalence follows from the standard identity

$$\|f_\alpha - f\|^2 = \|f_\alpha\|^2 + \|f\|^2 - 2\,\mathrm{Re}\,(f_\alpha \,|\, f) \qquad \square$$

Recall that an operator $T : H \to K$ is an **isometry** if $\|Tf\| = \|f\|$ for all $f \in H$. We obtain the following straightforward consequence of Lemma 13.28.

Corollary 13.29. *On the set*

$$\mathrm{Iso}(H) := \big\{T \in \mathscr{L}(H) : \|Tf\| = \|f\| \text{ for all } f \in H\big\}$$

of isometries, the weak and the strong operator topologies coincide.

Here is a useful characterisation of isometries.

Lemma 13.30. *For an operator $T \in \mathscr{L}(H;K)$ the following assertions are equivalent.*

(i) *T is an isometry, i.e., $\|Tf\| = \|f\|$ for all $f \in H$.*

(ii) *$(Tf \,|\, Tg) = (f \,|\, g)$ for all $f, g \in H$.*

(iii) *$T^*T = \mathrm{I}$.*

Proof. (i)$\Rightarrow$(ii): From the polarisation identity (see Section C.10) it follows that $\mathrm{Re}\,(Tf \,|\, Tg) = \mathrm{Re}\,(f \,|\, g)$ for all $f, g \in H$. Replacing g by ig we obtain (ii). The implications (ii)$\Rightarrow$(iii)$\Rightarrow$(i) are straightforward. □

Recall that an operator $P \in \mathscr{L}(H)$ is called an **orthogonal projection** if $P^2 = P$ and $\mathrm{ran}(P) \perp \mathrm{ran}(\mathrm{I} - P)$. Here is a characterisation.

Theorem 13.31. *Let H be a Hilbert space and $P \in \mathscr{L}(H)$. Then the following assertions are equivalent.*

(i) *$P = SS^*$ for some Hilbert space K and some linear isometry $S : K \to H$.*

(ii) *$P^2 = P$ and $\|P\| \leq 1$.*

(iii) *$\mathrm{ran}(\mathrm{I} - P) \perp \mathrm{ran}(P)$.*

(iv) *$P^2 = P = P^*$.*

(v) *$\mathrm{ran}(P)$ is closed, and $P = JJ^*$ with $J : \mathrm{ran}(P) \to H$ the canonical inclusion map.*

In this case, $\mathrm{ran}(S) = \mathrm{ran}(P)$ *in* (i).

Proof. (i)$\Rightarrow$(ii): By Lemma 13.30 above, $S^*S = \mathrm{I}$, whence $P^2 = (SS^*)(SS^*) = S(S^*S)S^* = SS^* = P$. Also, $\|P\| = \|S^*S\| \leq \|S^*\|\,\|S\| = 1$. Finally, $\mathrm{ran}(S) \subseteq \mathrm{ran}(P)$ is clear from $P = SS^*$, and the converse inclusion follows from $PS = SS^*S = S$.

(ii)$\Rightarrow$(iii): Let $f \in \mathrm{ran}(P)$, $g \in H$. Then for $c \in \mathbb{C}$ we have by (ii)

$$\|f - cPg\|^2 = \|P(f - cg)\|^2 \leq \|f - cg\|^2 .$$

Expanding the scalar products yields

$$2\,\mathrm{Re}\, c\,(f \,|\, g - Pg) \leq |c|^2\,(\|g\|^2 - \|Pg\|^2).$$

Now we specialise $c = t\overline{(f\,|\,g-Pg)}$ with $t>0$, and obtain

$$2t\big|(f\,|\,g-Pg)\big|^2 \le t^2\big|(f\,|\,g-Pg)\big|^2\big(\|g\|^2-\|Pg\|^2\big) \quad \text{for all } t>0,$$

and that implies $f \perp g-Pg$ as was to be proved.

(iii)$\Rightarrow$(iv): Let $f,g\in H$. Then by (iii)

$$\big\|Pf-P^2f\big\|^2 = (P(f-Pf)\,|\,(\mathrm{I}-P)Pf) = 0,$$

whence $P^2=P$. Moreover, by (ii) and symmetry

$$(Pf\,|\,g) = (Pf\,|\,g-Pg)+(Pf\,|\,Pg) = (Pf\,|\,Pg) = \cdots = (f\,|\,Pg)$$

which means that $P^*=P$.
(iv)$\Rightarrow$(v): Let $Q := JJ^*$. Since $\mathrm{ran}(J)=\mathrm{ran}(P)$ by construction, $PJ=J$, and hence $PQ=Q$ and then $QP=Q$ by (iv). On the other hand, since J is an isometry, $QJ = J(J^*J)=J$, whence $QP=P$. It follows that $Q=P$.

The implication (v)$\Rightarrow$(i) is trivial. □

Corollary 13.32. a) *For an orthogonal projection $P\in\mathscr{L}(H)$ and $f\in H$ we have*

$$Pf=f \quad\Leftrightarrow\quad \|Pf\|=\|f\|.$$

b) *For orthogonal projections $P,Q\in\mathscr{L}(H)$ we have*

$$\mathrm{ran}(P)\subseteq\mathrm{ran}(Q) \quad\Leftrightarrow\quad PQ=P \quad\Leftrightarrow\quad QP=P.$$

c) *Every orthogonal projection P on H is uniquely determined by its range* $\mathrm{ran}(P)$.

Proof. a) Since $\mathrm{ran}(\mathrm{I}-P)\perp\mathrm{ran}(P)$, by Pythagoras' lemma

$$\|f\|^2 = \|Pf\|^2+\|(\mathrm{I}-P)f\|^2$$

for every $f\in H$.

b) It is clear that $\mathrm{ran}(P)\subseteq\mathrm{ran}(Q)$ is equivalent with $QP=P$. But since Q and P are self-adjoint, the latter is equivalent with $P=P^*=(QP)^*=P^*Q^*=PQ$.

c) If P,Q are orthogonal projections on H with $\mathrm{ran}(P)=\mathrm{ran}(Q)$, then by b) we have $P=PQ=Q$. □

A self-adjoint operator $T\in\mathscr{L}(H)$ is called **positive semi-definite** if $(Tf\,|\,f)\ge 0$ for all $f\in H$. In this case, the mapping $(f,g)\to(Tf\,|\,g)$ defines a semi-inner product on H. The square of a self-adjoint operator is clearly positive semi-definite. More generally, if T is positive definite and $S\in\mathscr{L}(H)$, then STS^* is positive definite, too. If T is positive semi-definite, so is each power T^n. Moreover, each orthogonal projection is positive semi-definite since

$$(Pf\,|\,f) = \big(P^2 f\,\big|\,f\big) = (Pf\,|\,Pf) = \|Pf\|^2 \geq 0$$

for every $f \in H$.

Proposition 13.33. *Let $T \in \mathscr{L}(H)$ be a positive semi-definite contraction. Then*

$$Sf := \lim_{n\to\infty} T^n f$$

exists in H for each $f \in H$ and yields a positive semi-definite contraction S.

Proof. Since T is a contraction, we have

$$(Tf\,|\,f) \leq \|Tf\| \cdot \|f\| \leq (f\,|\,f)\,,$$

hence the operator $\mathrm{I} - T$ is positive semi-definite. From the identity

$$T(\mathrm{I}-T) = (\mathrm{I}-T)T(\mathrm{I}-T) + T(\mathrm{I}-T)T$$

and since T and $\mathrm{I}-T$ are self-adjoint, we conclude that $T(\mathrm{I}-T)$ is positive semi-definite.

By what is proved above, for $k \in \mathbb{N}_0$ and $f \in H$ we can write

$$\big(T^{2k+1} f\,\big|\,f\big) = \big(TT^k f\,\big|\,T^k f\big) \leq \big(T^k f\,\big|\,T^k f\big) = \big(T^{2k} f\,\big|\,f\big)\,,$$

and, since $T(\mathrm{I}-T)$ is positive semi-definite,

$$\big(T^{2k} f\,\big|\,f\big) = \big(T^2 T^{k-1} f\,\big|\,T^{k-1} f\big) \leq \big(TT^{k-1} f\,\big|\,T^{k-1} f\big) = \big(T^{2k-1} f\,\big|\,f\big)\,.$$

Hence for fixed $f \in H$ the sequence $((T^n f\,|\,f))_{n\in\mathbb{N}}$ decreasing and positive, so convergent. Moreover, since T is self-adjoint, we have

$$\|T^n f - T^m f\|^2 = \big(T^{2n} f\,\big|\,f\big) + \big(T^{2m} f\,\big|\,f\big) - 2\,\mathrm{Re}\,(T^{n+m} f\,|\,f) \to 0,$$

which shows that $(T^n f)_{n\in\mathbb{N}}$ is a Cauchy sequence. The properties of the limit operator S follow easily. □

If P, Q are orthogonal projections on H, then $F := \mathrm{ran}(P) \cap \mathrm{ran}(Q)$ is a closed subspace of H. Here is how to compute the orthogonal projection onto F.

Proposition 13.34. *Let $P, Q \in \mathscr{L}(H)$ be orthogonal projections. Then the strong limit*

$$Sf := \lim_{n\to\infty} (PQ)^n f$$

exists in H for every $f \in H$, and S is the orthogonal projection onto $\mathrm{ran}(P) \cap \mathrm{ran}(Q)$.

Proof. Let $f \in H$ and define $T := QPQ$. Then T is a positive semi-definite contraction and hence $Rf := \lim_{n\to\infty} T^n f$ exists by Proposition 13.33, and R is self-adjoint. By construction $QT^n = T^n Q = T^n$ for each $n \in \mathbb{N}$, whence $QR = RQ = R$ follows.

We can also conclude $QPR = R$. Now, since $(PQ)^n = PT^{n-1}$, we have that hence $(PQ)^n \to PR =: S$ strongly. Since P, Q, R are contractions, so is S. It also follows, that $(PQ)^{2n} \to S^2$, whence $S = S^2$. This shows that S is an orthogonal projection, and $RP = PR$. We have $PS = S = S^* = SP$ and $QS = QPR = QRP = RP = S$, and as a consequence $\mathrm{ran}(S) \subseteq \mathrm{ran}(P) \cap (Q)$. But if $f \in \mathrm{ran}(P) \cap \mathrm{ran}(Q)$, then $(QP)^n f = f$ for each f and hence $Sf = f$. It follows that $\mathrm{ran}(P) \cap \mathrm{ran}(Q) = \mathrm{ran}(S)$, which concludes the proof. □

Exercises

1. Prove the remaining statements of Theorem 13.2.

2. Prove Proposition 13.6.

3. For $T \in \mathscr{L}(H)$ a positive definite operator prove that its power T^n, $n \in \mathbb{N}$ are all positive definite.

4. Prove the claims made in Remark 13.22.

5. Let $(\mathrm{X}; \varphi)$ be an MDS, $1 \leq p < \infty$, and $E \subseteq \mathrm{L}^p(\mathrm{X})$ with $\mathbf{1} \in E$. Show that the following assertions are equivalent.

(i) $\mathrm{cl}_{\mathrm{L}^1}(E)$ is a factor of $(\mathrm{X}; \varphi)$.

(ii) E is a T_φ-invariant Banach sublattice of $\mathrm{L}^p(\mathrm{X})$.

(iii) $E = \mathrm{cl}_{\mathrm{L}^p}(A)$ for some T_φ-invariant C^*-subalgebra of $\mathrm{L}^\infty(\mathrm{X})$.

(iv) $E = \Phi(\mathrm{L}^p(\mathrm{Y}))$ for some embedding

$$\Phi : (\mathrm{L}^1(\mathrm{Y}); T_\psi) \to (\mathrm{L}^1(\mathrm{X}); T_\varphi)$$

of measure-preserving systems.

Chapter 14
Compact Semigroups and Groups

Compact groups were introduced already in Chapter 2 as nice examples of TDSs. In Chapter 5 they reappeared, endowed with their Haar measure, as simple examples of MDSs. As we shall see later in Chapter 17, these examples are by no means artificial: within a structure theory of general dynamical systems they form the fundamental building blocks.

Another reason to study groups (or even semigroups) is that they are present in the mere setting of general dynamical systems already. If $(K;\varphi)$ is an invertible TDS, then $\{\varphi^n : n \in \mathbb{Z}\}$ is an Abelian group of transformations of K and—passing to the Koopman operator—$\{T_\varphi^n : n \in \mathbb{Z}\}$ is a group of operators on $\mathrm{C}(K)$. Analogous remarks hold for MDSs, of course. If the dynamical system is not invertible, then one obtains *semi*groups instead of groups (see definition below), and indeed, many notions and results from Chapters 2 and 3 can be generalised to general (semi)group actions on compact spaces (see, e.g., [Gottschalk and Hedlund (1955)]).

In this chapter we shall present some facts about compact semigroups and groups. We concentrate on those that are relevant for a deeper analysis of dynamical systems, e.g., for the decomposition theorem to come in Chapter 16. Our treatment is therefore far from being complete and the reader is referred to [Hofmann and Mostert (1966)], [Hewitt and Ross (1979)], [Rudin (1990)] and [Hofmann and Morris (2006)] for further information.

14.1 Compact Semigroups

A **semigroup** is a nonempty set S with an operation $S \times S \to S$, $(s,t) \mapsto st := s \cdot t$—usually called *multiplication*—which is associative, i.e., one has $r(st) = (rs)t$ for all $r,s,t \in S$. If the multiplication is commutative (i.e., $st = ts$ for all $s,t \in S$), then the semigroup is called **Abelian**. An element e of a semigroup S is called an **idempotent** if $e^2 = e$, a **zero element** if $es = se = e$ for all $s \in S$, and a **neutral element** if $se = es = s$ for all $s \in S$. It is easy to see that, though there may be many idempotents, but at most one neutral/zero element in a semigroup. The semigroup S is called a

group if it has a neutral element e and for every $s \in S$ the equation $sx = xs = e$ has a solution x. It is easy to see that this solution is unique, called the **inverse** s^{-1} of s. (We assume the reader to be familiar with the fundamentals of group theory and refer, e.g., to [Lang (2005)].)

For subsets $A, B \subseteq S$ and elements $s \in S$ we write

$$sA = \{sa : a \in A\}, \quad As = \{as : a \in A\}, \quad AB := \{ab : a \in A, b \in B\}.$$

A **right ideal** of S is a *nonempty* subset J such that $JS \subseteq J$. Similarly one defines **left ideals**. A subset which is simultaneously a left and a right ideal is called a (two-sided) **ideal**. Obviously, in an Abelian semigroup this condition reduces to $JS \subseteq J$. A **subsemigroup** of S is a nonempty subset $H \subseteq S$ such that $HH \subseteq H$.

A right (left, two-sided) ideal of a semigroup S is called **minimal** if it is minimal with respect to set inclusion within the set of all right (left, two-sided) ideals. The intersection of all ideals

$$K(S) := \bigcap \{J : J \subseteq S,\ J \text{ ideal}\}$$

is called the **Sushkevich kernel** of S. It is either empty, or the (unique) minimal ideal. (Indeed, if J, I are minimal ideals of S, then $\emptyset \neq JI \subseteq I \cap J$ is an ideal, and hence $I = I \cap J = J$ by minimality.)

Minimal right (left) ideals are either disjoint or equal. An idempotent element in a minimal right (left) ideal is called **minimal idempotent** and has special properties.

Lemma 14.1 (Minimal Idempotents). *Let S be semigroup and $e \in S$ with $e^2 = e$. Then the following assertions are equivalent.*

(i) *eS (or Se) is a minimal right ideal (or left ideal).*

(ii) *e is contained in some minimal right ideal (left ideal).*

(iii) *eSe is a group (with neutral element e).*

In this case, e is minimal in the set of idempotents with respect to the ordering

$$p \leq q \quad \Leftrightarrow \quad pq = qp = p.$$

Proof. (i)$\Rightarrow$(ii) is trivial. For the proof of (ii)$\Rightarrow$(iii) let R be a minimal right ideal with $e \in R$, and left $G := eSe$. Clearly e is a neutral element in G. Let $s \in S$ be arbitrary. Then $eseR \subseteq R$ is also a right ideal, and by minimality $eseR = R$. In particular, there is $x \in R$ such that $esex = e$. It follows that $g := ese \in G$ has the right inverse $h := exe \in G$. Applying the same reasoning to x we find that also h has a right inverse, say k. But then $hg = hge = hghk = hk = e$, and hence h is also the left inverse of g.

(iii)$\Rightarrow$(i): Suppose that $G := eSe$ is a group, and let $R \subseteq eS$ be a right ideal. Take $r = es \in R$. Then there is $x \in S$ such that $e = (ese)(exe) = rexe \in R$. It follows $eS \subseteq R$, so $R = eS$.

Suppose f is an idempotent and $f \leq e$, i.e., $ef = fe = f = f^2$. Then $fS \subseteq eS$ and by minimality $fS = eS$. Hence there is $x \in S$ such that $fx = e^2 = e$. But then $f = fe = f^2x = fx = e$. □

As we are doing analysis, semigroups are interesting to us only if endowed with a topology which in some sense is related to the algebraic structure. One has different possibilities here.

Definition 14.2. A semigroup endowed with a topology is called a **left-topological semigroup** if for each $a \in S$ the *left multiplication* with a, i.e., the mapping

$$S \to S, \qquad s \mapsto as$$

is continuous. Similarly, S is called a **right-topological semigroup** if for each $a \in S$ the *right multiplication* with a, i.e., the mapping

$$S \to S, \qquad s \mapsto sa$$

is continuous. If both left and right multiplications are continuous, S is a **semitopological semigroup**. And S is a **topological semigroup** if the multiplication mapping

$$S \times S \to S, \qquad (s,t) \mapsto st$$

is continuous. A **topological group** is a topological semigroup that is algebraically a group and such that the inversion mapping $s \mapsto s^{-1}$ is continuous.

Clearly an Abelian semigroup is left-topological if and only if it is semitopological. In a later section we shall see that $\beta\mathbb{N}$ (familiar already from Chapter 4) is an example of a left-topological but not semitopological semigroup.

One says that in a semitopological semigroup the multiplication is *separately* continuous, whereas in a topological semigroup it is *jointly* continuous. Of course there are examples of semitopological semigroups that are not topological, i.e., such that the multiplication is not jointly continuous (see Exercise 3). The example particularly interesting to us is $\mathscr{L}(E)$, E a Banach space, endowed with the weak operator topology. By Proposition C.18 and Example C.18 this is a semitopological semigroup, which in general is not topological.

Any semigroup can be made into a topological one if endowed with the discrete topology. This indicates that such semigroups may be too general to study. However, **compact** left-topological semigroups (i.e., ones whose topology is compact) exhibit some amazing structure, which we shall study now.

Theorem 14.3 (Ellis). *In a compact left-topological semigroup, every right ideal contains a minimal right ideal. Every minimal right ideal is closed and contains at least one idempotent.*

Proof. Note that sS is a closed right ideal for any $s \in S$. If R is any right ideal and $x \in R$ then $xS \subseteq R$, and hence any right ideal contains a closed one. If R is even minimal, then we must have $R = xS$, and R is itself closed.

Now if J_0 is a given right ideal, then let $\mathscr{M}$ be the set of all closed right ideals of S contained in J_0. Then $\mathscr{M}$ is nonempty and partially ordered by set inclusion. Moreover, any chain $\mathscr{C}$ in $\mathscr{M}$ has a lower bound $\bigcap\mathscr{C}$, since this is nonempty by compactness. By Zorn's lemma, there is a minimal element R in $\mathscr{M}$. If $J \subseteq R$ is also a right ideal and $x \in J$, then $xS \subseteq J \subseteq R$, and xS is a closed right ideal. By construction $xS = R$, whence $J = R$, too.

Finally, we find by an application of Zorn's lemma as before a nonempty closed subsemigroup H of R which is minimal within all closed subsemigroups of R. If $e \in H$ then eH is a closed subsemigroup of R, contained in H, whence $eH = H$. The set $\{t \in H : et = e\}$ is then nonempty and closed, and a subsemigroup of R. By minimality, it must coincide with H, and hence contains e. This yields $e^2 = e$, concluding the proof. □

Lemma 14.4. *In a compact left-topological semigroup, the Sushkevich kernel satisfies*

$$K(S) = \bigcup\{R : R \text{ minimal right ideal}\} \neq \emptyset. \tag{14.1}$$

Moreover, an idempotent of S is minimal if and only if it is contained in $K(S)$.

Proof. To prove "$\supseteq$", let I be an ideal and J a minimal right ideal. Then $JI \subseteq J \cap I$, so $J \cap I$ is a nonempty right ideal. By minimality $J = J \cap I$, i.e., $J \subseteq I$. For "$\subseteq$" it suffices to show that the right-hand side of (14.1) is an ideal. Let R be any minimal right ideal, $x \in S$, and $R' \subseteq xR$ another right ideal. Then $\emptyset \neq \{y : xy \in R'\} \cap R$ is a right ideal, whence by minimality equal to R. But this means that $xR \subseteq R'$ and it follows that xR is also minimal, whence contained in the right-hand side of (14.1).

The remaining statement follows from (14.1) and Lemma 14.1. □

We can now state and prove the main result.

Theorem 14.5. *Let S be a compact semitopological semigroup. Then the following assertions are equivalent.*

(i) *Every minimal right ideal is a minimal left ideal.*

(ii) *There is a unique minimal right ideal and a unique minimal left ideal.*

(iii) *There is a unique minimal idempotent.*

(iv) *The Sushkevich kernel $K(S)$ is a group.*

The conditions (i)–(iv) *are satisfied in particular if S is Abelian.*

Proof. Clearly, (ii) implies (i), by Lemma 14.4 and symmetry.

(i)$\Rightarrow$(iv): Let R be a minimal right ideal and let $e \in R$ be an idempotent, which exists by Theorem 14.3. By hypothesis, R is also a left ideal, i.e., an ideal. Hence $K(S) \subseteq R \subseteq K(S)$ (by Lemma 14.4), yielding $R = K(S)$. Since R is minimal as a left ideal and a right ideal, $R = eS = Se$, from which it follows that $K(S) = R = eSe$, a group by Lemma 14.1, hence (iv).

(iv)$\Rightarrow$(iii): A minimal idempotent has to lie in $K(S)$ by Lemma 14.4, and so must coincide with the unique neutral element of $K(S)$.

(iii)$\Rightarrow$(ii): Since different minimal right ideals are disjoint, and each one contains an idempotent, there can be only one. By symmetry, the same is true for left ideals, whence (ii) follows. □

Suppose that S is a compact semitopological semigroup that satisfies the conditions (i)–(iv) of Theorem 14.5, so that $K(S)$ is a group. Then $K(S)$ is also compact, as it coincides with the unique right (left) ideal, closed by Theorem 14.3. The following celebrated result of Ellis (see [Ellis (1957)] or [Hindman and Strauss (1998), Section 2.5]) states that in this case the group $K(S)$ is already topological.

Theorem 14.6 (Ellis). *Let G be a semitopological group whose topology is locally compact. Then G is a topological group.*

As mentioned, combining Theorem 14.5 with Ellis' result we obtain the following.

Corollary 14.7. *Let S be a compact semitopological semigroup that satisfies the equivalent conditions of Theorem 14.5, e.g., suppose S Abelian. Then its minimal ideal $K(S)$ is a compact (topological) group.*

One can ask for which non-Abelian semigroups S the minimal ideal $K(S)$ is a group. One class of examples is given by the so-called **amenable** semigroups, for details see [Day (1957a)] or [Paterson (1988)]. Another one, and very important for operator theory, are the weakly closed semigroups of contractions on certain Banach spaces, see Section 16.1 below.

Proof of Ellis' Theorem in the Metrisable Case

The proof of Theorem 14.6 is fairly involved. The major difficulty is deducing joint continuity from separate continuity. By making use of the group structure it is enough to find *one single* point in $G \times G$ at which multiplication is continuous. This step can be achieved by applying a more general result of Namioka [Namioka (1974)] that asserts that under appropriate assumptions a separately continuous function has many points of joint continuity, see also [Kechris (1995)] or [Todorcevic (1997)]. For our purpose the following formulation suffices. Recall from Appendix A that in a Baire space, Baire's Category Theorem A.10 is valid.

Proposition 14.8. *Let Z, Y be metric spaces, X a Baire space and let $f : X \times Y \to Z$ be a separately continuous function. For every $b \in Y$, the set*

$$\big\{x \in X : \ f \text{ is not continuous at } (x,b)\big\}$$

is of first category in X.

Proof. a) Let $b \in Y$ be fixed. For $n,k \in \mathbb{N}$ define

$$X_{b,n,k} := \bigcap_{y \in \mathrm{B}(b,\frac{1}{k})} \big\{x \in X : d(f(x,y),f(x,b)) \le \tfrac{1}{n}\big\}.$$

By the continuity of f in the first variable, we obtain that the sets $X_{b,n,k}$ are closed. From the continuity of f in the second variable, we infer that for all $n \in \mathbb{N}$ the sets $X_{b,n,k}$ cover X, i.e.,

$$X = \bigcup_{k\in\mathbb{N}} X_{b,n,k} = \bigcup_{k\in\mathbb{N}} X^\circ_{b,n,k} \cup \bigcup_{k\in\mathbb{N}} \partial X_{b,n,k}.$$

(Here $X^\circ_{b,n,k}$ denotes the interior of $X_{b,n,k}$.) Since X is Baire space and $\partial X_{b,n,k}$ are nowhere dense, it follows by Theorem A.11 that the open set

$$\bigcup_{k\in\mathbb{N}} X^\circ_{b,n,k}$$

is dense in X. This yields, again by Theorem A.11, that

$$X_b := \bigcap_{n\in\mathbb{N}} \bigcup_{k\in\mathbb{N}} X^\circ_{b,n,k}$$

is a dense G_δ set. If $a \in X_b$, then for all $n \in \mathbb{N}$ there is $k \in \mathbb{N}$ with $a \in X^\circ_{b,n,k}$. So with an appropriate open neighbourhood U of a we even obtain $U \subseteq X_{b,n,k}$. Since f is continuous in the first variable, we can take the neighbourhood U such that $d(f(a,b),d(x,b)) \le \frac{1}{n}$ holds for all $x \in U$. Now, if $x \in U \subseteq X_{b,n,k}$ and $y \in \mathrm{B}(b,\frac{1}{k})$, we obtain

$$d(f(x,y),f(a,b)) \le d(f(x,y),f(x,b)) + d(f(x,b),f(a,b)) \le \frac{2}{n}.$$

Hence f is continuous at $(a,b) \in X \times Y$ if $a \in X_b$. Since X_b is a dense G_δ set, the assertion follows. □

Proof of Ellis' Theorem (compact, metrisable case). Apply Proposition 14.8 to the multiplication to obtain joint continuity at *one point* in $G \times G$. Then, by translation, the joint continuity *everywhere* follows.

We prove that the inversion mapping $g \mapsto g^{-1}$ is continuous on G. Let $g_n \to g$. It suffices to show that there is a subsequence of $(g_n^{-1})_{n\in\mathbb{N}}$ convergent to g^{-1}. By compactness, we find a subsequence $(g_{n_k}^{-1})_{k\in\mathbb{N}}$ convergent to some h as $k \to \infty$. By the joint continuity of the multiplication $1 = g_{n_k} g_{n_k}^{-1} \to gh$, whence $h = g^{-1}$. □

Remark 14.9. A completely different proof of Ellis' theorem in the compact (but not necessarily metrisable) case is based on *Grothendieck's theorem* about weak compactness in $\mathrm{C}(K)$-spaces. Roughly sketched, one first constructs Haar measure m on the compact semitopological group G, then represents G by the right regular representation on the Hilbert space $H = \mathrm{L}^2(G,\mathrm{m})$, and finally uses the fact that on the set of unitary operators on H the weak and strong operator topologies coincide. As multiplication is jointly continuous for the latter topology and the representation is faithful, G is a topological group. A detailed account can be found in Appendix E, see in particular Theorem E.12.

As we have seen, by passing to the minimal ideal we are naturally led from compact semitopological semigroups to compact *topological* groups. Their study is the subject of the next section.

14.2 Compact Groups

In the following, G always denotes a topological group with neutral element $1 = 1_G$. For a set $A \subseteq G$ we write $A^{-1} := \{a^{-1} : a \in A\}$ and call it **symmetric** if $A^{-1} = A$. Note that A is open if and only if A^{-1} is open since the inversion mapping is a homeomorphism of G. In particular, every open neighbourhood U of 1 contains a symmetric open neighbourhood of 1, namely $U \cap U^{-1}$.

Since the left and right multiplications by any given element $g \in G$ are homeomorphisms of G, the set of open neighbourhoods of the neutral element 1 completely determines the topology of G. Indeed, U is an open neighbourhood of 1 if and only if Ug (or gU) is an open neighbourhood of $g \in G$.

The following are the basic properties of topological groups needed later.

Lemma 14.10. *For a topological group G the following statements hold.*

a) *If V is an open neighbourhood of 1, then there is $1 \in W$ open, symmetric with $WW \subseteq V$.*

b) *If H is a topological group and $\varphi : G \to H$ is a group homomorphism, then φ is continuous if and only if it is continuous at 1_G.*

c) *If H is a normal subgroup of G, then the factor group G/H is a topological group with respect to the quotient topology.*

d) *If H is a (normal) subgroup of G, then $\overline{H}$ is also a (normal) subgroup of G.*

Proof. To prove a) we use that the multiplication is (jointly) continuous to find $1 \in U$ open with $UU \subseteq V$. Then $W := U \cap U^{-1}$ has the desired property. For b), c) and d) see Exercise 2 and Exercise 4. □

Recall that a topological group is called a (locally) **compact group** if its topology is (locally) compact. Important (locally) compact groups are the additive groups $\mathbb{R}^d$ and the (multiplicative) tori $\mathbb{T}^d$, $d \in \mathbb{N}$. Furthermore, every group G can be made into a locally compact group, denoted by G_d, by endowing it with the *discrete topology*.

Proposition 14.11. *Let G be a compact group, and let $f \in \mathrm{C}(G)$. Then f is* **uniformly continuous***, i.e., for any $\varepsilon > 0$ there is a symmetric open neighbourhood U of 1 such that $gh^{-1} \in U$ implies $|f(g) - f(h)| < \varepsilon$.*

Proof. Fix $\varepsilon > 0$. By continuity of f, for each $g \in G$ the set

$$V_g := \big\{z \in G : |f(zg) - f(g)| < \varepsilon/2\big\}$$

is an open neighbourhood of 1. By Lemma 14.10.a) there is a symmetric open neighbourhood W_g of 1 such that $W_g W_g \subseteq V_g$. Now $G = \bigcup_{g\in G} W_g g$, and hence by compactness $G = \bigcup_{g\in F} W_g g$ for some finite set $F \subseteq G$. Define

$$U := \bigcap_{g\in F} W_g,$$

which is a symmetric open neighbourhood of 1. For $xy^{-1} \in U$ we find $g \in F$ such that $y \in W_g g \subseteq V_g g$. Hence $x \in Uy \subseteq W_g W_g g \subseteq V_g g$. Therefore $xg^{-1}, yg^{-1} \in V_g$, which implies that

$$|f(x) - f(y)| \leq |f(x) - f(g)| + |f(y) - f(g)| < \varepsilon. \qquad \square$$

Proposition 14.11 yields a second proof for a fact already established in Proposition 10.11. Recall that for an element $a \in G$ we denote by L_a and R_a the Koopman operators associated with the left and right rotations by a, respectively. That is

$$(L_a f)(x) := f(ax), \quad (R_a f)(x) := f(xa) \qquad (x \in G, f : G \to \mathbb{C}).$$

Corollary 14.12. *Let G be a compact group and let $f \in \mathrm{C}(G)$. Then the mappings*

$$G \to \mathrm{C}(G), \quad a \mapsto L_a f \tag{14.2}$$
$$G \to \mathrm{C}(G), \quad a \mapsto R_a f \tag{14.3}$$

are continuous.

Proof. Let $\varepsilon > 0$ and choose U as in Proposition 14.11. If $ab^{-1} \in U$, then $(ax)(bx)^{-1} = ab^{-1} \in U$ for every $x \in G$. Hence $\|L_a f - L_b f\|_\infty = \sup_{x\in G} |f(ax) - f(bx)| \leq \varepsilon$. This proves the first statement. To prove the second, apply the first to the function $f^\sim(x) := f(x^{-1})$ and note that $x \mapsto x^{-1}$ is a homeomorphism of G. $\square$

We remark that Corollary 14.12 allows to prove that the operators L_a and R_a on $\mathrm{C}(G)$ are mean ergodic, see Corollary 10.12.

The Haar Measure

As explained in Chapter 5, on any compact group there is a unique rotation invariant (Baire) probability measure m, called the **Haar measure**. In this section, we shall prove this for the Abelian case; a proof in the general case is in Appendix E.4.

Theorem 14.13 (Haar Measure). *On a compact group G there is a unique left invariant Baire probability measure. This measure is also right invariant and inversion invariant and is strictly positive.*

Proof. It is easy to see that if $\mu \in \mathrm{M}^1(G)$ is left (right) invariant, then $\tilde{\mu} \in \mathrm{M}^1(G)$, defined by

$$\langle f, \tilde{\mu} \rangle = \int_G f(x^{-1})\,\mathrm{d}\mu(x) \qquad (f \in \mathrm{C}(G))$$

is right (left) invariant. On the other hand, if μ is left invariant and ν is right invariant, then by Fubini's theorem for $f \in \mathrm{C}(G)$

$$\begin{aligned}\int_G f(x)\,\mathrm{d}\mu(x) &= \int_G \int_G f(x)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) = \int_G \int_G f(xy)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) \\ &= \int_G \int_G f(xy)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x) = \int_G \int_G f(y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x) = \int_G f(y)\,\mathrm{d}\nu(y),\end{aligned}$$

and hence $\mu = \nu$. This proves uniqueness, and that every left invariant measure is also right invariant and inversion invariant.

Suppose that m is such an invariant probability measure. To see that it is strictly positive, let $0 \neq f \geq 0$. Then the sets $U_a := [L_a f > 0]$ cover G, hence by compactness there is a finite set $F \subseteq G$ such that $G \subseteq \bigcup_{a \in F} [L_a f > 0]$. Define $g := \sum_{a \in F} L_a f$ and $c := \inf_{x \in G} h(x) > 0$. Then $g \geq c\mathbf{1}$, whence

$$c = \int_G c\mathbf{1}\,\mathrm{dm} \leq \int_G g\,\mathrm{dm} = \sum_{a \in F} \int_K L_a f\,\mathrm{dm} = \sum_{a \in F} \int_K f\,\mathrm{dm}.$$

It follows that $\langle f, \mathrm{m} \rangle > 0$.

Only existence is left to show. As announced, we treat here only the case that G is Abelian. Then the family $(L'_a)_{a \in G}$ of adjoints of left rotations is a commuting family of continuous affine mappings on $\mathrm{M}^1(K)$, which is convex and weakly* compact, see Section 10.1. Hence the Markov–Kakutani Theorem 10.1 yields a common fixed point m, which is what we were looking for. □

Remark 14.14. A proof for the non-Abelian case can be based on a more refined fixed-point theorem, say of Ryll-Nardzewski, see [Ryll-Nardzewski (1967)] or [Dugundji and Granas (2003), §7.9]. Our proof in Appendix E.4 is based on the classical construction, see e.g. [Rudin (1991)]. However, we start there from slightly more general hypotheses in order to prove Ellis' theorem, see Theorem E.10.

14.3 The Character Group

An important tool in the study of locally compact Abelian groups is its character group. A **character** of a locally compact group G is a continuous homomorphism $\chi : G \to \mathbb{T}$. The set of characters

$$G^* := \{\chi : \chi \text{ character of } G\}$$

is an Abelian group with respect to pointwise multiplication and neutral element $\mathbf{1}$; it is called the **character group** (or **dual group**).

For many locally compact Abelian groups the dual group can be described explicitly. For example, the character group of $\mathbb{R}$ is given by

$$\mathbb{R}^* = \left\{ t \mapsto \mathrm{e}^{2\pi \mathrm{i} \alpha t} : \alpha \in \mathbb{R} \right\} \simeq \mathbb{R},$$

and the character group of $\mathbb{T} \simeq [0,1)$ by

$$\mathbb{T}^* = \{ z \mapsto z^n : n \in \mathbb{Z} \} \simeq \left\{ t \mapsto \mathrm{e}^{2\pi \mathrm{i} n t} : n \in \mathbb{Z} \right\} \simeq \mathbb{Z}$$

(see Examples 14.28 and Exercise 7). In these examples, the character group seems to be quite "rich", and an important result says that this is actually a general fact.

Theorem 14.15. *Let G be a locally compact Abelian group. Then its dual group G^* separates the points of G, i.e., for every $1 \neq a \in G$ there is a character $\chi \in G^*$ such that $\chi(a) \neq 1$.*

We shall not prove this theorem in its full generality, but shall treat (eventually) the two cases most interesting to us, namely that G is discrete and that G is compact. For discrete groups, the proof reduces to pure algebra, see Proposition 14.24 at the end of this section. The other case is treated in Chapter 16 where we develop the representation theory of compact groups to some extent. There we shall prove the more general fact that the finite-dimensional unitary representations of a compact group separate the points (Theorem 16.39). Theorem 14.15 for compact Abelian groups is then a straightforward consequence, see Theorem 16.43.

We proceed with exploring the consequences of Theorem 14.15. The proof of the first one is Exercise 5.

Corollary 14.16. *Let G be a locally compact Abelian group and let H be a closed subgroup of G. If $g \notin H$, then there is a character $\chi \in G^*$ with $\chi(g) \neq 1$ and $\chi|_H = 1$.*

We now turn to compact groups. For such groups one has $G^* \subseteq \mathrm{C}(G) \subseteq \mathrm{L}^2(G)$. The following is a preliminary result.

Proposition 14.17 (Orthogonality). *Let G be a compact Abelian group. Then G^* is an orthonormal set in $\mathrm{L}^2(G)$.*

Proof. Let $\chi_1, \chi_2 \in G^*$ be two characters and consider $\chi := \chi_1 \overline{\chi_2}$. Then

$$\alpha := (\, \chi_1 \,|\, \chi_2 \,) = \int_G \chi_1 \overline{\chi_2} \, \mathrm{dm} = \int_G \chi(g) \, \mathrm{dm}(g) = \int_G \chi(gh) \, \mathrm{dm}(g) = \chi(h) \alpha$$

for every $h \in G$. Hence either $\chi = \mathbf{1}$ (in which case $\chi_1 = \chi_2$) or $\alpha = 0$. □

By the orthogonality of the characters we obtain the following.

Proposition 14.18. *Let G be a compact Abelian group and let $X \subseteq G^*$ be a subset separating the points of G. Then the subgroup $\langle X \rangle$ generated by X is equal to G^*. Moreover, $\mathrm{lin}\langle X \rangle$ is dense in the Banach space $\mathrm{C}(G)$.*

Proof. Consider $A = \operatorname{lin}\langle X\rangle$ which is a conjugation invariant subalgebra(!) of $\mathrm{C}(G)$ separating the points of G. So by the Stone–Weierstraß Theorem 4.4 it is dense in $\mathrm{C}(G)$. If there is $\chi \in G^* \setminus \langle X\rangle$, take $f \in A$ with $\|\chi - f\|_\infty < 1$. Then by Proposition 14.17 χ is orthogonal to A and we obtain the following contradiction

$$1 > \|f-\chi\|_{\mathrm{L}^2}^2 = \|f\|_{\mathrm{L}^2}^2 - (f\,|\,\chi) - (\chi\,|\,f) + \|\chi\|_{\mathrm{L}^2}^2 = 1 + \|f\|_{\mathrm{L}^2}^2 \geq 1. \qquad \square$$

For a compact Abelian group, any linear combinations of characters is called a **trigonometric polynomial**.

Corollary 14.19. *For a compact Abelian group G the following statements hold.*

a) *The set* $\operatorname{lin} G^*$ *of trigonometric polynomials is dense in* $\mathrm{C}(G)$.

b) *The dual group G^* forms an orthonormal basis of* $\mathrm{L}^2(G)$.

Proof. a) follows from Proposition 14.18 with $X = G^*$, since G^* separates the points of G by Theorem 14.15. A fortiori, $\operatorname{lin} G^*$ is dense in $\mathrm{L}^2(G)$, and hence b) follows from Proposition 14.17. $\square$

Topology on the Character Group

Let G be a locally compact Abelian group. In this section we describe a topology turning G^* into a locally compact group as well.

Consider the product space $\mathbb{T}^G$ of all functions from G to $\mathbb{T}$ and endow it with the product topology (Appendix A.5). Then by Tychonov's Theorem A.5, this is a compact space. It is an easy exercise to show that it is actually a *compact group* with respect to the pointwise operations. The subspace topology on $G^* \subseteq \mathbb{T}^G$, called the **pointwise topology**, makes G^* a topological group. However, only in exceptional cases this topology is (locally) compact.

It turns out that a better choice is the *topology of uniform convergence on compact sets*, called the **compact-open topology**. This also turns G^* into a topological group, and unless otherwise specified, we always take this topology on G^*.

Examples 14.20. 1) If the group G is compact, the compact-open topology is the same as the topology inherited from the norm topology on $\mathrm{C}(G)$.

2) If the group G is discrete, then the compact-open topology is just the pointwise topology.

One has the following nice duality.

Proposition 14.21. *Let G be a locally compact Abelian group. If G is compact, then G^* is discrete; and if G is discrete, then G^* is compact.*

Proof. Suppose that G is compact. Take $\chi \in G^*$, then $\chi(G) \subseteq \mathbb{T}$ is a compact subgroup of $\mathbb{T}$. Therefore, if χ satisfies

$$\|\mathbf{1}-\chi\|_\infty = \sup_{g\in G}|1-\chi(g)| < \operatorname{dist}(1,\mathrm{e}^{2\pi i/3}) = \sqrt{3}$$

then we must have $\chi(G)=\{1\}$, and hence $\chi=\mathbf{1}$. Consequently, $\{\mathbf{1}\}$ is an open neighbourhood of $\mathbf{1}$, and G^* is discrete.

Suppose that G is discrete. Then by Example 14.20.2 the dual group G^* carries the pointwise topology. But it is clear that a pointwise limit of homomorphims is again a homomorphism. This means that G^* is closed in the compact product space $\mathbb{T}^G$, hence compact. □

It is actually true that G^*, endowed with the compact-open topology, is a locally compact Abelian group whenever G is a locally compact Abelian group, see [Hewitt and Ross (1979), §23].

The next is an auxiliary result, whose proof is left as Exercise 9. Note that an algebraic isomorphism $\Phi: G\to H$ between topological groups G,H is a **topological isomorphism** if Φ and Φ^{-1} are continuous.

Proposition 14.22. *Let G and H be locally compact Abelian groups. Then the product group $G\times H$ with the product topology is a locally compact topological group. For any $\chi\in G^*$ and $\psi\in H^*$ we have $\chi\otimes\psi\in(G\times H)^*$, where $\chi\otimes\psi(x,y):=\chi(x)\psi(y)$. Moreover, the mapping*

$$\Psi: G^*\times H^*\to (G\times H)^*,\quad (\chi,\psi)\mapsto \chi\otimes\psi$$

is a topological isomorphism.

Supplement: Characters of Discrete Abelian Groups

In this supplement we shall examine the character group of a discrete Abelian group. This is, of course, pure algebra. Our first result states, in algebraic terms, that the group $\mathbb{T}$ is an injective $\mathbb{Z}$-module.

Proposition 14.23. *Let G be an Abelian group, $H\subseteq G$ a subgroup, and let $\psi: H\to\mathbb{T}$ be a homomorphism. Then there is a homomorphism $\chi: G\to\mathbb{T}$ extending ψ.*

Proof. Consider the following collection of pairs

$$\mathscr{M} := \big\{(K,\varphi)\,:\, K\subseteq G \text{ subgroup}, \varphi: K\to\mathbb{T} \text{ homomorphism}, \varphi|_H=\psi\big\}.$$

This set is partially ordered by the following relation: $(K_1,\varphi_1)\le(K_2,\varphi_2)$ if and only if $K_1\subseteq K_2$ and $\varphi_2|_{K_1}=\varphi_1$. By Zorn's lemma, as the chain condition is clearly satisfied, there is a maximal element $(K,\varphi)\in\mathscr{M}$. We prove that $K=G$.

If $x\in G\setminus K$, then we construct an extension of φ to $K_1:=\langle K\cup\{x\}\rangle$ contradicting maximality. We have $K_1=\{x^n h\,:\, n\in\mathbb{Z}, h\in K\}$. If no power x^n for $n\ge 2$ belongs to K, then the mapping

$$\varphi_1 : K_1 \to \mathbb{T}, \qquad \varphi_1(x^i h) := \varphi(h), \qquad (i \in \mathbb{Z}, h \in K)$$

is a well-defined group homomorphism. If $x^n \in K$ for some $n \geq 2$, then take the smallest such n. Let $\alpha \in \mathbb{T}$ be some n^{th} root of $\varphi(x^n)$ and set $\varphi_1(x^i h) := \alpha^i \varphi(h)$ for $i \in \mathbb{Z}$, $h \in K$. This mapping is well-defined: If $x^i h = x^j h'$ and $i > j$, then $x^{i-j} h = h' \in K$ and $x^{i-j} \in K$, so $\varphi(h') = \varphi(x^{i-j} h) = \varphi(x^{i-j})\varphi(h)$. By assumption, n divides $i - j$, so $\varphi(x^{i-j} h) = \alpha^{i-j}\varphi(h)$. Of course, $\varphi_1 : K_1 \to \mathbb{T}$ is a homomorphism extending φ (hence ψ), so $(K_1, \varphi_1) \in \mathscr{M}$ with $(K_1, \varphi_1) > (K, \varphi)$, a contradiction. □

As a corollary we obtain the version of Theorem 14.15 for the discrete case.

Proposition 14.24. *If G is a (discrete) Abelian group, then the characters separate the points of G.*

Proof. Let $x \in G$, $x \neq 1$. If there is $n \geq 2$ with $x^n = 1$, then take $1 \neq \alpha \in \mathbb{T}$ an n^{th} root of unity. If the order of x is infinite, then take any $1 \neq \alpha \in \mathbb{T}$. For $k \in \mathbb{Z}$ set $\psi(x^k) = \alpha^k$. Then ψ is a nontrivial character of $\langle x \rangle$. By Proposition 14.23 we can extend it to a character χ on G that separates 1 and x. □

14.4 The Pontryagin Duality Theorem

For a locally compact Abelian group G consider the compact space $\mathbb{T}^{G^*}$ which endowed with the pointwise multiplication is a compact group. Define the mapping

$$\Phi : G \to \mathbb{T}^{G^*}, \quad \Phi(g)\chi = \chi(g). \tag{14.4}$$

It is easy to see that Φ is a continuous homomorphism, and by Theorem 14.15 it is injective. By Exercise 10, $\Phi(g)$ is actually a character of the dual group G^* for any $g \in G$.

Lemma 14.25. *The mapping*

$$\Phi : G \to G^{**} := (G^*)^*$$

is continuous.

We prove this lemma for discrete or compact groups only. The general case requires some more efforts, see [Hewitt and Ross (1979), §23].

Proof. If G is discrete, then Φ is trivially a continuous mapping. If G is compact, then G^* is discrete. So the compact sets in G^* are just the finite ones, and the topology on G^{**} is the topology of pointwise convergence. So Φ is continuous. □

The following fundamental theorem of Pontryagin asserts that G is topologically isomorphic to its **bi-dual group** G^{**} under the mapping Φ. Again, we prove this theorem only for discrete or compact groups; for the general case see [Hewitt and Ross (1979), Theorem 24.8] or [Folland (1995), (4.31)].

Theorem 14.26 (Pontryagin Duality Theorem). *For a locally compact Abelian group G the mapping $\Phi : G \to G^{**}$ defined in* (14.4) *is a topological isomorphism.*

Proof. It remains to prove that Φ is surjective and its inverse in continuous.

First, suppose that G is discrete. Then, by Proposition 14.21, G^* is compact and the subgroup $\operatorname{ran}\Phi \subseteq G^{**}$ clearly separates the points of G^*. By Proposition 14.18 we have $\operatorname{ran}\Phi = G^{**}$. Since both G and G^{**} are discrete, the mapping Φ is a homeomorphism.

Let now G be compact. Then G^* is discrete and G^{**} is compact by Proposition 14.21. Since Φ is continuous and injective, $\operatorname{ran}\Phi$ is a compact, hence closed subgroup of G^{**}. If $\operatorname{ran}\Phi \neq G^{**}$, then by Corollary 14.16 there is a character $\gamma \neq \mathbf{1}$ of G^{**} with $\gamma|_{\operatorname{ran}\Phi} = 1$. Now, since G^* is discrete, by the first part of the proof we see that there is $\chi \in G^*$ with $\gamma(\varphi) = \varphi(\chi)$ for all $\varphi \in G^{**}$. In particular, for $g \in G$ and $\varphi = \Phi(g)$ we have $\chi(g) = \Phi(g)\chi = \gamma(\Phi(g)) = 1$, a contradiction. Since G is compact and Φ continuous onto G^{**}, it is actually a homeomorphism (see Appendix A.7). □

An important message of Pontryagin's Theorem is that the dual group determines the group itself. This is stated as follows.

Corollary 14.27. *Two locally compact Abelian groups are topologically isomorphic if and only if their duals are topologically isomorphic.*

Examples 14.28. 1) For $n \in \mathbb{Z}$ define

$$\chi_n : \mathbb{T} \to \mathbb{T}, \qquad \chi_n(z) := z^n \qquad (z \in \mathbb{T}).$$

Clearly, each χ_n is a charachter of $\mathbb{T}$, and $(n \mapsto \chi_n)$ is a (continuous) homomorphism of $\mathbb{Z}$ into $\mathbb{T}^*$. Since χ_1 separates the points of $\mathbb{T}$, Proposition 14.18 tells that this homomorphism is surjective, i.e.,

$$\mathbb{T}^* = \{\chi_n : n \in \mathbb{Z}\} \simeq \mathbb{Z}.$$

2) By inductive application of Proposition 14.22 and by part 1) we obtain

$$(\mathbb{T}^d)^* = \{(z \mapsto z_1^{k_1} z_2^{k_2} \cdots z_d^{k_d}) : k_i \in \mathbb{Z},\ i = 1, \ldots, d\} \simeq \mathbb{Z}^d.$$

3) The dual group $\mathbb{Z}^*$ is topologically isomorphic to $\mathbb{T}$. This follows from part 1) and Pontryagin's Theorem 14.26.

4) The mapping $\Psi : \mathbb{R} \to \mathbb{R}^*$, $\Psi(\alpha) = (t \mapsto \mathrm{e}^{2\pi \mathrm{i} \alpha t})$ is a topological isomorphism. We leave the proof as Exercise 7.

Let us return to the mapping Φ from G into $\mathbb{T}^{G^*}$ defined in (14.4). Since this latter space is compact, the closure $\mathrm{b}G$ of $\operatorname{ran}\Phi$ in $\mathbb{T}^{G^*}$ is compact. Then $\mathrm{b}G$ is a compact group (see Exercise 2.b) and G is densely and continuously embedded in $\mathrm{b}G$ by Φ; $\mathrm{b}G$ is called the **Bohr-compactification** of G.

Proposition 14.29. *Let G be a locally compact Abelian group and* bG *its Bohr-compactification. Then the following assertions hold.*

a) *If G is compact, then*
$$\mathrm{b}G = (G^*)^* \simeq G.$$

b) *The dual group of* bG *is topologically isomorphic to the dual group of G, but endowed with the discrete topology, i.e.,*
$$(\mathrm{b}G)^* \simeq (G^*)_{\mathrm{d}}.$$

c) *The Bohr-compactification* bG *is obtained by taking the dual of G, endowing it with the discrete topology and taking the dual again, i.e.,*
$$\mathrm{b}G \simeq (G^*)^*_{\mathrm{d}}.$$

Proof. a) Since Φ is continuous, $\Phi(G)$ is compact, and hence closed in $\mathbb{T}^{G^*}$. So $\mathrm{b}G = \Phi(G) = (G^*)^*$.
b) For $\chi \in G^*$ the projection $\pi_\chi : \mathbb{T}^{G^*} \to \mathbb{T}$, $\pi_\chi(\gamma) = \gamma(\chi)$, is a character of $\mathbb{T}^{G^*}$ and hence of bG, by restriction. We claim that the mapping
$$(G^*)_{\mathrm{d}} \to (\mathrm{b}G)^*, \qquad \chi \mapsto \pi_\chi$$
is a (topological) isomorphism. Since both spaces carry the discrete topology, it suffices to show that it is an algebraic isomorphism. Now, if $\chi, \chi' \in G^*$ and $g \in G$, then
$$\begin{aligned}(\pi_\chi \pi_{\chi'})(\Phi(g)) &= \pi_\chi(\Phi(g)) \cdot \pi_{\chi'}(\Phi(g)) = \Phi(g)\chi \cdot \Phi(g)\chi' = \chi(g) \cdot \chi'(g) \\ &= (\chi\chi')(g) = \Phi(g)(\chi\chi') = \pi_{\chi\chi'}(\Phi(g)).\end{aligned}$$
Hence, $\pi_\chi \pi_{\chi'} = \pi_{\chi\chi'}$ on $\Phi(G)$, and therefore on bG, by continuity. If $\pi_\chi = \mathbf{1}$, then
$$1 = \pi_\chi(\Phi(g)) = \cdots = \chi(g) \quad \text{for all } g \in G,$$
whence $\chi = \mathbf{1}$. Finally, note that $X := \{\pi_\chi : \chi \in G^*\}$ is a subgroup of $(\mathrm{b}G)^*$ that separates the points (trivially). Hence, by Proposition 14.18 it must be all of $(\mathrm{b}G)^*$.
c) follows from b) and Pontryagin's Theorem 14.26. □

14.5 Application: Ergodic Rotations and Kronecker's Theorem

We now use the duality theory of compact/discrete groups in order to characterise dynamical properties of group rotations.

Theorem 14.30. *Let G be a compact Abelian group with Haar measure* m*, let $a \in G$ and consider the rotation MDS $(G, \mathrm{m}; a)$ with Koopman operator T on $\mathrm{L}^2(G)$. Then the peripheral point spectrum of T is*

$$\sigma_p(T) = \{\chi(a) : \chi \in G^*\}.$$

Moreover the following assertions are equivalent.

(i) *The rotation system* $(G, \mathrm{m}; a)$ *is ergodic.*

(ii) $\chi(a) = 1$ *implies* $\chi = \mathbf{1}$ *for each* $\chi \in G^*$*, i.e.,* $\{a\}$ *separates* G^*.

(iii) *The rotation system* $(G; a)$ *is topologically transitive/minimal.*

Proof. If $\chi \in G^*$, then $T\chi = \chi(a)\chi$, whence $\chi(a) \in \sigma_p(T)$. Conversely, suppose that $\lambda \in \sigma_p(T)$ and $f \in \mathrm{L}^2(G)$ is such that $Tf = \lambda f$. By Corollary 14.19 we can write $f = \sum_{\chi \in G^*} (f \,|\, \chi)\,\chi$, apply T and obtain

$$\sum_{\chi \in G^*} \lambda\,(f \,|\, \chi)\,\chi = \lambda f = Tf = \sum_{\chi \in G^*} \chi(a)\,(f \,|\, \chi)\,\chi.$$

Therefore, $(\lambda - \chi(a))\,(f \,|\, \chi) = 0$ for all $\chi \in G^*$. If $f \neq 0$ then at least one $(f \,|\, \chi) \neq 0$, whence $\chi(a) = \lambda$.

Suppose that (ii) holds. Then if $Tf = f$ it follows from the previous argument with $\lambda = 1$ that $f \perp \chi$ for each $\chi \neq \mathbf{1}$. Hence $\dim \mathrm{fix}(T) = 1$, i.e., (i). On the other hand if (i) holds and $\chi(a) = 1$, then $\chi \in \mathrm{fix}(T)$, whence $\chi = \mathbf{1}$. This proves (ii).

By Theorem 10.13 (i) is equivalent to (iii). □

We now are in the position to give the hitherto postponed proof of Kronecker's theorem (Theorem 2.35) Actually we present two proofs: One exploiting the fact that the characters form an orthonormal basis and another making use of the Bohr-compactification.

Theorem 14.31 (Kronecker). *For* $a = (a_1, \ldots, a_d) \in \mathbb{T}^d$ *the rotation system* $(\mathbb{T}^d; a)$ *is topologically transitive (=minimal) if and only if* $a_1, a_2, \ldots, a_d$ *are linearly independent in the* $\mathbb{Z}$*-module* $\mathbb{T}$*, i.e.,*

$$k_1, k_2, \ldots, k_d \in \mathbb{Z}, \quad a_1^{k_1} a_2^{k_2} \cdots a_n^{k_d} = 1 \quad \Rightarrow \quad k_1 = k_2 = \cdots = k_d = 0.$$

First Proof. By Example 14.28.2) each character of $\mathbb{T}^d$ has the form $z \mapsto z^{k_1} \cdots z^{k_d}$. Hence the $\mathbb{Z}$-independence of $a_1, \ldots, a_d$ is precisely assertion (ii) in Theorem 14.30 for $G = \mathbb{T}^d$. □

Second Proof. If $a_1^{k_1} a_2^{k_2} \cdots a_n^{k_d} = 1$ for $0 \neq (k_1, k_2, \ldots, k_d) \in \mathbb{Z}^d$, then the set

$$A := \left\{x \in \mathbb{T}^d : x_1^{k_1} x_2^{k_2} \cdots x_d^{k_d} = 1\right\}$$

is nonempty, closed, and invariant under the rotation by a. If $k_i \neq 0$, there are exactly k_i points in A of the form $(1, \ldots, 1, x_i, 1, \ldots, 1)$, whence $A \neq \mathbb{T}^d$. Hence, the TDS $(\mathbb{T}^d; a)$ is not minimal.

For the converse implication suppose that $a_1, a_2, \ldots, a_d$ are linearly independent in the $\mathbb{Z}$-module $\mathbb{T}$, and recall from Example 14.28 that $\mathbb{Z}^*$ and $\mathbb{T}$ are topologically isomorphic, so we may identify them.

Let $b := (b_1, \dots, b_d) \in \mathbb{T}^d$, and consider $G = \langle a_1, a_2, \dots, a_d \rangle$. Since $\{a_1, \dots, a_n\}$ is linearly independent in the $\mathbb{Z}$-module (= Abelian group) $\mathbb{T}$, there exists a $\mathbb{Z}$-linear mapping (= group homomorphism)

$$\psi : G \to \mathbb{T} \quad \text{with } \psi(a_i) = b_i \text{ for } i = 1, \dots, D.$$

By Proposition 14.23 below, there is a group homomorphism $\chi : \mathbb{T} \to \mathbb{T}$ extending ψ. Obviously $\chi \in (\mathbb{T}_\mathrm{d})^*$, which by Proposition 14.29 is topologically isomorphic to $\mathrm{b}\mathbb{Z}$. Now take $\varepsilon > 0$. By the definition of $\mathrm{b}\mathbb{Z}$ and its topology, we find $k \in \mathbb{Z}$ such that

$$|b_i - a_i^k| = \big|\chi(a_i) - \Phi(k)(a_i)| < \varepsilon$$

for $i = 1, \dots, n$ (recall that $\Phi(k)(z) = z^k$ for $z \in \mathbb{T} \simeq \mathbb{Z}^*$). Therefore $\mathrm{orb}(1) = \{(a_1^k, a_2^k, \dots, a_d^k) : k \in \mathbb{Z}\}$ is dense in $\mathbb{T}^d$. □

Exercises

1. Show that in a compact left-topological semigroup S an element $s \in S$ is idempotent if and only if it generates a minimal closed subsemigroup.

2. a) Let S be a semitopological semigroup and let $H \subseteq S$ be a subsemigroup. Show that if H is a subsemigroup (ideal) then $\overline{H}$ is a subsemigroup (ideal), too. Show that if H is Abelian then also $\overline{H}$ is.

b) Show that if G is a topological group and H is a (normal) subgroup of G then $\overline{H}$ is a (normal) subgroup as well.

3. Consider $S := \mathbb{R} \cup \{\infty\}$, the one-point compactification of $\mathbb{R}$, and define thereon the addition $t + \infty := \infty + t := \infty$, $\infty + \infty := \infty$ for $t \in \mathbb{R}$. Show that S is a compact semitopological but not a topological semigroup. Determine the minimal ideal.

4. Prove assertions b), c) in Lemma 14.10. Prove also the following generalisation of Proposition 14.11: If G and H be topological groups, G is compact and $f : G \to H$ is continuous, then f is even uniformly continuous in the sense that for any V open neighbourhood of 1_H there is an open neighbourhood U of 1_G such that $gh^{-1} \in U$ implies $f(g)f(h)^{-1} \in V$.

5. Prove Corollary 14.16 for

a) G a compact or discrete Abelian group,

*b) G a general locally compact Abelian group.

(Hint: Prove that the factor group G/H is a locally compact (in particular Hausdorff) Abelian group and apply Theorem 14.15 to this group.)

6. Let G be a topological group and $H \subseteq G$ an open topological subgroup. Show that H is closed. Show that the connected component G_0 of $1 \in G$ is a closed normal subgroup.

7. Show that the dual group of $\mathbb{R}$ is as claimed in Section 14.3.

8. Determine the dual of the group $\mathbb{A}_2$ of dyadic integers. (Hint: Consider the TDS from Exercise 3.8. Show that $\varphi^{2^n}(0) \to 0$, $n \to \infty$ (0 denotes the neutral element in $\mathbb{A}_2$). For $\chi \in \mathbb{A}_2^*$, set $a = \chi(\mathbf{1})$ and prove that $a^{2^n} \to 1$ as $n \to \infty$.)

9. Prove Proposition 14.22. (Hint: To see surjectivity of Ψ, for given $\varphi(G \times H)^*$ define the characters $\chi(g) = \varphi(g, 1_H)$ and $\psi(h) = \varphi(1_G, h)$ for $g \in G$ and $h \in H$.)

10. For $\Phi : G \to \mathbb{T}^{G^*}$ defined in (14.4) prove the following assertions.

a) Φ maps into G^{**}.

b) For general locally compact Abelian groups $\Phi : G \to G^{**}$ is continuous.

Chapter 15
Topological Dynamics Revisited

15.1 The Čech–Stone Compactification

15.2 The Space of Ultrafilters

15.3 Multiple Recurrence

15.4 Applications to Ramsey Theory

Exercises

Chapter 16
The Jacobs–de Leeuw–Glicksberg Decomposition

In what follows we shall apply the results of Chapter 14 to compact semigroups of *bounded linear operators* on a Banach space E. We recall from Appendix C.8 the following definitions: The **strong operator topology** on $\mathscr{L}(E)$ is the coarsest topology that renders all evaluation mappings

$$\mathscr{L}(E) \to E, \qquad T \mapsto Tx \qquad (x \in E)$$

continuous. The **weak operator topology** is the coarsest topology that renders all mappings

$$\mathscr{L}(E) \to \mathbb{C}, \qquad T \mapsto \langle Tx, x' \rangle \qquad (x \in E, x' \in E')$$

continuous, where E' is the dual space of E. Note that if $E = H$ is a Hilbert space, in the latter definition we could have written

$$\mathscr{L}(H) \to \mathbb{C}, \qquad T \mapsto (Tx \,|\, y) \qquad (x, y \in H),$$

these two properties being equivalent by virtue of the Riesz–Fréchet Theorem C.20.

We write $\mathscr{L}_s(E)$ and $\mathscr{L}_w(E)$ to denote the space $\mathscr{L}(E)$ endowed with the strong and the weak operator topology, respectively. To simplify terminology, we shall speak of weakly/strongly closed, open, relatively compact, compact etc. sets when we intend the weak/strong operator topology.[2] In order to render the notation more feasible, we shall write $\mathrm{cl}_w \mathscr{T}$, $\mathrm{cl}_s \mathscr{T}$ for the closure of a set $\mathscr{T} \subseteq \mathscr{L}(E)$ with respect to the weak/strong operator topology. For a subset $A \subseteq E$ we shall write $\mathrm{cl}_\sigma A$ to denote its closure in the $\sigma(E, E')$-topology.

Note that a subset $\mathscr{T} \subseteq \mathscr{L}(E)$ is a semigroup (with respect to operator multiplication) if it is closed under this operation. The operator multiplication is separately continuous with respect to both the strong and the weak operator topologies (cf. Appendix C.8). Hence—in the terminology of Chapter 14—both $\mathscr{L}_s(E)$ and $\mathscr{L}_w(E)$ are semitopological semigroups with respect to the operator multiplication. The set

[2] This should not be confused with the corresponding notion for the weak topology of $\mathscr{L}(E)$ as a Banach space, a topology which will not appear in this book.

of *contractions*

$$\mathscr{B}_1 = \mathscr{B}_1(E) := \big\{T \in \mathscr{L}(E) : \|T\| \le 1\big\}$$

is a closed subsemigroup (in both topologies); and this subsemigroup has jointly continuous multiplication in the strong operator topology. In the following we shall concentrate mainly on $\mathscr{L}_w(E)$.

16.1 Weakly Compact Operator Semigroups

Our main tool in identifying (relatively) weakly compact subsets of operators is the following result.

Theorem 16.1. *Let E be a Hilbert space, or, more general, a reflexive Banach space. Then for each $M \ge 0$ the closed norm-ball*

$$\mathscr{B}_M(E) := \big\{T \in \mathscr{L}(E) : \|T\| \le M\big\}$$

is compact in the weak operator topology.

Proof. We shall give the proof in the Hilbert space case and leave the general case to Exercise 1. Without loss of generality we may take $M = 1$. Let $\mathrm{B}_H := \{x \in H : \|x\| \le 1\}$ be the closed unit ball of H. Consider the injective mapping

$$\Phi : \mathscr{B}_1 \to \prod_{x,y\in \mathrm{B}_H} \mathbb{C}, \qquad \Phi(T) := \big((Tx|y)\big)_{x,y\in \mathrm{B}_H}.$$

We endow the target space with the product topology (Appendix A.5), and $\mathscr{B}_1$ with the weak operator topology, so Φ is a homeomorphism onto its image $\Phi(\mathscr{B}_1)$. We claim that this image is actually closed. Indeed, suppose $(T_\alpha)_\alpha$ to be a net in $\mathscr{B}_1$ such that

$$c(x,y) := \lim_\alpha (T_\alpha x|y) \quad \text{exists for all } x,y \in \mathrm{B}_H.$$

Then this limit actually exists for *all* $x,y \in H$, the mapping $c : H \times H \to \mathbb{C}$ is sesquilinear and $|c(x,y)| \le M\|x\|\,\|y\|$ for all $x,y \in H$. By Corollary C.22 there is $T \in \mathscr{L}(H)$ such that $c(x,y) = (Tx|y)$ for all $x,y \in H$. It follows that $\|T\| \le M$, and hence $T \in \mathscr{B}_1$ and $\Phi(T) = \lim_\alpha \Phi(T_\alpha)$.

Finally, note that $\Phi(\mathscr{B}_1) \subseteq \prod_{x,y\in \mathrm{B}_H}\{z \in \mathbb{C} : |z| \le 1\}$, which is compact by Tychonoff's theorem A.5. Since $\Phi(\mathscr{B}_1)$ is closed, it is compact as well, and since Φ is a homeomorphism onto its image, $\mathscr{B}_1$ is compact as claimed. □

If we restrict to the case $M = 1$ in the previous result, we obtain the following.

Corollary 16.2. *Let E be a Hilbert space or, more general, a reflexive Banach space. Then the set of contractions $\mathscr{B}_1 := \{T \in \mathscr{L}(E) : \|T\| \le 1\}$ is a compact semitopological semigroup in the weak operator topology, with unit element* I.

Let H be a Hilbert space. Recall from Corollary 13.29 that on the set

$$\mathrm{Iso}(H) = \big\{T \in \mathscr{L}(H) : \ T \text{ is an isometry}\big\} \subseteq \mathscr{B}_1(H)$$

the weak and the strong operator topologies coincide. Since the operator multiplication is jointly continuous on $\mathscr{B}_1(H)$ for the strong operator topology, we obtain that $\mathrm{Iso}(H)$ is a *topological* semigroup and the subgroup of **unitary** operators

$$\mathrm{U}(H) := \big\{T \in \mathscr{L}(H) : T \text{ is unitary}\big\}$$

is a topological group (with identity I) with respect to the weak (=strong) operator topology. Hence we obtain almost for free the following special case of Ellis' Theorem 14.6.

Theorem 16.3. *Let $\mathscr{S} \subseteq \mathscr{B}_1(H)$ be a weakly compact subsemigroup of contractions on a Hilbert space H. If $\mathscr{S}$ is algebraically a group, then it is a compact topological group, and the strong and weak operator topologies coincide on $\mathscr{S}$.*

Proof. Let P be the unit element of $\mathscr{S}$. Then $P^2 = P$ and $\|P\| \leq 1$, hence P is an orthogonal projection onto the Hilbert space $K := \mathrm{ran}(P)$. Since $S = PS$ for each $S \in \mathscr{S}$, the space K is invariant under $\mathscr{S}$ and hence

$$\Phi : \mathscr{S} \to \mathrm{U}(K), \qquad S \mapsto S|_K$$

is a well defined homomorphism of groups. Since $S = SP$, this homomorphism is injective. Clearly Φ is also continuous, and since $\mathscr{S}$ is compact, Φ is a homeomorphism onto its image $\Phi(\mathscr{S})$. But on $\Phi(\mathscr{S})$ as a subgroup of $\mathrm{U}(K)$ the strong and weak operator topologies coincide, and since $S = SP$ for each $S \in \mathscr{S}$, this holds on $\mathscr{S}$ as well. In particular, the multiplication is jointly continuous. Since the inversion mapping is continuous on $\Phi(\mathscr{S})$, this must be so on $\mathscr{S}$ as well, and the proof is complete. □

Let now $\mathrm{X} = (X, \Sigma, \mu)$ be a probability space. Then the semitopological semigroup of contractions on $\mathrm{L}^1(\mathrm{X})$ is, in general, not weakly compact. However, if we restrict to those operators which are also L^2-contractive, we can use the Hilbert space results from above.

Lemma 16.4. *Let* X *be a probability space, and let $E := \mathrm{L}^1(\mathrm{X})$. Then the semigroup*

$$\mathscr{S} := \big\{T \in \mathscr{L}(E) : \|T\| \leq 1 \quad \textit{and} \quad \|Tf\|_2 \leq \|f\|_2 \ \forall\, f \in \mathrm{L}^2(\mathrm{X})\big\}$$

of simultaneous L^1- and L^2-contractions is weakly compact. Moreover, on $\mathscr{S}$ the weak (strong) operator topologies of $\mathscr{L}(\mathrm{L}^1)$ and $\mathscr{L}(\mathrm{L}^2)$ coincide.

Proof. We can consider $\mathscr{S}$ as a subset of $\mathscr{B}_1(\mathrm{L}^2)$. As such, it is weakly closed since the L^1-contractivity of an L^2-contraction T is expressible in weak terms by

$$\left|\int_{\mathrm{X}} (Tf) \cdot g\right| \leq \|f\|_1 \, \|g\|_\infty$$

for all $f, g \in \mathrm{L}^\infty$. By Corollary 16.2 $\mathscr{S}$ is weakly compact. The weak (strong) operator topologies on $\mathscr{L}(\mathrm{L}^1)$ and $\mathscr{L}(\mathrm{L}^2)$ coincide on $\mathscr{S}$, since $\mathscr{S}$ is norm-bounded in both spaces and L^2 is dense in L^1. Hence the claim is proved. □

Trivially, if one intersects a weakly compact semigroup $\mathscr{S} \subseteq \mathscr{L}(E)$ with any weakly closed subsemigroup of $\mathscr{L}(E)$, then the result is again a weakly compact semigroup. With the help of Exercises 2 and 3 and the interpolation Theorem 8.23, the following result is immediate.

Corollary 16.5. *Let* X *be a finite measure space. Then the following sets are weakly compact subsemigroups of* $\mathscr{L}(\mathrm{L}^1)$.

1) *The set of all Dunford–Schwartz operators.*
2) *The set of all positive Dunford–Schwartz operators.*
3) *The set of Markov operators*

$$\mathrm{M}(\mathrm{L}^1) := \big\{T \in \mathscr{L}(\mathrm{L}^1) : T \geq 0,\, T\mathbf{1} = \mathbf{1},\, T'\mathbf{1} = \mathbf{1}\big\}.$$

Moreover, on each of these sets the weak operator topologies of $\mathscr{L}(\mathrm{L}^1)$ *and* $\mathscr{L}(\mathrm{L}^2)$ *coincide. In particular, the semigroup*

$$\mathrm{Emb}(\mathrm{X}) = \big\{S \in \mathrm{M}(\mathrm{X}) : S \text{ Markov embedding}\big\}$$

of Markov embeddings (see Definition 12.10) is a topological semigroup with respect to the weak (=strong) operator topology.

By combining Corollary 16.5 with Theorem 16.3 above, we obtain the following important corollary.

Corollary 16.6. *Let* $\mathscr{S} \subseteq \mathrm{M}(\mathrm{L}^1(\mathrm{X}))$ *be a weakly compact subsemigroup of Markov operators on some probability space* X. *If* $\mathscr{S}$ *is algebraically a group, then it is a compact topological group. Moreover, the weak and strong operator topologies coincide on* $\mathscr{S}$.

By actually looking at the *proof* of Theorem 16.3, we can say much more. In the situation of Corollary 16.6 let Q be the unit element of $\mathscr{S}$. Then Q is a Markov projection, and we can choose freely a model for $F := \mathrm{ran}(Q)$, i.e., a probability space Y and a Markov embedding $J \in \mathrm{M}(\mathrm{X};\mathrm{Y})$ such that $\mathrm{ran}(J) = F$ (Theorem 13.18). Then we can consider the mapping

$$\Phi : \mathscr{S} \to \mathrm{M}(\mathrm{Y}), \qquad S \mapsto J^{-1}(S|_F)J$$

which is a topological isomorphism onto its image $\Phi(\mathscr{S})$. The latter is a compact topological subgroup of $\mathrm{Aut}\,(\mathrm{Y})$, the group of Markov automorphisms on Y.

Finally, let us state a rather trivial case of weakly compact operator (semi)groups.

Proposition 16.7. *Let G be a compact group. Then the sets of left and right rotations*

$$\mathscr{L} = \{L_a : a \in G\} \quad \textit{and} \quad \mathscr{R} = \{R_a : a \in G\}$$

are strongly compact subgroups of $\mathscr{B}_1(\mathrm{C}(K))$*, topologically isomorphic to* G*.*

Proof. The map $a \mapsto R_a$, $G \to \mathscr{B}_1(\mathrm{C}(K))$ is a group homomorphism, continuous for the strong operator topology by Corollary 14.12. Since G is compact, R is a topological isomorphism onto its image $\mathscr{R}$. The case of the left rotations is similar, by looking at the group homomorphism $a \mapsto L_{a^{-1}}$. □

More on Compact Operator Semigroups

Let E be a Banach space. Then the following result is useful to recognise relatively weakly compact subsets of $\mathscr{L}(E)$.

Lemma 16.8. *Let* E *be a Banach space, let* $\mathscr{T} \subseteq \mathscr{L}(E)$*, and let* $D \subseteq E$ *be a subset such that* $\operatorname{lin} D$ *is dense in* E*. Then the following assertions are equivalent.*

(i) $\mathscr{T}$ *is relatively weakly compact, i.e, relatively compact in the weak operator topology.*

(ii) $\mathscr{T}x = \{Tx : T \in \mathscr{T}\}$ *is relatively weakly compact in* E *for each* $x \in E$*.*

(iii) $\mathscr{T}$ *is norm-bounded and* $\mathscr{T}x = \{Tx : T \in \mathscr{T}\}$ *is relatively weakly compact in* E *for each* $x \in D$*.*

Proof. (i)⇒(ii): For fixed $x \in E$ the mapping $\mathscr{L}_\sigma(E) \to (E, \sigma(E,E'))$, $T \mapsto Tx$, is continuous and hence carries relatively compact subsets of $\mathscr{L}_\sigma(E)$ to relatively weakly compact subsets of E.

(ii)⇒(iii): The first statement follows from the Principle of Uniform Boundedness, the second is a trivial consequence of (ii).

(iii)⇒(i): Let $(T_\alpha)_\alpha \subseteq \mathscr{T}$ be some net. We have to show that it has a subnet, converging weakly to some bounded operator $T \in \mathscr{L}(E)$. Consider the product space

$$K := \prod_{x \in D} \overline{\mathscr{T}x}^{\sigma},$$

which is compact by Tychonov's Theorem A.5. Then

$$\alpha \mapsto (T_\alpha x)_{x \in D}$$

is a net in K, hence there is a subnet $(T_\beta)_\beta$, say, such that $Tx := T_\beta x$ exists in the weak topology for every $x \in D$. By linearity, we may suppose without loss of generality that D is a linear subspace of E, whence $T : D \to E$ is linear. From the norm-boundedness of $\mathscr{T}$ it now follows that $\|Tx\| \leq M\|x\|$ for all $x \in D$, where $M := \sup_{S \in \mathscr{T}} \|S\|$. Since D is norm-dense in E, T extends uniquely to a bounded linear operator on E, i.e., $T \in \mathscr{L}(E)$.

It remains to show that $T_\beta \to T$ in $\mathscr{L}_\sigma(E)$. Take $y \in E, x \in D$ and $x' \in E'$. Then from the estimate

$$\begin{aligned}\big|\langle T_\beta y - T_\beta y, x'\rangle\big| &\le \big|\langle T_\beta(y-x), x'\rangle\big| + \big|\langle T_\beta x - Tx, x'\rangle\big| + \big|\langle T(x-y), x'\rangle\big| \\ &\le 2M\,\|x'\| \cdot \|y-x\| + \big|\langle T_\beta x - Tx, x'\rangle\big|\end{aligned}$$

we obtain

$$\limsup_\beta \big|\langle T_\beta y - T_\beta y, x'\rangle\big| \le 2M\,\|x'\| \cdot \|y-x\|.$$

Since for fixed y the latter can be made arbitrarily small, the proof is complete. □

It is clear that Theorem 16.1 about reflexive spaces follows from Lemma 16.8, as the closed unit ball of a reflexive space is weakly compact.

Remarks 16.9. 1) For historical reasons, a relatively weakly compact semigroup $\mathscr{T} \subseteq \mathscr{L}(E)$ is sometimes called **weakly almost periodic** (e.g., in [Krengel (1985)]). Analogously, relatively strongly compact semigroups are termed **(strongly) almost periodic**.

2) If $\mathscr{T} \subseteq \mathscr{L}(E)$ is relatively strongly compact, then it is relatively weakly compact, and both topologies coincide on the (strong=weak) closure of $\mathscr{T}$. In Exercise 4 it is asked to establish the analogue of Lemma 16.8 for relative strong compactness.

3) A semigroup $\mathscr{T} \subseteq \mathscr{L}(E)$ is relatively weakly compact if and only if $\operatorname{conv}(\mathscr{T})$ is relatively weakly compact. (This follows from Lemma 16.8 together with the Kreĭn–Šmulian Theorem C.11.) The same is true for the strong operator topology.

4) Let $T \in \mathscr{L}(E)$ be a single operator. The semigroup **generated** by T is

$$\mathscr{T}_T := \big\{T^n : n \in \mathbb{N}_0\big\} = \big\{\mathrm{I}, T, T^2, \dots\big\}$$

Suppose that $\mathscr{T}_T$ is relatively weakly compact. Equivalently, by Lemma 16.8,

$$\operatorname{cl}_\sigma\big\{T^n f : n \in \mathbb{N}_0\big\} \quad \text{is weakly compact for each } f \in E.$$

Then T is certainly power-bounded. By 3), i.e., by the Kreĭn–Šmulian Theorem C.11,

$$\overline{\operatorname{conv}}\big\{T^n f : n \in \mathbb{N}_0\big\} \quad \text{is weakly compact for each } f \in E.$$

In particular, for each $f \in E$ the sequence $(\mathrm{A}_n[T]f)_{n\in\mathbb{N}}$ has a weak cluster point, whence by Proposition 8.18 it actually converges in norm. To sum up, we have shown that if T generates a relatively weakly compact semigroup, then T is mean ergodic.
The converse statement fails to hold. In fact, Example 8.27 exhibits a multiplication operator T on c, the space of convergent sequences, such that T is

mean ergodic but T^2 is not. However, T^2 generates a relatively weakly compact semigroup as soon as T does so.

We close this section with a classical result. If S is a semigroup and $a \in S$, then as in the group case we denote by L_a, R_a the left and the right rotation by a defined as $(L_a f)(x) = f(ax)$ and $(R_a f)(x) = f(xa)$ for $f : S \to \mathbb{C}$ and $x \in S$.

Theorem 16.10. *Let S be a compact semitopological semigroup. Then the semigroup of left rotations $\mathscr{L} = \{L_a : a \in S\}$ is a weakly compact semigroup of operators on $\mathrm{C}(S)$.*

Proof. By Lemma 16.8 we only have to check whether for fixed $f \in \mathrm{C}(S)$ the orbit $M := \{L_a f : a \in S\}$ is weakly compact in $\mathrm{C}(S)$. By Grothendieck's Theorem E.5, this is equivalent to M being compact in $\mathrm{C}_p(S)$, the space $\mathrm{C}(S)$ endowed with the pointwise topology. But by separate continuity of the multiplication of S, the mapping

$$S \to \mathrm{C}_p(S), \qquad a \mapsto L_a f$$

is continuous, and since S is compact, its image is compact. □

16.2 The Jacobs–de Leeuw–Glicksberg Decomposition

Let E be a Banach space and let $\mathscr{T} \subseteq \mathscr{L}(E)$ be a semigroup of operators on E. Then

$$\mathscr{S} := \mathrm{cl}_w(\mathscr{T}),$$

the weak operator closure of $\mathscr{S}$, is a semitopological semigroup. We call the semigroup $\mathscr{T}$ **JdLG-admissible** if $\mathscr{S}$ is weakly compact and contains a *unique* minimal idempotent, i.e., satisfies the equivalent conditions of Theorem 14.5. (Note that by Theorem 14.3 every weakly compact operator semigroup contains at least one minimal idempotent.)

The following gives a list of important examples.

Theorem 16.11. *Let E be a Banach space and $\mathscr{T} \subseteq \mathscr{L}_w(E)$ a semigroup of operators on E. Then $\mathscr{T}$ is JdLG-admissible in the following cases.*

a) *$\mathscr{T}$ is Abelian and relatively weakly compact.*

b) *E is a Hilbert space and $\mathscr{T}$ consists of contractions.*

c) *$E = \mathrm{L}^1(\mathrm{X})$, X a probability space, and $\mathscr{T} \subseteq \mathrm{M}(\mathrm{X})$, i.e., $\mathscr{T}$ consists of Markov operators.*

Proof. The case a) was already mentioned in Theorem 14.5. (Note that by Exercise 14.2 the closure $\mathscr{S}$ of $\mathscr{T}$ is Abelian, too.) The case c) can be reduced to b) via Corollary 16.5. To prove b) suppose that P and Q are minimal idempotents in $\mathscr{S} = \mathrm{cl}_w(\mathscr{T})$. Then P and Q are contractive projections, hence orthogonal projections

(Theorem 13.31). Since $\mathscr{S}Q$ is a minimal left ideal (Lemma 14.1) and $\mathscr{S}PQ$ is a left ideal contained in $\mathscr{S}Q$, it follows that these left ideals are identical. Hence there is an operator $S \in \mathscr{S}$ such that $SPQ = Q$. Now let $f \in H$ be arbitrary. Since S and P are contractions, it follows that

$$\|Qf\| = \|SPQf\| \leq \|PQf\| \leq \|Qf\|.$$

So $\|PQf\| = \|Qf\|$, and since P is an orthogonal projection, by Corollary 13.32 we obtain $PQf = Qf$. Since $f \in H$ was arbitrary, we conclude that $PQ = Q$, i.e., $\mathrm{ran}(Q) \subseteq \mathrm{ran}(P)$. By symmetry $\mathrm{ran}(P) = \mathrm{ran}(Q)$, and since orthogonal projections are uniquely determined by their range, it follows that $P = Q$. □

With a little more effort one can show the following.

Theorem 16.12. *If E and its dual space E' both are strictly convex, then every relatively weakly compact semigroup of contractions on E is JdLG-admissible.*

Proof. (Sketch) Recall that a Banach space E is **strictly convex** if $\|f\| = \|g\| = 1$ and $f \neq g$ implies that $\|f+g\| < 2$. If P is a contractive projection on a strictly convex space E, then it has the property a) of Corollary 13.32, see Exercise 5. Hence in the proof above we conclude that $PQ = Q$. By taking right ideals in place of left ideals, we find an operator $T \in \mathscr{S}$ such that $PQT = P$, and taking adjoints yields $P' = T'Q'P'$. If E' is also strictly convex, it follows as before that $P' = Q'P'$, i.e., $P = PQ = Q$. □

The theorem is applicable to spaces $\mathrm{L}^p(\mathrm{X})$ with $1 < p < \infty$, since these are strictly convex and $\mathrm{L}^{p'} = \mathrm{L}^q$ with $p^{-1} + q^{-1} = 1$. Note that these spaces are reflexive, whence contraction semigroups are automatically relatively weakly compact.

Corollary 16.13. *Every semigroup of contractions on a space $\mathrm{L}^p(\mathrm{X})$, X some measure space and $1 < p < \infty$, is JdLG-admissible.*

Remark 16.14. Theorem 16.12 is due to de Leeuw and Glicksberg [de Leeuw and Glicksberg (1961), Corollary 4.14]. There one can find even a *characterisation* of JdLG-admissible semigroups in terms of the **amenability** of their weak closures. See [Day (1957b)] and [Paterson (1988)] for more about this notion.

JdLG-admissible Semigroups

Let again E be a Banach space, $\mathscr{T} \subseteq \mathscr{L}(E)$ a semigroup of operators and

$$\mathscr{S} = \mathrm{cl}_w(\mathscr{T})$$

its weak operator closure. In the following we suppose that $\mathscr{T}$ is JdLG-admissible, i.e., $\mathscr{S}$ is weakly compact and contains a unique minimal idempotent Q.

By Theorem 14.5, $\mathscr{S}$ has a unique nonempty minimal ideal

$$\mathscr{G} := K(\mathscr{S}) = Q\mathscr{S} = \mathscr{S}Q = Q\mathscr{S}Q \tag{16.1}$$

which is a group with unit element Q. Since $Q^2 = Q$, Q is a projection, and hence the space E decomposes into a direct sum

$$E = E_{\mathrm{r}}(\mathscr{T}) \oplus E_{\mathrm{s}}(\mathscr{T}) = \operatorname{ran} Q \oplus \ker Q,$$

called the **Jacobs–de Leeuw–Glicksberg decomposition** (JdLG-decomposition for short) of E associated with $\mathscr{T}$. This decomposition goes back to [Jacobs (1956)], [de Leeuw and Glicksberg (1959)] and [de Leeuw and Glicksberg (1961)]. The space $E_{\mathrm{r}} := E_{\mathrm{r}}(\mathscr{S})$ is called the **reversible part** of E and its elements the **reversible vectors**; the space $E_{\mathrm{s}} := E_{\mathrm{s}}(\mathscr{T})$ is called the **(almost weakly) stable** part and its element the **almost weakly stable** vectors. This terminology will become clear shortly. Moreover, we shall see below that for $\mathscr{T} = \mathscr{T}_T = \{T^n : n \in \mathbb{N}_0\}$ consisting of the iterates of one single operator $T \in \mathscr{L}(E)$, we have $E_{\mathrm{s}}(\mathscr{T}) = E_{\mathrm{s}}(T)$, as introduced in Section 9.2.

Proposition 16.15. *In the situation above, E_{r} and E_{s} are invariant under $\mathscr{S}$.*

Proof. For $S \in \mathscr{S}$ we have $Q\mathscr{S} = \mathscr{S}Q$, hence there are operators $S_1, S_2 \in \mathscr{S}$ such that $QS = S_1Q$ and $SQ = QS_2$. Therefore, for $f = Qf \in E_{\mathrm{r}} = \operatorname{ran}(Q)$ we have $Sf = SQf = QS_2f \in E_{\mathrm{r}}$, and for $f \in E_{\mathrm{s}} = \ker(Q)$ we have $QSf = S_1Qf = 0$, whence $Sf \in E_{\mathrm{s}}$. □

By this proposition we may restrict the operators from $\mathscr{S}$ to the invariant subspaces E_{r} and E_{s} and obtain

$$\mathscr{T}_{\mathrm{r}} := \big\{T|_{E_{\mathrm{r}}} : T \in \mathscr{T}\big\}, \quad \mathscr{S}_{\mathrm{r}} := \big\{S|_{E_{\mathrm{r}}} : S \in \mathscr{S}\big\} \quad \subseteq \mathscr{L}(E_{\mathrm{r}})$$

and

$$\mathscr{T}_{\mathrm{s}} := \big\{T|_{E_{\mathrm{s}}} : T \in \mathscr{T}\big\}, \quad \mathscr{S}_{\mathrm{s}} := \big\{S|_{E_{\mathrm{s}}} : S \in \mathscr{S}\big\} \quad \subseteq \mathscr{L}(E_{\mathrm{s}}).$$

By Exercise 6, the restriction maps

$$S \mapsto S|_{E_{\mathrm{r}}} \quad \text{and} \quad S \mapsto S|_{E_{\mathrm{s}}}$$

are continuous for the weak operator topologies. Since $\mathscr{S}$ is compact and $\mathscr{S} = \mathrm{cl}_w(\mathscr{T})$, it follows that

$$\mathscr{S}_{\mathrm{r}} = \mathrm{cl}_w(\mathscr{T}_{\mathrm{r}}) \quad \text{and} \quad \mathscr{S}_{\mathrm{s}} = \mathrm{cl}_w(\mathscr{T}_{\mathrm{s}}).$$

Let us look at an example.

Example 16.16 (Mean Ergodic Decomposition). Let $T \in \mathscr{L}(E)$ and suppose that the semigroup

$$\mathscr{T}_T = \big\{\mathrm{I}, T, T^2, T^3 \dots\big\}$$

generated by T is relatively compact in $\mathscr{L}_w(E)$. This implies in particular that T is power-bounded. Moreover, by Remark 16.9.3 the semigroup

$$\mathscr{S} := \overline{\mathrm{conv}}(\mathscr{T}_T) = \overline{\mathrm{conv}}\{\mathrm{I}, T, T^2, T^3 \dots\}$$

is weakly compact. By Remark 16.9.4, the operator T is mean ergodic, and one has the mean ergodic decomposition

$$E = \mathrm{fix}(T) \oplus \overline{\mathrm{ran}(\mathrm{I}-T)}$$

with ergodic projection $P = P_T$. Since P is the strong limit of the Cesàro means $\mathrm{A}_n[T]$, it belongs to $\mathscr{S}$. But since $PT = TP = P$, one has $\mathscr{S}P = \{P\}$. Hence $\{P\}$ is the minimal ideal of $\mathscr{S}$, and the JdLG-decomposition for $\mathscr{S} = \overline{\mathrm{conv}}(\mathscr{T}_T)$ coincides with the mean ergodic decomposition for T.

Example 16.16 shows that we do not obtain new information when applying the JdLG-decomposition to the semigroup $\mathrm{conv}(\mathscr{T}_T)$. This is different when we apply it to the semigroup $\mathscr{T}_T$ itself.

Example 16.17 (The Kronecker Factor). Let $(\mathrm{X};\varphi)$ be an MDS with Koopman operator $T = T_\varphi$ on $E := \mathrm{L}^1(\mathrm{X})$. Since the convex semigroup $\mathrm{M}(\mathrm{X})$ of Markov operators is weakly compact (Theorem 13.2.a)), the semigroup $\mathscr{T}_T = \{T^n : n \in \mathbb{N}_0\}$ generated by T is relatively weakly compact. Since this semigroup is Abelian, its closure

$$\mathscr{S}_T := \mathrm{cl}_w(\mathscr{T}_T)$$

is Abelian, too, and hence $\mathscr{T}_T$ is JdLG-admissible, cf. Theorem 16.11. We obtain a corresponding JdLG-decomposition

$$E = E_\mathrm{r} \oplus E_\mathrm{s} = \mathrm{ran}(Q) \oplus \ker(Q),$$

where $Q \in \mathscr{S}$ is the unique minimal idempotent. Since Q is a Markov operator, it is a Markov projection, and hence $E_\mathrm{r} = \mathrm{ran}(Q)$ is a factor of $(\mathrm{X};\varphi)$, called the **Kronecker factor**. Clearly $\mathrm{fix}(T) \subseteq \mathrm{fix}(S)$ for each $S \in \mathscr{S}$, so in particular

$$\mathrm{fix}(T) \subseteq \mathrm{fix}(Q) = E_\mathrm{r}.$$

On the other hand, if $Qf = 0$ then $0 \in \mathrm{cl}_\sigma\{T^n f : n \in \mathbb{N}_0\}$ (see Proposition 16.18 below). Hence $0 \in \mathrm{fix}(T) \cap \overline{\mathrm{conv}}\{T^n f : n \in \mathbb{N}_0\}$, and by Proposition 8.18 it follows that $f \in \overline{\mathrm{ran}}(\mathrm{I}-T)$. That is to say

$$E_\mathrm{s} \subseteq \overline{\mathrm{ran}}(\mathrm{I}-T).$$

In particular, we have $\int_\mathrm{X} f = 0$ for each $f \in E_\mathrm{s}$.

16.3 The Almost Weakly Stable Part

We now investigate the two parts of the JdLG-decomposition separately. Let us first deal with the "(almost weakly) stable" part.

Proposition 16.18. *Let $E = E_{\mathrm{r}} \oplus E_{\mathrm{s}}$ be the JdLG-decomposition of a Banach space E with respect to a JdLG-admissible semigroup $\mathscr{T} \subseteq \mathscr{L}(E)$ with weak closure $\mathscr{S} := \mathrm{cl}_w(\mathscr{T})$. Then the following assertions hold.*

a) *$\mathscr{T}_{\mathrm{s}}$ is a JdLG-admissible semigroup on E_{s}, with weak operator closure $\mathscr{S}_{\mathrm{s}}$.*

b) *The minimal idempotent of $\mathscr{S}_{\mathrm{s}}$ is the zero operator $0 \in \mathscr{S}_{\mathrm{s}}$, and its minimal ideal is $K(\mathscr{S}_{\mathrm{s}}) = \{0\}$.*

c) *The elements of E_{s} are characterised by*

$$u \in E_{\mathrm{s}} \quad \Longleftrightarrow \quad 0 \in \mathrm{cl}_\sigma\{Tu : T \in \mathscr{T}\} \quad \Longleftrightarrow \quad Su = 0 \quad \textit{for some } S \in \mathscr{S}.$$

Proof. Let Q be the minimal idempotent of $\mathscr{S}$. Then Q restricts to 0 on E_{s}, and hence a) and b) are clear.

For c), note first that for each $u \in E$ the mapping $T \mapsto Tu$ is continuous from $\mathscr{L}_w(E)$ to $(E, \sigma(E,E'))$. Since $\mathscr{S} = \mathrm{cl}_w(\mathscr{T})$ is compact, one has $\mathrm{cl}_\sigma(\mathscr{T}u) = (\mathrm{cl}_w\mathscr{T})u = \mathscr{S}u$. This proves the second equivalence of c).

If $u \in E_{\mathrm{s}} = \ker Q$, then it follows immediately that $0 = Qu \in \mathscr{S}u$. Conversely, suppose that $Su = 0$ for some $S \in \mathscr{S}$. Since $QS \in \mathscr{G} = K(\mathscr{S})$ and $\mathscr{G}$ is a group with unit element Q, there is $R \in \mathscr{G}$ such that $RQS = Q$. Hence $Qu = RQSu = 0$, whence $u \in E_{\mathrm{s}}$. □

A JdLG-admissible semigroup $\mathscr{T}$ on a Banach space E is called **almost weakly stable** if $E = E_{\mathrm{s}}$. Equivalently, $\mathscr{T}$ is almost weakly stable if 0 is in the weak closure of each orbit $\{Tu : T \in \mathscr{T}\}$, $u \in E$.

In the following we shall investigate more closely the special case of the semigroup

$$\mathscr{T}_T = \{\mathrm{I}, T, T^2, T^3, \dots\}$$

generated by a single operator $T \in \mathscr{L}(E)$. We suppose that this semigroup is relatively weakly compact. (Recall that this is automatic if T is power bounded and E is reflexive, or if T is a Markov operator on some $\mathrm{L}^1(\mathrm{X})$.) Then $\mathscr{S} := \mathrm{cl}_w(\mathscr{T})$ is Abelian, whence $\mathscr{T}$ is JdLG-admissible, and we can form the JdLG-decomposition $E = E_{\mathrm{r}} \oplus E_{\mathrm{s}}$ associated with $\mathscr{T}_T$. The following theorem states in particular that the almost weakly stable subspace E_{s} coincides with $E_{\mathrm{s}}(T)$, introduced in Section 9.2.

Recall from Chapter 9, page 132, that a sequence $(x_n)_n$ in a topological space converges in density to x, in notation $\text{D-lim}\, n \to \infty x_n = x$, if there is a subsequence $J \subseteq \mathbb{N}$ with density $\mathrm{d}(J) = 1$ such that $\lim_{n \in J} x_n = x$.

Theorem 16.19. *Let T generate a relatively weakly compact semigroup $\mathscr{T}_T$ on a Banach space E, and let $M \subseteq E'$ be norm-dense in the dual space E'. Then for $x \in E$ the following assertions are equivalent.*

(i) $\underset{n\to\infty}{\text{D-lim}} \langle T^n x, x' \rangle = 0$ *for all* $x' \in M$.

(ii) $\underset{n\to\infty}{\text{D-lim}} \langle T^n x, x' \rangle = 0$ *for all* $x' \in E'$.

(iii) $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \left|\langle T^j x, x'\rangle\right| = 0$ *for all* $x' \in M$.

(iv) $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \left|\langle T^j x, x'\rangle\right| = 0$ *for all* $x' \in E'$.

(v) $\sup_{x'\in E', \|x'\|\le 1} \frac{1}{n}\sum_{k=0}^{n-1} \left|\left\langle T^k x, x'\right\rangle\right| \longrightarrow_{n\to\infty} 0$.

(vi) $0 \in \mathrm{cl}_\sigma\{T^n x : n \in \mathbb{N}\}$, *i.e.,* $x \in E_{\mathrm{s}}(\mathscr{T})$.

(vii) *There exists a subsequence* $(n_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ *such that* $T^{n_j}x \to 0$ *weakly as* $j \to \infty$.

(viii) $\text{D-lim}_n\ T^n x = 0$ *weakly.*

Proof. We note that T is necessarily power-bounded. For such operators the pairwise equivalence of (i)–(v) has been shown already in Theorem 9.12 and Proposition 9.14. For the remaining part note first that the implications (viii)$\Rightarrow$(vii)$\Rightarrow$(vi) are trivial.

For the proof of the remaining assertions we need some preparatory remarks. Define $F := \overline{\mathrm{lin}}\{T^n x : n \ge 0\}$, which is then a closed separable subspace of E. Since F is also weakly closed, the set $K := \mathrm{cl}_\sigma\{T^n x : n \in \mathbb{N}_0\}$ (closure in E) is contained in F and hence a weakly compact subset of F. (By the Hahn–Banach theorem the weak topology of F is induced by the weak topology on E.) Hence, in the following we suppose that E is separable.

Then, by the Hahn–Banach theorem we can find a sequence $(x'_j)_{j\in\mathbb{N}} \subseteq E'$ with $\|x'_j\| = 1$ for all $j \in \mathbb{N}$ and such that $\{x_j : j \in \mathbb{N}\}$ separates the points of E (see Exercise 8), and define

$$d(y,z) := \sum_{j=1}^{\infty} 2^{-j} \left|\langle y - z, x'_j\rangle\right| \qquad (y,z \in E). \tag{16.2}$$

Then d is a metric on E and continuous for the weak topology. Since K is weakly compact, d is a metric for the weak topology on K.

We can now turn to the proof of the remaining implications.
(v)$\Rightarrow$(viii): It follows from (v) and the particular form (16.2) of the metric on K that

$$\frac{1}{n}\sum_{j=0}^{n} d(T^j x, 0) \to 0 \quad \text{as } n \to \infty.$$

The Koopman–von Neumann Lemma 9.13 yields $\text{D-lim}_n\, d(T^n x, 0) = 0$, which is (viii).

(vi)$\Rightarrow$(vii): Note that

$$0 \in \mathrm{cl}_\sigma\{T^n x : n \in \mathbb{N}_0\} = \{x, Tx, \ldots, T^{k-1}x\} \cup \mathrm{cl}_\sigma\{T^n x : n \ge k\}$$

for every $k \in \mathbb{N}$. Since $T^k x = 0$ implies $T^m x = 0$ for all $m \ge k$, one has

$$0 \in \bigcap_{k\in\mathbb{N}} \mathrm{cl}_\sigma\{T^n x : n \geq k\},$$

that is, 0 is a weak cluster point of $(T^n x)_{n\in\mathbb{N}_0}$. Since K is metrisable, (vii) follows.

(vii)$\Rightarrow$(v): This is similar to the proof of Proposition 9.14. By passing to the equivalent norm $|||y||| := \sup_{n\in\mathbb{N}_0} \|T^n y\|$ (Exercise 9.3) one may suppose that T is a contraction. Hence the (weakly*) compact set $\mathrm{B}' := \{x' \in E' : \|x'\| \leq 1\}$ is invariant under T', and we obtain a TDS $(\mathrm{B}'; T')$ with associated Koopman operator S on $\mathrm{C}(\mathrm{B}')$. Let $f(x') := |\langle x, x'\rangle|$. Then by hypothesis, there is a subsequence $(n_j)_j$ such that $S^{n_j} f \to 0$ pointwise on B'. By the dominated convergence theorem, $S^{n_j} f \to 0$ weakly. Hence $0 \in \mathrm{fix}(S) \cap \overline{\mathrm{conv}}\{S^n f : n \geq 1\}$ and by Proposition 8.18, implication (iv)$\Rightarrow$(i), it follows that $\mathrm{A}_n[S] f \to 0$ in the norm of $\mathrm{C}(\mathrm{B}')$, and this is (v). □

Weakly Mixing Systems Revisited

Let $(\mathrm{X}; \varphi)$ be an MDS with Koopman operator $T = T_\varphi$ on $E = \mathrm{L}^1(\mathrm{X})$, and recall that the semigroup $\mathscr{T}_T$ generated by T is JdLG-admissible, cf. Example 16.17. The corresponding JdLG-decomposition is

$$E = E_{\mathrm{r}} \oplus E_{\mathrm{s}},$$

where E_{r} is the Kronecker factor and $E_{\mathrm{s}} = E_{\mathrm{s}}(T)$ by Theorem 16.19. Therefore, we can extend the list of characterisations of weakly mixing systems given in Chapter 9. For convenience we include also the equivalences obtained there.

Theorem 16.20. *Let $(\mathrm{X}; \varphi)$ by an MDS with associated Koopman operator $T = T_\varphi$ on $E := \mathrm{L}^p(\mathrm{X})$, $p \in [1, \infty)$. Then the following assertions are equivalent.*

(i) *The MDS $(\mathrm{X}; \varphi)$ is weakly mixing.*

(ii) *For every $f \in \mathrm{L}^p(\mathrm{X})$, $g \in \mathrm{L}^q(\mathrm{X})$ there is a subsequence $J \subseteq \mathbb{N}$ of density $\mathrm{d}(J) = 1$ with*

$$\lim_{n\in J} \int_{\mathrm{X}} (T^n f) \cdot g \;=\; \Big(\int_{\mathrm{X}} f\Big) \cdot \Big(\int_{\mathrm{X}} g\Big).$$

(iii) $\dfrac{1}{n} \displaystyle\sum_{k=0}^{n-1} \Big| \Big\langle T^k f, g \Big\rangle - \langle f, \mathbf{1}\rangle \cdot \langle \mathbf{1}, g\rangle \Big| \underset{n\to\infty}{\longrightarrow} 0$ *for all $f \in \mathrm{L}^p$, $g \in \mathrm{L}^q$.*

(iv) *For each $f \in E$ with $\int_{\mathrm{X}} f = 0$ one has $f \in E_{\mathrm{s}}(T)$.*

(v) $\mathrm{fix}(T) = \mathrm{lin}\{\mathbf{1}\}$ *and* $\mathrm{ran}(\mathrm{I} - T) \subseteq E_{\mathrm{s}}(T)$.

(vi) $E = \mathrm{lin}\{\mathbf{1}\} \oplus E_{\mathrm{s}}(T)$.

(vii) *The k^{th} iterate MDS $(\mathrm{X}; \varphi^k)$ is weakly mixing for some/each $k \in \mathbb{N}$.*

(viii) $(\mathrm{X} \otimes \mathrm{Y}; \varphi \times \psi)$ *is weakly mixing for some/each weakly mixing system $(\mathrm{Y}; \psi)$.*

(ix) $(\mathrm{X} \otimes \mathrm{X}; \varphi \times \varphi)$ *is weakly mixing.*

(x) $\displaystyle\int_{\mathrm{X}} f \cdot \mathbf{1} \in \mathrm{cl}_\sigma\{T^n f : n \in \mathbb{N}_0\}$ *for each $f \in E$.*

(xi) *For each* $f \in \mathrm{L}^p(\mathrm{X})$ *there is subsequence* $J \subseteq \mathbb{N}$ *such that*

$$\lim_{n\in J} \int_{\mathrm{X}} (T^n f)\cdot g \to \Big(\int_{\mathrm{X}} f\Big)\cdot\Big(\int_{\mathrm{X}} g\Big)$$

for all $g \in \mathrm{L}^q(\mathrm{X})$.

(xii) *For each* $f \in \mathrm{L}^p(\mathrm{X})$ *there is subsequence* $J \subseteq \mathbb{N}$ *of density* $d(J) = 1$ *such that*

$$\lim_{n\in J} \int_{\mathrm{X}} (T^n f)\cdot g \to \Big(\int_{\mathrm{X}} f\Big)\cdot\Big(\int_{\mathrm{X}} g\Big)$$

for all $g \in \mathrm{L}^q(\mathrm{X})$.

(xiii) *The Kronecker factor of* $(\mathrm{X};\varphi)$ *is trivial, i.e., equals* $\mathrm{lin}\{\mathbf{1}\}$.

Proof. The assertions (i)–(ix) are equivalent by Theorem 9.16 and Corollaries 9.18 and 9.21.

Now note that with $h := f - \langle f, \mathbf{1}\rangle \cdot \mathbf{1}$, assertions (x)–(xii) can be rewritten equivalently as:

(x) $0 \in \mathrm{cl}_\sigma\{T^n h : n \geq 0\}$.

(xi) $\lim_{n\in J} T^n h = 0$ weakly for some subsequence $J \subseteq \mathbb{N}$.

(xii) $\text{D-}\lim_{n\in\mathbb{N}} T^n h = 0$ weakly.

By Theorem 16.19, each of these assertions is equivalent to (ii). Finally, (xiii) is equivalent to (vi) since $E_{\mathrm{s}}(\mathscr{T}_T) = E_{\mathrm{s}}(T)$, by what was shown above. □

We shall now investigate the reversible part in the JdLG-decomposition. In the case of a Koopman operator, we shall obtain a characterisation of the Kronecker factor that ultimately leads to yet another (and finally: spectral) characterisation of weakly mixing systems in Corollary 16.35 below.

16.4 Compact Group Actions on Banach Spaces

We now turn to the reversible part of a Jacobs–de Leeuw–Glicksberg decomposition. Again let us suppose that $\mathscr{T}$ is a JdLG-admissible semigroup of operators on a Banach space E.

Proposition 16.21. *Let* $E = E_{\mathrm{r}} \oplus E_{\mathrm{s}}$ *be the JdLG-decomposition of a Banach space* E *with respect to a JdLG-admissible semigroup* $\mathscr{T} \subseteq \mathscr{L}(E)$ *with weak closure* $\mathscr{S} := \mathrm{cl}_w(\mathscr{T})$. *Then the following assertions hold.*

a) $\mathscr{T}_{\mathrm{r}}$ *is a JdLG-admissible semigroup on* E_{r}, *with weak operator closure* $\mathscr{S}_{\mathrm{r}}$.

b) *The minimal idempotent of* $\mathscr{S}_{\mathrm{r}}$ *is the identity operator* $\mathrm{I} \in \mathscr{S}_{\mathrm{r}}$, *and its minimal ideal is* $K(\mathscr{S}_{\mathrm{s}}) = \mathscr{S}_{\mathrm{r}}$.

c) *The reversible part* E_{r} *is given by*

$$E_{\mathrm{r}} = \big\{u \in E : v \in \mathrm{cl}_\sigma(\mathscr{T}u) \;\Rightarrow\; u \in \mathrm{cl}_\sigma(\mathscr{T}v)\big\}$$
$$= \big\{u \in E : v \in \mathscr{S}u \;\Rightarrow\; u \in \mathscr{S}v\big\}.$$

Hence a vector $u \in E$ is reversible if and only if whenever some vector v can be reached by the action of $\mathscr{S}$ starting from u, then one can return to u again by the action of $\mathscr{S}$.

Proof. Let Q be the unique minimal idempotent in $\mathscr{S}$, so $E_{\mathrm{r}} = \mathrm{ran}(Q)$. If $u \in E_{\mathrm{r}}$, then $\mathscr{S}u = \mathscr{S}Qu = \mathscr{G}u$. Hence if $v \in \mathscr{S}u$, then there is $R \in \mathscr{G}$ with $v = Ru$. Since $\mathscr{G}$ is a group, there is $S \in \mathscr{G}$ with $SR = Q$. Hence $u = Qu = SRu = Sv \in \mathscr{S}v$. Conversely, if $u \in E$ is such that there is $S \in \mathscr{S}$ with the property that $SQu = u$, then $u = SQu = QSu \in \mathrm{ran}\,Q$. □

The next result opens the door for the methods of harmonic analysis.

Theorem 16.22. *The group $\mathscr{G}$ is a compact topological group, the strong and the weak operator topology coincide on $\mathscr{G}$, and the restriction map*

$$\mathscr{G} \to \mathscr{S}_{\mathrm{r}}, \qquad S \mapsto S|_{E_{\mathrm{r}}}$$

is a topological isomorphism of compact groups.

Proof. For the special case that E is a Hilbert space and $\mathscr{T}$ is a semigroup of contractions, this is Theorem 16.3. In the case that $E = \mathrm{L}^1(\mathrm{X})$, X some probability space, and $\mathscr{T} \subseteq \mathrm{M}(\mathrm{X})$, it is Corollary 16.6. In the general case, that $\mathscr{G}$ is a compact group follows from Ellis' Theorem 14.6. The assertion about the topological isomorphism is then evident.

To prove that the strong and the weak operator topologies coincide on $\mathscr{G}$ we need some more sophisticated tools, to be developed below. □

Continuous Representations of Compact Groups

To complete the proof of Theorem 16.22 and to study the reversible part, it is convenient to pass to a little more abstract framework. Let G be a compact group with Haar measure m and neutral element 1. A **representation** of G on a Banach space E is any mapping $\pi : G \to \mathscr{L}(E)$ satisfying

1) $\pi_1 = \mathrm{I}$,
2) $\pi_{xy} = \pi_x \pi_y \quad (x, y \in G)$.

The representation π is called **weakly continuous**, if the mapping

$$G \to \mathbb{C}, \qquad x \mapsto \langle \pi_x u, u' \rangle$$

is continuous for all $u \in E$, $u' \in E'$. In other words, $\pi : G \to \mathscr{L}_w(E)$ is continuous. In contrast, π is called **(strongly) continuous** if $\pi : G \to \mathscr{L}_s(E)$ is continuous, i.e., each mapping

$$G \to E, \qquad x \mapsto \pi_x u \qquad (u \in E)$$

is continuous for the norm topology on E.

Lemma 16.23. *Let $\pi : G \to \mathscr{L}(E)$ be a representation of a compact group E on a Banach space E. Then the following assertions are equivalent.*

(i) *π is strongly continuous.*

(ii) *The mapping*
$$G \times E \to E, \qquad (x,u) \mapsto \pi_x u$$
is continuous.

(iii) *π is weakly continuous and on $\pi(G) = \{\pi_x : x \in G\} \subseteq \mathscr{L}(E)$ the weak and strong operator topologies coincide.*

Proof. (i)$\Leftrightarrow$(ii): Suppose that (i) holds. By the uniform boundedness principle $M := \sup_{x\in G} \|\pi_x\| < \infty$. Let $(x_\alpha)_\alpha \subseteq G$ and $(u_\alpha)_\alpha \subseteq E$ are nets with $x_\alpha \to x \in G$ and $u_\alpha \to u \in E$. Then

$$\|\pi_{x_\alpha} u_\alpha - \pi_x u\| \le M \|u_\alpha - u\| + \|\pi_{x_\alpha} u - \pi_x u\| \to 0.$$

The implication (ii)$\Rightarrow$(i) is trivial.

(i)$\Leftrightarrow$(iii): Suppose that (i) holds. Then clearly π is weakly continuous. Also, $K := \pi(G)$ is compact in the strong operator topology. Since the weak operator topology is still Hausdorff, the two topologies must coincide on K (Proposition A.4). The converse implication (iii)$\Rightarrow$(i) is trivial. □

In the following let $\pi : G \to \mathscr{L}(E)$ be a *weakly* continuous representation of the compact group G on a Banach space E. It follows from the uniform boundedness principle that $\sup_{x\in G} \|\pi_x\| < \infty$. For $f \in \mathrm{L}^1(G)$ we define the integral

$$\pi_f := \int_G f(x)\pi_x \,\mathrm{d}x \tag{16.3}$$

as an operator $E \to E''$ by

$$\langle \pi_f u, u' \rangle := \int_G f(x) \langle \pi_x u, u' \rangle \,\mathrm{d}x \qquad (u \in E, u' \in E').$$

Our first goal is to show that actually $\pi_f u \in E$ for each $u \in E$.

Lemma 16.24. *In the setting described above, one has $\pi_f u \in E$ for each $f \in \mathrm{L}^1(G)$ and $u \in E$.*

Proof. We may suppose that $f \ge 0$ and $\int_G f = 1$. By the Kreĭn–Šmulian Theorem C.11 the set

$$K := \overline{\mathrm{conv}}\big\{\pi_x u : x \in G\big\} \subseteq E$$

is weakly compact. Let $u'' \in E'' \setminus E$. Then, by the Hahn–Banach theorem, E' separates u'' from K. This means that there is $u' \in E'$ and $c \in \mathbb{R}$ such that

$$\operatorname{Re}\langle \pi_x u, u' \rangle \le c < \operatorname{Re}\langle u'', u' \rangle \qquad (x \in G).$$

Multiplying with $f(x)$ and integrating yields

$$\operatorname{Re}\langle \pi_f u, u' \rangle \le c < \operatorname{Re}\langle u'', u' \rangle,$$

which shows that $\pi_f u \neq u''$. □

Remark 16.25. We refer to [Rudin (1991), Theorem 3.27] for more information on weak integration. In special cases the use of the Kreĭn–Šmulian theorem in the proof of Lemma 16.24 can be avoided. For instance, if $E = H$ is a Hilbert space, one can appeal to Corollary C.22 from elementary Hilbert space theory. The same holds if E contains a densely embedded Hilbert space such that π restricts to a (weakly) continuous representation on H.

By Lemma 16.24, $\pi_f \in \mathscr{L}(E)$ for each $f \in \mathrm{L}^1(G)$. A simple estimate yields

$$\|\pi_f\| \le \Big(\sup_{x\in G} \|\pi_x\|\Big) \cdot \|f_1\|, \tag{16.4}$$

showing that $\pi : \mathrm{L}^1(G) \to \mathscr{L}(E)$ is a bounded linear mapping.

Lemma 16.26. *In the situation above, we have* $E = \overline{\operatorname{lin}}\{\pi_f u : f \in \mathrm{C}(G), u \in E\}$.

Proof. Let $u' \in E'$ such that for all $u \in E$ and $f \in \mathrm{C}(G)$ one has $\langle \pi_f u, u' \rangle = 0$. For fixed $u \in E$ this means that

$$\int_G f(x) \langle \pi_x u, u' \rangle \, \mathrm{d}x = 0 \quad \text{for all } f \in \mathrm{C}(G),$$

whence $\langle \pi_x u, u' \rangle = 0$ m-almost everywhere. But the Haar measure is strictly positive, whence $\langle \pi_x u, u' \rangle = 0$ for all $x \in G$ and $u \in E$. Letting $x = 1$ yields $u' = 0$, and the claim follows by a standard application of the Hahn–Banach theorem. □

Corollary 16.27. *Every weakly continuous representation $\pi : G \to \mathscr{L}(E)$ of a compact group G on a Banach space E is strongly continuous.*

Proof. We define

$$F := \big\{ u \in E : \text{the map} \quad G \to E,\ x \mapsto \pi_x u \quad \text{is continuous} \big\}.$$

Then F is a closed subspace of E by the uniform boundedness of the operators π_x, $x \in G$. If $f \in \mathrm{L}^1(G)$ we have

$$\begin{aligned} \langle \pi_x \pi_f u, u' \rangle = \langle \pi_f u, \pi_x' u' \rangle &= \int_G f(y) \langle \pi_y u, \pi_x' u' \rangle \, \mathrm{d}y = \int_G f(y) \langle \pi_{xy} u, u' \rangle \, \mathrm{d}y \\ &= \int_G f(x^{-1} y) \langle \pi_y u, u' \rangle \, \mathrm{d}y. \end{aligned}$$

This shows that $\pi_x \pi_f u = \pi_{\tau_x f} u$, where

$$(\tau_x f)(y) = f(x^{-1}y) \qquad (x,y \in G, f \in \mathrm{L}^1(G)).$$

But the mapping

$$G \to \mathrm{L}^1(G), \quad x \to \tau_x f$$

is continuous, and hence $\mathrm{ran}(\pi_f) \subseteq F$. It follows that $F = E$ by Lemma 16.26. □

We are now in the position to conclude the proof we have left unfinished.

Proof of Theorem 16.22. Let $\mathscr{G}$ be the minimal ideal of the weak operator closure $\mathscr{S} = \mathrm{cl}_w(\mathscr{T})$ of a JdLG-admissible semigroup $\mathscr{T} \subseteq \mathscr{L}(E)$. We know already that $\mathscr{G}$ is a compact topological group, by Ellis' theorem. Let Q be the unit element in $\mathscr{G}$, i.e., the unique minimal idempotent in $\mathscr{S}$, and let $E_\mathrm{r} := \mathrm{ran}(Q)$ be the reversible part. The restriction mapping

$$\pi : \mathscr{G} \to \mathscr{L}(E_\mathrm{r}), \quad \pi_S := S|_{E_\mathrm{r}}$$

is a weakly and hence, by Corollary 16.27, a strongly continuous representation of $\mathscr{G}$. By Lemma 16.23 on $\pi(\mathscr{G}) = \mathscr{S}_\mathrm{r}$ the strong and weak operator topologies coincide. Since $S = SQ$ for each $S \in \mathscr{G}$, the strong and weak operator topologies coincide also on $\mathscr{G}$, and that remained to be proved. □

To obtain further insight into the structure of E_r we shall employ some facts from the theory of unitary representations of compact groups. (For convenience, these facts are presented in the supplement below.) A (finite-dimensional) **unitary representation** of a compact group G is any continuous group homomorphism

$$\chi : G \to \mathrm{U}(n)$$

where $\mathrm{U}(n)$ is the group of unitary $n \times n$-matrices. We write $\chi = (\chi_{ij})_{i,j}$ and call the $\chi_{ij} \in \mathrm{C}(G)$ the associated **coordinate functions**. A unitary representation is **irreducible** if the canonical action of $\chi(G)$ on $\mathbb{C}^n$ has no invariant subspaces apart from $\{0\}$ and $\mathbb{C}^n$.

Let $\pi : G \to \mathscr{L}(E)$ be any continuous representation of G on some Banach space E. A tuple $(e_j)_{j=1}^n$ is called a (finite) **unitary system** for (π, E) if the $e_1, \ldots, e_n \in E$ are linearly independent, $F := \mathrm{lin}\{e_1, \ldots, e_n\}$ is $\pi(G)$-invariant and the corresponding matrix representation $\chi : G \to \mathbb{C}^{n \times n}$ defined by

$$\pi_x e_i = \sum_{j=1}^n \chi_{ij}(x) e_j \qquad (j = 1, \ldots, n)$$

is unitary, i.e., satisfies $\chi(x) \in \mathrm{U}(n)$ for all $x \in G$. A unitary system $(e_j)_{j=1}^n$ is called **irreducible** if F does not have any nontrivial $\pi(G)$-invariant subspaces.

Lemma 16.28. *Let $\pi : G \to \mathscr{L}(E)$ be a continuous representation of the compact group G on some Banach space E.*

a) *Let $\chi : G \to \mathrm{U}(n)$ be a finite-dimensional unitary representation of G and $u \in E$. Then the finite-dimensional space $\mathrm{lin}\{\pi_{\chi_{ij}} u : i = 1, \ldots, n\}$ is $\pi(G)$-invariant.*

b) *Let $F \subseteq E$ be a finite-dimensional $\pi(G)$-invariant subspace. Then there is a unitary system $(e_j)_{j=1}^n$ for (π, E) with $F = \operatorname{lin}\{e_1, \dots, e_n\}$.*

Proof. a) We note first that

$$(\tau_x \chi_{ij})(y) = \chi_{ij}(x^{-1}y) = \sum_{k=1}^n \chi_{ik}(x^{-1})\chi_{kj}(y) \qquad (x, y \in G).$$

Hence $\pi_x \pi_{\chi_{ij}} u = \pi_{\tau_x \chi_{ij}} u = \sum_{k=1}^n \chi_{ik}(x^{-1}) \pi_{\chi_{kj}} u$.
b) Take any inner product $(\cdot\,|\,\cdot)_F$ on F, and define

$$(u\,|\,v) := \int_G (\pi_x u\,|\,\pi_x v)_F \, \mathrm{d}x \qquad (u, v \in F).$$

Then $(\cdot\,|\,\cdot)$ is an inner product since

$$(u\,|\,u) = \int_G \|\pi_x u\|_F^2 \, \mathrm{d}x = 0 \quad \Rightarrow \quad \|u\|_F = \|\pi_1 u\|_F = 0.$$

With respect to this new inner product, each π_x acts isometrically by the rotation invariance of the Haar measure. Hence any orthonormal basis $e_1, \dots, e_n$ of F with respect to this inner product is a unitary system for (π, E) spanning F. □

We now arrive at the main result of this section.

Theorem 16.29. *Let $\pi : G \to \mathscr{L}(E)$ be a continuous representation of the compact group G on some Banach space E. Then*

$$\begin{aligned} E &= \operatorname{cl} \bigcup \big\{ F : F \text{ is a } \pi(G)\text{-invariant subspace of } E, \ \dim F < \infty \big\} \\ &= \overline{\operatorname{lin}} \big\{ e_j : (e_j)_{j=1}^n \ \text{(irreducible) unitary system}, \ j = 1, \dots, n, n \in \mathbb{N} \big\}. \end{aligned}$$

Proof. By Lemma 16.28 each $\pi(G)$-invariant finite-dimensional subspace F of E is the span of a unitary system. Since one can decompose each unitary representation orthogonally into irreducible subspaces (see supplement below), the second set equality is clear. To prove the first, note that by Theorem 16.36 from the supplement below, the linear span of the coordinate functions of finite-dimensional unitary representations of G form a dense set in $\mathrm{C}(G)$. Since by (16.4) the mapping $\mathrm{L}^1(G) \to \mathscr{L}(E)$, $f \mapsto \pi_f$, is bounded we have

$$\pi_f u \in \overline{\operatorname{lin}} \big\{ \pi_{\chi_{ij}} u : n \in \mathbb{N}, \, \chi \text{ unitary representation of } G \text{ on } \mathbb{C}^n, \ i, j = 1, \dots, n \big\}$$

for every $f \in \mathrm{C}(G)$. Hence, by Lemma 16.28 we obtain

$$\pi_f u \in \operatorname{cl} \bigcup \big\{ F : F \text{ is a } \pi(G)\text{-invariant subspace of } E, \ \dim F < \infty \big\}$$

for each $f \in \mathrm{C}(G)$ and $u \in E$, and we conclude the proof by an appeal to Lemma 16.26 above. □

By Proposition 16.42 below, the irreducible unitary representations of a compact *Abelian* group coincides with its character group G^*. Hence we arrive at the following corollary.

Corollary 16.30. *If $\pi : G \to \mathscr{L}(E)$ is a continuous representation of a compact Abelian group G, then*

$$E = \overline{\operatorname{lin}}\big\{u \in E : \exists \chi \in G^* \text{ with } \pi_x u = \chi(x)u \,\forall\, x \in G\big\}.$$

We refer to Exercise 14 for a more information about representations of compact Abelian groups.

16.5 The Reversible Part Revisited

The preceding considerations can now be used to obtain a new characterisation of the reversible part of a JdLG-admissible semigroup $\mathscr{T}$ of operators on a Banach space E.

Similar to group representations, let us call a tuple $(e_j)_{j=1}^n$ of vectors of E a (finite) **unitary system** for a semigroup $\mathscr{T} \subseteq \mathscr{L}(E)$ if $(e_1,\dots,e_n)$ is linearly independent, $F := \operatorname{lin}\{e_1,\dots,e_n\}$ is $\mathscr{T}$-invariant, and the corresponding matrix representation of $\mathscr{T}$ on F defined by

$$Te_i = \sum_{j=1}^n \chi_{ij}(T)e_j \qquad (i = 1,\dots,n,\, T \in \mathscr{T}).$$

satisfies $\chi(T) \in \mathrm{U}(n)$ for each $T \in \mathscr{T}$. Furthermore, a unitary system $(e_j)_{j=1}^n$ for $\mathscr{T}$ is called **irreducible** if F does not have any nontrivial subspaces invariant under the action of $\mathscr{T}$.

Theorem 16.31 (Jacobs–de Leeuw–Glicksberg). *Let $E = E_{\mathrm{r}} \oplus E_{\mathrm{s}}$ be the JdLG-decomposition of a Banach space E with respect to a JdLG-admissible semigroup $\mathscr{T} \subseteq \mathscr{L}(E)$. Then*

$$E_{\mathrm{r}} = \overline{\operatorname{lin}}\{u :\ (e_j)_j \text{ (irred.) unitary sys. for } \mathscr{T},\ u = e_j \text{ for some } j\}.$$

Proof. We use the previous terminology, i.e., $\mathscr{S} = \mathrm{cl}_w(\mathscr{T})$ is the weak operator closure of $\mathscr{T}$, Q is the minimal idempotent, $E_{\mathrm{r}} = \operatorname{ran}(Q)$ and $\mathscr{G} := \mathscr{S}Q = Q\mathscr{S}$.

Let $B := (e_1,\dots,e_n)$ be some unitary system for $\mathscr{T}$ and denote by $F := \operatorname{lin}\{e_1,\dots,e_n\}$ the generated subspace. As F is closed and invariant under $\mathscr{T}$, it must be invariant under $\mathscr{S}$ as well. Let $\chi(S) \in \mathbb{C}^{n\times n}$ be the matrix representation of $S|_F$ with respect to the basis B. Then χ is continuous and a homomorphism of semigroups. By hypothesis, $\chi(\mathscr{T}) \subseteq \mathrm{U}(n)$, and since the latter is closed, $\chi(S) \in \mathrm{U}(n)$ for all $S \in \mathscr{S}$. In particular, $\chi(Q) \in \mathrm{U}(n)$, and since Q is an idempotent, $\chi(Q) = \mathrm{I}_n$, the $n \times n$ identity matrix. This means that Q acts as the identity on F, whence $F \subseteq E_{\mathrm{r}}$.

For the converse we first note that the compact group $\mathscr{G}$ acts continuously on E_{r}. Hence, by Theorem 16.29, E_{r} is the closed linear span of all irreducible unitary systems for this representation. But it is clear that each unitary system $(e_j)_{j=1}^n$ in E_{r} for $\mathscr{G}$ is a unitary system for $\mathscr{T}$, since $T = TQ$ on E_{r} for each $T \in \mathscr{T}$. Further, if it is irreducible for the $\mathscr{G}$-action, then it is irreducible for the $\mathscr{T}$-action as well, since $\mathscr{G} \subseteq \mathrm{cl}_w(\mathscr{T})$. □

We refer to Exercise 9 for a converse to Theorem 16.31. In the case of Abelian semigroups we obtain the following corollary.

Corollary 16.32. *Let E be a Banach space, and let $\mathscr{T} \subseteq \mathscr{L}(E)$ be an Abelian relatively weakly compact semigroup. Then*

$$E_{\mathrm{r}} = \overline{\mathrm{lin}}\big\{u \in E : \forall T \in \mathscr{T}\ \exists \lambda \in \mathbb{T} : Tu = \lambda u\big\}.$$

Example 16.33. Let H be a Hilbert space and let $\mathscr{T} \subseteq \mathscr{L}(H)$ be semigroup of contractions. Since the set of contractions is compact in $\mathscr{L}_w(H)$ by Corollary 16.2, $\mathscr{T}$ is relatively weakly compact. The associated JdLG-decomposition $H = H_{\mathrm{r}} \oplus H_{\mathrm{s}}$ is orthogonal since the JdLG-projection Q is a contraction, see Proposition 13.31. The group $\mathscr{S}_{\mathrm{r}}$ is hence a group of invertible contractions, i.e., of unitary operators. In particular, each $T \in \mathscr{T}$ restricts to a unitary operator on H_{r}.

However, the reversible part can be trivial even if $\mathscr{T}$ is itself a unitary group. To see this, take $H := \ell^2(\mathbb{Z})$ and T the (right or left) shift. Then T is a unitary operator and hence $\mathscr{T} := \{T^n : n \in \mathbb{Z}\}$ is a relatively weakly compact unitary group. But $T^n f \to 0$ weakly for every $f \in H$, whence $H = H_{\mathrm{s}}$ and $H_{\mathrm{r}} = \{0\}$.

Finally, we apply our results to the semigroup $\mathscr{T}_T$ generated by a single operator $T \in \mathscr{L}(E)$.

Theorem 16.34. *Let $T \in \mathscr{L}(E)$ generate a relatively weakly compact semigroup $\mathscr{T}_T$. Then the following assertions hold.*

a) *$E_{\mathrm{r}}(T)$ is spanned by the eigenvectors associated with unimodular eigenvalues of T, i.e.,*

$$E_{\mathrm{r}}(T) = \overline{\mathrm{lin}}\big\{u \in E : \exists \lambda \in \mathbb{T} \text{ with } Tu = \lambda u\big\}.$$

b) $\mathrm{fix}(T) \subseteq E_{\mathrm{r}}(T)$ *and* $E_{\mathrm{s}}(T) \subseteq \overline{\mathrm{ran}}(\mathrm{I} - T)$.

c) $E_{\mathrm{s}}(T) = \overline{\mathrm{ran}}(\mathrm{I} - T)$ *if and only if* $\sigma_p(T) \cap \mathbb{T} = \{1\}$.

d) $E = E_{\mathrm{s}}(T)$ *if and only if* $\sigma_p(T) \cap \mathbb{T} = \emptyset$.

Proof. a) is an immediate consequence of Corollary 16.32 since if $Tu = \lambda u$, then $T^n u = \lambda^n u$ for all $n \in \mathbb{N}_0$.

b) The inclusion $\mathrm{fix}(T) \subseteq E_{\mathrm{r}}$ follows from a), but is actually elementary since the JdLG-projection Q is in $\mathrm{cl}_w \mathscr{T}_T$. The second inclusion can be seen from Corollary 9.15 together with Theorem 16.19. However, one can use a more direct argument:

If $u \in E_s$ then $0 \in \mathrm{cl}_\sigma\{T^n u : n \in \mathbb{N}_0\}$. But then $0 \in \overline{\mathrm{conv}}\{T^n u : n \in \mathbb{N}_0\}$ and hence $u = u - 0 \in \overline{\mathrm{ran}}(\mathrm{I} - T)$ by Proposition 8.18.

c) follows from b) and a); and d) follows from a) directly. □

Let us apply Theorem 16.34 to the Koopman operator of an MDS $(\mathrm{X}; \varphi)$.

Corollary 16.35. *Let T be the Koopman operator on $\mathrm{L}^1(\mathrm{X})$ of an MDS $(\mathrm{X}; \varphi)$. Then the following assertions hold.*

a) *The Kronecker factor is spanned by the eigenfunctions of T associated with unimodular eigenvalues.*

b) *If the system is ergodic, each such eigenfunction is contained in $\mathrm{L}^\infty(G)$, in fact even unimodular up to a constant.*

c) *The system is weakly mixing if and only if it is ergodic and $\sigma_p(T) \cap \mathbb{T} = \{1\}$.*

Proof. a) holds by Theorem 16.34 and the definition of the Kronecker factor in Example 16.17. b) has been treated in Proposition 7.17. To prove c), note that by Theorem 16.20.(v) a system is weakly mixing if and only if $\mathrm{fix}(T) = \mathrm{lin}\{\mathbf{1}\}$ and $\mathrm{ran}(\mathrm{I} - T) \subseteq E_s(T)$. Hence c) follows from Theorem 16.34, part b) and c). □

We note that c) is in fact a spectral characterisation of weakly mixing systems, since the system is ergodic if and only if $\mathrm{fix}(T) = \mathrm{lin}\{\mathbf{1}\}$.

Supplement: Unitary Representations of Compact Groups

In the following G is a compact group with Haar measure m. For $E = H$ a Hilbert space a representation $\pi : G \to \mathscr{L}(E)$ is called **unitary** if it is strongly continuous, and $\pi_{x^{-1}} = \pi_x^*$ for each $x \in G$. That is to say, each operator π_x is a unitary operator on H. A unitary representation (π, H) is called **finite-dimensional** if $\dim H < \infty$.

If (π, H_π) and (ρ, H_ρ) are two unitary representations of G, an operator $A : H_\pi \to H_\rho$ is called **intertwining** if $A\pi_x = \rho_x A$ for all $x \in G$. A closed subspace F of H is called **reducing** if it is invariant under the action of G. Since $\pi(G) \subseteq \mathscr{L}(H)$ is invariant under taking adjoints, a closed subspace F is reducing if and only its orthogonal complement $F^\perp$ is reducing, and this happens precisely then, when the orthogonal projection P onto F is intertwining.

Every reducing subspace F gives rise to a **subrepresentation** by restriction: $\pi_x^F := \pi_x|_F$. A unitary representation is called **irreducible** if $\{0\}$ and H are the only reducing subspaces. By virtue of an induction argument one sees easily that a finite-dimensional unitary representation decomposes orthogonally into irreducible ones (however not in a unique manner).

Coordinate Functions

After a choice of orthonormal basis, a unitary representation on a finite-dimensional Hilbert space H is the same as a group homomorphism

$$\psi : G \to \mathrm{U}(n)$$

the matrix group of unitary $n \times n$ matrices. A continuous function $f \in \mathrm{C}(K)$ is called a **coordinate function** if there is a continuous unitary representation $\psi : G \to \mathrm{U}(n)$ and some indices i, j with $f = \psi_{ij}$. Equivalently, a coordinate function is any function of the form $f(x) = \big(\pi_x e_i \,\big|\, e_j\big)$, where $\pi : G \to \mathscr{L}(H)$ is a continuous unitary representation and H is a finite-dimensional Hilbert space with orthonormal basis $(e_j)_j$. If the representation is irreducible, f is called an **irreducible coordinate function**. The aim of the present section is mainly to prove the following central result.

Theorem 16.36. *Let G be a compact group. Then the linear span of irreducible coordinate functions is dense in* $\mathrm{C}(G)$.

The proof of this theorem is rather lengthy. First, we recall from the above that any finite-dimensional unitary representation decomposes orthogonally into irreducible representations. By picking an orthonormal basis subordinate to this decomposition, we see that any coordinate function is a linear combination of irreducible coordinate functions. So it remains to prove that the linear span of coordinate functions is dense in $\mathrm{C}(G)$. The strategy is, of course, to employ the Stone–Weierstrass Theorem 4.4. The following auxiliary result takes the first step.

Lemma 16.37. *The product of two coordinate functions is again a coordinate function. The pointwise conjugate of a coordinate function is a coordinate function.*

Proof. Let $\pi : G \to \mathrm{U}(n)$ and $\rho : G \to \mathrm{U}(m)$ be two unitary representations of the compact group G. Denote $\pi_x = (\pi_{ij}(x))_{i,j}$ and $\rho_x = (\rho_{kl}(x))_{k,l}$. Define $\chi : G \to \mathbb{C}^{nm \times nm}$ by

$$\chi_x = (\chi_{\alpha\beta}(x))_{\alpha,\beta}, \qquad \chi_{\alpha\beta}(x) := \pi_{ij}(x)\rho_{kl}(x) \qquad (\alpha = (i,k), \beta = (k,j));$$

and $\overline{\pi} : G \to \mathbb{C}^{n \times n}$ by

$$\overline{\pi}_x := (\overline{\pi_{ij}(x)})_{i,j}.$$

Then, by direct verification, $\chi : G \to \mathrm{U}(nm)$ and $\overline{\pi} : G \to \mathrm{U}(n)$ are again unitary representations of G (Exercise 10). □

Remark 16.38. The constructions in the preceding proof can be formulated more perspicuously in the language of abstract representation theory. Namely, let (π, H) and (ρ, K) be unitary representations of G then one can form the tensor product representation $(W, H \otimes K)$ by $W_x := \pi_x \otimes \rho_x$, and the contragredient representation $(\overline{\pi}, H')$ on the dual space H' of H by $\overline{\pi}_x := \pi'_{x^{-1}}$. The matrix representations of these constructions lead to the formulation above.

It follows from Lemma 16.37 that the linear span of the coordinate functions is a conjugation invariant subalgebra of $\mathrm{C}(G)$. Since the constant function $\mathbf{1}$ is a coordinate functions (of the trivial representation $\pi_x := \mathrm{I}$ on $\mathbb{C}^1$), it remains to show that the coordinate functions separate the points.

Theorem 16.39. *The finite-dimensional unitary representations (equivalently, the set of coordinate functions) of a compact group G separate the points of G.*

The proof of Theorem 16.39 needs some technical preparations.

Convolution and the Left Regular Representation

For $f \in \mathrm{L}^1(G)$ and $x \in G$ we define

$$f^*(y) := \overline{f(y^{-1})} \text{and} \quad (\tau_x f)(y) := f(x^{-1}y) \qquad (y \in G). \tag{16.5}$$

By invariance of the Haar measure and simple algebra

$$\tau : G \to \mathscr{L}(\mathrm{L}^p(G)), \qquad x \mapsto \tau_x$$

is a representation of G, called the **left regular representation** of G on $\mathrm{L}^p(G)$. In the case $p = 2$ this representations is unitary. Moreover, it follows easily from Corollary 14.12 that this representation is continuous. Hence, for $f \in \mathrm{L}^1(G)$ we can define as in (16.3) the integral

$$\tau_f := \int_G f(x)\tau_x \,\mathrm{d}x$$

in the weak sense, see Remark 16.25. Applying this operator to $g \in \mathrm{L}^1(G)$ yields the **convolution**

$$f * g := \tau_f(g) = \int_G f(x)\tau_x g \,\mathrm{d}x.$$

If $g \in \mathrm{C}(G)$, then the map $x \mapsto \tau_x g$ is even continuous in $\mathrm{C}(G)$ (Corollary 14.12), whence we can view τ as a continuous representation of G on $\mathrm{C}(G)$. It follows that in this case $f * g = \tau_f(g) \in \mathrm{C}(G)$ again, and we can evaluate pointwise to obtain

$$(f * g)(y) = \int_G f(x)(\tau_x g)(y) \,\mathrm{d}x = \int_G f(x) g(x^{-1}y) \,\mathrm{d}x \tag{16.6}$$

for $y \in G$, by the invariance of the Haar measure. We note the following useful facts.

Lemma 16.40. *For $f, g, h \in \mathrm{L}^1(G)$ and $x \in G$ the following statements are true.*

a) $\|f * g\|_1 \le \|f\|_1 \|g\|_1$.

b) $(f * g)^* = g^* * f^*$.

c) $\tau_x \tau_f g = \tau_{\tau_x f} g$

d) $(h * f) * g = h * (f * g)$.

Furthermore, if $f,g \in \mathrm{L}^2(G)$ then $f*g \in \mathrm{C}(G)$ and

$$\|f*g\|_\infty \le \|f\|_2 \|g\|_2, \quad \|f*g\|_2 \le \|f\|_2 \|g\|_1, \quad (f*f^*)(1) = \|f\|_2^2.$$

Proof. a) By (16.4) we have

$$\|f*g\|_1 = \|\tau_f g\| \le \|\tau_f\| \, \|g\|_1 \le \|f\|_1 \|g\|_1,$$

since each τ_x, $x \in G$, is an isometry on $\mathrm{L}^1(G)$.

b) For $f,g \in \mathrm{C}(G)$ this is a simple computation; for general $f,g \in \mathrm{L}^1(G)$ use an approximation argument and a).

c) This has been established already in the proof of Corollary 16.27 for a general representation.

d) Take $u \in \mathrm{L}^\infty(G)$. Then by c)

$$\langle \tau_h \tau_f g, u \rangle = \int_G h(x) \langle \tau_x \tau_f g, u \rangle \, \mathrm{d}x = \int_G h(x) \langle \tau_{\tau_x f} g, u \rangle \, \mathrm{d}x = \langle \tau_{\tau_h f} g, u \rangle$$

since $v \mapsto \langle \tau_v g, u \rangle$ is a continuous functional on $\mathrm{L}^1(G)$.

For the remaining part we note first that $f*g \in \mathrm{C}(G)$ if f,g are both continuous. The estimate $\|f*g\|_\infty \le \|f\|_2 \|g\|_2$ is an easy consequence of the Cauchy–Schwarz inequality and the invariance of the Haar measure. The other norm estimate follows from

$$\|f*g\|_2 = \|(f*g)^*\|_2 = \|g^**f^*\|_2 = \|\tau_{g^*} f^*\|_2 \le \|g^*\|_1 \|f^*\|_2 = \|f\|_2 \|g\|_1$$

where we used again (16.4), the fact that that τ is a unitary representation on $\mathrm{L}^2(G)$, and the reflection invariance of the Haar measure. The statements for general $f,g \in \mathrm{L}^2(G)$ follow by approximation with continuous functions. □

Our major tool for proving Theorem 16.39 is the following result.

Proposition 16.41. *Let $0 \neq f \in \mathrm{L}^2(G)$. Then there is a finite-dimensional unitary representation (π, H) of G such that $\pi_f \neq 0$.*

Proof. Let $h := f*f^*$. Then $h^* = h$, $h \in \mathrm{C}(G)$ and $h(1) = \|f\|_2^2 \neq 0$, by Lemma 16.40. Since h is continuous and the Haar measure is strictly positive, $\|h\|_2 \neq 0$.

Define the operator A on $\mathrm{L}^2(G)$ by $Au := u*h$. By Lemma 16.40, A is bounded with $\|A\| \le \|h\|_1$. Moreover, it is easy to see that A is self-adjoint (Exercise 11). We note that

$$\tau_f A f^* = \tau_f (f^* * h) = (f*(f^**h)) = (f*f^**h = h*h.$$

Since $h^* = h$, $(h*h)(1) = \|h\|_2^2 > 0$, and as before it follows that $\tau_f A f^* \neq 0$. In particular, $A \neq 0$.

Next, we note that A is an integral operator with kernel $k(y,x) := h(x^{-1}y)$, since

$$(Au)(y) = (u * h)(y) = \int_G u(x)h(x^{-1}y)\,\mathrm{d}x = \int_G k(y,x)u(x)\,\mathrm{d}x \qquad (y \in G).$$

It follows that A is compact. (One can employ different reasonings here, see Exercise 12.) By the Spectral Theorem C.23, A can be written as a convergent sum

$$A = \sum\nolimits_{\lambda \in \Lambda} \lambda P_\lambda$$

where $\Lambda \subseteq \mathbb{C} \setminus \{0\}$ is a finite or countable set, for each $\lambda \in \Lambda$ the eigenspace $E_\lambda := \ker(\lambda \mathrm{I} - A)$ satisfies $0 < \dim E_\lambda < \infty$, and P_λ is the orthogonal projection onto E_λ. (Note that since $A \neq 0$, $\Lambda \neq \emptyset$.)

Finally, A intertwines τ with itself, since by Lemma 16.40.c) we have

$$\tau_x A u = \tau_x (u * h) = (\tau_x u) * h = A \tau_x u \qquad (u \in \mathrm{L}^2(G)).$$

This implies that each eigenspace E_λ is invariant under τ_x, and hence gives rise to the (finite-dimensional) subrepresentation $\pi^\lambda_x := \tau_x|_{E_\lambda}$. We claim that *there is at least one* $\lambda \in \Lambda$ *with* $\pi^\lambda_f \neq 0$. Indeed, since $\tau_f A f^* \neq 0$ we must have $\lambda \tau_f P_\lambda f^* \neq 0$ for some λ. But on E_λ we have $\tau_f = \pi^\lambda_f$, and hence the claim is proved. Taking $(\pi, H) := (\pi^\lambda, E_\lambda)$ concludes the proof of the proposition. □

Proof of Theorem 16.39. Let $a, b \in G$ with $a \neq b$. We have to devise a finite-dimensional unitary representation (π, H) of G with $\pi_a \neq \pi_b$. By passing to ab^{-1} we may suppose $a \neq 1$ and we aim for a representation with $\pi_a \neq \mathrm{I}$. We can pick a neighbourhood U of 1 with $aU \cap U = \emptyset$ and a continuous function $0 \leq u \in \mathrm{C}(G)$ with $u(1) > 0$ and $\mathrm{supp}(u) \subseteq U$. Hence the function $f := u - \tau_a u$ is nonzero. By Proposition 16.41 there is a finite-dimensional unitary representation (π, H) of G with $\pi_f \neq 0$. This implies that

$$0 \neq \pi_f = \pi_{u - \tau_a u} = \pi_u - \pi_{\tau_a u} = \pi_u - \pi_a \pi_u,$$

whence $\pi_a \neq \mathrm{I}$. □

Compact Abelian Groups

Let us consider the special situation that the compact group G is Abelian. Let (π, H) be a finite-dimensional irreducible unitary representation of G and let $x \in G$. Since $\dim H < \infty$, π_x must have an eigenvalue $\lambda(x) \in \mathbb{T}$. Since G is Abelian, π_x intertwines π with itself, and hence $\ker(\lambda(x)\mathrm{I} - \pi_x)$ is a reducing subspace. By irreducibility, it must be all of H and hence $\pi_x = \lambda(x)\mathrm{I}$. But then every one-dimensional subspace is reducing, whence $\dim H = 1$. We have proved the following.

Proposition 16.42. *A (finite-dimensional) unitary represention of a compact Abelian group is irreducible if and only if it is one-dimensional.*

Note that the unitary group of $\mathbb{C}^1$ is $\mathrm{U}(1) = \mathbb{T}$, and a one-dimensional representation of G is the same as a continuous group homomorphism $G \to \mathbb{T}$, i.e., a **character** of G. The set of characters is G^*, the dual group of G, introduced in Section 14.3. The following is hence an immediate consequence of Proposition 16.42 and Theorem 16.36.

Theorem 16.43. *For a compact Abelian group G the dual group G^* separates the points of G. The set* $\operatorname{lin} G^*$ *of trigonometric polynomials is a dense subalgebra of* $\mathrm{C}(G)$.

Remarks 16.44. 1) It follows from Theorem 16.29 that each infinite-dimensional (continuous) representation of a compact group G has a nontrivial finite-dimensional invariant subspace. In particular each irreducible unitary representation of G is finite-dimensional.

2) Two unitary representations (π, H) and (ρ, K) are called **equivalent** if there is a unitary operator $U : H \to K$ that intertwines the representations. The **Peter–Weyl theorem** states that $\mathrm{L}^2(G)$ decomposes orthogonally

$$\mathrm{L}^2(G) \cong \bigoplus\nolimits_{\alpha} \underbrace{E_\alpha \oplus \cdots \oplus E_\alpha}_{\dim(E_\alpha)\text{-times}}$$

into finite-dimensional subspaces E_α such that (1) each E_α reduces the left regular representation, (2) the induced subrepresentation (τ, E_α) is irreducible, (3) each irreducible representation of G equivalent to precisely one (τ, E_α). We refer to [Deitmar and Echterhoff (2009), Chapter 7] for further details.

Exercises

1. Give a proof of Theorem 16.1 for a reflexive Banach space E. (Hint: Mimic the Hilbert space proof.)

2. Let E be a Banach space. Show that the following sets are closed subsemigroups of $\mathscr{L}_w(E)$.

a) $\{T \in \mathscr{L}(E) : TQ = QT\}$, for some fixed operator $Q \in \mathscr{L}(E)$.

b) $\{T \in \mathscr{L}(E) : Tf = f\}$, for some fixed element $f \in E$.

3. Let $E = \mathrm{C}(K)$ or $E = \mathrm{L}^p(\mathrm{X})$ for some $1 \le p \le \infty$. Show that the following sets are closed subsemigroups of $\mathscr{L}_w(E)$.

a) $\{T \in \mathscr{L}(E) : T \ge 0\}$.

b) $\{T \in \mathscr{L}(E) : \overline{Tf} = T\overline{f} \ \forall\, f \in E\}$.

c) $\{T \in \mathscr{L}(E) : T \ge 0,\ Th \le h\}$ for some fixed element $h \in E_{\mathbb{R}}$.

4. Prove the analogue of Lemma 16.8 for the strong operator topology, i.e., the following.

Let E be a Banach space, let $\mathscr{T} \subseteq \mathscr{L}(E)$, and let $D \subseteq E$ be a subset such that $\operatorname{lin} D$ *is dense in E. Then the following assertions are equivalent.*

(i) *$\mathscr{T}$ is relatively strongly compact, i.e, relatively compact in the strong operator topology.*

(ii) *$\mathscr{T}x = \{Tx : T \in \mathscr{T}\}$ is relatively compact in E for each $x \in E$.*

(iii) *$\mathscr{T}$ is norm-bounded and $\mathscr{T}x = \{Tx : T \in \mathscr{T}\}$ is relatively compact in E for each $x \in D$.*

5. Let E be a strictly convex Banach space, and let $P \in \mathscr{L}(E)$ be a projection with $\|P\| \leq 1$. Show that for any $f \in E$

$$\|Pf\| = \|f\| \quad \Longleftrightarrow \quad Pf = f.$$

(Hint: Write $Pf = P\big(\frac{f+Pf}{2}\big)$.)

6. Let E be a Banach space, and $F \subseteq E$ a closed subspace. Show that the set

$$\mathscr{L}_F(E) = \big\{T \in \mathscr{L}(E) : T(F) \subseteq F\big\}$$

is weakly closed and the restriction mapping

$$\mathscr{L}_F(E) \to \mathscr{L}(F), \qquad T \mapsto T|_F$$

is continuous for the weak and strong operator topologies. (Hint: Use the Hahn–Banach theorem.)

7. Let A be an algebra over $\mathbb{R}$ and let $S \subseteq A$ be multiplicative, i.e., $S \cdot S \subseteq S$. Show that $\operatorname{conv}(S)$ is multiplicative, too. Show that if $st = ts$ holds for all $s,t \in S$ then it also holds for all $s,t \in \operatorname{conv}(S)$.

8. Let E be a separable Banach space. Show that there is a sequence $(x'_j)_{j\in\mathbb{N}} \subseteq E'$ with $\|x'_j\| = 1$ for all $j \in \mathbb{N}$ and such that $\{x_j : j \in \mathbb{N}\}$ separates the points of E. (Hint: Use the Hahn–Banach theorem.)

9. Let E be a Banach space, and let $\mathscr{T} \subseteq \mathscr{L}(E)$ be a norm-bounded subsemigroup. Recall from page 250 the notion of a (finite) unitary system for $\mathscr{T}$. Suppose that

$$E = \overline{\operatorname{lin}}\big\{u \in E : \exists \text{ fin. unit. sys. } (e_j)_j \text{ for } \mathscr{T} \text{ with } u \in \operatorname{lin}\{e_1,\dots,e_j\}\big\}$$

Show that $\mathscr{T}$ is relatively strongly compact, and that $\operatorname{cl}_w\mathscr{T} = \operatorname{cl}_s\mathscr{T}$ is a compact subgroup of $\mathscr{L}(E)$ with the identity operator as neutral element (cf. Proposition 17.1).

10. Let $\pi : G \to \mathrm{U}(n)$ and $\rho : G \to \mathrm{U}(m)$ be two unitary representations of the compact group G. Denote $\pi_x = (\pi_{ij}(x))_{i,j}$ and $\rho_x = (\rho_{kl}(x))_{k,l}$. Define $\chi : G \to \mathbb{C}^{nm\times nm}$ by

$$\chi_x = (\chi_{\alpha\beta}(x))_{\alpha,\beta}, \qquad \chi_{\alpha\beta}(x) := \pi_{ij}(x)\rho_{kl}(x) \qquad (\alpha = (i,k), \beta = (k,j));$$

and $\overline{\pi} : G \to \mathbb{C}^{n\times n}$ by

$$\overline{\pi}_x := (\overline{\pi_{ij}(x)})_{i,j}.$$

Show that $\chi : G \to \mathrm{U}(nm)$ and $\overline{\pi} : G \to \mathrm{U}(n)$ are again unitary representations of G.

11. Let $u, v \in \mathrm{L}^2(G)$ and $h \in \mathrm{L}^1(G)$. Show that

$$(u^* \,|\, v^*)_{\mathrm{L}^2} = (v \,|\, u)_{\mathrm{L}^2} \quad \text{and} \quad (u * h \,|\, v) = (u \,|\, v * h^*).$$

Conclude that the operator A in the proof of Proposition 16.41 is self-adjoint.
(Hint: Show the assertion for $h \in \mathrm{C}(G)$ first and then employ an approximation argument.)

12. Let K be a compact space, and let $\mu \in \mathrm{M}(K)$ be a positive measure. Further, let $k \in \mathrm{C}(K \times K)$ and define

$$(Tf)(y) := \int_K k(y,x) f(x)\, \mu(\mathrm{d}x) \qquad (f \in \mathrm{L}^2(K,\mu), y \in K).$$

Show that T is a compact operator through the following ways:

a) k can be approximated uniformly on $K \times K$ by finite linear combinations of functions of the form $f \otimes g$, $f, g \in \mathrm{C}(K)$, cf. the proof of Proposition 10.11. Show that this leads to an approximation in operator norm of T by finite rank operators.

b) Realise that T is a Hilbert–Schmidt operator and as such is compact, see [Young (1988), Sec. 8.1] or [Deitmar and Echterhoff (2009), Sec. 5.3].

13. Let $\pi : G \to \mathscr{L}(E)$ be a continuous representation of a compact group G on a Banach space E. For $f \in \mathrm{L}^1(G)$ define π_f as in (16.3). Show that

$$\pi_f \pi_g = \pi_{f*g}$$

where $f * g$ denotes convolution of f, g as in Section (16.6). Then show that if $E = H$ is a Hilbert space,

$$(\pi_f)^* = \pi_{f^*}$$

where $f^*(x) = \overline{f(x^{-1})}$ as in (16.5).

14. Let G be a compact Abelian group and let $\pi : G \to \mathscr{L}(E)$ be a continuous representation of G. For a character $\chi \in G^*$ consider the operator $P_\chi := \pi_{\overline{\chi}}$, i.e.,

$$P_\chi u = \int_G \overline{\chi(x)}\, \pi_x u \,\mathrm{d}x \qquad (u \in E).$$

a) Show that $\chi * \eta = (\chi \,|\, \eta)\, \eta$ for any two characters $\chi, \eta \in G^*$. Then show that

$$P_\chi P_\eta = \delta_{\chi\eta} P_\eta \qquad (\chi, \eta \in G^*).$$

b) Show that $\operatorname{ran}(P_\chi) = \{u \in E : \pi_x u = \chi(x)u \text{ for all } x \in G\}$ and

$$\mathrm{I} = \sum\nolimits_{\chi \in G^*} P_\chi$$

the sum being convergent in the strong sense.

c) Show that if $E = H$ is a Hilbert space, then each P_χ is self-adjoint and

$$H = \bigoplus\nolimits_{\chi \in G^*} \operatorname{ran} P_\chi$$

is an orthogonal decomposition of H.

Chapter 17
Dynamical Systems with Discrete Spectrum

In the previous chapter we obtained important information on the structure of weakly compact semigroups of linear operators on Banach spaces. The central example is the closure $\mathscr{S}$ in the weak operator topology of a semigroup $\mathscr{T}_T := \{T^n : n \in \mathbb{N}_0\}$ generated by a single operator $T \in \mathscr{L}(E)$, E a Banach space. In such a compact semitopological semigroup we have found a smallest ideal $K(\mathscr{S})$, which is actually a group. The unit element Q of this group is a projection inducing the JdLG-decomposition $E = E_{\mathrm{r}} \oplus E_{\mathrm{s}}$.

In this chapter, with applications to dynamical systems in mind, we confine our attention to the case when $E = E_{\mathrm{r}}$, i.e., $Q = \mathrm{I}$ is the identity operator. Equivalently, we suppose that $K(\mathscr{S}) = \mathscr{S}$ consists of invertible operators only. By Theorem 16.34 this implies that E is spanned by the eigenvectors of T corresponding to unimodular eigenvalues. In the first section we study this property.

17.1 Operators with Discrete Spectrum

Let E be a Banach space. An operator $T \in \mathscr{L}(E)$ is said to have **discrete spectrum** on a Banach space E if it is power-bounded and E is spanned by the eigenvectors of T corresponding to unimodular eigenvalues, i.e.,

$$E = \overline{\mathrm{lin}}\big\{x : \exists\, \lambda \in \mathbb{T} : Tx = \lambda x\big\}. \tag{17.1}$$

The first result states that an operator with discrete spectrum generates a relatively compact semigroup.

Proposition 17.1. *Suppose $T \in \mathscr{L}(E)$ has discrete spectrum on the Banach space E. Then the semigroup*

$$\mathscr{T} := \big\{T^n : n \in \mathbb{N}_0\big\}$$

is relatively strongly compact. Moreover, on the (strong=weak) closure $\mathscr{S}$ of $\mathscr{T}$ the two topologies coincide, and $\mathscr{S}$ is a topological group with the identity operator as neutral element.

Proof. Note that Exercise 16.9 contains a more general statement, and the following proof works also there. By Exercise 16.4 we only have to prove that for all $u \in E$, u a linear combination of eigenvectors corresponding to unimodular eigenvalues, the orbit $\{T^n u : n \in \mathbb{N}\}$ is relatively compact in E. For such elements, however, the orbit is a bounded subset of a finite dimensional subspace and, as such, is relatively compact. The remaining assertions follow from the results in Section 16.4. □

As a consequence an operator with discrete spectrum is *mean ergodic*, by Example 16.16. The next proposition provides some detail about operators acting on the L^p scale.

Proposition 17.2. *Let* X *be a probability space and $T \in \mathrm{L}^1(\mathrm{X})$ be a Dunford–Schwartz operator. If T has discrete spectrum on $\mathrm{L}^p(\mathrm{X})$ for some $p \in [1,\infty)$, then it has discrete spectrum on spaces in the L^p-scale, $p \in [1,\infty)$.*

Proof. If T has discrete spectrum on L^p for some $p \in [1,\infty)$, then it has discrete spectrum on L^r for all $r \in [1,p]$, since for such r the space L^r is densely and continuously embedded in L^p. Therefore we may suppose that T has discrete spectrum on L^1. Then we have to prove that it has discrete spectrum on L^r for all $r \in (1,\infty)$. By Theorem 16.1 we know that T generates a relatively weakly compact semigroup in $\mathscr{L}(E)$, $E := \mathrm{L}^r(\mathrm{X})$. Hence, by the results from Section 16.2 we obtain the JdLG-decomposition $E = E_{\mathrm{s}} \oplus E_{\mathrm{r}}$. If $f \in E_{\mathrm{s}}$, then by Theorem 16.19 there is a sequence such that $T^{n_j} f \to 0$ weakly in E, but this implies $T^{n_j} f \to 0$ weakly in $\mathrm{L}^1(\mathrm{X})$, proving that T cannot have discrete spectrum on L^1. □

The aim is now to examine dynamical systems whose Koopman operator has discrete spectrum. The chief example for such dynamical systems is again the group rotation.

Example 17.3. Let G be a compact Abelian group and let $T := L_a$ be the Koopman operator induced by the rotation with $a \in G$. Since every character $\chi \in G^*$ is an eigenfunction of T corresponding to the eigenvalue $\chi(a) \in \mathbb{T}$ and since $\operatorname{lin} G^*$ is dense in $\mathrm{C}(G)$ (Proposition 14.18), T has discrete spectrum on $\mathrm{C}(G)$. A fortiori, T has discrete spectrum also on $\mathrm{L}^p(G)$ for every $1 \le p < \infty$, where μ is the Haar measure on G.

A TDS $(K;\varphi)$ is said to have **discrete spectrum** if its Koopman operator T on $\mathrm{C}(K)$ has discrete spectrum. If $(\mathrm{X};\varphi)$ is an MDS, then by Proposition 17.2 above, its Koopman operator has discrete spectrum on $\mathrm{L}^p(\mathrm{X})$ for some $p \in [1,\infty)$ if the Koopman operator has this property for all p in this range. In this case, we say that the MDS has discrete spectrum.

If we have a topological or a measure-preserving dynamical system with discrete spectrum, its Koopman operator T is a bijective isometry and hence satisfies $\sigma(T) \subseteq$

$\mathbb{T}$ by Proposition C.19. If in addition the TDS is minimal (or the MDS is ergodic), then the eigenspaces of T are all one-dimensional, by Theorem 4.19 (or Proposition 7.17). We shall show that the dynamical system is completely determined (up to isomorphism) by this spectral information. Indeed, it will turn out that the dynamical system is essentially a minimal/ergodic rotation on a compact group. To begin with, let us have a closer look at this model situation.

17.2 Monothetic Groups

A topological group G is called **monothetic** if there is an element $a \in G$ such that the cyclic subgroup $\langle a \rangle = \{a^n : n \in \mathbb{Z}\}$ is dense in G; in this case a is called a **generating element** of G. By Exercise 14.2 monothetic groups are necessarily Abelian. The next proposition yields some information about locally compact monothetic groups.

Proposition 17.4 (Weil's Lemma). *Let G be a locally compact monothetic group with generating element $a \in G$ (hence G is Abelian). Then G is either topologically isomorphic to $\mathbb{Z}$ or compact. In the latter case (and only then) the set $\{a^n : n \in \mathbb{N}_0\}$ is dense in G.*

Proof. Suppose that G and $\mathbb{Z}$ are not isomorphic. We prove that the set $\{a^n : n \in \mathbb{N}\}$ is dense in G. Let U be a nonempty open subset in G. By assumption there is $n \in \mathbb{Z}$ with $a^n \in U$. Take a symmetric neighbourhood of 1 with $a^n V \subseteq U$. If $a^k \in V$ holds only for finitely many $k \in \mathbb{Z}$, then $\langle a \rangle$ would be discrete and infinite, and hence G would be isomorphic to $\mathbb{Z}$, a contradiction. This shows that there is $k \in \mathbb{N}$ such that $n+k \geq 0$ and $a^{n+k} \in a^n V \subseteq U$, yielding the required denseness assertion.

Still assuming that G and $\mathbb{Z}$ are not isomorphic, we show that G is compact. Let U be a symmetric open neighbourhood of $1 \in G$ such that $\overline{U}$ is compact. Since we already saw that $A := \{a^n : n \in \mathbb{N}_0\}$ is dense in G, we obtain that $G = AU = A\overline{U}$. This means that for every $g \in G$ there is $n \in \mathbb{N}_0$ with $a^n \in gU$. Take the minimal n with this property and denote it by n_g. By the compactness of $\overline{U}$ there is $m \in \mathbb{N}_0$ such that for the finite set $F := \{a^0, a^1, \ldots, a^m\}$ one has $\overline{U} \subseteq FU$. By definition we have $a^{n_g} \in gU \subseteq gFU$, i.e., one has $a^{n_g - j} \in gU$ for some $j \leq m$. From the minimality of $n_g \in \mathbb{N}_0$ we obtain $n_g - j \leq 0$, i.e., $n_g \leq k$. This proves that $g \in a^{j-n_g}U^{-1} = a^{j-n_g}U \subseteq F\overline{U}$, so $G = F\overline{U}$, which is a compact set. The assertion is thus proved. □

For *compact groups* we have the next characterisations of being monothetic.

Proposition 17.5. *For a compact group G with Haar measure* m *and an element $a \in G$ the following statements are equivalent.*

(i) *The group G is monothetic with generating element a.*

(ii) *The cyclic subgroup $\langle a \rangle := \{a^n : n \in \mathbb{Z}\}$ is dense in G.*

(iii) *The TDS $(G; a)$ is minimal.*

(iv) *The MDS $(G, \mathrm{m}; a)$ is ergodic.*

(v) *The element a separates G^*.*

Proof. The equivalence of the first four statement was already stated in Theorem 10.13 (together with a longer list of equivalences). By continuity, two characters are equal if they coincide on a dense subset, hence (ii) implies (v). By Corollary 14.16 if a separates G^*, then $\langle a\rangle \subseteq G$ must be dense. □

An immediate consequence of this and Proposition 17.1 is the following.

Corollary 17.6. *Suppose $T \in \mathscr{L}(E)$ has discrete spectrum on the Banach space E. Then the weak closure*

$$\mathscr{G} := \mathrm{cl}_w\big\{T^n : n \in \mathbb{N}_0\big\} \subseteq \mathscr{L}_w(E)$$

is a compact monothetic group and the rotation TDS $(\mathscr{G};T)$ is minimal.

We now characterise the duals of compact monothetic groups.

Proposition 17.7. *Let G be a compact group. If G is monothetic with generating element $a \in G$, then G^* is (algebraically) isomorphic to the group*

$$G^*(a) := \big\{\chi(a) : \chi \in G^*\big\} \subseteq \mathbb{T}.$$

Conversely, if G^* is algebraically isomorphic to a subgroup of $\mathbb{T}$, then G is monothetic.

Proof. Suppose that a generates G. Then the set $G^*(a)$ is a subgroup of $\mathbb{T}$ and the evaluation mapping $G^* \ni \chi \mapsto \chi(a)$ is a surjective group homomorphism. Furthermore, it is even injective by Proposition 17.5.

For the converse, suppose that $\gamma\colon G^* \to \mathbb{T}$ is an injective homomorphism, hence γ is a character of G^*. By Pontryagin's Theorem 14.26 there is $a \in G$ with $\gamma(\chi) = \chi(a)$ for all $\chi \in G^*$. We claim that $\langle a\rangle$ is dense in G. Indeed, if $H := \mathrm{cl}\langle a\rangle \neq G$, then by Corollary 14.16 there exists a character $\chi \neq \mathbf{1}$ of G being constant 1 on $\langle a\rangle$. But then γ is not injective, a contradiction. □

Recall from Section 14.5 the following interpretation of the set $G^*(a)$.

Proposition 17.8. *Let G be a compact Abelian group with Haar measure* m *and fix $a \in G$. Consider $T := L_a$ the Koopman operator of the rotation by a considered either as an operator on $E = \mathrm{C}(G)$ or on $E = \mathrm{L}^p(G)$, $p \in [1,\infty)$. Then*

$$G^*(a) = \sigma_p(T).$$

Proof. For the case when T is considered on $E = \mathrm{L}^2(G)$, this was proved in Theorem 14.30. Since the characters are continuous and are eigenvectors of T, the statement is also proved for the case $E = \mathrm{C}(G)$. Hence also the inclusion $G^*(a) \subseteq \sigma_p(T)$ is evident for any of the spaces L^p. Suppose $\lambda \notin G^*(a)$ is an eigenvalue with an L^p-eigenvector $f \neq 0$. Since the characters belong to $\mathrm{C}(G) \subseteq \mathrm{L}^q(G)$, q the conjugate

exponent, we obtain $0 = \langle Tf - \lambda f, \chi \rangle = (\overline{\chi(a)} - \lambda)\langle f, \chi \rangle$. The assumptions imply $\langle f, \chi \rangle = 0$ for all $\chi \in G^*$ which in turn yields $f = 0$, a contradiction (use that $\operatorname{lin} G^*$ is dense in L^p). □

The next result characterises isomorphic minimal/ergodic group rotations.

Proposition 17.9. *For two monothetic compact groups G and H with generating elements $a \in G$ and $b \in H$ the following statements are equivalent.*

(i) *The TDSs $(G;a)$ and $(H;b)$ are isomorphic.*

(ii) *The MDSs $(G, \mathrm{m}_G; a)$ and $(H, \mathrm{m}_H; b)$ are isomorphic.*

(iii) *There is a topological group isomorphism $\Phi : G \to H$ with $\Phi(a) = b$.*

(iv) $G^*(a) = H^*(b)$.

Proof. (iii)$\Rightarrow$(i), (ii): If Φ is a topological group isomorphism with $\Phi(a) = b$, then Φ is also an isomorphism of the dynamical systems.

(i) or (ii)$\Rightarrow$(iv): If the dynamical systems are isomorphic, the Koopman operators on the corresponding spaces are similar. Hence they have the same spectral properties, in particular they have the same point spectrum. So (iv) follows from Proposition 17.8.

(iv)$\Rightarrow$(iii): If $G^*(a) = H^*(b)$, then by Proposition 17.7 the discrete groups G^* and H^* are isomorphic. Recall that this isomorphism is given by

$$G^* \ni \chi \mapsto \chi(a) = \eta(b) \mapsto \eta \in H^*.$$

By Pontryagin's Theorem 14.26 this implies that G and H are topologically isomorphic under $\Phi : G \to H$, where $\Phi(g) \in H$ is the unique element with $\eta(\Phi(g)) = \chi(g)$ if $\chi(a) = \eta(b)$, $\chi \in G^*$, $\eta \in H^*$. Now $\Phi(a) = b$, and the assertion follows. □

Example 17.10. Let H be an arbitrary subgroup of $\mathbb{T}_\mathrm{d}$. By Proposition 17.7, the compact group $G := H^*$ is monothetic with some generating element $a \in G$. As shown before, the group rotation $(G;a)$ is minimal and has discrete spectrum (see Proposition 17.5 and Example 17.3). Its Koopman operator T on $\mathrm{C}(G)$ has point spectrum

$$\sigma_p(T) = \big\{\chi(a) : \chi \in G^* = H^{**} = H\big\} = H,$$

by Theorem 14.30 and Proposition 17.7.

The next section shows that this example is quite general.

17.3 Dynamical Systems with Discrete Spectrum

In this section, we study dynamical systems with discrete spectrum, both in the topological and in the measure-preserving situation.

Minimal TDSs with Discrete Spectrum

We first confine our attention to the topological case. Our aim is to give a complete description (up to isomorphism) of minimal TDSs with discrete spectrum.

If two TDSs are isomorphic, the Koopman operators on the corresponding $\mathrm{C}(K)$-spaces are similar and hence have the same spectral properties. In particular, they have the same point spectrum. This shows that the point spectrum of the Koopman operator is an **isomorphism invariant** of TDSs. On the other hand, Proposition 17.9 and Proposition 17.8 show that within the class of minimal group rotations, the point spectrum completely determines the system up to isomorphism. Therefore we say that it is a **complete isomorphism invariant** for this class. Our aim in this section is to show that it is a complete isomorphism invariant even for the (larger) class of minimal TDS with discrete spectrum. This is done by showing that every minimal TDS with discrete spectrum is essentially a minimal group rotation.

Theorem 17.11. *Let $(K;\varphi)$ be a minimal TDS with discrete spectrum. Then it is isomorphic to a minimal rotation TDS on a compact monothetic group.*

Proof. By Corollary 17.6 the Koopman operator $T = T_\varphi$ on $\mathrm{C}(K)$ generates a compact group

$$\mathscr{G} := \mathrm{cl}_w\{T^n : n \in \mathbb{N}_0\} = \mathrm{cl}_s\{T^n : n \in \mathbb{N}_0\} \subseteq \mathscr{L}_s(\mathrm{C}(K)).$$

We now show that $(K;\varphi)$ is isomorphic to the rotation TDS $(\mathscr{G};T)$. The operator T and hence every $S \in \mathscr{G}$ is a Banach algebra homomorphism on $\mathrm{C}(K)$. Therefore, by Theorem 4.13, there exist *unique* continuous mappings

$$\psi_S : K \to K$$

such that $\quad Sf = f \circ \psi_S \qquad$ for every $S \in \mathscr{G}$, $f \in \mathrm{C}(K)$.

We also obtain

$$\varphi = \psi_T \quad \text{and} \quad \psi_{S_1 S_2} = \psi_{S_1} \circ \psi_{S_2} \qquad \text{for all } S_1, S_2 \in \mathscr{G}.$$

Pick $x_0 \in K$ and define

$$\Psi : \mathscr{G} \to K \quad \text{by} \quad \Psi(S) := \psi_S(x_0) \text{ for } S \in \mathscr{G}.$$

We show that this map yields an *isomorphism* between $(\mathscr{G};T)$ and $(K;\varphi)$. Indeed, let $S, R \in \mathscr{G}$, $f \in \mathrm{C}(K)$ and $\varepsilon > 0$ be given. Then

$$|f(\Psi(S)) - f(\Psi(R))| = |f(\psi_S(x_0)) - f(\psi_R(x_0))| = |Sf(x_0) - Rf(x_0)| \le \varepsilon,$$

whenever $\|Sf - Rf\|_\infty \le \varepsilon$. This shows the continuity of $f \circ \Psi$ for every $f \in \mathrm{C}(K)$ and, by Lemma 4.12, also the continuity of Ψ.

Observe that $\Psi(\mathscr{G})$ is a closed φ-invariant subset of K. Since $(K;\varphi)$ is minimal, it follows that $\Psi(\mathscr{G}) = K$, i.e., Ψ is surjective.

To prove injectivity, suppose that $\Psi(S_1) = \Psi(S_2)$ for some $S_1, S_2 \in \mathscr{G}$. Then

$$\psi_{S_1}(\varphi^n(x_0)) = \psi_{S_1}(\psi_T^n(x_0)) = \psi_T^n(\psi_{S_1}(x_0)) = \psi_T^n(\psi_{S_2}(x_0)) = \psi_{S_2}(\varphi^n(x_0))$$

for all $n \in \mathbb{N}_0$. By minimality, $\overline{\mathrm{orb}}_+(x_0) = K$. From the continuity of ψ_{S_1} and ψ_{S_2} we conclude that $\psi_{S_1} = \psi_{S_2}$, which yields $S_1 = S_2$.

Since $\varphi = \psi_T$, the diagram

$$\begin{array}{ccc} \mathscr{G} & \xrightarrow{S \mapsto TS} & \mathscr{G} \\ {\scriptstyle\Psi}\downarrow & & \downarrow{\scriptstyle\Psi} \\ K & \xrightarrow[\varphi]{} & K \end{array}$$

commutes, and the proof is hence complete. □

The above representation theorem and the characterisation of isomorphic minimal group rotations in Proposition 17.9 yield the next theorem.

Theorem 17.12. *Two minimal TDSs $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ with discrete spectrum are isomorphic if and only if the corresponding Koopman operators have the* same *point spectrum.*

Proof. Of course, operators induced by isomorphic TDSs always have the same point spectrum. For the converse, suppose that $(K_1;\varphi_1)$ and $(K_2;\varphi_2)$ are minimal TDSs having Koopman operators $T_1 := T_{\varphi_1}$, $T_2 := T_{\varphi_2}$ with discrete spectrum and the same point spectrum. The two TDSs are isomorphic to group rotations $(G_1;a_1)$ and $(G_2;a_2)$ by Theorem 17.11. But the Koopman operators of these TDSs have the same point spectrum, so by Proposition 17.9 they are isomorphic. □

Combining this theorem with Example 17.10 we see that for each subgroup $H \subseteq \mathbb{T}$ there is a (up to isomorphism) unique minimal TDS $(K;\varphi)$ with discrete spectrum whose Koopman operator has exactly H as its point spectrum.

Ergodic MDSs with Discrete Spectrum

If $(\mathrm{X};\varphi)$, $\mathrm{X} = (X,\Sigma,\mu)$, is an MDS with discrete spectrum, then the Koopman operator $T := T_\varphi$ is a bijective isometry on $\mathrm{L}^2(\mathrm{X})$, hence unitary. Analogously to what we have done above, we show that the isomorphism of two ergodic MDSs with discrete spectrum depends only on the point spectrum of the Koopman operators, and each ergodic MDS with discrete spectrum is isomorphic to an ergodic rotation on a compact group.

The approach is the following: Given an MDS $(\mathrm{X};\varphi)$ with discrete spectrum, we construct a topological model $(K,\nu;\psi)$ with corresponding Koopman operator having discrete spectrum in the space $\mathrm{C}(K)$. Then we can use the already established

TDS-result. The main point is to choose the model (i.e., the corresponding subalgebra $A \subseteq \mathrm{L}^\infty(\mathrm{X})$, see Section 12.3) carefully so that the TDS becomes minimal and has discrete spectrum.

The following result from [Halmos and von Neumann (1942)] is the fundamental step in the "classification" of ergodic MDSs with discrete spectrum.

Theorem 17.13 (Halmos–von Neumann). *Let $(\mathrm{X};\varphi)$ be an ergodic MDS such that the Koopman operator $T := T_\varphi$ has discrete spectrum in $\mathrm{L}^p(\mathrm{X})$ for some $1 \le p < \infty$. Then $(\mathrm{X};\varphi)$ is isomorphic to an ergodic rotation on a compact monothetic group endowed with the Haar measure.*

Proof. Every eigenfunction $f \in \mathrm{L}^p(\mathrm{X})$ of T belongs to $\mathrm{L}^\infty(\mathrm{X})$ by Corollary 16.35.b). Since the product of two eigenfunctions is again an eigenfunction, the linear hull of

$$\big\{ f \in \mathrm{L}^p(\mathrm{X}) : Tf = \lambda f \text{ for some } \lambda \in \mathbb{T} \big\}$$

is a conjugation invariant subalgebra of $\mathrm{L}^\infty(\mathrm{X})$, and its closure A in $\mathrm{L}^\infty(\mathrm{X})$ is a commutative C^*-algebra. By the Gelfand–Naimark Theorem 4.21 there exists a compact space K and an isometric $*$-isomorphism $\Phi : A \to \mathrm{C}(K)$. The restriction of T to A is an algebra homomorphism on A, therefore so is $S := \Phi \circ T \circ \Phi^{-1}$ on $\mathrm{C}(K)$. By Theorem 4.13, S is induced by some continuous mapping $\psi : K \to K$.

We show that $(K;\psi)$ is a minimal TDS with discrete spectrum. Since $T|_A$ has discrete spectrum on the space A, also S has discrete spectrum on the space $\mathrm{C}(K)$ and is mean ergodic thereon. We have $1 = \dim \mathrm{fix}(T) = \dim \mathrm{fix}(S)$, i.e., ψ is uniquely ergodic. The measure ν on K defined by

$$\langle f, \nu \rangle := \big\langle \Phi^{-1} f, \mu \big\rangle \quad (f \in \mathrm{C}(K))$$

is ψ-invariant and strictly positive, because μ is φ-invariant and strictly positive. This means that ψ is strictly ergodic, and by Proposition 10.9 the TDS $(K;\psi)$ is minimal.

We can now apply Theorem 17.12 to $(K;\psi)$ and obtain the existence of a compact monothetic group G with generating element a such that $(K;\psi)$ is isomorphic under some homeomorphism $\theta : G \to K$ to the rotation TDS $(G;a)$. This isomorphism induces an algebra isomorphism $T_\theta : \mathrm{C}(K) \to \mathrm{C}(G)$. We can sum up the above in the following commutative diagram.

$$\begin{array}{ccccc}
A & \xrightarrow{\Phi} & \mathrm{C}(K) & \xrightarrow{T_\theta} & \mathrm{C}(G) \\
\Big\downarrow{\scriptstyle T} & & \Big\downarrow{\scriptstyle S} & & \Big\downarrow{\scriptstyle L_a} \\
A & \xrightarrow[\Phi]{} & \mathrm{C}(K) & \xrightarrow[T_\theta]{} & \mathrm{C}(G)
\end{array}$$

where $(L_a)f(g) := f(ag)$ for $f \in \mathrm{C}(G)$ (the Koopman operator of the rotation by a). Now A, $\mathrm{C}(K)$ and $\mathrm{C}(G)$ are dense subspaces in $\mathrm{L}^p(\mathrm{X})$, $\mathrm{L}^p(K,\nu)$ and $\mathrm{L}^p(G)$, re-

spectively (where we take the Haar measure on G). Since $(T'_\theta)^{-1}\nu$ is an invariant measure for the minimal TDS $(G;a)$, it coincides with the Haar measure. Therefore T_θ is isometric for the corresponding L^p-norms. The Koopman operators T, L_a and T_θ are all isometric lattice isomorphisms if considered on the corresponding L^p-spaces (see Chapter 7). Similarly, Φ is isometric by the construction of ν. By Theorem 7.18, A is Banach lattice and $\Phi : A \to \mathrm{C}(K)$ is a lattice isomorphism. By denseness we can extend Φ to a positive isometry (hence a lattice isomorphism) $\widetilde{\Phi} : \mathrm{L}^p(\mathrm{X}) \to \mathrm{L}^p(K,\mu)$. We obtain that the diagram

$$\begin{array}{ccc} \mathrm{L}^p(\mathrm{X}) & \xrightarrow{T_\theta \circ \widetilde{\Phi}} & \mathrm{L}^p(G) \\ {\scriptstyle T_\varphi}\downarrow & & \downarrow{\scriptstyle T_\psi} \\ \mathrm{L}^p(\mathrm{X}) & \xrightarrow{T_\theta \circ \widetilde{\Phi}} & \mathrm{L}^p(G) \end{array}$$

is commutative, hence by Corollary 12.13 and its subsequent paragraph the claim is proved. □

As in the topological case we deduce from the above theorem that ergodic MDSs with discrete spectrum are completely determined by their point spectrum, i.e., point spectrum is a complete isomorphism invariant for this class of MDSs. The Halmos–von Neumann theorem is a crucial step in the study of the **isomorphism problem** of MDSs, i.e., in the quest for complete isomorphism invariants for dynamical systems. A historical account can be found in [Rédei and Werndl (2012)].

Corollary 17.14. *Two ergodic MDSs with Koopman operators having discrete spectrum in* L^p, $1 \leq p < \infty$, *are isomorphic if and only if the Koopman operators have the same point spectrum.*

A further type of isomorphism of MDSs can be defined as follows. Two MDSs $(\mathrm{X};\varphi)$ and $(\mathrm{Y};\psi)$ are called **spectrally isomorphic** if there is a Hilbert space isomorphism $S : \mathrm{L}^2(\mathrm{X}) \to \mathrm{L}^2(\mathrm{Y})$ intertwining the Koopman operators, i.e., $ST_\varphi = T_\psi S$ with the corresponding Koopman operators on the L^2-space. Isomorphic MDSs are spectrally isomorphic by Corollary 12.13 and in particular by the remark following it. Since Koopman operators of spectrally isomorphic MDSs have the same point spectrum the next result is an immediate consequence.

Corollary 17.15. *Two ergodic MDSs with discrete spectrum are isomorphic if and only if they are spectrally isomorphic.*

17.4 Examples

1. Rotations on Tori

The d-tori $G := \mathbb{T}^d$ are monothetic groups for $d \in \mathbb{N}$. Indeed, Kronecker's Theorem 14.31 describes precisely their generating elements. However, we can even form

"infinite dimensional" tori: Let I be a nonempty set, then the product $G := \mathbb{T}^I$ with the product topology and the pointwise operations is a compact group. The next theorem shows when this group is monothetic. The proof relies on the existence of sufficiently many rationally independent elements in $\mathbb{T}$.

Theorem 17.16. *Consider the compact group $G := \mathbb{T}^I$, I a nonempty set. Then G is monothetic if and only if* $\operatorname{card}(I) \le \operatorname{card}(\mathbb{T})$.

Proof. If G is monothetic, then by Proposition 17.7 we have that $\operatorname{card}(G^*) \le \operatorname{card}(\mathbb{T})$. Since the projections $\pi_\iota : \mathbb{T}^I \to \mathbb{T}$, $\iota \in I$, are all different characters of G, we obtain $\operatorname{card}(I) \le \operatorname{card}(G^*) \le \operatorname{card}(\mathbb{T})$.

Now suppose that $\operatorname{card}(I) \le \operatorname{card}(\mathbb{T})$. Take a Hamel basis $\mathscr{B}$ of $\mathbb{R}$ over $\mathbb{Q}$ such that $1 \in \mathscr{B}$. Then $\operatorname{card}(\mathscr{B}) = \operatorname{card}(\mathbb{T})$, so there is an injective function $f : I \to \mathscr{B} \setminus \{1\}$. Define $a := (a_\iota)_{\iota\in I} := (\mathrm{e}^{2\pi \mathrm{i} f(\iota)})_{\iota\in I} \in \mathbb{T}^I$. If $U \subseteq \mathbb{T}^I$ is a nonempty open rectangle, then by Kronecker's Theorem 14.31 we obtain $a^n \in U$ for some $n \in \mathbb{Z}$. □

The above product construction is very special and makes heavy use of the structure of $\mathbb{T}$. In fact, products of monothetic groups in general may not be monothetic (see Exercise 1). Notice also that these monothetic tori $\mathbb{T}^I$ are all connected (as the product of connected spaces).

2. Dyadic Integers

Consider the measure space $([0,1), \mathrm{Bo}, \lambda)$, λ the Lebesgue measure on $[0,1)$, and define

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

Then $([0,1), \mathrm{Bo}, \lambda; \varphi)$ is an MDS (see Exercise 5.6). Let $T := T_\varphi$ be its Koopman operator which is unitary on $\mathrm{L}^2([0,1), \mathrm{Bo}, \lambda)$. Define

$$g_m := \mathbf{1}_{[0,2^{-m})}$$

and

$$f_m := \sum_{k=0}^{2^m-1} \mathrm{e}^{-\frac{2\pi \mathrm{i} k}{2^m}} T^k g_m \qquad (m \in \mathbb{N}_0).$$

Notice that $T^{2^m} g_m = g_m$ and that $f_m \ne 0$ is an eigenvector of T, since

$$T f_m = \sum_{k=0}^{2^m-1} \mathrm{e}^{-\frac{2\pi \mathrm{i} k}{2^m}} T^{k+1} g_m = \mathrm{e}^{\frac{2\pi \mathrm{i}}{2^m}} \sum_{k=1}^{2^m} \mathrm{e}^{-\frac{2\pi \mathrm{i} k}{2^m}} T^k g_m = \mathrm{e}^{\frac{2\pi \mathrm{i}}{2^m}} f_m,$$

that is $\mathrm{e}^{\frac{2\pi \mathrm{i}}{2^m}} \in \sigma_p(T)$. Since T is multiplicative, it also follows that $T f_m^n = \mathrm{e}^{\frac{2\pi \mathrm{i} n}{2^m}} f_m^n$. Moreover, as T is unitary, f_m^n and $f_{m'}^{n'}$ are orthogonal in L^2 whenever $m \ne m'$ or $0 \le n \ne n' \le 2^m - 1$. Some computation gives

$$g_m = \frac{1}{2^m} \sum_{k=0}^{2^m-1} f_m^k.$$

So, as $\operatorname{lin}\{g_m : m \in \mathbb{N}_0\}$ is dense in L^2, $\operatorname{lin}\{f_m^k : m \in \mathbb{N}_0, 0 \le k < 2^m - 1\}$ is also dense, hence $\{f_m^k : m \in \mathbb{N}_0, 0 \le k < 2^m - 1\}$ forms an orthonormal basis in L^2. This shows that T has discrete spectrum and that

$$\sigma_p(T) = \left\{ \mathrm{e}^{\frac{2\pi i k}{2^m}} : m \in \mathbb{N}_0, 0 \le k \le 2^m - 1 \right\}.$$

By the usual argument of expanding a vector $f \in \operatorname{fix}(T)$ in the above orthonormal basis, we can prove that f is constant, hence by Theorem 7.14 the MDS $([0,1), \mathrm{Bo}, \lambda; \varphi)$ is ergodic (cf. proof of Theorem 14.30). Now the Halmos–von Neumann Theorem 17.13 tells that there is compact group such that the MDS is isomorphic to a group rotation. In our case, however, this group can be described concretely. Those who solved Exercise 14.8 might have noticed that $\sigma_p(T)$ is actually algebraically isomorphic to the dual group of the *dyadic integers* $\mathbb{A}$ (see Example 2.11). In fact, for $\mathbf{1} = (1,0,0,\dots) \in \mathbb{A}$ one has

$$\sigma_p(T) = \left\{ \mathrm{e}^{\frac{2\pi i k}{2^m}} : m \in \mathbb{N}_0, 0 \le k \le 2^m - 1 \right\} = \left\{ \chi(\mathbf{1}) : \chi \in \mathbb{A}^* \right\} = \sigma_p(T_{\varphi_\mathbf{1}}).$$

So by Corollary 17.14 $([0,1), \mathrm{Bo}, \lambda; \varphi)$ is isomorphic to $(\mathbb{A}, \Sigma, \mu; \mathbf{1})$, Σ the (product) Borel algebra, μ the Haar measure on $\mathbb{A}$ and $\varphi_\mathbf{1}$ the rotation by $\mathbf{1}$ (notice that the rotation on $\mathbb{A}$ is ergodic because it is minimal as $\operatorname{orb}_+(\mathbf{1})$ is dense in $\mathbb{A}$).

In contrast to the tori in Example 17.4.1, this group $\mathbb{A}$ is **totally disconnected**, i.e., every nonempty connected subset of $\mathbb{A}$ is a singleton (because it is a product of discrete spaces).

3. Bernoulli Shifts

The Bernoulli shift $B(p_0, p_1, \dots, p_{k-1}) = (\mathscr{W}_k^+, \Sigma, \mu; \tau)$, discussed in Example 5.1.5, is ergodic, but—except the trivial case—does not have discrete spectrum. This can be seen as follows. Denote by $T := T_\tau$ the Koopman operator on L^2 and recall from Proposition 6.16 that

$$\lim_{n\to\infty} \int_{\mathscr{W}_k^+} T^n \mathbf{1}_A \cdot \mathbf{1}_B \,\mathrm{d}\mu = \lim_{n\to\infty} \mu(\tau^{*n} A \cap B) = \mu(A)\mu(B) = \int_{\mathscr{W}_k^+} \mathbf{1}_A \cdot \mathbf{1}_B \,\mathrm{d}\mu$$

holds for all $A, B \in \Sigma$. It is now a routine argument to show that this holds for any $f, g \in \mathrm{L}^2(\mathscr{W}_k^+, \Sigma, \mu)$ replacing $\mathbf{1}_A$ and $\mathbf{1}_B$. If $f \in \ker(\lambda \mathrm{I} - T)$ for some $\lambda \in \mathbb{T}$, then

$$\lim_{n\to\infty} \lambda^n \langle f, g \rangle = \langle f, g \rangle$$

follows for all $g \in \mathrm{L}^2$. But this can only happen if $f = 0$ or $\lambda = 1$. This shows that $\sigma_p(T) \cap \mathbb{T} = \{1\}$, and the ergodicity implies $\dim \operatorname{fix}(T) = 1$, so T cannot have

discrete spectrum (except $\dim \mathrm{L}^2 = 1$). The next example puts the present one in a more general perspective.

4. Weak mixing vs. Discrete Spectrum

Let $(\mathrm{X};\varphi)$ be an MDS and consider its Koopman operator $T := T_\varphi$, say, on $E := \mathrm{L}^2(\mathrm{X})$. Since E is a Hilbert space and T is a contraction, the semigroup generated by T is JdLG-admissible, hence the space can be decomposed as

$$E := E_{\mathrm{r}} \oplus E_{\mathrm{s}}.$$

The restriction $T_{\mathrm{r}} := T|_{E_{\mathrm{r}}}$ has discrete spectrum on E_{r}, whereas $T_{\mathrm{s}} := T|_{E_{\mathrm{s}}}$ is almost weakly stable on E_{s}. Furthermore, by Theorem 16.20, the MDS is weakly mixing if and only if

$$E = \mathrm{lin}\{\mathbf{1}\} \oplus E_{\mathrm{s}},$$

i.e., if the reversible part is $\mathrm{lin}\{\mathbf{1}\}$. These yield the following evident fact.

Proposition 17.17. *Let $(\mathrm{X};\varphi)$ be an MDS that has discrete spectrum and is weakly mixing, then $\mathrm{L}^2(\mathrm{X}) = \mathrm{lin}\{\mathbf{1}\}$, hence this space is one-dimensional.*

The above proposition is an instance of what is called by T. Tao the *"dichotomy between structure and randomness"*, see [Tao (2007)].

Exercises

1. Show that a monothetic group is commutative. Describe all finite monothetic groups.

2. Prove that if G is a closed subgroup of $\mathbb{T}$, then $G = \mathbb{T}$ or G is finite cyclic group.

3. Prove that $(\mathbb{T};a)$ and $(\mathbb{T};a^{-1})$, $a \in \mathbb{T}$, are isomorphic.

4. Show that if $T \in \mathscr{L}(E)$, E a Banach space, is Cesàro-bounded, then T is mean ergodic with

$$\overline{\mathrm{ran}}(\mathrm{I}-T) = \overline{\mathrm{lin}}\{x \in E : \exists\, 1 \neq \lambda \in \mathbb{T} : Tx = \lambda x\}$$

5. Let T be a power-bounded operator on a Banach space E and let $A \subseteq \mathbb{T}$ such that the linear hull of

$$\bigcup_{\lambda \in A} \ker(\lambda \mathrm{I} - T)$$

is dense in E. Show that $\sigma_p(T) \subseteq A$.

6. Let $(\mathrm{X};\varphi)$ be an ergodic MDS whose Koopman operator $T := T_\varphi$ has discrete spectrum on $H := \mathrm{L}^2(\mathrm{X})$. Prove that if $\dim H = \infty$, then $\sigma(T) = \mathbb{T}$. (We note that the same is true even without the assumption of "discrete spectrum".)

7. Determine the isomorphism classes of rotations on $\mathbb{T}$.

8. Give an example of two nonisomorphic but spectrally isomorphic MDSs $(X;\varphi)$ and $(Y;\psi)$.

Chapter 18
A Glimpse at Arithmetic Progressions

In this chapter we shall discuss some of the many applications of ergodic theory to (combinatorial) number theory. An **arithmetic progression** of length $k \in \mathbb{N}$ is a set A of natural numbers of the form

$$\{a, a+n, a+2n, \ldots, a+(k-1)n\} = \{a+jn : 0 \le j < k\}$$

for some $a, n \in \mathbb{N}$. The number a is called the **starting point** and n is called the **distance** of this arithmetic progression. An arithmetic progression of length k is also called a k-term arithmetic progression.

One aspect of combinatorial number theory deals with the question how "big" a subset $A \subseteq \mathbb{N}$ must be in order to ensure that it contains infinitely many arithmetic progressions of a certain length, or even of arbitrary length. The following result of van der Waerden [van der Waerden (1927)] is the classical example for a statement of this kind.

Theorem 18.1 (Van der Waerden). *If $\mathbb{N} = A_1 \cup \ldots \cup A_m$, then there is $j \in \{1, \ldots, m\}$ such that*

$$\forall k\ \exists a\ \exists n : \quad a, a+n, a+2n, \ldots, a+(k-1)n \in A_j.$$

In other words, if you colour the natural numbers with finitely many colours, then there are arbitrarily long *monochromatic* arithmetic progressions, i.e., progressions all of whose members carry the same colour. (However, in general one cannot find a monochromatic progression of *infinite* length.)

Many years later, a major extension was proved by Szemerédi in [Szemerédi (1975)] using the concept of the *upper density*

$$\overline{\mathrm{d}}(A) = \limsup_{n\to\infty} \frac{\mathrm{card}(A \cap \{1, \ldots, n\})}{n}$$

of a subset $A \subseteq \mathbb{N}$ (cf. Section 9.2 and Exercise 9.1).

Theorem 18.2 (Szemerédi). *Any subset $A \subseteq \mathbb{N}$ with $\overline{\mathrm{d}}(A) > 0$ contains arbitrarily long arithmetic progressions.*

In the 1970's it was discovered that the results of van der Waerden and Szemerédi (and many others) could be obtained also by "dynamical" methods, i.e., by transferring them to statements about dynamical systems. Major contributions came from Furstenberg who was able to recover van der Waerden's theorem by methods from topological dynamics (see [Furstenberg (1981)], [Pollicott and Yuri (1998)]). Later on he used ergodic theory to obtain and extend the result of Szemerédi.

His ideas building heavily on the structure theory of ergodic MDSs were further developed by many authors and finally led to one of the most striking results in this area so far.

Theorem 18.3 (Green–Tao). *The set of prime numbers contains arbitrarily long arithmetic progressions.*

We refer to [Green and Tao (2008)] and in particular to Tao's ICM lecture [Tao (2007)] for the connection with ergodic theory.

The discovery of such a beautiful structure within the chaos of prime numbers was preceded and accompanied by various numerical experiments. Starting from the arithmetic progression $7, 37, 67, 97, 127, 157$ (of length 6 and distance 30) the actual world record (since 16 March, 2012 by J. Fry[2] is a progression of length 26 starting at

$$3\ 486\ 107\ 472\ 997\ 423$$

with distance

$$371\ 891\ 575\ 525\ 470.$$

However striking, the Green–Tao theorem still falls short of the following audacious conjecture formulated by Erdős and Turán in 1936.

Conjecture 18.4 (Erdős–Turán). Let $A \subseteq \mathbb{N}$ be such that $\sum_{a\in A}(1/a) = +\infty$. Then A contains arbitrarily long arithmetic progressions.

The Green–Tao theorem is—even in its ergodic theoretic parts—far beyond the scope of this book. However, we shall describe the major link between "density results" (like Szemerédi's) and ergodic theory and apply it to obtain weaker, but nevertheless stunning results (the theorems of Roth and Furstenberg–Sárközy, see below). Again, our major operator theoretic tool is the JdLG-decomposition.

18.1 The Furstenberg Correspondence Principle

Let us fix a subset $A \subseteq \mathbb{N}$ with positive upper density $\overline{\mathrm{d}}(A) > 0$. For given $k \in \mathbb{N}$ consider the following statement.

[2] http://users.cybercity.dk/~dsl522332/math/aprecords.htm
The quest for a 27-term arithmetic progression still continues.

$$\text{There exist } a,n \in \mathbb{N} \text{ such that } a, a+n, a+2n, \dots a+(k-1)n \in A. \qquad (\mathrm{AP}_k)$$

Our goal is to construct an associated dynamical system that allows us to reformulate (AP_k) in ergodic theoretic terms.

Consider the compact metric space $\mathscr{W}_2^+ = \{0,1\}^{\mathbb{N}_0}$ and the left shift τ on it. A subset $B \subseteq \mathbb{N}_0$ can be identified with a point in $\mathscr{W}_2^+$ via its characteristic function:

$$B \subseteq \mathbb{N}_0 \quad \longleftrightarrow \quad \mathbf{1}_B \in \mathscr{W}_2^+.$$

We further define

$$K := \overline{\{\tau^n \mathbf{1}_A : n \in \mathbb{N}_0\}} \subseteq \mathscr{W}_2^+.$$

Then K is a closed τ-invariant subset of $\mathscr{W}_2^+$, i.e., $(K;\tau)$ is a (forward transitive) TDS. The set

$$M := \left\{(x_n)_{n=0}^{\infty} \in K : \ x_0 = 1\right\}$$

is open and closed in K, and hence so is every set $\tau^{-j}(M)$ $(j \in \mathbb{N}_0)$. Note that we have

$$n \in A \quad \text{ if and only if } \quad \tau^n(\mathbf{1}_A) \in M. \qquad (18.1)$$

We now translate (AP_k) into a property of this dynamical system. By (18.1), we obtain for $a \in \mathbb{N}$ and $n \in \mathbb{N}_0$ the following equivalences.

$$\begin{aligned} &a, a+n, \dots, a+(k-1)n \in A \\ &\qquad \Longleftrightarrow \tau^a \mathbf{1}_A \in M, \ \tau^n \tau^a \mathbf{1}_A \in M, \dots, \ \tau^{(k-1)n} \tau^a \mathbf{1}_A \in M \\ &\qquad \Longleftrightarrow \tau^a \mathbf{1}_A \in M \cap \tau^{-n}(M) \cap \dots \cap \tau^{-(k-1)n}(M). \end{aligned}$$

Since each $\tau^{-j}(M)$ is open and $\{\tau^a \mathbf{1}_A : a \in \mathbb{N}\}$ is dense in K, we have that (AP_k) is equivalent to

$$\exists n \in \mathbb{N} : \ M \cap \tau^{-n}(M) \cap \dots \cap \tau^{-(k-1)n}(M) \neq \emptyset. \qquad (18.2)$$

The strategy to prove (18.2) is now to turn the TDS into an MDS by choosing an invariant measure μ in such a way that for some $n \in \mathbb{N}$ the set

$$M \cap \tau^{-n}(M) \cap \dots \cap \tau^{-(k-1)n}(M)$$

has positive measure. Now, since $\overline{\mathrm{d}}(A) > 0$, there is a subsequence $(n_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ such that

$$\lim_{j\to\infty} \frac{1}{n_j} \sum_{k=1}^{n_j} \mathbf{1}_A(k) = \overline{\mathrm{d}}(A) > 0.$$

With this subsequence we define a sequence of probability measures $(\mu_j)_{j\in\mathbb{N}}$ on K by

$$\mu_j(B) := \frac{1}{n_j} \sum_{k=1}^{n_j} \delta_{\tau^k \mathbf{1}_A}(B) \qquad (B \in \mathrm{Ba}(K), j \in \mathbb{N}),$$

where $\delta_{\tau^k \mathbf{1}_A}$ stands for the Dirac measure at $\tau^k \mathbf{1}_A$. By using (18.1) we have

$$\mu_j(M) = \frac{1}{n_j} \sum_{k=1}^{n_j} \mathbf{1}_A(k) \to \overline{d}(A) \quad \text{as } j \to \infty. \tag{18.3}$$

The metrisability of the compact set $\mathrm{M}^1(K)$ for the weak*-topology implies that there exists a subsequence (which we again denote by $(\mu_j)_{j\in\mathbb{N}}$) weakly* converging to some probability measure μ on K. Since M is open and closed, the characteristic function $\mathbf{1}_M : K \to \mathbb{R}$ is continuous, hence (18.3) implies that $\mu(M) = \overline{d}(A)$.

To show that μ is τ-invariant, take $f \in \mathrm{C}(K)$ and note that

$$\begin{aligned}\int_K (f\circ\tau)\,\mathrm{d}\mu_j - \int_K f\,\mathrm{d}\mu_j &= \frac{1}{n_j}\sum_{k=1}^{n_j}\Big(f(\tau^{k+1}\mathbf{1}_A) - f(\tau^k\mathbf{1}_A)\Big)\\ &= \frac{1}{n_j}\big(f(\tau^{n_j+1}\mathbf{1}_A) - f(\tau\mathbf{1}_A)\big)\end{aligned}$$

for every $j \in \mathbb{N}$. By construction of μ, the first term in this chain of equations converges to

$$\int_K (f\circ\tau)\,\mathrm{d}\mu - \int_K f\,\mathrm{d}\mu$$

as $j \to \infty$, the last one converges to 0 since $n_j \to \infty$, implying that μ is τ-invariant (cf. Remark 10.3.3).

In this way we obtain a topological MDS $(K,\mu;\tau)$ such that $\mu(M) = \overline{\mathrm{d}}(A) > 0$. Furthermore, we may suppose without loss of generality that this MDS is *ergodic*. Indeed,

$$\mathrm{M}^1_\tau(K) = \overline{\mathrm{conv}}\{\mu \in \mathrm{M}^1_\tau(K) \,:\, \mu \text{ is ergodic}\}$$

by (10.1). Since $\mathbf{1}_M \in \mathrm{C}(K)$ and since by the above there is at least one $\mu \in \mathrm{M}^1_\tau(K)$ with $\int_K \mathbf{1}_M\,\mathrm{d}\mu = \mu(M) > 0$, there must also be an ergodic measure with this property.

Consider now the associated Koopman operator $T := T_\tau$ on $\mathrm{L}^2(K,\mu)$. The existence of a k-term arithmetic progression in A, i.e., (AP_k), is certainly ensured by

$$\exists n \in \mathbb{N}: \;\; \mathbf{1}_M \cdot T^n\mathbf{1}_M \cdots T^{(k-1)n}\mathbf{1}_M \neq 0$$

or the even stronger statement

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \int_K \mathbf{1}_M \cdot (T^n\mathbf{1}_M)\cdots(T^{(k-1)n}\mathbf{1}_M)\,\mathrm{d}\mu \;>\; 0. \tag{18.4}$$

Hence we have established the following result from [Furstenberg (1977)].

Theorem 18.5 (Furstenberg Correspondence Principle). *If for every ergodic MDS* $(\mathrm{X};\varphi)$ *the Koopman operator* $T := T_\varphi$ *on* $\mathrm{L}^2(\mathrm{X})$ *satisfies the condition*

$$\limsup_{n\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_{\mathrm{X}} f \cdot (T^n f) \cdots (T^{(k-1)n} f) \; > 0 \tag{18.5}$$

for all $0 < f \in \mathrm{L}^\infty(\mathrm{X})$*, then* (AP_k) *holds for every* $A \subseteq \mathbb{N}$ *with* $\bar{\mathrm{d}}(A) > 0$*.*

(Here and in the following we write $f > 0$ when we mean $f \geq 0$ and $f \neq 0$.) In [Furstenberg (1977)] it was shown that (18.5) is indeed true for every ergodic MDS and every $k \in \mathbb{N}$. This completed the ergodic theoretic proof of Szemerédi's theorem. As a matter of fact, Furstenberg proved the truth of the apparently stronger version

$$\liminf_{n\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_{\mathrm{X}} f \cdot (T^n f) \cdots (T^{(k-1)n} f) \; > 0 \tag{18.6}$$

of (18.5) As we shall see in a moment, this makes no difference because, actually, the *limit exists*.

18.2 A Multiple Ergodic Theorem

Furstenberg, in his correspondence principle, worked with expressions of the form

$$\frac{1}{N} \sum_{n=0}^{N-1} \int_{\mathrm{X}} f \cdot (T^n f) \cdots (T^{(k-1)n} f) = \int_{\mathrm{X}} f \cdot \frac{1}{N} \sum_{n=0}^{N-1} (T^n f) \cdots (T^{(k-1)n} f).$$

Recently, B. Host and B. Kra in [Host and Kra (2005b)] discovered a surprising fact regarding the multiple Cesàro sums on the right-hand side.

Theorem 18.6 (Host–Kra). *Let* $(\mathrm{X}; \varphi)$ *be an ergodic MDS, and consider the Koopman operator* $T := T_\varphi$ *on* $\mathrm{L}^2(\mathrm{X})$*. Then the limit of the multiple ergodic averages*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (T^n f_1) \cdot (T^{2n} f_2) \cdots (T^{(k-1)n} f_{k-1}) \tag{18.7}$$

exists in $\mathrm{L}^2(\mathrm{X})$ *for all* $f_1, \ldots, f_{k-1} \in \mathrm{L}^\infty(\mathrm{X})$ *and* $k \geq 2$*.*

Note that for $k = 2$ the Host–Kra theorem is just von Neumann's Mean Ergodic Theorem 8.1 (and holds even for arbitrary $f_1 \in \mathrm{L}^2(\mathrm{X})$). We shall give a proof of the Host–Kra theorem for the case $k = 3$ here. The general case is the subject of Chapter 20.

Our major tool is the JdLG-decomposition of $E := \mathrm{L}^2(\mathrm{X})$ with respect to the semigroup $\mathscr{T}_T$ into

$$\mathrm{L}^2(\mathrm{X}) = \mathrm{ran}(Q) \oplus \ker(Q) = E_{\mathrm{r}} \oplus E_{\mathrm{s}},$$

see Example 16.17. The space E_{r} is the *Kronecker factor* and the projection

$$Q : \mathrm{L}^2(\mathrm{X}) \to E_{\mathrm{r}}$$

is a Markov projection. Let us recall some facts from Chapters 13 and 16.

Lemma 18.7 (Properties of the Kronecker Factor). *The space $E_{\mathrm{r}} \cap \mathrm{L}^\infty(\mathrm{X})$ is a closed subalgebra of $\mathrm{L}^\infty(\mathrm{X})$ and*

$$Q(f \cdot g) = (Qf) \cdot g \tag{18.8}$$

for all $f \in \mathrm{L}^\infty(\mathrm{X})$ and $g \in E_{\mathrm{r}} \cap \mathrm{L}^\infty(\mathrm{X})$. Moreover, E_{r} is generated by eigenfunctions corresponding to unimodular eigenvalues of T.

In order to prove that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (T^n f) \cdot (T^{2n} g) \tag{18.9}$$

exists for every $f, g \in \mathrm{L}^\infty(\mathrm{X})$, we may split f according to the JdLG-decomposition and consider the cases

$$(1) \quad f \in E_{\mathrm{s}} \qquad \text{and} \qquad (2) \quad f \in E_{\mathrm{r}}$$

separately. Case (1) is covered by the following lemma, which actually yields slightly more information. Notice that the proof is very similar to that of "weakly mixing of all orders", i.e., Theorem 9.25.

Lemma 18.8. *Let $f, g \in \mathrm{L}^\infty(\mathrm{X})$ such that $f \in E_{\mathrm{s}}$ or $g \in E_{\mathrm{s}}$. Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (T^n f) \cdot (T^{2n} g) = 0$$

in $\mathrm{L}^2(\mathrm{X})$.

Proof. The proof is an application of the van der Corput Lemma. We let $u_n := (T^n f) \cdot (T^{2n} g)$ and write

$$\begin{aligned}
(u_n \,|\, u_{n+m}) &= \int_{\mathrm{X}} (T^n f) \cdot (T^{2n} g) \cdot (T^{n+m} \overline{f}) \cdot (T^{2n+2m} \overline{g}) \\
&= \int_{\mathrm{X}} T^n \big[(f T^m \overline{f}) \cdot T^n (g T^{2m} \overline{g}) \big] \\
&= \int_{\mathrm{X}} (f T^m \overline{f}) \cdot T^n (g T^{2m} \overline{g}).
\end{aligned}$$

Hence, for fixed $m \in \mathbb{N}$

$$\frac{1}{N} \sum_{n=0}^{N-1} (u_n \,|\, u_{n+m}) = \int_{\mathrm{X}} (f T^m \overline{f}) \frac{1}{N} \sum_{n=0}^{N-1} T^n (g T^{2m} \overline{g}) \to \int_{\mathrm{X}} f T^m \overline{f} \cdot \int_{\mathrm{X}} g T^{2m} \overline{g}$$

as $N \to \infty$ since $(\mathrm{X};\varphi)$ is ergodic. Therefore we have

$$\begin{aligned}\gamma_m := \lim_{N\to\infty}\Big|\frac{1}{N}\sum_{n=0}^{N-1}(u_n\,|\,u_{n+m})\Big| &= \Big|\int_{\mathrm{X}}(fT^m\overline{f})\cdot\int_{\mathrm{X}}(gT^{2m}\overline{g})\Big| \\ &= \Big|(f\,|\,T^m f)\,\big(g\,\big|\,T^{2m}g\big)\Big|.\end{aligned}$$

If $f \in E_{\mathrm{s}}$, by Theorem 16.19, we obtain that

$$0 \le \frac{1}{N}\sum_{m=0}^{N-1}\gamma_m \le \frac{1}{N}\sum_{m=0}^{N-1}\big|(f\,|\,T^m f)\big|\cdot\|g\|_\infty^2 \to 0 \quad \text{as} \quad N\to\infty.$$

In the case that $g \in E_{\mathrm{s}}$, the reasoning is similar. Now, the van der Corput Lemma 9.24 implies that

$$\frac{1}{N}\sum_{n=0}^{N-1}(T^n f)\cdot(T^{2n}g) = \frac{1}{N}\sum_{n=0}^{N-1}u_n \to 0. \qquad \square$$

Proof of Theorem 18.6, case $k=3$: By Lemma 18.8 we may suppose that $f \in E_{\mathrm{r}}$. Let $g \in \mathrm{L}^\infty(\mathrm{X})$ be fixed, and define $(S_N)_{N\in\mathbb{N}} \subseteq \mathscr{L}(E)$ by

$$S_N f := \frac{1}{N}\sum_{n=0}^{N-1}(T^n f)\cdot(T^{2n}g).$$

If f is an eigenfunction of T with eigenvalue $\lambda \in \mathbb{T}$, then

$$S_N f = f\frac{1}{N}\sum_{n=0}^{N-1}(\lambda T^2)^n g$$

holds. Since the operator λT^2 is mean ergodic as well (see Theorem 8.6), we have that $(S_N f)_{N\in\mathbb{N}}$ converges as $N\to\infty$. This yields that $(S_N f)_{N\in\mathbb{N}}$ converges for every

$$f \in \operatorname{lin}\big\{h : Th = \lambda h \text{ for some } \lambda \in \mathbb{T}\big\},$$

which is a dense subset of E_{r} by Theorem 16.34. Since the estimate $\|S_N\| \le \|g\|_\infty$ is valid for all $N \in \mathbb{N}$, we obtain the convergence of $(S_N)_{N\in\mathbb{N}}$ on the whole E_{r}. $\square$

In order to obtain Theorem 18.2 it remains to show that the limit in the Host–Kra theorem is positive whenever $0 < f = f_1 = \ldots = f_{k-1}$.

Proposition 18.9. *In the setting of Theorem 18.6*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\cdot(T^n f)\cdot(T^{2n}f)\cdots(T^{(k-1)n}f) > 0$$

whenever $0 < f \in \mathrm{L}^\infty(\mathrm{X})$.

Again, we prove this result for $k = 3$ only. Our argument follows Furstenberg [Furstenberg (1981), Theorem 4.27].

Proof. It suffices to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot (T^n f) \cdot (T^{2n} f) \; > 0. \tag{18.10}$$

We have $f = f_s + f_r$ with $f_r := Qf \in E_r$ and $f_s = f - Qf \in E_s$. Since Q is a Markov operator and $f > 0$, it follows that $f_r > 0$ as well. Let $g, h \in L^\infty(X)$ be arbitrary. Since $E_r \perp E_s$ in $L^2(X)$ it follows from Lemma 18.7 that

$$\int_X f \cdot (T^n g) \cdot (T^{2n} h) \; = 0$$

whenever two of the functions f, g, h are in E_r and the remaining one is from E_s. By Lemma 18.8 we have

$$\frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot (T^n g) \cdot (T^{2n} h) \; \to 0$$

if at least one of the functions g, h is from E_s. So we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f \cdot (T^n f) \cdot (T^{2n} f) \; = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f_r \cdot (T^n f_r) \cdot (T^{2n} f_r).$$

We may therefore suppose without loss of generality that $f = f_r$.

Take now $0 < \varepsilon < \int_X f^3$ and suppose that $\|f\|_\infty \le 1$, again without loss of generality. We show that there is some subset $B \subseteq \mathbb{N}$ with bounded gaps (also called **relatively dense** or **syndetic**; see Definition 3.9.b) such that

$$\int_X f \cdot (T^n f) \cdot (T^{2n} f) \; > \int_X f^3 \; - \varepsilon \qquad \text{for all } n \in B. \tag{18.11}$$

Recall that this property of B means that there is $L \in \mathbb{N}$ such that every interval of length $L-1$ intersects B.

Since $f \in E_r$, there exists ψ of the form $\psi = c_1 g_1 + \ldots + c_j g_j$ for some eigenfunctions $g_1, \ldots, g_j$ of T corresponding to eigenvalues $\lambda_1, \ldots, \lambda_j \in \mathbb{T}$ such that $\|f - \psi\|_2 < \frac{\varepsilon}{6}$. Observe first that for every $\delta > 0$ there is a relatively dense set $B \subseteq \mathbb{N}$ such that

$$|\lambda_1^n - 1| < \delta, \; \ldots, \; |\lambda_j^n - 1| < \delta \quad \text{for all } n \in B.$$

Indeed, if we consider the rotation by $(\lambda_1, \ldots, \lambda_j)$ on $\mathbb{T}^j$, this is precisely the almost periodicity of the point $(1, \ldots, 1) \in \mathbb{T}^j$ (see Proposition 3.12). Now, by taking a suitable $\delta > 0$ depending on the coefficients $c_1, \ldots c_j$, we obtain

$$\|T^n\psi - \psi\|_2 < \frac{\varepsilon}{12}$$

and $$\|T^n f - f\|_2 \le \|T^n f - T^n\psi\|_2 + \|T^n\psi - \psi\|_2 + \|\psi - f\|_2 < \frac{\varepsilon}{2}$$

for all $n \in B$. Moreover, for $n \in B$ one has

$$\begin{aligned}\|T^{2n} f - f\|_2 &\le \|T^{2n} f - T^{2n}\psi\|_2 + \|T^{2n}\psi - T^n\psi\|_2 + \|T^n\psi - \psi\|_2 + \|\psi - f\|_2 \\ &\le 2\|f - \psi\|_2 + 2\|T^n\psi - \psi\|_2 < \frac{\varepsilon}{2}.\end{aligned}$$

Altogether we obtain

$$\begin{aligned}\Big|\int_X f\cdot T^n f\cdot T^{2n} f - \int_X f^3\Big| &\le \int_X f\cdot T^n f\cdot |T^{2n} f - f| + \int_X f^2\cdot |T^n f - f| \\ &\le \|f\|_\infty^2\big(\|T^{2n} f - f\|_2 + \|T^n f - f\|_2\big) < \varepsilon,\end{aligned}$$

and (18.11) is proved. Since $B \cap [jL,(j+1)L) \neq \emptyset$ for all $j \in \mathbb{N}_0$, it follows that

$$\begin{aligned}\frac{1}{LN}\sum_{n=0}^{LN-1}\int_X f\cdot T^n f\cdot T^{2n} f &\ge \frac{1}{LN}\sum_{n\in B, n=0}^{LN-1}\int_X f\cdot T^n f\cdot T^{2n} f \\ &\ge \frac{1}{L}\Big(\int_X f^3 - \varepsilon\Big) > 0.\end{aligned}$$

□

Combining the above results leads to the classical theorem of Roth, a precursor of Szemerédi's Theorem 18.2.

Theorem 18.10 (Roth). *Let $A \subseteq \mathbb{N}$ have positive upper density. Then A contains infinitely many arithmetic progressions of length* 3.

The full proof of Szemerédi's theorem is based on an inductive procedure involving "higher-oder" decompositions of a JdLG-flavour. See for instance [Furstenberg (1977), Furstenberg (1981)], [Petersen (1989)], [Tao (2008a)], [Green (2008)] or [Einsiedler and Ward (2011)].

18.3 The Furstenberg–Sárközy Theorem

We continue in the same spirit, but instead of arithmetic progressions we now consider pairs of the form $\{a, a+n^2\}$ for $a, n \in \mathbb{N}$.

Theorem 18.11 (Furstenberg–Sárközy). *Let $A \subseteq \mathbb{N}$ have positive upper density. Then there exist $a, n \in \mathbb{N}$ such that $a, a+n^2 \in A$.*

For a more general statement we refer to [Furstenberg (1981), Proposition 3.19]. In order to prove the above theorem we first need an appropriate correspondence principle.

Theorem 18.12 (Furstenberg Correspondence Principle for Squares). *If for every ergodic MDS* $(\mathrm{X};\varphi)$ *and its Koopman operator* $T := T_\varphi$ *on* $\mathrm{L}^2(\mathrm{X})$ *the condition*

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_{\mathrm{X}} f \cdot (T^{n^2} f) \; > 0 \tag{18.12}$$

is satisfied for all $0 < f \in \mathrm{L}^\infty(\mathrm{X})$*, then Theorem 18.11 holds.*

The proof is analogous to the one of Theorem 18.5. To show (18.12) we begin with the following "nonconventional" ergodic theorem.

Theorem 18.13 (Mean Ergodic Theorem for Squares). *Let* T *be a contraction on a Hilbert space* H*. Then the Cesàro square averages*

$$S_N u := \frac{1}{N} \sum_{n=0}^{N-1} T^{n^2} u$$

converge for every $u \in H$ *as* $N \to \infty$*.*

Proof. We suppose here in addition that T is an isometry and leave the proof of the general case to the reader. By the JdLG-decomposition it suffices to prove the assertion for $u \in H_{\mathrm{r}}$ and $u \in H_{\mathrm{s}}$ separately.

Take first $u \in H_{\mathrm{r}}$ satisfying $Tu = \mathrm{e}^{2\pi \mathrm{i}\alpha} u$ for some $\alpha \in [0,1)$. Then we have

$$S_N u = \frac{1}{N} \sum_{n=0}^{N-1} \mathrm{e}^{2\pi \mathrm{i}\alpha n^2} u.$$

If α is irrational, this sequence converges to 0 by the equidistribution of $(\alpha n^2)_n$ (see Propositions 10.20 and 10.21). If $\alpha \in \mathbb{Q}$, say $d\alpha \in \mathbb{N}_0$ for some $d \in \mathbb{Z} \setminus \{0\}$, then one can see that

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathrm{e}^{2\pi \mathrm{i}\alpha n^2} \to \frac{1}{d} \sum_{j=0}^{d-1} \mathrm{e}^{2\pi \mathrm{i}\alpha j^2} \quad (N \to \infty).$$

Since each S_N ise contractive and since the linear span of all eigenvectors corresponding to unimodular eigenvalues is dense in H_{r} (see Theorem 16.34), we obtain convergence for every $u \in H_{\mathrm{r}}$.

Take now $u \in H_{\mathrm{s}}$ and define $u_n := T^{n^2} u$. In order to use the van der Corput lemma we calculate

$$(u_n \,|\, u_{n+m}) = \left(T^{n^2} u \,\middle|\, T^{(n+m)^2} u \right) = \left(u \,\middle|\, T^{2nm+m^2} u \right) = \left(T^{*m^2} u \,\middle|\, (T^{2m})^n u \right),$$

where we have used that T is an isometry. This implies for any fixed $m \in \mathbb{N}$ that

$$\gamma_m = \limsup_{N\to\infty} \Big| \frac{1}{N} \sum_{n=0}^{N-1} (u_n \,|\, u_{n+m}) \Big| \le \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \Big| \Big(T^{*m^2} u \,\Big|\, (T^{2m})^n u \Big) \Big|$$
$$\le 2m \limsup_{N\to\infty} \frac{1}{2mN} \sum_{j=0}^{2mN-1} \Big| \Big(T^{*m^2} u \,\Big|\, T^j u \Big) \Big| = 0,$$

since $u \in H_{\mathrm{s}}$, where the last equality follows from Theorem 16.19. The van der Corput Lemma 9.24 concludes the proof. □

So we see that, in particular, for an MDS $(\mathrm{X};\varphi)$ and its Koopman operator $T := T_\varphi$ the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f \cdot T^{n^2} f \tag{18.13}$$

exists in $\mathrm{L}^2(\mathrm{X})$ for every $f \in \mathrm{L}^\infty(\mathrm{X})$. To finish the proof of Theorem 18.11 it remains to show that it is positive whenever f is.

Proposition 18.14. *Let* $(\mathrm{X};\varphi)$ *be an ergodic MDS. Then the limit* (18.13) *is positive for every* $0 < f \in \mathrm{L}^\infty(\mathrm{X})$.

The proof is analogous to that of Proposition 18.9 and again uses the decomposition $\mathrm{L}^2(\mathrm{X}) = E_{\mathrm{s}} \oplus E_{\mathrm{r}}$, the vanishing of the above limit on E_{s} and a relative density argument on E_{r}.

Remarks 18.15. 1) A sequence $(n_k)_{k\in\mathbb{N}}$ is called a **Poincaré sequence** (see, e.g., [Furstenberg (1981), Def. 3.6]) if for every MDS $(\mathrm{X};\varphi)$, $\mathrm{X} = (X,\Sigma,\mu)$, and $A \in \Sigma$ with $\mu(A) > 0$ one has

$$\mu(A \cap \varphi^{-n_k}(A)) > 0 \quad \text{for some } k \in \mathbb{N}.$$

Poincaré's Theorem 6.10 tells that $(n)_{n\in\mathbb{N}}$ is a Poincaré sequence, hence the terminology. One can prove that Proposition 18.14 remains valid even for not necessarily ergodic MDSs, so we obtain that $(n^2)_{n\in\mathbb{N}}$ is a Poincaré sequence.

2) The above results remain true if one replaces the polynomial $p(n) = n^2$ by an arbitrary polynomial p with integer coefficients and $p(0) = 0$, see Furstenberg [Furstenberg (1981), pp. 66–75].

Final remarks

There are many generalisations and modifications of the above results. We only quote a few examples. The first is from [Furstenberg and Weiss (1996)].

Theorem 18.16 (Furstenberg–Weiss). *Let* $(\mathrm{X};\varphi)$ *be an MDS with associated Koopman operator* $T := T_\varphi$ *on* $X = \mathrm{L}^2(\mathrm{X})$. *Then the limit of the multiple ergodic averages*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{an} f \cdot T^{bn} g \tag{18.14}$$

exists in $\mathrm{L}^2(\mathrm{X})$ *for all* $f,g \in \mathrm{L}^\infty(\mathrm{X})$ *and* $a,b \in \mathbb{N}$.

Remark 18.17. In [Furstenberg and Weiss (1996)] it is also shown that the averages

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^n f \cdot T^{n^2} g$$

converge in $\mathrm{L}^2(\mathrm{X})$ for all $f,g \in \mathrm{L}^\infty(\mathrm{X})$. Together with the positivity of this limit for $f = g > 0$ and the Furstenberg correspondence principle this yields the existence of (many) triples $a, a+n, a+n^2$ in any $A \subseteq \mathbb{N}$ with $\overline{\mathrm{d}}(A) > 0$.

Host and Kra [Host and Kra (2005a)] proved convergence of multiple ergodic averages for strongly ergodic systems with powers $p_1(n), \ldots, p_{k-1}(n)$ replacing $n, \ldots, (k-1)n$ in (18.7), where $p_1, \ldots, p_{k-1}$ are arbitrary polynomials with integer coefficients. This leads to the existence of a subset of the form $\{a, a+p_1(n), \ldots, a+p_{k-1}(n)\}$ in every set $A \subseteq \mathbb{N}$ with positive upper density.

Remark 18.18. The results of Szemerédi and Furstenberg–Sárközy presented in this Chapter remain true in a slightly strengthened form. One can replace the upper density $\overline{\mathrm{d}}$ by the so-called **upper Banach-density** defined as

$$\overline{B\mathrm{d}}(A) := \limsup_{m,n\to\infty} \frac{\mathrm{card}(A \cap \{m, \ldots, m+n\})}{n+1}.$$

The actual change to be carried out in order to obtain these generalisations is in Furstenberg's Correspondence Principle(s), see [Furstenberg (1981)].

Exercises

Chapter 19
Joinings

19.1 Dynamical Systems and their Joinings

19.2 Inductive Limits and the Minimal Invertible Extension

19.3 Relatively Independent Joinings

Supplement: Tensor Products and Projective Limits

Exercises

Chapter 20
The Host–Kra–Tao Theorem

20.1 Idempotent Classes

20.2 The Existence of Sated Extensions

20.3 The Furstenberg Self-Joining

20.4 Proof of the Host–Kra–Tao Theorem

Exercises

Chapter 21
More Ergodic Theorems

21.1 Polynomial Ergodic Theorems

21.2 Weighted Ergodic Theorems

21.3 Wiener–Wintner Theorems

21.4 Even More Ergodic Theorems

1. Subsequential ergodic theorems

2. Entangled ergodic theorems

3. Multiple ergodic theorems

4. Bourgain return time theorem

Exercises

Appendix A
Topology

A.1 Metric Spaces

A **metric space** is a pair (Ω, d) consisting of a nonempty set Ω and a function $d : \Omega \times \Omega \to \mathbb{R}$ with the following properties:

1) $d(x,y) \geq 0$, and $d(x,y) = 0$ if and only if $x = y$.
2) $d(x,y) = d(y,x)$.
3) $d(x,y) \leq d(x,z) + d(z,y)$ **(triangle inequality)**.

Such a function d is called a **metric** on Ω. If instead of 1) we require only $d(x,x) = 0$ for all $x \in \Omega$, we have a **semi-metric**. For $A \subseteq \Omega$ and $x \in \Omega$ we define

$$d(x,A) := \inf\{d(x,y) : y \in A\},$$

called the **distance** of x from A. By a **ball** with center x and radius $r > 0$ we mean either of the sets

$$\mathrm{B}(x,r) := \{y \in \Omega : d(x,y) < r\},$$
$$\overline{\mathrm{B}}(x,r) := \{y \in \Omega : d(x,y) \leq r\}.$$

A set $O \subseteq \Omega$ is called **open** if for each $x \in O$ there is a ball $B \subseteq O$ with centre x and radius $r > 0$. A set $A \subseteq \Omega$ is called **closed** if $\Omega \setminus A$ is open. The ball $\mathrm{B}(x,r)$ is open, and the ball $\overline{\mathrm{B}}(x,r)$ is closed for any $x \in \Omega$ and $r > 0$.

A sequence $(x_n)_{n\in\mathbb{N}}$ in Ω is **convergent to the limit** $x \in \Omega$ (in short: $x_n \to x$), if for all $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ with $d(x_n,x) < \varepsilon$ for $n \geq n_0$. Limits are unique, i.e., if $x_n \to x$ and $x_n \to y$, then $x = y$. For a subset $A \subseteq \Omega$ the following assertions are equivalent.

(i) A is closed.
(ii) if $x \in \Omega$ and $d(x,A) = 0$, then $x \in A$.
(iii) if $(x_n)_{n\in\mathbb{N}} \subseteq A$ and $x_n \to x \in \Omega$, then $x \in A$.

A **Cauchy sequence** $(x_n)_{n\in\mathbb{N}}$ in Ω is a sequence with the property that for all $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ with $d(x_n, x_m) < \varepsilon$ for $n, m \geq n_0$. Each convergent sequence is a Cauchy sequence. If conversely every Cauchy sequence is convergent, then the metric d as well as the metric space (Ω, d) is called **complete**.

A.2 Topological Spaces

Let Ω be a set. A set system $\mathcal{O} \subseteq \mathcal{P}(\Omega)$ is called a **topology** on Ω if it has the following three properties:

1) $\emptyset, \Omega \in \mathcal{O}$;
2) $O_1, \ldots, O_n \in \mathcal{O}, n \in \mathbb{N} \quad \Rightarrow \quad O_1 \cap \cdots \cap O_n \in \mathcal{O}$;
3) $O_\iota \in \mathcal{O}, \iota \in I \quad \Rightarrow \quad \bigcup_{\iota\in I} O_\iota \in \mathcal{O}$.

A **topological space** is a pair $(\Omega, \mathcal{O})$, where Ω is a set and $\mathcal{O} \subseteq \mathcal{P}(\Omega)$ is a topology on Ω. A subset $O \subseteq \Omega$ is called **open** if $O \in \mathcal{O}$. A subset $A \subseteq \Omega$ is called **closed** if $A^c = \Omega \setminus A$ is open. By de Morgans' laws, finite unions and arbitrary intersections of closed sets are closed.

If (Ω, d) is a metric space, the family $\mathcal{O} := \{O \subseteq \Omega : O \text{ open}\}$ of open sets (as defined in A.1) is a topology, called the topology **induced by the metric** [Willard (2004), Thm. 2.6].

A topological space is called **metrisable** if there exists a metric that induces the topology. It is called **completely metrisable** if it is metrisable by a complete metric. Not every topological space is metrisable, cf. Section A.7 below. On the other hand, two different metrics may give rise to the same topology. In this case the metrics are called **equivalent**. It is possible in general that a complete metric is equivalent to one which is not complete.

If $\mathcal{O}, \mathcal{O}'$ are both topologies on Ω and $\mathcal{O}' \subseteq \mathcal{O}$, then $\mathcal{O}$ is called **finer** than $\mathcal{O}'$ and $\mathcal{O}'$ **coarser** than $\mathcal{O}$. On each set there is a coarsest topology, namely $\mathcal{O} = \{\emptyset, \Omega\}$, called the **trivial topology**; and a finest topology, namely $\mathcal{O} = \mathcal{P}(\Omega)$, called the **discrete topology**. The discrete topology is metrisable, e.g., by the **discrete metric** $d(x,y) := \mathbf{1}_{\{x\}}(y) = \delta_{xy}$ (Kronecker delta). Clearly, a topology on Ω is the discrete topology if and only if every singleton $\{x\}$ is an open set.

If $(\mathcal{O}_\iota)_\iota$ is a family of topologies on Ω, then $\bigcap_\iota \mathcal{O}_\iota$ is again a topology. Hence if $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ is any set system one can consider

$$\tau(\mathcal{E}) := \bigcap \big\{ \mathcal{O} : \mathcal{E} \subseteq \mathcal{O} \subseteq \mathcal{P}(\Omega),\, \mathcal{O} \text{ is a topology} \big\},$$

the coarsest topology in which each member of $\mathcal{E}$ is open. It is called the topology **generated** by $\mathcal{E}$.

Let $(\Omega, \mathcal{O})$ be a topological space. A **neighbourhood** in Ω of a point $x \in \Omega$ is a subset $U \subseteq \Omega$ such that there is an open set $O \subseteq \Omega$ with $x \in O \subseteq U$. A set is open

if and only if it is a neighbourhood of all of its points. If A is a neighbourhood of x, then x is called an **interior point** of A. The set of all interior points of a set A is denoted by A° and is called the **interior** of A. The **closure** of a subset $A \subseteq \Omega$ is

$$\overline{A} := \bigcap \{F : A \subseteq F \subseteq \Omega,\ F \text{ closed}\},$$

which is obviously the smallest closed set that contains A. Alternatively, the closure of A is sometimes denoted by

$$\mathrm{cl}A \quad \text{or} \quad \mathrm{cl}_{\mathcal{O}} A$$

especially when one wants to stress the particular topology considered.

The (topological) **boundary** of A is the set $\partial A := \overline{A} \setminus A^\circ$. For $x \in \Omega$ one has $x \in \overline{A}$ if and only if every neighbourhood of x has nonempty intersection with A. If (Ω, d) is a metric space and $A \subseteq \Omega$, then $\overline{A} = \{x : d(x,A) = 0\}$, and $x \in \overline{A}$ if and only if x is the limit of a sequence in A.

A subset A of a topological space $(\Omega, \mathcal{O})$ is called **dense** in Ω if $\overline{A} = \Omega$. A topological space Ω is called **separable** if there is a countable set $A \subseteq \Omega$ which is dense in Ω.

If $\Omega' \subseteq \Omega$ is a subset, then the **subspace topology** on Ω' (induced by the topology $\mathcal{O}$ on Ω) is given by $\mathcal{O}_{\Omega'} := \{\Omega' \cap O : O \in \mathcal{O}\}$. If $\mathcal{O}$ is induced by a metric d on Ω, then the restriction of d to $\Omega' \times \Omega'$ is a metric on Ω' and this metric induces the subspace topology there.

A subspace of a separable metric space is again separable, but the analogous statement for general topological spaces is false [Willard (2004), §16].

A topological space Ω is called **Hausdorff** if any two points $x, y \in \Omega$ can be separated by disjoint open neighbourhoods, i.e., there are $U, V \in \mathcal{O}$ with $U \cap V = \emptyset$ and $x \in U$, $y \in V$. A subspace of a Hausdorff space is Hausdorff and each metric space is Hausdorff. If $\mathcal{O}$, $\mathcal{O}'$ are two topologies on Ω, $\mathcal{O}$ finer than $\mathcal{O}'$ and $\mathcal{O}'$ Hausdorff, then also $\mathcal{O}$ is Hausdorff. In a Hausdorff space each singleton $\{x\}$ is a closed set.

If $(\Omega, \mathcal{O})$ is a topological space and $x \in \Omega$, then one calls

$$\mathcal{U}(x) := \{U \subseteq \Omega : U \text{ is a neighbourhood of } x\}$$

the **neighbourhood filter** of $x \in \Omega$. A system $\mathcal{U} \subseteq \mathcal{U}(x)$ is called an (open) **neighbourhood base** for x or a **fundamental system of (open) neighbourhoods** of x, if for each $O \in \mathcal{U}(x)$ there is some $U \in \mathcal{U}$ with $x \in U \subseteq O$ (and each $U \in \mathcal{U}$ is open). In a metric space (Ω, d), the family of open balls $\mathrm{B}(x, 1/n)$, $n \in \mathbb{N}$, is an open neighbourhood base of $x \in \Omega$.

A **base** $\mathcal{B} \subseteq \mathcal{O}$ for the topology $\mathcal{O}$ on Ω is a system such that each open set can be written as a union of elements of $\mathcal{B}$. If $\mathcal{B}$ is a base for the topology of Ω, then $\{U \in \mathcal{B} : x \in U\}$ is an open neighbourhood base for x, and if $\mathcal{U}_x$ is an open neighbourhood base for x, for each $x \in \Omega$, then $\mathcal{B} := \bigcup_{x \in \Omega} \mathcal{U}_x$ is a base for the

topology. For example, in a metric space (Ω, d) the family of open balls is a base for the topology.

A topological space is called **second countable** if it has a countable base for its topology.

Lemma A.1. *Every second countable space is separable, and every separable metric space is second countable.*

A proof is in [Willard (2004), Thm. 16.11]. A separable topological space need not be second countable [Willard (2004), §16].

Let $(\Omega, \mathcal{O})$ be a topological space and let $\Omega' \subseteq \Omega$. A point $x \in \Omega'$ is called an **isolated point** of Ω' if the singleton $\{x\}$ is open. Further, $x \in \Omega$ is called an **accumulation point** of Ω' if it is not isolated in $\Omega' \cup \{x\}$ for the subspace topology. Equivalently, every neighbourhood of x in Ω contains points of Ω' different from x. A point $x \in \Omega$ is a **cluster point** of the sequence $(x_n)_{n\in\mathbb{N}} \subseteq \Omega$ if any neighbourhood of x contains infinitely many members of the sequence, i.e., if

$$x \in \bigcap\nolimits_{n\in\mathbb{N}} \overline{\{x_k : k \geq n\}}.$$

If Ω is a metric space, then x a cluster point of a sequence $(x_n)_{n\in\mathbb{N}} \subseteq \Omega$ if and only if x is the limit of a subsequence $(x_{k_n})_{n\in\mathbb{N}}$.

A.3 Continuity

Let $(\Omega, \mathcal{O})$, $(\Omega', \mathcal{O}')$ be topological spaces. A mapping $f : \Omega \to \Omega'$ is called **continuous at** $x \in \Omega$, if for each (open) neighbourhood V of $f(x)$ in Ω' there is an (open) neighbourhood U of x such that $f(U) \subseteq V$.

The mapping f is called **continuous** if it is continuous at every point. Equivalently, f is continuous if and only if the inverse image $f^{-1}(O)$ of each open (closed) set $O \in \mathcal{O}'$ is open (closed) in Ω. If we want to indicate the used topologies, we say that $f : (\Omega, \mathcal{O}) \to (\Omega', \mathcal{O}')$ is continuous.

If Ω is endowed with the discrete topology or Ω' is endowed with the trivial topology, then every mapping $f : \Omega \to \Omega'$ is continuous.

For metric spaces, the continuity of $f : \Omega \to \Omega'$ (at x) is the same as **sequential continuity** (at x), i.e., f is continuous at $x \in \Omega$ if and only if $f(x_n) \to f(x)$ whenever $x_n \to x$ in Ω.

A mapping $f : \Omega \to \Omega'$ is called a **homeomorphism** if f is bijective and both f and f^{-1} are continuous. Two topological spaces are called **homeomorphic** if there is a homeomorphism that maps one onto the other. The Hausdorff property, separability, metrisability, complete metrisability and the property of having a countable base are all preserved under homeomorphisms.

A mapping $f : (\Omega, d) \to (\Omega', d')$ between two metric spaces is called **uniformly continuous** if for each $\varepsilon > 0$ there is $\delta > 0$ such that $d(x,y) < \delta$ implies

$d'(f(x),f(y)) < \varepsilon$ for all $x,y \in \Omega$. Clearly, a uniformly continuous mapping is continuous. If $A \subseteq \Omega$ then the distance function $d(\cdot,A)$ from Ω to $\mathbb{R}$ is uniformly continuous since the triangle inequality implies that

$$|d(x,A) - d(y,A)| \leq d(x,y) \qquad (x,y \in \Omega).$$

A.4 Inductive and Projective Topologies

Let $(\Omega_\iota, \mathcal{O}_\iota)$, $\iota \in I$, be a family of topological spaces, let Ω be a nonempty set, and let $f_\iota : \Omega_\iota \to \Omega$, $\iota \in I$, be given mappings. Then

$$\mathcal{O}_{\text{ind}} := \big\{A \subseteq \Omega \,:\, f_\iota^{-1}(A) \in \mathcal{O}_\iota \text{ for all } \iota \in I\big\}.$$

is a topology on Ω, called the **inductive topology** with respect to the family $(f_\iota)_{\iota\in I}$. It is the finest topology such that all the mappings f_ι become continuous. A mapping $g : (\Omega, \mathcal{O}_{\text{ind}}) \to (\Omega', \mathcal{O}')$ is continuous if and only if all the mappings $g \circ f_\iota : \Omega_\iota \to \Omega'$, $\iota \in I$, are continuous.

As an example of an inductive topology we consider a topological space $(\Omega, \mathcal{O})$ and a *surjective* map $f : \Omega \to \Omega'$. Then the inductive topology on Ω' with respect to f is called the **quotient topology**. In this case Ω' is called a **quotient space** of Ω with respect to f, and f is called a **quotient mapping**.

A common instance of this situation is when $\Omega' = \Omega/\!\sim$ is the set of equivalence classes with respect to some equivalence relation $\sim$ on Ω, and $f : \Omega \to \Omega/\!\sim$ is the natural mapping, i.e., f maps each point to its equivalence class. A set $A \subseteq \Omega/\!\sim$ is open in $\Omega/\!\sim$ if and only if $\bigcup A$, the union of the elements in A, is open in Ω.

Let $(\Omega_\iota, \mathcal{O}_\iota)$, $\iota \in I$, be a family of topological spaces, let Ω be a nonempty set, and let $f_\iota : \Omega \to \Omega_\iota$, $\iota \in I$, be given mappings. The **projective topology** on Ω with respect to the family $(f_\iota)_{\iota\in I}$ is

$$\mathcal{O}_{\text{proj}} := \tau\big\{f_\iota^{-1}(O_\iota) \,:\, O_\iota \in \mathcal{O}_\iota \text{ for all } \iota \in I\big\} \subseteq \mathcal{P}(\Omega).$$

This is the coarsest topology for which all the functions f_ι become continuous. A *base* for this topology is

$$\mathcal{B}_{\text{proj}} := \{f_{\iota_1}^{-1}(O_{\iota_1}) \cap \cdots \cap f_{\iota_n}^{-1}(O_{\iota_n}) \,:\, n \in \mathbb{N}, O_{\iota_k} \in \mathcal{O}_{\iota_k} \text{ for all } k = 1,\ldots,n\}.$$

A mapping $g : (\Omega', \mathcal{O}') \to (\Omega, \mathcal{O}_{\text{proj}})$ is continuous if and only if all the functions $f_\iota \circ g : \Omega' \to \Omega_\iota$, $\iota \in I$, are continuous.

An example of a projective topology is the *subspace topology*. Indeed, if Ω is a topological space and $\Omega' \subseteq \Omega$ is a subset, then the subspace topology on Ω' is the projective topology with respect to the inclusion mapping $\Omega' \to \Omega$.

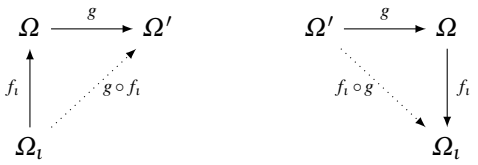


Fig. A.1 Continuity for the inductive respectively for the projective topology

A.5 Product Spaces

Let $(\Omega_\iota)_{\iota\in I}$ be a nonempty family of nonempty topological spaces. The **product topology** on

$$\Omega := \prod_{\iota\in I}\Omega_\iota = \left\{x : I \to \bigcup \Omega_\iota \,:\, x(\iota)\in\Omega_\iota \ \forall\, \iota\in I\right\}$$

is the projective topology with respect to the canonical projections $\pi_\iota : \Omega \to \Omega_\iota$. Instead of $x(\iota)$ we usually write x_ι. A base for this topology is formed by the open **cylinder sets**

$$A_{\iota_1,\dots,\iota_n} := \{x = (x_\iota)_{\iota\in I} \,:\, x_{\iota_k}\in O_{\iota_k} \text{ for } k = 1,\dots,n\}$$

for $\iota_1,\dots,\iota_n \in I$, $n\in\mathbb{N}$ and O_{ι_k} open in Ω_{ι_k}. For the product of two (or finitely many) spaces we also use the notation $\Omega\times\Omega'$ and the like. Convergence in the product space is the same as coordinatewise convergence. If $\Omega_\iota = \Omega$ for all $\iota\in I$, then we use the notation Ω^I for the product space.

If (Ω_j, d_j), $j\in I$ is a family of *at most countably* many (complete) metric spaces, then their product $\prod_{j\in I}\Omega_j$ is (completely) metrisable. E.g., if $I\subseteq\mathbb{N}$ one can use the the metric

$$d(x,y) := \sum_{j\in I}\frac{d(x_j,y_j)}{2^j(1+d(x_j,y_j))} \qquad (x,y\in\prod\nolimits_{j\in I}\Omega_j).$$

It follows from the triangle inequality that in a metric space (Ω,d) one has

$$|d(x,y)-d(u,v)| \le d(x,u)+d(y,v) \qquad (x,y,u,v\in\Omega).$$

This shows that the metric $d:\Omega\times\Omega\to\mathbb{R}$ is uniformly continuous.

A.6 Spaces of Continuous Functions

For Ω a topological space the sets $\mathrm{C}(\Omega)$ and $\mathrm{C}^{\mathrm{b}}(\Omega)$ of all continuous, respectively bounded and continuous functions $f : \Omega \to \mathbb{K}$ are algebras over $\mathbb{K}$ with respect to pointwise multiplication and addition ($\mathbb{K}$ stands for $\mathbb{R}$ or $\mathbb{C}$). The unit element is $\mathbf{1}$, the function which is constant with value 1. For $f \in \mathrm{C}^{\mathrm{b}}(\Omega)$ one defines

$$\|f\|_\infty := \sup\big\{|f(x)| : x \in \Omega\big\}.$$

Then $\|\cdot\|_\infty$ is a norm on $\mathrm{C}^{\mathrm{b}}(\Omega)$, turning it into a commutative unital Banach algebra. Convergence with respect to $\|\cdot\|_\infty$ is called **uniform convergence**. In the case that Ω is a metric space, the set

$$\mathrm{BUC}(\Omega) := \big\{f \in \mathrm{C}^{\mathrm{b}}(\Omega) : f \text{ is uniformly continuous}\big\}$$

is a closed subalgebra of $\mathrm{C}^{\mathrm{b}}(\Omega)$.

For a general topological space Ω, the space $\mathrm{C}(\Omega)$ may be quite "small". For example, if Ω carries the trivial topology, the only continuous functions thereon are the constant ones. In "good" topological spaces, the continuous functions **separate the points**, i.e., for every $x, y \in \Omega$ such that $x \neq y$ there is $f \in \mathrm{C}(\Omega)$ such that $f(x) \neq f(y)$. (Such spaces are necessarily Hausdorff.) An even stronger property is when continuous functions **separate closed sets**. This means that for every pair of disjoint closed subsets $A, B \subseteq \Omega$ there is a function $f \in \mathrm{C}(\Omega)$ such that

$$0 \leq f \leq 1, \quad f(A) \subseteq \{0\}, \quad f(B) \subseteq \{1\}.$$

If (Ω, d) is a metric space and $A \subseteq \Omega$, then the function $d(\cdot, A)$ is uniformly continuous; moreover, it is zero precisely on $\overline{A}$. Hence by considering functions of the type

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

for disjoint closed sets $A, B \subseteq \Omega$, one obtains the following.

Lemma A.2. *On a metric space (Ω, d), bounded uniformly continuous functions separate closed sets.*

A.7 Compactness

Let Ω be a set. A collection $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ of subsets of Ω is said to have the **finite intersection property** if

$$F_1 \cap \cdots \cap F_n \neq \emptyset$$

for every finite subcollection $F_1, \ldots, F_n \in \mathcal{F}$, $n \in \mathbb{N}$.

A topological space $(\Omega, \mathcal{O})$ has the **Heine–Borel property** if every collection of closed sets with the finite intersection property has a nonempty intersection. Equivalently, Ω has the Heine–Borel property if every open cover of Ω has a finite subcover.

A topological space is called **compact** if it is Hausdorff and has the Heine–Borel property. (Note that this deviates slightly from the definition in [Willard (2004), Def. 17.1].) A subset $\Omega' \subseteq \Omega$ of a topological space is called **compact** if it is compact with respect to the subspace topology. A compact set in a Hausdorff space is closed, and a closed subset in a compact space is compact [Willard (2004), Thm. 17.5]. A **relatively compact** set is a set whose closure is compact.

Lemma A.3. a) *Let Ω be a Hausdorff topological space, and let $A, B \subseteq \Omega$ be disjoint compact subsets of Ω. Then there are disjoint open subsets $U, V \subseteq \Omega$ such that $A \subseteq U$ and $B \subseteq V$.*

b) *Let Ω be a compact topological space, and let $O \subseteq \Omega$ be open and $x \in O$. Then there is an open set $U \subseteq \Omega$ such that $x \in U \subseteq \overline{U} \subseteq O$.*

A proof of a) is in [Lang (1993), Prop. 3.5]. For b) apply a) to $A = \{x\}$ and $B = O^{\mathrm{c}}$.

Proposition A.4. *Let $(\Omega, \mathcal{O})$ be a compact space.*

a) *If $f : \Omega \to \Omega'$ is continuous and Ω' Hausdorff, then $\overline{f(A)} = f(\overline{A})$ is compact for every $A \subseteq \Omega$. If in addition f is bijective, then f is a homeomorphism.*

b) *If $\mathcal{O}'$ is another topology on Ω, coarser than $\mathcal{O}$ but still Hausdorff, then $\mathcal{O} = \mathcal{O}'$.*

c) *Every continuous function on Ω is bounded, i.e., $\mathrm{C}^{\mathrm{b}}(\Omega) = \mathrm{C}(\Omega)$.*

For the proof see [Willard (2004), Thm.s 17.7 and 17.14].

Theorem A.5 (Tychonov). *Suppose $(\Omega_\iota)_{\iota \in I}$ is a family of nonempty topological spaces. Then the product space $\Omega = \prod_{\iota \in I} \Omega_\iota$ is compact if and only if each Ω_ι, $\iota \in I$ is compact.*

The proof rests on Zorn's lemma and can be found in [Willard (2004), Thm. 17.8] or [Lang (1993), Thm. 3.12].

Theorem A.6. *For a metric space (Ω, d) the following assertions are equivalent.*

(i) *Ω is compact.*

(ii) *Ω is complete and* **totally bounded**, *i.e., for each $\varepsilon > 0$ there is a finite set $F \subseteq \Omega$ such that $\Omega \subseteq \bigcup_{x \in F} \mathrm{B}(x, \varepsilon)$.*

(iii) *Ω is* **sequentially compact**, *i.e., each sequence $(x_n)_{n \in \mathbb{N}} \subseteq \Omega$ has a convergent subsequence.*

Moreover, if Ω is compact, it is separable and has a countable base, and $\mathrm{C}(\Omega) = \mathrm{C}^{\mathrm{b}}(\Omega) = \mathrm{BUC}(\Omega)$.

For the proof see [Lang (1993), Thm. 3.8–Prop. 3.11].

Theorem A.7 (Heine–Borel). *A subset of $\mathbb{R}^d$ is compact if and only if it is closed and bounded.*

A proof can be found in [Lang (1993), Cor. 3.13].
if each of

Theorem A.8. *For a compact topological space K the following assertions are equivalent.*

(i) *K is metrisable.*

(ii) *K is second countable.*

The proof is based on Urysohn's metrisation theorem [Willard (2004), Thm. 23.1].

A Hausdorff topological space $(\Omega, \mathcal{O})$ is **locally compact** if each of its points has a compact neighbourhood. Equivalently, a Hausdorff space is locally compact if the relatively compact open sets form a base for its topology.

If Ω is locally compact, $A \subseteq \Omega$ is closed and $O \subseteq \Omega$ is open, then $A \cap O$ is locally compact [Willard (2004), Thm. 18.4].

A compact space is (trivially) locally compact. If Ω is locally compact but not compact, and if p is some point not contained in Ω, then there is a unique compact topology on $\Omega^* := \Omega \cup \{p\}$ such that Ω is a dense open set in Ω^* and the subspace topology on Ω coincides with its original topology. This construction is called the **one-point compactification** of Ω [Willard (2004), Def. 19.2].

A.8 Connectedness

A topological space $(\Omega, \mathcal{O})$ is **connected** if $\Omega = A \cup B$ with disjoint open (closed) sets $A, B \subseteq \Omega$ implies that $A = \Omega$ or $B = \Omega$. Equivalently, a space is connected if and only if the only closed and open (abbreviated **clopen**) sets are $\emptyset$ and Ω itself. A subset $A \subseteq \Omega$ is called **connected** if it is connected with respect to the subspace topology. A **disconnected** set (space) is one which is not connected.

A continuous image of a connected space is connected. The connected subsets of $\mathbb{R}$ are precisely the intervals (finite or infinite).

A topological space $(\Omega, \mathcal{O})$ is called **totally disconnected** if the only connected subsets in Ω are the trivial ones that is $\emptyset$ and the singletons $\{x\}$, $x \in \Omega$.

Subspaces and products of totally disconnected spaces are totally disconnected [Willard (2004), Thm. 29.3]. Every discrete space is totally disconnected. Consequently, products of discrete spaces are totally disconnected. An example of a not discrete, compact, uncountable space is the **Cantor set** C. One way to construct C is to remove the open middle third from the unit interval, and then in each step remove the open middle third of each of the remaining intervals. In the limit, one obtains

$$C = \left\{ x \in [0,1] : x = \sum_{n=1}^{\infty} \frac{a_j}{3^j},\ j \in \{0,2\} \right\}.$$

The space C is homeomorphic to $\{0,1\}^{\mathbb{N}}$ via the map $x \mapsto (a_j/2)_{j\in\mathbb{N}}$ [Willard (2004), Ex. 19.9.c].

Theorem A.9. *A compact space is totally disconnected if and only if the set of clopen subset of Ω is a base for its topology.*

The proof is in [Willard (2004), Thm. 29.7].

A topological space $(\Omega, \mathcal{O})$ is called **extremally disconnected** if the closure of every open set $G \in \mathcal{O}$ is open. Equivalently, a space is extremally disconnected if each pair of disjoint open sets have disjoint closures.

Open subspaces of extremally disconnected spaces are extremally disconnected. Every discrete space is extremally disconnected, and an extremally disconnected metric space is discrete. In particular, the Cantor set is *not* extremally disconnected, and this shows that products of extremally disconnected spaces need not be extremally disconnected [Willard (2004), §15G].

A.9 Category

A subset A of a topological space Ω is called **nowhere dense** if its closure has empty interior, i.e., $(\overline{A})^\circ = \emptyset$. A set A is called of **first category** in Ω if it is the union of countably many nowhere dense subsets of Ω. Clearly, countable unions of sets of first category are of first category.

A set A is called of **second category** in Ω if it is not of first category. Equivalently, A is of second category if, whenever $A \subseteq \bigcup_{n\in\mathbb{N}} A_n$ then one of the sets $\overline{A_n \cap A}$ must contain an interior point.

Sets of first category are considered to be "small", whereas sets of second category are "large". A topological space $(\Omega, \mathcal{O})$ is called a **Baire space** if every nonempty open subset of Ω is "large", i.e., of second category in Ω. The proof of the following result is in [Willard (2004), Cor. 25.4].

Theorem A.10. *Each locally compact space and each complete metric space is a Baire space.*

A countable intersection of open sets in a topological space Ω is called a **$\mathbf{G}_\delta$-set**; and a countable union of closed sets of Ω is called a **$\mathbf{F}_\sigma$-set**. In a metric space, every closed set A is a G_δ-set since $A = \bigcap_{n\in\mathbb{N}} \{x \in \Omega : d(x,A) < 1/n\}$. The following result follows easily from the definitions.

Theorem A.11. *Let Ω be a Baire-space.*

a) *An F_σ-set is of first category if and only if it has empty interior.*

b) *A G_δ-set is of first category if and only if it is nowhere dense.*

c) *A countable intersection of dense G_δ-sets is dense.*

d) *A countable union of F_σ-sets with empty interior has empty interior.*

A.10 Polish Spaces

A topological space Ω is called a **Polish space** if it is separable and completely metrisable.

Lemma A.12. *Any G_δ-subset of a Polish space is again a Polish space.*

The proof is in [Willard (2004), Thm. 24.12].

Theorem A.13. *Let Ω be a locally compact space. Then Ω is Polish if and only if it is second countable.*

Proof. If Ω is Polish, then it is separable and metrisable, so it is second countable. Conversely, if it is second countable, then so is its one-point compactification Ω^*. By Theorem A.8, this is metrisable, and hence a Polish space by Theorem A.6. Then Ω is also Polish because it is an open subset of a Polish space. □

A.11 Nets

A **directed set** is a set Λ together with a relation $\leq$ on Λ with the following properties:

1) $\alpha \leq \alpha$,
2) $\alpha \leq \beta$ and $\beta \leq \gamma$ imply that $\alpha \leq \gamma$,
3) for every $\alpha, \beta \in \Lambda$ there is $\gamma \in \Lambda$ such that $\alpha, \beta \leq \gamma$.

The relation $\leq$ is called a **direction** on Λ. Typical examples of directed sets include

a) $\mathbb{N}$ with the natural meaning of $\leq$.
b) $\mathbb{N}$ with the direction given by divisibility: $n \leq m \Leftrightarrow n|m$.
c) $\mathcal{U}(x)$, the neighbourhood filter of a given point x in a topological space Ω, with the direction given by reversed set inclusion: $U \leq V \Leftrightarrow V \subseteq U$.

Let $(\Omega, \mathcal{O})$ be a topological space. A **net** in Ω is a map $x : \Lambda \to \Omega$, where $(\Lambda, \leq)$ is a directed set. In the context of nets one usually writes x_α in place of $x(\alpha)$, and $x = (x_\alpha)_{\alpha \in \Lambda} \subseteq \Omega$.

A net $(x_\alpha)_{\alpha \in \Lambda} \subseteq \Omega$ is **convergent** to $x \in \Omega$, written $x_\alpha \to x$, if

$$\forall U \in \mathcal{U}(x)\ \exists \alpha_0 \in \Lambda\ \forall \alpha \geq \alpha_0 : x_\alpha \in U.$$

In this case the point $x \in \Omega$ is called the **limit** of the net $(x_\alpha)_\alpha$. (Note that if $\Lambda = \mathbb{N}$ with the natural direction and Ω is a metric space, this coincides with the usual notion of a limit of a sequence.) If Ω is Hausdorff, then a net in Ω can have at most one limit.

A **subnet** of a net $(x_\alpha)_{\alpha \in \Lambda} \subseteq \Omega$ is each net $(x_{\lambda(\alpha')})_{\alpha' \in \Lambda'}$, where $(\Lambda', \leq)$ is a directed set and $\lambda : \Lambda' \to \Lambda$ is increasing and **cofinal**. The latter means that for each

$\alpha \in \Lambda$ there is $\alpha' \in \Lambda'$ such that $\alpha \leq \lambda(\alpha')$. A subnet of a sequence need not be a sequence any more [Willard (2004), Ex. 11B].

All the fundamental notions of set theoretic topology can be characterized in terms of nets, see [Willard (2004), Chap. 11] and [Willard (2004), Thm. 17.4].

Theorem A.14. a) *If $x \in \Omega$ and $A \subseteq \Omega$, then $x \in \overline{A}$ if and only there is a net $(x_\alpha)_{\alpha\in\Lambda} \subseteq A$ such that $x_\alpha \to x$.*

b) *If $A \subseteq \Omega$, then A is closed if an only if it contains the limit of each net in A, convergent in Ω.*

c) *For a net $(x_\alpha)_{\alpha\in\Lambda} \subseteq \Omega$ and $x \in \Omega$ the following assertions are equivalent.*

(i) *Some subnet of $(x_\alpha)_{\alpha\in\Lambda}$ is convergent to x.*

(ii) *x is a* **cluster point** *of $(x_\alpha)_{\alpha\in\Lambda}$, i.e.,*

$$x \in \bigcap\nolimits_{\alpha\in\Lambda} \overline{\{x_\beta : \beta \geq \alpha\}}.$$

d) *A Hausdorff space Ω is compact if and only if each net in Ω has a convergent subnet.*

e) *The map $f : \Omega \to \Omega'$ is continuous at $x \in \Omega$ if and only if $x_\alpha \to x$ implies $f(x_\alpha) \to f(x)$ for every net $(x_\alpha)_{\alpha\in\Lambda} \subseteq \Omega$.*

f) *Let $f_\iota : \Omega \to \Omega_\iota$, $\iota \in I$. Then Ω carries the projective topology with respect to the family $(f_\iota)_{\iota\in I}$ if and only if for each net $(x_\alpha)_{\alpha\in\Lambda} \subseteq \Omega$ and $x \in \Omega$ one has*

$$x_\alpha \to x \;\Leftrightarrow\; f_\iota(x_\alpha) \to f_\iota(x) \quad \forall\, \iota \in I.$$

In particular, f) implies that in a product space $\Omega = \prod_\iota \Omega_\iota$ a net $(x_\alpha)_{\alpha\in\Lambda} \subseteq \Omega$ converges if and only if it converges in each coordinate $\iota \in I$.

Let (Ω, d) be a metric space. A **Cauchy net** in Ω is a net $(x_\alpha)_{\alpha\in\Lambda} \subseteq \Omega$ such that

$$\forall\, \varepsilon > 0 \;\; \exists\, \alpha \in \Lambda : \; \alpha \leq \beta, \gamma \;\Rightarrow\; d(x_\beta, x_\gamma) < \varepsilon.$$

Every subnet of a Cauchy net is a Cauchy net. Every convergent net is a Cauchy net, and if a Cauchy net has a convergent subnet, then it is convergent.

Theorem A.15. *In a complete metric space every Cauchy net converges.*

See [Willard (2004), Thm. 39.4] for a proof.

Appendix B
Measure and Integration Theory

We begin with some general set theoretic notions. Let X be a set. Then its **power set** is denoted by

$$\mathcal{P}(X) := \{A : A \subseteq X\}.$$

The **complement** of $A \subseteq X$ is denoted by $A^{\mathrm{c}} := X \setminus A$, and its **characteristic function** is

$$\mathbf{1}_A(x) := \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$$

for $x \in X$. One often writes $\mathbf{1}$ in place of $\mathbf{1}_X$ if the reference set X is understood. For a sequence $(A_n)_n \subseteq \mathcal{P}(X)$ we write $A_n \searrow A$ if

$$A_n \supseteq A_{n+1} \quad \text{and} \quad \bigcap\nolimits_{n\in\mathbb{N}} A_n = A.$$

Similarly, $A_n \nearrow A$ is short for

$$A_n \subseteq A_{n+1} \quad \text{and} \quad \bigcup\nolimits_{n\in\mathbb{N}} A_n = A.$$

A family $(A_\iota)_\iota \subseteq \mathcal{P}(X)$ is called **pairwise disjoint** if $\iota \neq \eta$ implies that $A_\iota \cap A_\eta = \emptyset$. A subset $\mathcal{E} \subseteq \mathcal{P}(X)$ is often called a **set system** over X. A set system is called $\cap$-**stable** ($\cup$-stable, $\setminus$-stable) if $A, B \in \mathcal{E}$ implies that $A \cap B$ (and $A \cup B$, $A \setminus B$) belongs to $\mathcal{E}$ as well. If $\mathcal{E}$ is a set system, then any mapping $\mu : \mathcal{E} \to [0,\infty]$ is called a (positive) **set function**. Such a set function is called σ**-additive** if

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

whenever $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{E}$ is pairwise disjoint and $\bigcup_n A_n \in \mathcal{E}$. Here we adopt the convention that

$$a + \infty = \infty + a = \infty \qquad (-\infty < a \leq \infty).$$

A similar rule holds for sums $a+(-\infty)$ where $a \in [-\infty,\infty)$. The sum $\infty+(-\infty)$ is *not defined*. Other conventions for computations with the values $\pm\infty$ are:

$$0\cdot(\pm\infty)=(\pm\infty)\cdot 0=0, \quad \alpha\cdot(\pm\infty)=(\pm\infty)\cdot\alpha=\pm\infty \quad \beta\cdot(\pm\infty)=(\pm\infty)\cdot\beta=\mp\infty$$

for $-\infty<\beta<0<\alpha$. If $f: X\to Y$ is a mapping and $B\subseteq Y$ then we denote

$$[f\in B] := f^{-1}(B) := \big\{x\in X : f(x)\in B\big\}.$$

Likewise, if $P(x_1,\ldots,x_n)$ is a property of n-tuples $(x_1,\ldots,x_n)\in(Y)^n$ and $f_1,\ldots,f_n : X\to Y$ are mappings, then we write

$$[P(f_1,\ldots,f_n)] := \big\{x\in X : P(f_1(x),\ldots,f_n(x)) \text{ holds}\big\}.$$

For example, for $f,g: X\to Y$ we abbreviate $[f=g] := \{x\in X : f(x)=g(x)\}$.

B.1 σ-Algebras

Let X be any set. A **σ-algebra** over X is a set system $\Sigma\subseteq\mathcal{P}(X)$, such that the following hold.

1) $\emptyset, X\in\Sigma$.
2) If $A,B\in\Sigma$ then $A\cup B, A\cap B, A\setminus B\in\Sigma$.
3) If $(A_n)_{n\in\mathbb{N}}\subseteq\Sigma$, then $\bigcup_{n\in\mathbb{N}}A_n, \bigcap_{n\in\mathbb{N}}A_n\in\Sigma$.

If a set system Σ satisfies merely 1) and 2), it is called an **algebra**, and if Σ satisfies just 2) and $\emptyset\in\Sigma$, then it is called a **ring**. A pair (X,Σ) with Σ being a σ-algebra over X is called a **measurable space**.

Clearly, $\{\emptyset,X\}$ is the smallest and $\mathcal{P}(X)$ is the largest σ-algebra over X. An arbitrary intersection of σ-algebras over the same set X is again a σ-algebra. Hence for $\mathcal{E}\subseteq\mathcal{P}(X)$ one can form

$$\sigma(\mathcal{E}) := \bigcap\big\{\Sigma : \mathcal{E}\subseteq\Sigma\subseteq\mathcal{P}(X), \Sigma \text{ a } \sigma\text{-algebra}\big\},$$

the σ-algebra **generated** by $\mathcal{E}$. It is the smallest σ-algebra that contains all sets from $\mathcal{E}$. If $\Sigma=\sigma(\mathcal{E})$, we call $\mathcal{E}$ a **generator** of Σ.

If X is a topological space, the σ-algebra generated by all open sets is called the **Borel σ-algebra** $\mathrm{Bo}(X)$. By 1) and 2), $\mathrm{Bo}(X)$ contains all closed sets as well. A set belonging to $\mathrm{Bo}(X)$ is called a **Borel set** or **Borel measurable**.

As an example consider the **extended real line** $X=[-\infty,\infty]$. This becomes a compact metric space via the (order-preserving) homeomorphism $\arctan : [-\infty,\infty]\to[-\pi/2,\pi/2]$. The Borel algebra $\mathrm{Bo}([-\infty,\infty])$ is generated by $\{(\alpha,\infty] : \alpha\in\mathbb{R}\}$.

A **Dynkin system** (also called **λ-system**) on a set X is a subset $\mathcal{D} \subseteq \mathcal{P}(X)$ with the following properties:

1) $X \in \mathcal{D}$.
2) If $A, B \in \mathcal{D}$ and $A \subseteq B$ then $B \setminus A \in \mathcal{D}$.
3) If $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}$ and $A_n \nearrow A$, then $A \in \mathcal{D}$.

Dynkin systems play an important technical role in measure theory due to the following result, see [Bauer (1990), p.8] or [Billingsley (1979), Thm.3.2].

Theorem B.1 (Dynkin). *If $\mathcal{D}$ is a Dynkin system and $\mathcal{E} \subseteq \mathcal{D}$ is $\cap$-stable, then $\sigma(\mathcal{E}) \subseteq \mathcal{D}$.*

3

B.2 Measures

Let X be a set and $\Sigma \subseteq \mathcal{P}(X)$ a σ-algebra over X. A (positive) **measure** is a σ-additive set function

$$\mu : \Sigma \to [0, \infty].$$

In this case the triple (X, Σ, μ) is called a **measure space** and the sets in Σ are called **measurable sets**. If $\mu(X) < \infty$, the measure is called **finite**. If $\mu(X) = 1$, it is called a **probability measure** and (X, Σ, μ) is called a **probability space**. Suppose $\mathcal{E} \subseteq \Sigma$ is given and there is a sequence $(A_n)_n \subseteq \mathcal{E}$ such that

$$\mu(A_n) < \infty \quad (n \in \mathbb{N}) \quad \text{and} \quad X = \bigcup\nolimits_{n\in\mathbb{N}} A_n,$$

then the measure μ is called **σ-finite with respect to** $\mathcal{E}$. If $\mathcal{E} = \Sigma$, we simply call it σ-finite.

From the σ-additivity of the measure one derives the following properties:

a) **(Finite Additivity)** $\mu(\emptyset) = 0$ and

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B) \qquad (A, B \in \Sigma).$$

b) **(Monotonicity)** $A, B \in \Sigma, \quad A \subseteq B \quad \Rightarrow \quad \mu(A) \leq \mu(B)$.

c) **(σ-Subadditivity)** $(A_n)_n \subseteq \Sigma \quad \Rightarrow \quad \mu(\bigcup_{n\in\mathbb{N}} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

See [Billingsley (1979), p.134] for the elementary proofs. The following is an application of Dynkin's theorem, see [Billingsley (1979), Thm.10.3].

Theorem B.2 (Uniqueness Theorem). *Let $\Sigma = \sigma(\mathcal{E})$ with $\mathcal{E}$ being $\cap$-stable. Let μ, ν be two measures on Σ, both σ-finite with respect to $\mathcal{E}$. If μ and ν coincide on $\mathcal{E}$, they are equal.*

B.3 Construction of Measures

An **outer measure** on a set X is a mapping

$$\mu^* : \mathcal{P}(X) \to [0,\infty]$$

such that $\mu^*(\emptyset) = 0$ and μ^* is monotone and σ-subadditive. The following result allows to obtain a measure from an outer measure, see [Billingsley (1979), Thm.11.1].

Theorem B.3 (Carathéodory). *Let μ^* be an outer measure on the set X. Define*

$$\mathcal{M}(\mu^*) := \{A \subseteq X : \mu^*(H) = \mu^*(H \cap A) + \mu^*(H \setminus A) \quad \forall H \subseteq X\}.$$

Then $\mathcal{M}(\mu^*)$ is a σ-algebra and $\mu^*\big|_{\mathcal{M}(\mu^*)}$ is a measure on it.

The set system $\mathcal{E} \subseteq \mathcal{P}(X)$ is called a **semi-ring** if it satisfies the following two conditions:

(i) $\mathcal{E}$ is $\cap$-stable and $\emptyset \in \mathcal{E}$.

(ii) If $A, B \in \mathcal{E}$, then $A \setminus B$ is a disjoint union of members of $\mathcal{E}$.

An example of such a system is $\mathcal{E} = \{(a,b] : a \leq b\} \subseteq \mathcal{P}(\mathbb{R})$. If $\mathcal{E}$ is a semi-ring, then the system of all disjoint unions of members of $\mathcal{E}$ is a ring.

Theorem B.4 (Hahn). *Let $\mathcal{E} \subseteq \mathcal{P}(X)$ and let $\mu : \mathcal{E} \to [0,\infty]$ satisfy $\emptyset \in \mathcal{E}$ and $\mu(\emptyset) = 0$. Then $\mu^* : \mathcal{P}(X) \to [0,\infty]$ defined by*

$$\mu^*(A) := \inf\Big\{ \sum_{n\in\mathbb{N}} \mu(E_n) : (E_n)_n \subseteq \mathcal{E},\, A \subseteq \bigcup_n E_n \Big\} \qquad (A \in \mathcal{P}(X))$$

is an outer measure. Moreover, if $\mathcal{E}$ is a semi-ring and $\mu : \mathcal{E} \to [0,\infty]$ is σ-additive on $\mathcal{E}$, then $\sigma(\mathcal{E}) \subseteq \mathcal{M}(\mu^*)$ and $\mu^*|_{\mathcal{E}} = \mu$.

See [Billingsley (1979), p.140] for a proof. One may summarise these results in the following way: If a set function on a semi-ring $\mathcal{E}$ is σ-additive on $\mathcal{E}$ then it has a extension to a measure on $\sigma(\mathcal{E})$. If in addition X is σ-finite with respect to $\mathcal{E}$, then this extension is unique.

Sometimes, for instance in the construction of infinite products, it is convenient to work with the following criterion from [Billingsley (1979), Thm.10.2].

Lemma B.5. *Let $\mathcal{E}$ be an algebra over a set X, and let $\mu : \mathcal{E} \to [0,\infty)$ be a finitely additive set function with $\mu(X) < \infty$. Then μ is σ-additive on $\mathcal{E}$ if and only if for each decreasing sequence $(A_n)_n \subseteq \mathcal{E}$, $A_n \searrow \emptyset$, one has $\mu(A_n) \to 0$.*

B.4 Measurable Functions and Mappings

Let (X,Σ) and (Y,Σ') be measurable spaces. A mapping $\varphi : X \to Y$ is called **measurable** if

$$[\varphi \in A] \in \Sigma \quad \text{for all} \quad A \in \Sigma'.$$

(It suffices to check this condition for each A from a generator of Σ'.) We denote by

$$\mathfrak{M}(X;Y) = \mathfrak{M}(X,\Sigma;X,\Sigma')$$

the set of all measurable mappings between X and Y. For the special case $Y = [0,\infty]$ we write

$$\mathfrak{M}_+(X) := \big\{ f : X \to [0,\infty] \,:\, f \text{ is measurable} \big\}.$$

A composition of measurable mappings is measurable. For $A \in \Sigma$ its characteristic function $\mathbf{1}_A$ is measurable. If X, Y are topological spaces and $\varphi : X \to Y$ is continuous, then it is $\mathrm{Bo}(X)$-$\mathrm{Bo}(Y)$ measurable. The following lemma summarises the basic properties of positive measurable functions, see [Billingsley (1979), Sec.13].

Lemma B.6. *Let (X,Σ,μ) be a measure space. Then the following assertions hold.*

a) *If $f,g \in \mathfrak{M}_+(X), \alpha \geq 0$, then $fg, f+g, \alpha f \in \mathfrak{M}_+(X)$.*

b) *If $f,g \in \mathfrak{M}(X;\mathbb{R})$ and $\alpha,\beta \in \mathbb{R}$, then $fg, \alpha f + \beta g \in \mathfrak{M}(X;\mathbb{R})$.*

c) *If $f,g : X \to [-\infty,\infty]$ are measurable, then $-f, \min\{f,g\}, \max\{f,g\}$ are measurable.*

d) *If $f_n : X \to [-\infty,\infty]$ is measurable for each $n \in \mathbb{N}$ then $\sup_n f_n, \inf_n f_n$ are measurable.*

A **simple function** on a measure space (X,Σ,μ) is a linear combination of characteristic functions of measurable sets. [Billingsley (1979), Thm.13.5].

Lemma B.7. *Let $f : X \to [0,\infty]$ be measurable. Then there exists a sequence of simple functions $(f_n)_n$ such that*

$$0 \leq f_n \leq f_{n+1} \nearrow f \qquad \textit{(pointwise as } n \to \infty),$$

and such that the convergence is uniform if f is bounded.

B.5 The Integral of Positive Measurable Functions

Given a measure space (X,Σ,μ) there is a unique mapping

$$\mathfrak{M}_+(X) \to [0,\infty], \qquad f \mapsto \int_X f \,\mathrm{d}\mu,$$

called the **integral**, such that the following assertions hold for $f, f_n, g \in \mathfrak{M}_+(X)$, $\alpha \geq 0$, and $A \in \Sigma$:

a) $\displaystyle\int_X \mathbf{1}_A \,\mathrm{d}\mu = \mu(A)$;

b) $f \le g \quad \Rightarrow \quad \int_X f\,\mathrm{d}\mu \le \int_X g\,\mathrm{d}\mu$;

c) $\int_X (f+\alpha g)\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu + \alpha \int_X g\,\mathrm{d}\mu$;

d) If $0 \le f_1 \le f_2 \le \ldots$ and $f_n \to f$ pointwise, then $\int_X f\,\mathrm{d}\mu = \lim_{n\to\infty} \int_X f_n\,\mathrm{d}\mu$.

See [Rana (2002), Sec.5.2] or [Billingsley (1979), Sec.15]. Assertion d) is called the **Monotone Convergence Theorem** or the theorem of **Beppo Levi**.

If (X,Σ,μ) is a measure space, (Y,Σ') a measurable space and $\varphi : X \to Y$ measurable, then a measure is defined on Σ' by

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

The measure $\varphi_*\mu$ is called the **image** of μ under φ, or the **push-forward** of μ along φ. If μ is finite or a probability measure, so is $\varphi_*\mu$. If $f \in \mathfrak{M}_+(Y)$, then

$$\int_Y f\,\mathrm{d}(\varphi_*\mu) = \int_X (f\circ\varphi)\,\mathrm{d}\mu.$$

B.6 Product Spaces

If (X_1,Σ_1) and (X_2,Σ_2) are measurable spaces, then the **product σ-algebra** on the product space $X_1 \times X_2$ is

$$\Sigma_1 \otimes \Sigma_2 := \sigma\big\{A \times B : A \in \Sigma_1, B \in \Sigma_2\big\}.$$

If $\mathcal{E}_j$ is a generator of Σ_j with $X_j \in \mathcal{E}_j$ for $j = 1,2$, then

$$\mathcal{E}_1 \times \mathcal{E}_2 := \big\{A \times B : A \in \mathcal{E}_1, B \in \mathcal{E}_2\big\}$$

is a generator of $\Sigma_1 \otimes \Sigma_2$. If (X,Σ) is another measurable space, then a mapping $f = (f_1,f_2) : X \to X_1 \times X_2$ is measurable if and only if the projections $f_1 = \pi_1 \circ f, f_2 = \pi_2 \circ f$ are both measurable. If $f : (X_1 \times X_2, \Sigma_1 \otimes \Sigma_2) \to (Y,\Sigma')$ is measurable, then $f(x,\cdot) : X_2 \to Y$ is measurable for every $x \in X_1$, see [Billingsley (1979), Thm.18.1].

Theorem B.8 (Tonelli). *Let (X_j,Σ_j,μ_j), $j = 1,2$, be σ-finite measure spaces and $f \in \mathfrak{M}_+(X_1 \times X_2)$. Then the functions*

$$F_1 : X_1 \to [0,\infty], \quad x \mapsto \int_{X_2} f(x,\cdot)\,\mathrm{d}\mu_2$$
$$F_2 : X_2 \to [0,\infty], \quad y \mapsto \int_{X_1} f(\cdot,y)\,\mathrm{d}\mu_1$$

are measurable and there is a unique measure $\mu_1 \otimes \mu_2$ such that

$$\int_{X_1} F_1 \,\mathrm{d}\mu_1 = \int_{X_1\times X_2} f \,\mathrm{d}(\mu_1\otimes\mu_2) = \int_{X_2} F_2 \,\mathrm{d}\mu_2.$$

For a proof see [Billingsley (1979), Thm.18.3]. The measure $\mu_1\otimes\mu_2$ is called the **product measure** of μ_1 and μ_2. Note that for the particular case $F = f_1\otimes f_2$ with

$$(f_1\otimes f_2)(x_1,x_2) := f_1(x_1)\cdot f_2(x_2) \qquad (f_j\in\mathfrak{M}_+(X_j), x_j\in X_j \quad (j=1,2)),$$

we obtain $\displaystyle\int_{X_1\times X_2}(f_1\otimes f_2)\,\mathrm{d}(\mu_1\otimes\mu_2) = \left(\int_{X_1} f_1\,\mathrm{d}\mu_1\right)\,\left(\int_{X_2} f_2\,\mathrm{d}\mu_2\right).$

Let us turn to infinite products. Suppose $((X_\iota,\Sigma_\iota))_{\iota\in I}$ is a collection of measurable spaces. Consider the product space

$$X := \prod_{\iota\in I} X_\iota$$

with corresponding projections $\pi_\iota : X\to X_\iota$. A set of the form

$$\bigcap_{\iota\in F}[\pi_\iota\in A_\iota] = \prod_{\iota\in F} A_\iota \times \prod_{\iota\notin F} X_\iota$$

with $F\subseteq I$ finite and sets $A_\iota\in\Sigma_\iota$ for $\iota\in F$, is called a **(measurable) cylinder**. Let

$$\mathcal{E} := \left\{\textstyle\bigcap_{\iota\in F}[\pi_\iota\in A_\iota] : F\subseteq I \text{ finite}, A_\iota\in\Sigma_\iota, (\iota\in F)\right\}$$

be the semi-ring(!) of all cylinders. Then

$$\bigotimes_{\iota\in I}\Sigma_\iota := \sigma(\mathcal{E})$$

is called the **product σ-algebra** of the Σ_ι.

Theorem B.9 (Infinite Products). *Let (X_ι,Σ_ι) with $\iota\in I$ be a collection of probability spaces. Then there is a unique probability measure $\mu := \bigotimes_\iota \mu_\iota$ on $\Sigma = \bigotimes_\iota\Sigma_\iota$ such that*

$$\mu\left(\textstyle\bigcap_{\iota\in F}[\pi_\iota\in A_\iota]\right) = \prod_{\iota\in F}\mu_\iota(A_\iota)$$

for every finite subset $F\subseteq I$ and sets $A_\iota\in\Sigma_\iota$, $\iota\in F$.

The proof is based on Lemma B.5, see [Hewitt and Stromberg (1969), Chap.22] or [Halmos (1950), Sec.38]. In the case that $I=\mathbb{N}$ is countable, Theorem B.9 is a consequence of the Ionescu-Tulcea theorem, to be treated next.

The Ionescu Tulcea's Theorem

For a measurable space (X,Σ) we denote by $\mathrm{M}^+(X,\Sigma)$ the set of all positive and by $\mathrm{M}^1(X,\Sigma)$ the set of all probability measures on (X,Σ). There is a natural σ-algebra $\widetilde{\Sigma}$ on $\mathrm{M}^+(X,\Sigma)$, the smallest such that each mapping

$$\mathrm{M}^+(X,\Sigma) \to [0,\infty], \quad \nu \mapsto \nu(A) \qquad (A \in \Sigma)$$

is measurable.

Let (X_j,Σ_j), $j=1,2$ be measurable spaces. A **measure kernel** from X_1 to X_2 is a measurable mapping $\mu : X_2 \to \mathrm{M}^+(X_1,\Sigma_1)$. Such a kernel μ can also be interpreted as a mapping of two variables

$$\mu : X_2 \times \Sigma \to [0,\infty],$$

and we shall do so when it seems convenient. If $\mu(y,\cdot) \in \mathrm{M}^1(X_1,\Sigma_1)$ for each $y \in X_2$ then μ is called a **probability kernel**.

Let (X,Σ) be another measurable space, and let $\mu : X_2 \to \mathrm{M}^+(X_1,\Sigma_1)$ be a kernel. Then there is a Koopman operator

$$\begin{aligned} &T_\mu : \mathfrak{M}_+(X \times X_1) \to \mathfrak{M}_+(X \times X_2) \\ &(T_\mu f)(x,x_2) := \int_{X_1} f(x,x_1)\,\mu(x_2,\mathrm{d}x_1) \quad (x \in X, x_2 \in X_2). \end{aligned}$$

The operator T_μ is additive and positively homogeneous, and if $f_n \nearrow f$ pointwise on X_1 then $T_\mu f_n \nearrow T_\mu f$ pointwise on X_2. Moreover,

$$T_\mu(f \otimes g) = (f \otimes \mathbf{1}) \cdot T_\mu(\mathbf{1} \otimes g) \qquad (f \in \mathfrak{M}_+(X), g \in \mathfrak{M}_+(X_2)).$$

Conversely, each operator $T : \mathfrak{M}_+(X \times X_1) \to \mathfrak{M}_+(X \times X_2)$ with these properties is of the form T_μ, for some kernel μ.

If $\mu : X_2 \to \mathrm{M}^+(X_1)$ and $\nu : X_3 \to \mathrm{M}^+(X_2)$ are kernels, then $T_\nu \circ T_\mu = T_\eta$ for

$$\eta(x_3,A) := \int_{X_2} \mu(x_2,A)\nu(x_3,\mathrm{d}x_2) \qquad (x_3 \in X_3, A \in \Sigma_1).$$

Kernels can be used to construct measures on infinite products. Let (X_n,Σ_n), $n \in \mathbb{N}$, be measurable spaces, and let $X := \prod_{n\in\mathbb{N}} X_n$ be the Cartesian product, with the projections $\pi_n : X \to X_n$. The natural σ-algebra over X is

$$\bigotimes_n \Sigma_n := \sigma\big\{\pi_n^{-1}(A_n) : n \in \mathbb{N}, A_n \in \Sigma_n\big\}.$$

A generating algebra is

$$A := \Big\{A_n \times \prod_{k>n} X_k : n \in \mathbb{N}, A_n \in \Sigma_1 \otimes \ldots \otimes \Sigma_n\Big\},$$

the algebra of **cylinder sets**.

Theorem B.10 (Ionescu Tulcea). [Ethier and Kurtz (1986), p.504] *Let (X_n,Σ_n), $n \in \mathbb{N}$, be measurable spaces, let*

$$\mu_n : X_1 \times \cdots \times X_{n-1} \to \mathrm{M}^1(X_n) \qquad (n \in \mathbb{N}, n \geq 2)$$

be probability kernels, and let μ_1 be a probability measure on X_1. Let

$$X^{(n)} := X_1 \times \cdots \times X_n \qquad \textit{with} \qquad \Sigma^{(n)} = \Sigma_1 \otimes \ldots \otimes \Sigma_n.$$

Let, for $n \geq 1$, $\quad T_n : \mathfrak{M}_+(X^{(n)}, \Sigma^{(n)}) \to \mathfrak{M}_+(X^{(n-1)}, \Sigma^{(n-1)})$ be given by

$$(T_n f)(x^{(n-1)}) = \int_{X_n} f(x^{(n-1)}, x_n)\, \mu_n(x^{(n-1)}, \mathrm{d}x_n) \qquad (x^{(n-1)} \in X^{(n-1)}).$$

Then there is a unique probability measure ν on $X^{(\infty)} := \prod_{n\in\mathbb{N}} X_n$ such that

$$\int_{X^{(\infty)}} f(x_1, \ldots, x_n)\, \mathrm{d}\nu(x_1, \ldots) = T_1 T_2 \ldots T_n f \qquad (f \in \mathfrak{M}_+(X_1 \times \cdots \times X_n))$$

for every $n \in \mathbb{N}$.

An important special case of the Ionescu Tulcea theorem is the construction of the infinite product measure in the case that the index set I is countable. Indeed, if one has a probability measure ν_n on (X_n, Σ_n), for each $n \in \mathbb{N}$ and applies the Ionescu Tulcea theorem with $\mu_n \equiv \nu_n$, then the measure ν of the theorem satisfies

$$(\pi_1, \ldots, \pi_n)_* \nu = \nu_1 \otimes \ldots \otimes \nu_n \qquad (n \in \mathbb{N}).$$

That means, $\nu = \bigotimes_n \nu_n$ is the product of the ν_n.

B.7 Null Sets

Let (X, Σ, μ) be a measure space. A set $A \subseteq X$ is called a **null set** if there is a set $N \in \Sigma$ such that $A \subseteq N$ and $\mu(N) = 0$. (In general a null set need not be measurable). Null sets have the following properties:

a) If A is a null set and $B \subseteq A$, then B is also a null set.

b) If each A_n is a null set, then $\bigcup_{n\in\mathbb{N}} A_n$ is a null set.

The following results shows the connection between null sets and the integral, see [Billingsley (1979), Thm.15.2].

Lemma B.11. *Let (X, Σ, μ) be a measure space and let $f : X \to [-\infty, \infty]$ be measurable. Then the following assertions hold.*

a) $\int_X |f|\, \mathrm{d}\mu = 0$ *if and only if the set* $[f \neq 0] = [|f| > 0]$ *is a null set.*

b) *If* $\int_X |f|\, \mathrm{d}\mu < \infty$*, then the set* $[|f| = \infty]$ *is a null set.*

One says that two functions f, g are equal **μ-almost everywhere** (abbreviated by "$f = g$ a.e." or "$f \sim_\mu g$") if the set $[f \neq g]$ is a null set. More generally, let P be a property of points of X. Then P holds **almost everywhere** or for **μ-almost all** $x \in X$ if the set

$$\{x \in X : P \text{ does not hold for } x\}$$

is a μ-null set. If μ is understood, we leave out the reference to it.

For each set E, the relation $\sim_\mu$ ("is equal μ-almost everywhere to") is an equivalence relation on the space of mappings from X to E. For such a mapping f we sometimes denote by $[f]$ its equivalence class, in situations when notational clarity is needed. If μ is understood, we write simply $\sim$ instead of $\sim_\mu$.

If (E,d) is a separable metric space with its Borel σ-algebra, we denote by

$$\mathrm{L}^0(X;E) := \mathrm{L}^0(X,\Sigma,\mu;E) := \mathfrak{M}(X;E)/_\sim$$

the space of equivalence classes of measurable mappings modulo equality almost everywhere. Prominent examples here are $E = \mathbb{R}$, $E = [0,\infty]$ or $E = [-\infty,\infty]$. In the case $E = \mathbb{C}$ we abbreviate $\mathrm{L}^0(X) := \mathrm{L}^0(X,\Sigma,\mu;\mathbb{C})$ if Σ and μ are understood.

B.8 The Lebesgue Spaces

Let $\mathrm{X} := (X,\Sigma,\mu)$ be a measure space. For $f \in \mathrm{L}^0(\mathrm{X})$ we define

$$\|f\|_\infty := \inf\{t > 0 : \mu[\,|f(\cdot)| > t\,] = 0\}$$

and, if $1 \le p < \infty$,

$$\|f\|_p := \left(\int_X |f(\cdot)|^p \,\mathrm{d}\mu\right)^{\frac{1}{p}}.$$

Then we let

$$\mathrm{L}^p(\mathrm{X}) := \mathrm{L}^p(X,\Sigma,\mu;\mathbb{C}) := \left\{f \in \mathrm{L}^0(\mathrm{X}) : \|f\|_p < \infty\right\}$$

for $1 \le p \le \infty$. The space $\mathrm{L}^1(\mathrm{X})$ is called the space of (equivalence classes of) **integrable** functions. The following is in [Rudin (1987), Chap.3].

Theorem B.12. *Let (X,Σ,μ) be a measure space and $1 \le p \le \infty$. Then the following assertions hold.*

a) *The space $\mathrm{L}^p(\mathrm{X})$ is a Banach space with respect to the norm $\|\cdot\|_p$.*

b) *If $1 \le p < \infty$ and $f_n \to f$ in $\mathrm{L}^p(\mathrm{X})$, then there is $g \in \mathrm{L}^p(\mathrm{X})$ and a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ such that $|f_{n_k}| \le g$ a.e. for all $k \in \mathbb{N}$ and $f_{n_k} \to f$ pointwise a.e..*

c) *If $1 \le p < \infty$ and if $(f_n)_n \subseteq \mathrm{L}^p(\mathrm{X})$ is such that $f_n \to f$ a.e. and there is $g \in \mathrm{L}^p(\mathrm{X})$ such that $|f_n| \le g$ a.e. for all $n \in \mathbb{N}$, then $f \in \mathrm{L}^p(\mathrm{X})$, and $\|f_n - f\|_p \to 0$.*

Assertion c) is called the **Dominated Convergence Theorem** or **Lebesgue's theorem**.

The Integral

If $f \in \mathrm{L}^1(X,\Sigma,\mu)$ then the functions $(\mathrm{Re}\, f)^{\pm}, (\mathrm{Im}\, f)^{\pm}$ are positive and have finite integral. Hence one can define the **integral** of f by

$$\int_X f\,\mathrm{d}\mu := \int_X (\mathrm{Re}\, f)^+\,\mathrm{d}\mu - \int_X (\mathrm{Re}\, f)^-\,\mathrm{d}\mu + \mathrm{i}\left[\int_X (\mathrm{Im}\, f)^+\,\mathrm{d}\mu - \int_X (\mathrm{Im}\, f)^-\,\mathrm{d}\mu\right].$$

The integral is a linear mapping $\mathrm{L}^1(X,\Sigma,\mu) \to \mathbb{C}$ and satisfies

$$\left|\int_X f\,\mathrm{d}\mu\right| \le \int_X |f|\,\mathrm{d}\mu,$$

see [Rudin (1987), Chap.1].

Theorem B.13 (Averaging Theorem). *Let $S \subseteq \mathbb{C}$ be a closed subset and suppose that either (X,Σ,μ) is σ-finite or $0 \in S$. Let $f \in \mathrm{L}^1(\mathrm{X})$ such that*

$$\frac{1}{\mu(A)}\int_A f\,\mathrm{d}\mu \in S$$

for all $A \in \Sigma$ such that $0 < \mu(A) < \infty$. Then $f(\cdot) \in S$ almost everywhere.

A proof is in [Lang (1993), Thm.IV.5.15]. As a corollary one obtains that if $\int_A f = 0$ for all A of finite measure, then $f = 0$ almost everywhere.

Approximation

The following is an immediate consequence of Lemma B.7.

Theorem B.14. *Let (X,Σ,μ) be a measure space. Then the space* $\mathrm{lin}\{\mathbf{1}_A : A \in \Sigma\}$ *of simple functions is dense in* $\mathrm{L}^\infty(\mathrm{X})$.

Let (X,Σ,μ) be a measure space, and let $\mathcal{E} \subseteq \Sigma$. The space of **step functions** (with respect to $\mathcal{E}$) is

$$\mathrm{Step}(X,\mathcal{E},\mu) := \mathrm{lin}\big\{\mathbf{1}_B : B \in \mathcal{E},\ \mu(B) < \infty\big\}.$$

We abbreviate $\mathrm{Step}(\mathrm{X}) := \mathrm{Step}(X,\Sigma,\mu)$. By using the approximation $f\mathbf{1}_{[|f|\le n]} \to f$ and Theorem B.14 we obtain an approximation result for L^p.

Theorem B.15. *For any measure space (X,Σ,μ) and $1 \le p < \infty$, the space* $\mathrm{Step}(\mathrm{X})$ *is dense in* $\mathrm{L}^p(\mathrm{X})$.

Theorem B.15 combined with a Dynkin system argument yields the following.

Lemma B.16. *Let (X,Σ,μ) be a finite measure space. Let $\mathcal{E} \subseteq \Sigma$ be $\cap$-stable with $X \in \mathcal{E}$ and $\sigma(\mathcal{E}) = \Sigma$. Then* $\mathrm{Step}(X,\mathcal{E},\mu)$ *is dense in* $\mathrm{L}^p(\mathrm{X})$ *for each $1 \le p < \infty$.*

Theorem B.17. *Let (X,Σ,μ) be any measure space, and let $\mathcal{E}\subseteq\Sigma$ be $\cap$-stable with $\sigma(\mathcal{E})=\Sigma$. Suppose further that X is σ-finite with respect to E. Then* $\mathrm{Step}(X,\mathcal{E},\mu)$ *is dense in* $\mathrm{L}^p(\mathrm{X})$ *for each* $1\leq p<\infty$.

A related result, proved by a Dynkin system argument, holds on the level of sets.

Lemma B.18. *Let (X,Σ,μ) be a finite measure space, and let $\mathcal{E}\subseteq\Sigma$ be an algebra with $\sigma(\mathcal{E})=\Sigma$. Then for each $A\in\Sigma$ and each $\varepsilon>0$ there is $B\in\mathcal{E}$ such that $\mu(A\triangle B)<\varepsilon$.*

See also [Billingsley (1979), Thm.11.4] and [Lang (1993), Section VI.§6].

Fubini's Theorem

Consider two σ-finite measure spaces (X_j,Σ_j,μ_j), $j=1,2$, and their product

$$(X,\Sigma,\mu)=(X_1\times X_2,\Sigma_1\otimes\Sigma_2,\mu_1\otimes\mu_2).$$

Let $\mathcal{E}:=\{A_1\times A_2 : A_j\in\Sigma_j,\mu(A_j)<\infty\,(j=1,2)\}$ be the set of **measurable rectangles**. Then $\mathcal{E}$ satisfies the conditions of Theorem B.17.

Corollary B.19. *The space* $\mathrm{lin}\{\mathbf{1}_{A_1}\otimes\mathbf{1}_{A_2} : A_j\in\Sigma_j,\ \mu(A_j)<\infty\,(j=1,2)\}$ *is dense in* $\mathrm{L}^p(\mathrm{X})$ *for* $1\leq p<\infty$.

Combining this with Tonelli's theorem yields the following famous theorem [Lang (1993), Thm.8.4].

Theorem B.20 (Fubini). *Let $f\in\mathrm{L}^1(X_1\times X_2)$. Then for μ_1-almost every $x\in X_1$, $f(x,\cdot)\in\mathrm{L}^1(X_2)$ and with*

$$F:=\left(x\mapsto\int_{X_2}f(x,\cdot)\,\mathrm{d}\mu_2\right)$$

(defined almost everywhere on X_1) one has $F\in\mathrm{L}^1(X_1)$; moreover,

$$\int_{X_1}F\,\mathrm{d}\mu_1=\int_{X_1}\int_{X_2}f(x,y)\,\mathrm{d}\mu_2(y)\,\mathrm{d}\mu_1(x)=\int_{X_1\otimes X_2}f\,\mathrm{d}(\mu_1\otimes\mu_2).$$

B.9 Complex Measures

A **complex measure** on a measurable space (X,Σ) is a mapping $\mu:\Sigma\to\mathbb{C}$ which is σ-additive and satisfies $\mu(\emptyset)=0$. If the range of μ is contained in $\mathbb{R}$, μ is called a **signed measure**. For a complex measure μ one defines its **total variation** $|\mu|$ by

$$|\mu|(A):=\inf\Big\{\sum_{n=1}^{\infty}|\mu(A_n)| : (A_n)_{n\in\mathbb{N}}\subseteq\Sigma\text{ pairwise disjoint},A=\textstyle\bigcup_n A_n\Big\}$$

for $A \in \Sigma$. Then $|\mu|$ is a positive finite measure, see [Rudin (1987), Thm.6.2]. With respect to the norm $\|\mu\|_1 := |\mu|(X)$, the space of complex measures on (X,Σ) is a Banach space.

Let μ be a complex measure on a measurable space (X,Σ). For a step function $f = \sum_{j=1}^{n} x_j \mathbf{1}_{A_j} \in \mathrm{Step}(X,\Sigma,|\mu|)$ one defines

$$\int_X f \,\mathrm{d}\mu = \sum_{j=1}^{n} x_j \mu(A_j)$$

as usual, and shows (using finite additivity) that this does not depend on the representation of f. Moreover, one obtains

$$\left| \int_X f \,\mathrm{d}\mu \right| \leq \int_X |f| \,\mathrm{d}|\mu| = \|f\|_{\mathrm{L}^1(|\mu|)},$$

whence the integral has a continuous linear extension to all of $\mathrm{L}^1(X,\Sigma,|\mu|)$.

Appendix C
Functional Analysis

In this appendix we review some notions and facts from functional analysis but refer to [Conway (1990)], [Dunford and Schwartz (1958)], [Rudin (1987)], [Rudin (1991)], [Megginson (1998)] or [Schaefer (1980)] for more information.

C.1 Banach Spaces

Let E be a vector space over $\mathbb{K}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. A **semi-norm** on E is a mapping $\|\cdot\| : E \to \mathbb{R}$ satisfying

$$\|x\| \geq 0, \quad \|\lambda x\| = |\lambda| \, \|x\|, \quad \text{and} \quad \|x+y\| \leq \|x\| + \|y\|$$

for all $x, y \in E$, $\lambda \in \mathbb{K}$.

A semi-norm is a **norm** if $\|x\| = 0$ happens if and only if $x = 0$. A norm on a vector space defines a metric via

$$d(x,y) := \|x-y\| \qquad (x,y \in E)$$

and hence a topology, called the **norm topology**. It follows directly from the axioms of the norm that the mappings

$$E \times E \to E, \quad (x,y) \mapsto x+y \quad \text{and} \quad \mathbb{K} \times E \to E, \quad (\lambda, x) \mapsto \lambda x$$

are continuous. From the triangle inequality it follows that

$$\big| \|x\| - \|y\| \big| \leq \|x-y\| \qquad (x,y \in E),$$

which implies that also the norm mapping itself is continuous.

A **normed space** is a vector space together with a norm on it, and a **Banach space** if the metric induced by the norm is complete. If E is a normed space, its **closed unit ball** is

$$\mathrm{B}_E := \{x \in E : \|x\| \leq 1\}.$$

The set B_E is compact if and only if E is finite-dimensional.

Recall that a metric space is **separable** if it contains a countable dense set. A normed space E is separable if and only if there is a countable set A such that its **linear span**

$$\operatorname{lin} A = \left\{ \sum_{j=1}^{n} \lambda_j a_j : n \in \mathbb{N},\ \lambda_j \in \mathbb{K},\ a_j \in A\ (j = 1, \ldots, n) \right\}$$

is dense in E.

A subset $A \subseteq E$ of a normed space is called **(norm) bounded** if there is $c > 0$ such that $\|a\| \leq c$ for all $a \in A$. Every compact subset is norm-bounded since the norm mapping is continuous.

If E is a Banach space and $F \subseteq E$ is a closed subspace, then the **quotient space** $X := E/F$ becomes a Banach space under the **quotient norm**

$$\|x + F\|_Y := \inf\{\|x + f\| : f \in F\}.$$

The linear mapping $q : E \to X$, $q(x) := x + F$, is called the **canonical surjection** or **quotient map**.

C.2 Banach Algebras

An **algebra** is a (real or complex) linear space A together with an associative bilinear mapping

$$A \times A \to A \qquad (f, g) \mapsto fg$$

called **multiplication** and a distinguished element $\mathrm{e} \in A$ satisfying

$$ea = ae = a \quad \text{for all } a \in A,$$

called the **unit element**. (Actually, one should call A a "unital" algebra then, but all the algebras we have reason to consider are unital, so we include the existence of a unit element into our notion of "algebra" for convenience.) An algebra A is **nondegenerate** if $A \neq \{0\}$, if and only if $\mathrm{e} \neq 0$. An algebra A is **commutative** if $fg = gf$ for all $f, g \in A$.

An algebra A which is also a Banach space is called a **Banach algebra** if

$$\|ab\| \leq \|a\|\,\|b\| \qquad \text{for all } a, b \in A.$$

In particular, multiplication is continuous. If $\mathrm{e} \neq 0$ then $\|\mathrm{e}\| \geq 1$.

An element $a \in A$ of an algebra A is called **invertible** if and only if there is an element b such that $ab = ba = \mathrm{e}$. This element, necessarily unique, is then called the **inverse** of a and denoted by a^{-1}.

A linear map $T : A \to B$ between two algebras A, B with unit elements $\mathrm{e}_A, \mathrm{e}_B$, respectively, is called **multiplicative** if

$$T(ab) = (Ta)(Tb) \qquad \text{for all } a, b \in A,$$

and an **algebra homomorphism** if in addition $T(\mathrm{e}_A) = \mathrm{e}_B$. An algebra homomorphism is an **algebra isomorphism** if it is bijective. In this case, its inverse is also an algebra homomorphism.

Let A be a commutative Banach algebra with unit e. An **(algebra) ideal** of A is a linear subspace $I \subseteq A$ satisfying

$$f \in I,\ g \in A \quad \Rightarrow \quad fg \in I.$$

Since the multiplication is continuous, the closure of an ideal is again an ideal. An ideal I of A is called **proper** if $I \neq A$, if and only if $\mathrm{e} \notin I$. For an element $a \in A$, the ideal Aa is called the **principal ideal** generated by a. It is the smallest ideal that contains a. Then a is not invertible if and only if aA is proper. A proper ideal I of A is called **maximal** if

$$I \subseteq J \subseteq A \quad \Rightarrow J = I \text{ or } J = A$$

for any ideal J of A.

If $T : A \to B$ is a continuous algebra homomorphism, then $\ker T$ is a closed ideal. Conversely, if I is closed ideal of A, then the quotient space A/I becomes a Banach algebra with respect to the multiplication

$$(x+I)(y+I) := (xy+I) \qquad (x, y \in A)$$

and unit element $\mathrm{e}+I$. Moreover, the canonical surjection $q : A \to A/I$ is a continuous algebra homomorphism with $\ker q = I$.

An **involution** on a complex Banach algebra A is a map $x \mapsto x^*$ satisfying

$$(x^*)^* = x, \quad (x+y)^* = x^* + y^*, \quad (\lambda x)^* = \overline{\lambda} x^*, \quad (xy)^* = y^* x^*$$

for all $x, y \in A$, $\lambda \in \mathbb{C}$. In an algebra with involution one has $\mathrm{e}^* = \mathrm{e}$ since

$$\mathrm{e}^* = \mathrm{e}^*\mathrm{e} = (\mathrm{e}^*\mathrm{e})^{**} = (\mathrm{e}^*\mathrm{e}^{**})^* = (\mathrm{e}^*\mathrm{e})^* = (\mathrm{e}^*)^* = \mathrm{e}.$$

An algebra homomorphism $T : A \to B$ between algebras with involution is called a $*$-homomorphism if

$$T(x^*) = (Tx)^* \qquad \text{for all } x \in A.$$

A bijective $*$-homomorphism is called a $*$**-isomorphism**.

A Banach algebra with involution is a C^***-algebra** if $\|x\|^2 = \|x^* x\|$ for all $x \in A$. It follows that $\|x^*\| = \|x\|$ for every $x \in A$. If A is a nondegenerate C^*-algebra, then $\|\mathrm{e}\| = 1$.

C.3 Linear Operators

A linear mapping $T : E \to F$ between two normed spaces is called **bounded** if $T(\mathrm{B}_E)$ is a norm-bounded set in F. The linear mapping T is bounded if and only if it is continuous if and only if it satisfies a norm estimate of the form

$$\|Tx\| \le c\,\|x\| \qquad (x \in E) \tag{C.1}$$

for some $c \ge 0$ independent of $x \in E$. The smallest c such that (C.1) holds is called the **operator norm** of T, denoted by $\|T\|$. One has

$$\|T\| = \sup\big\{\,\|Tx\| \,:\, x \in \mathrm{B}_E\big\}.$$

This defines a norm on the space

$$\mathscr{L}(E;F) := \big\{T : E \to F \,:\, T \text{ is a bounded linear mapping}\big\},$$

which is a Banach space if F is a Banach space. If E, F, G are normed spaces and $T \in \mathscr{L}(E;F)$, $S \in \mathscr{L}(F;G)$, then $ST := S \circ T \in \mathscr{L}(E;G)$ with $\|ST\| \le \|S\|\,\|T\|$. The identity operator I is neutral with respect to multiplication of operators, and clearly $\|\mathrm{I}\| = 1$ if $E \ne \{0\}$. In particular, if E is a Banach space, the space

$$\mathscr{L}(E) := \mathscr{L}(E;E)$$

is a Banach algebra with unit element I.

Linear mappings are also called **operators**. Associated with an operator $T : E \to F$ are its **kernel** and its **range**

$$\ker T := \big\{x \in E \,:\, Tx = 0\big\}, \qquad \operatorname{ran} T := \big\{Tx \,:\, x \in E\big\}.$$

One has $\ker T = \{0\}$ if and only if T is injective. If T is bounded, its kernel $\ker T$ is a closed subspace of E.

A subset $\mathscr{T} \subseteq \mathscr{L}(E;F)$ that is a bounded set in the normed space $\mathscr{L}(E;F)$ is often called **uniformly (norm) bounded**. The following result—sometimes called the Banach–Steinhaus theorem—gives an important characterisation. (See [Rudin (1987), Thm. 5.8] for a proof.)

Theorem C.1 (Principle of Uniform Boundedness). *Let E, F be Banach spaces and $\mathscr{T} \subseteq \mathscr{L}(E;F)$. Then $\mathscr{T}$ is uniformly bounded if and only if for each $x \in E$ the set $\{Tx : T \in \mathscr{T}\}$ is a bounded subset of F.*

A bounded linear mapping $T \in \mathscr{L}(E;F)$ is called **contractive** or a **contraction** if $\|T\| \le 1$. It is called **isometric** an **isometry** if $\|Tx\| = \|x\|$ holds for all $x \in E$. Isometries are injective contractions. An operator $P \in \mathscr{L}(E)$ is called a **projection** if $P^2 = P$. In this case, $Q := \mathrm{I} - P$ is also a projection, and P induces a direct sum decomposition

$$E = \operatorname{ran} P \oplus \operatorname{ran}(\mathrm{I} - P) = \operatorname{ran} P \oplus \ker P$$

of E into closed subspaces. Conversely, whenever E is a Banach space and one has a direct sum decomposition $X = F \oplus E$ into *closed* linear subspaces, then the associated projections $X \to F$, $X \to E$ are bounded. This is a consequence of the **Closed Graph Theorem** [Rudin (1987), p.114, Ex.16].

An operator $T \in \mathscr{L}(E;F)$ is called **invertible** or an **isomorphism** (of normed spaces) if T is bijective and T^{-1} is also bounded. If E, F are Banach spaces, the boundedness of T^{-1} is automatic by the following important theorem [Rudin (1987), Thm.5.10].

Theorem C.2 (Inverse Mapping Theorem). *If E, F are Banach spaces and $T \in \mathscr{L}(E;F)$ is bijective, then T is an isomorphism.*

A linear operator $T \in \mathscr{L}(E;F)$ is **compact** if it maps bounded sets in E to relatively compact sets in F. Equivalently, whenever $(x_n)_{n\in\mathbb{N}} \subseteq E$ is a bounded sequence in E, the sequence $(Tx_n)_{n\in\mathbb{N}} \subseteq F$ has a convergent subsequence. The set of compact operators

$$\mathscr{C}(E;F) := \{T \in \mathscr{L}(E;F) \,:\, T \text{ is compact}\}$$

is an (operator-)norm closed linear subspace of $\mathscr{L}(E;F)$ that contains all finite-rank operators. Moreover, a product of operators ST is compact whenever one of its factors is compact. See [Conway (1990), VI.3].

C.4 Duals, Bi-Duals and Adjoints

For a normed space E its **dual space** is $E' := \mathscr{L}(E;\mathbb{K})$. Since $\mathbb{K}$ is complete, this is always a Banach space. Elements of E' are called (bounded linear) **functionals**. One frequently writes $\langle x, x' \rangle$ in place of $x'(x)$ for $x \in E, x' \in E'$ and calls the mapping

$$E \times E' \to \mathbb{K}, \qquad (x, x') \mapsto \langle x, x' \rangle$$

the **canonical duality**.

Theorem C.3 (Hahn–Banach). *Let E be a Banach space, $F \subseteq E$ a linear subspace and $f' \in F'$. Then there exists $x' \in E'$ such that $x' = f'$ on F and $\|x'\| = \|f'\|$.*

For a proof see [Rudin (1987), Thm. 5.16]. A consequence of the Hahn–Banach theorem is that E' *separates the points* of E. Another consequence is that the norm of E can be computed as

$$\|x\| = \sup\{\,|\langle x, x' \rangle| \,:\, \|x'\| \le 1\,\}. \tag{C.2}$$

A third consequence is that E is separable whenever E' is.

Given normed spaces E, F and an operator $T \in \mathscr{L}(E;F)$ we define its **adjoint** operator $T' \in \mathscr{L}(F';E')$ by $T'y' := y' \circ T$, $y' \in F'$. Using the canonical duality this just means

$$\langle Tx, y' \rangle = \langle x, T'y' \rangle \qquad (x \in E, y' \in F').$$

The map $(T \mapsto T') : \mathscr{L}(E;F) \to \mathscr{L}(F';E')$ is linear and isometric, and one has $(ST)' = T'S'$ for $T \in \mathscr{L}(E;F)$ and $S \in \mathscr{L}(F;G)$.

The space E'' is called the **bi-dual** of E. The mapping

$$E \to E'', \qquad x \mapsto \langle x, \cdot \rangle = \big(x' \mapsto \langle x, x' \rangle\big)$$

is called the **canonical embedding**. The canonical embedding is a linear isometry (by (C.2)), and hence one can consider E to be a subspace of E''. Given $T \in \mathscr{L}(E;F)$ one has $T''|_E = T$. If the canonical embedding is surjective, in sloppy notation: $E = E''$, the space E is called **reflexive**.

C.5 The Weak* Topology

Let E be a Banach space and E' its dual. The coarsest topology on E' that makes all mappings

$$x' \mapsto \langle x, x' \rangle \qquad (x \in E)$$

continuous, is called the **weak* topology** (or: $\sigma(E', E)$-topology) on E' (cf. Appendix A.4). A fundamental system of open (and convex) neighbourhoods for a point $y' \in E'$ is given by the sets

$$\Big\{ x' \in E' : \max_{1 \le j \le d} \big| \langle x_j, x' - y' \rangle \big| < \varepsilon \Big\} \qquad (d \in \mathbb{N},\ x_1, \dots, x_d \in E,\ \varepsilon > 0).$$

Since the weak* topology is the projective topology with respect to the functions $(\langle x, \cdot \rangle)_{x \in E}$, a net $(x'_\alpha)_\alpha \subseteq E'$ converges weakly* (that is, in the weak* topology) to $x' \in E'$ if and only if $\langle x, x'_\alpha \rangle \to \langle x, x' \rangle$ for every $x' \in E'$ (cf. Theorem A.14).

Theorem C.4 (Banach–Alaoglu). *Let E be a Banach space. Then the dual unit ball $\mathrm{B}_{E'} = \{x' \in E' : \|x'\| \le 1\}$ is weakly* compact.*

The proof is a more or less straightforward application of Tychonov's Theorem A.5, see [Rudin (1991), Thm. 3.15]. Since the weak* topology is usually not metrisable, one cannot in general test continuity of mappings or compactness of sets via criteria using sequences. Therefore, the following theorem is often useful, see [Dunford and Schwartz (1958), V.5.1].

Theorem C.5. *Let E be a Banach space. Then the weak* topology on the dual unit ball $\mathrm{B}_{E'}$ is metrisable if and only if E is separable.*

Recall that E can be considered (via the canonical embedding) as a norm closed subspace of E''. The next theorem shows in particular that E is $\sigma(E'',E')$-dense in E'', see [Dunford and Schwartz (1958), V.4.5].

Theorem C.6 (Goldstine). *Let E be a Banach space. Then the closed unit ball B_E of E is $\sigma(E'',E')$-dense in the closed unit ball of E''.*

C.6 The Weak Topology

Let E be a Banach space and E' its dual. The coarsest topology on E that makes all mappings

$$x \mapsto \langle x, x' \rangle \qquad (x' \in E')$$

continuous is called the **weak topology** (or: $\sigma(E,E')$-topology) on E (cf. Appendix A.4). A fundamental system of open (and convex) neighbourhoods of the point $y \in E$ is given by the sets

$$\left\{ x \in E : \max_{1 \le j \le d} \left| \langle x-y, x'_j \rangle \right| < \varepsilon \right\} \qquad (d \in \mathbb{N},\ x'_1, \ldots, x'_d \in E',\ \varepsilon > 0).$$

Since the weak topology is the projective topology with respect to the functions $(\langle \cdot, x' \rangle)_{x' \in E'}$, a net $(x_\alpha)_\alpha \subseteq E$ converges weakly (that is, in the weak topology) to $x \in E$ if and only if $\langle x_\alpha, x' \rangle \to \langle x, x' \rangle$ for every $x' \in E'$ (Theorem A.14). When we view $E \subseteq E''$ via the canonical embedding as a subspace and endow E'' with the weak* topology, then the weak topology on E coincides with the subspace topology. If $A \subseteq E$ is a subset of E then its closure in the weak topology is denoted by $\mathrm{cl}_\sigma A$.

Theorem C.7 (Mazur). *Let E be a Banach space and let $A \subseteq E$ be convex. Then the norm closure of A coincides with its weak closure.*

See [Rudin (1991), Thm. 3.12] or [Dunford and Schwartz (1958), V.3.18].

The weak topology is not metrisable in general. Fortunately, there is a series of deep results on metrisability and compactness for the weak topology, facilitating its use. The first is the analogue of Theorem C.5, see [Dunford and Schwartz (1958), V.5.2].

Theorem C.8. *The weak topology on the closed unit ball of a Banach space E is metrisable if and only if the dual space E' is norm separable.*

The next theorem allows using sequences in testing of weak compactness, see Appendix E, Theorem E.17 below, or [Dunford and Schwartz (1958), V.6.1] or the more recent [Vogt (2010)].

Theorem C.9 (Eberlein–Šmulian). *Let E be a Banach space. Then $A \subseteq E$ is (relatively) weakly compact if and only if A is (relatively) weakly sequentially compact.*

Although the weak topology is usually far from being metrisable, the following holds, see the proof of Theorem 16.19 in the main text or [Dunford and Schwartz (1958), V.6.3].

Theorem C.10. *The weak topology on a weakly compact subset of a separable Banach space is metrisable.*

The next result is often useful when considering Banach-space valued (weak) integrals.

Theorem C.11 (Kreĭn–Šmulian). *The closed convex hull of a weakly compact subset of a Banach space is weakly compact.*

For the definition of the closed convex hull see Section C.7 below. For a proof see Appendix E, Theorem E.7 below or [Dunford and Schwartz (1958), V.6.4].

Finally we state a characterisation of reflexivity, which is a mere combination of the Banach–Alaoglu theorem and Goldstine's theorem.

Theorem C.12. *A Banach space is reflexive if and only if its closed unit ball is weakly compact.*

For a proof see [Dunford and Schwartz (1958), V.4.7]. In particular, every bounded set is relatively weakly compact if and only if the Banach space is reflexive.

C.7 Convex Sets and their Extreme Points

A subset $A \subseteq E$ of a real vector space E is called **convex** if, whenever $a, b \in A$ then also $ta + (1-t)b \in A$ for all $t \in [0,1]$. Geometrically this means that the straight line between a and b is contained in A whenever a and b are. Inductively one shows that a convex set contains arbitrary **convex combinations**

$$t_1 a_1 + \cdots + t_n a_n \qquad (0 \le t_1, \ldots, t_n \le 1,\ t_1 + \cdots + t_n = 1)$$

whenever $a_1, \ldots, a_n \in A$. Intersecting convex sets yields a convex set, and so for any set $B \subseteq E$ there is a smallest convex set containing B, its **convex hull**

$$\operatorname{conv}(B) = \bigcap \big\{ A : B \subseteq A,\ A \text{ convex} \big\}.$$

Alternatively, $\operatorname{conv}(B)$ is the set of all convex combinations of elements of B.

A vector space E endowed with a topology such that the operations of addition $E \times E \to E$ and scalar multiplication $\mathbb{K} \times E \to E$ are continuous, is called a **topological vector space**. In a topological vector space, the closure of a convex set is convex. We denote by

$$\overline{\operatorname{conv}}(B) := \overline{\operatorname{conv}(B)}$$

the **closed convex hull** of $B \subseteq E$, which is the smallest closed convex set containing B.

A topological vector space is called **locally convex** if it is Hausdorff and the topology has a base consisting of convex sets. Examples are the norm topology of a Banach space E, the weak topology on E and the weak* topology on E'.

Let $A \subseteq E$ be a convex set. A point $p \in A$ is called an **extreme point** of A if it cannot be written as a convex combinations of two points of A *distinct from* p. In other words, whenever $a, b \in A$ and $t \in [0,1]$ such that $p = ta + (1-t)b$, then $a = b = p$. Let us denote the set of extreme points of a convex set A with $\operatorname{ex} A$.

Theorem C.13 (Kreĭn–Milman). *Let E be a locally convex space and let $\emptyset \neq K \subseteq E$ be compact and convex. Then K is the closed convex hull of its extreme points: $K = \overline{\operatorname{conv}}(\operatorname{ex} K)$. In particular, $\operatorname{ex} K$ is not empty.*

For a proof see [Rudin (1991), Thm. 3.23]. Since for a Banach space E the weak* topology on E' is locally convex, the Kreĭn–Milman theorem applies in particular to weakly* compact convex subsets of E'.

Theorem C.14 (Milman). *Let E be a locally convex space and let $\emptyset \neq K \subseteq E$ be compact. If $\overline{\operatorname{conv}} K$ is also compact, then K contains the extreme points of $\overline{\operatorname{conv}} K$.*

A proof is in [Rudin (1991), Thm. 3.25]. Combining Milman's result with the Kreĭn–Šmulian Theorem C.11 yields the following.

Corollary C.15. *Let E be a Banach space and let $K \subseteq E$ be weakly compact. Then K contains the extreme points of $\overline{\operatorname{conv}} K$.*

C.8 The Strong and the Weak Operator Topology on $\mathscr{L}(E)$

Let E be a Banach space. Besides the norm topology there are two other canonical topologies on $\mathscr{L}(E)$. The **strong operator topology** is the coarsest topology such that all evaluation mappings

$$\mathscr{L}(E) \to E, \quad T \mapsto Tx \qquad (x \in E)$$

are continuous. Equivalently, it is the projective topology with respect to the family $(T \mapsto Tx)_{x \in E}$. A fundamental system of open (and convex) neighbourhoods of $T \in \mathscr{L}(E)$ is given by the sets

$$\Big\{ S \in \mathscr{L}(E) : \max_{1 \le j \le d} \big\| Tx_j - Sx_j \big\| < \varepsilon \Big\} \qquad (d \in \mathbb{N},\, x_1, \ldots, x_d \in E,\, \varepsilon > 0).$$

A net of operators $(T_\alpha)_\alpha \subseteq \mathscr{L}(E)$ **converges strongly** (i.e., in the strong operator topology) to T if and only if

$$T_\alpha x \to Tx \quad \text{for every } x \in E.$$

We denote by $\mathscr{L}_s(E)$ the space $\mathscr{L}(E)$ endowed with the strong operator topology, and the closure of a subset $A \subseteq \mathscr{L}(E)$ in this topology is denoted by $\mathrm{cl}_s A$. The strong operator topology is metrisable on (norm-)bounded sets if E is separable.

The **weak operator topology** is the coarsest topology such that all evaluation mappings

$$\mathscr{L}(E) \to \mathbb{K}, \quad T \mapsto \langle Tx, x' \rangle \qquad (x \in E, x' \in E')$$

are continuous. Equivalently, it is the projective topology with respect to the family $(T \mapsto \langle Tx, x' \rangle)_{x \in E, x' \in E'}$. A fundamental system of open (and convex) neighbourhoods of $T \in \mathscr{L}(E)$ is given by the sets

$$\Big\{ S \in \mathscr{L}(E) : \max_{1 \le j \le d} \big| \langle (T-S)x_j, x_j' \rangle \big| < \varepsilon \Big\}$$
$$(d \in \mathbb{N},\ x_1, \dots, x_d \in E,\ x_1', \dots, x_d' \in E',\ \varepsilon > 0).$$

A net of operators $(T_\alpha)_\alpha$ converges to T in the weak operator topology if and only if

$$\langle T_\alpha x, x' \rangle \to \langle Tx, x' \rangle \quad \text{for every } x \in E,\ x' \in E'.$$

The space $\mathscr{L}(E)$ considered with the weak operator topology is denoted by $\mathscr{L}_\sigma(E)$, and the closure of a subset $A \subseteq \mathscr{L}(E)$ in this topology is denoted by $\mathrm{cl}_\sigma A$. The weak operator topology is metrisable on norm-bounded sets if E' is separable.

The following is a consequence of the Uniform Boundedness Principle.

Proposition C.16. *Let E be a Banach space and let $\mathscr{T} \subseteq \mathscr{L}(E)$. Then the following assertions are equivalent.*

(i) *$\mathscr{T}$ is bounded for the weak operator topology, i.e., $\sup_{T \in \mathscr{T}} |\langle Tx, x' \rangle| < \infty$ for all $x \in E$ and $x' \in E'$.*

(ii) *$\mathscr{T}$ is bounded for the strong operator topology, i.e., $\sup_{T \in \mathscr{T}} \|Tx\| < \infty$ for all $x \in E$.*

(iii) *$\mathscr{T}$ is uniformly bounded, i.e., $\sup_{T \in \mathscr{T}} \|T\| < \infty$.*

The following simple property of strong operator convergence is very useful.

Proposition C.17. *For a uniformly bounded net $(T_\alpha)_\alpha \subseteq \mathscr{L}(E)$ and $T \in \mathscr{L}(E)$ the following assertions are equivalent.*

(i) *$T_\alpha \to T$ in $\mathscr{L}_s(E)$.*

(ii) *$T_\alpha x \to Tx$ for all x from a (norm-)dense set $D \subseteq E$.*

(iii) *$T_\alpha x \to Tx$ uniformly in x from (norm-)compact subsets of E.*

We now consider continuity of the multiplication

$$(S, T) \mapsto ST \qquad (S, T \in \mathscr{L}(E))$$

for the operator topologies considered above.

Proposition C.18. *The multiplication on $\mathscr{L}(E)$ is*

a) *jointly continuous for the norm topology in $\mathscr{L}(E)$;*

b) *jointly continuous on bounded sets for the strong operator topology;*

c) *separately continuous for the weak and strong operator topology.*

The multiplication is in general not jointly continuous for the weak operator topology. Indeed, if T is the **right shift** on $\ell^2(\mathbb{Z})$ given by $T((x_k)_{k\in\mathbb{Z}}) := (x_{k-1})_{k\in\mathbb{Z}}$. Then $(T^n)_n$ converges to 0 weakly, and the same holds for the left shift T^{-1}. But $(TT^{-1})^n = \mathrm{I}$ does not converge to 0.

C.9 Spectral Theory

Let E be an at least one-dimensional complex Banach space and $T \in \mathscr{L}(E)$. The **resolvent set** $\rho(T)$ of T is the set of all $\lambda \in \mathbb{C}$ such that the operator $\lambda\mathrm{I} - T$ is invertible. The function

$$\rho(T) \to \mathscr{L}(E), \quad \lambda \mapsto R(\lambda, T) := (\lambda\mathrm{I} - T)^{-1}$$

is called the **resolvent** of T. Its complement $\sigma(T) := \mathbb{C} \setminus \rho(T)$ is called the **spectrum** of T. The resolvent set is an open subset of $\mathbb{C}$, and given $\lambda_0 \in \rho(T)$ one has

$$R(\lambda, A) = \sum_{n=0}^{\infty} (\lambda_0 - \lambda)^n R(\lambda_0, T)^{n+1} \quad \text{for} \quad |\lambda - \lambda_0| < \|R(\lambda_0, T)\|^{-1}.$$

In particular, $\operatorname{dist}(\lambda_0, \sigma(T)) \geq \|R(\lambda_0, T)\|^{-1}$, showing that the norm of the resolvent blows up when λ_0 approaches a spectral point.

Every $\lambda \in \mathbb{C}$ with $|\lambda| > \|T\|$ is contained in $\rho(T)$, and in this case $R(\lambda, T)$ is given by the Neumann series

$$R(\lambda, T) := \sum_{n=0}^{\infty} \lambda^{-(n+1)} T^n.$$

In particular, one has $r(T) \leq \|T\|$, where

$$r(T) := \sup\{|\lambda| \,:\, \lambda \in \sigma(T)\}$$

is the **spectral radius** of T. An important result states that the spectrum is always a *nonempty* compact subset of $\mathbb{C}$ and the spectral radius can be computed by the formula

$$r(T) = \lim_{n\to\infty} \|T^n\|^{1/n} = \inf_{n\in\mathbb{N}} \|T^n\|^{1/n}.$$

(See [Rudin (1991), Thm 10.13] for a proof.) An operator is called **power-bounded** if $\sup_n \|T^n\| < \infty$. From the spectral radius formula it follows that the spectrum of a power-bounded operator is contained in the closed unit disc. For isometries one has more precise information.

Proposition C.19. *If E is a complex Banach space and $T \in \mathscr{L}(E)$ is an isometry, then exactly one of the following two cases holds.*

1) *T is not surjective and $\sigma(T) = \{z : |z| \leq 1\}$ is the full closed unit disc.*
2) *T is an isomorphism and $\sigma(T) \subseteq \mathbb{T}$.*

An important part of the spectrum is the **point spectrum**

$$\sigma_p(T) := \big\{\lambda \in \mathbb{C} : (\lambda \mathrm{I} - T) \text{ is not injective } \big\}.$$

Each $\lambda \in \sigma_p(T)$ is called an **eigenvalue** of T. For $\lambda \in \sigma_p(T)$ the closed subspace

$$\ker(\lambda \mathrm{I} - T)$$

is called the corresponding **eigenspace**, and every nonzero element $0 \neq x \in \ker(\lambda \mathrm{I} - T)$ is a corresponding **eigenvector**. An eigenvalue λ is called **simple** if its eigenspace is one-dimensional.

C.10 Hilbert Spaces

An **inner product** on a complex vector space E is a mapping $(\cdot\,|\,\cdot) : E \times E \to \mathbb{C}$ that is **sequilinear**, i.e., satisfies

$$(f + \lambda g \,|\, h) = (f \,|\, h) + \lambda\,(g \,|\, h) \tag{C.3}$$

$$(h \,|\, f + \lambda g) = (h \,|\, f) + \overline{\lambda}\,(h \,|\, g) \qquad (f, g, h \in E, \lambda \in \mathbb{C}), \tag{C.4}$$

and **symmetric**, i.e., satisfies

$$(f \,|\, g) = \overline{(g \,|\, f)} \qquad (f, g \in E)$$

and **positive definite**, i.e., satisfies

$$0 \neq f \quad \Rightarrow \quad (f \,|\, f) > 0.$$

The **norm** associated with an inner product on E is

$$\|f\| := \sqrt{(f \,|\, f)}. \tag{C.5}$$

For inner product and norm the the following elementary identities hold:

a) $\|f + g\|^2 = \|f\|^2 + \|g\|^2 + 2\,\mathrm{Re}\,(f \,|\, g)$;
b) $4\,\mathrm{Re}\,(f \,|\, g) = \|f + g\|^2 - \|f - g\|^2$;
c) $2\,\|f\|^2 + 2\,\|g\|^2 = \|f + g\|^2 + \|f - g\|^2$.

Here b) is called the **polarisation identity** and a) is called the **parallelogram identity**.

In an inner product space two vectors f,g are called **orthogonal**, in symbols: $f \perp g$, if $(f\,|\,g) = 0$. For a set $A \subseteq E$ one defines

$$A^{\perp} := \big\{ f \in E : f \perp g \,\forall\, g \in A \big\}.$$

Two sets $A, B \subseteq E$ are called orthogonal, in symbols $A \perp B$ if $f \perp g$ for all $f \in A, g \in B$. Property a) above accounts for **Pythagoras' lemma**

$$f \perp g \quad \Longleftrightarrow \quad \|f+g\|^2 = \|f\|^2 + \|g\|^2 .$$

That (C.5) indeed defines a norm on E is due to the **Cauchy–Schwarz inequality**

$$|(f\,|\,g)| \le \|f\|\,\|g\| \qquad (f,g \in E).$$

A **inner product space** is a vector space E together with an inner product on it. An inner product space $(E, (\cdot\,|\,\cdot))$ is a **Hilbert space** if E is a Banach space with respect to the norm (C.5).

The following result is fundamental in the theory of Hilbert spaces. For a proof see [Rudin (1991), Theorem 4.12] or [Conway (1990), Theorem I.3.4].

Theorem C.20 (Riesz–Fréchet). *Let H be a Hilbert space and let $\psi : H \to \mathbb{C}$ be a bounded linear functional on H. Then there is a unique $g \in H$ such that $\psi(f) = (f\,|\,g)$ for all $f \in H$.*

An important consequence of the Riesz–Fréchet theorem is that a net $(f_\alpha)_\alpha$ in H converges weakly to a vector $f \in H$ if and only if $(f_\alpha - f\,|\,g) \to 0$ for each $g \in H$.

Corollary C.21. *Each Hilbert space is reflexive. In particular, the closed unit ball of a Hilbert space is weakly compact.*

Also the following result is based on the Riesz–Fréchet theorem.

Corollary C.22. *Let H, K be Hilbert spaces and let $b : H \times K \to \mathbb{C}$ be a sesquilinear map (see* (C.3),(C.4)*) such that there is $M \ge 0$ with*

$$|b(f,g)| \le M\|f\|\,\|g\| \qquad (f \in H, g \in K).$$

Then there is a unique linear operator $S : H \to K$ with

$$b(f,g) = (Sf\,|\,g) \qquad (f \in H, g \in K).$$

For a given bounded linear operator $T : K \to H$ there is hence a unique linear operator $T^* : H \to K$ such that

$$(T^* f\,|\,g)_K = (f\,|\,Tg)_H \qquad (f \in H, g \in K).$$

The operator T^* is called the **(Hilbert space) adjoint** of T.

An operator $T \in \mathscr{L}(H)$ is called **self-adjoint** if $T^* = T$. Self-adjoint operators have a nice spectral theory, in particular if they are compact.

Theorem C.23 (Spectral Theorem). *Let T be a compact self-adjoint operator on a Hilbert space H. Then H has an orthonormal basis consisting of eigenvectors of T, and every eigenspace corresponding to a nonzero eigenvector is finite-dimensional. More precisely, A is the norm convergent sum*

$$A = \sum_{\lambda \in \Lambda} \lambda P_\lambda \tag{C.6}$$

where $\Lambda \subseteq \mathbb{C} \setminus \{0\}$ is a finite or countable set, for each $\lambda \in \Lambda$ the eigenspace $E_\lambda := \ker(\lambda \mathrm{I} - A)$ satisfies $0 < \dim E_\lambda < \infty$ and P_λ is the orthogonal projection onto E_λ.

It is easy to see that the eigenspaces E_λ are pairwise orthogonal. A proof of the Spectral Theorem can be found in [Conway (1990), II.5.1].

Appendix D
The Riesz Representation Theorem

In this appendix we give a proof for the Riesz Representation Theorem 5.7. In the whole section, K denotes a nonempty compact topological space and $\mathrm{Ba}(K)$ and $\mathrm{Bo}(K)$ denote the σ-algebra of Baire and Borel sets of K, respectively, i.e.,

$$\mathrm{Ba}(K) = \sigma\big\{[f > 0] : 0 \le f \in \mathrm{C}(K)\big\} \quad \text{and} \quad \mathrm{Bo}(K) = \sigma\big\{O : O \subseteq K \text{ open}\big\}.$$

Moreover, $\mathrm{M}(K)$ denotes the space of complex Baire measures on K. It will eventually become clear that each such Baire measure has a unique extension to a regular Borel measure.

Each $\mu \in \mathrm{M}(K)$ induces a linear functional

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

This functional is bounded (by $\|\mu\|_{\mathrm{M}}$) since

$$|\langle f, \mu \rangle| \le \int_K |f| \,\mathrm{d}|\mu| \le \|f\|_\infty |\mu|(K) = \|f\|_\infty \|\mu\|_{\mathrm{M}}$$

for each $f \in \mathrm{C}(K)$. In this way we obtain a linear and contractive map

$$\mathrm{d} : \mathrm{M}(K) \to \mathrm{C}(K)', \qquad \mu \mapsto \mathrm{d}\mu = \langle \cdot, \mu \rangle \tag{D.1}$$

of $\mathrm{M}(K)$ into the dual space of $\mathrm{C}(K)$. The statement of the Riesz representation theorem is that this map is an isometric isomorphism.

D.1 Uniqueness

Let us introduce the space

$$\mathrm{BM}(K) := \big\{f : K \to \mathbb{C} : f \text{ is bounded and Baire measurable}\big\}.$$

This is a commutative C^*-algebra with respect to the sup-norm. Moreover, it is closed under the so-called bp-convergence. We say that a sequence $(f_n)_{n\in\mathbb{N}}$ of functions on K converges **boundedly and pointwise** (or: bp-converges) to a function f if

$$(1)\quad \sup_{n\in\mathbb{N}} \|f_n\|_\infty < \infty \quad \text{and} \qquad (2)\quad f_n(x) \to f(x) \quad \text{for each } x \in K.$$

The following result is very useful in extending results from continuous to general Baire measurable functions.

Theorem D.1. *Let K be a compact topological space, and let $E \subseteq \mathrm{BM}(K)$ such that* (1) $\mathrm{C}(K) \subseteq E$ *and* (2) E *is closed under bp-convergence. Then* $E = \mathrm{BM}(K)$.

Proof. By taking the intersection of all subsets E of $\mathrm{BM}(K)$ with the properties (1) and (2), we may suppose that E is the smallest one, i.e., is contained in each of them. We show first that E is a linear subspace of $\mathrm{BM}(K)$. For $f \in \mathrm{BM}(K)$ we define

$$E_f := \big\{g \in E : f+g \in E\big\}.$$

If $f \in \mathrm{C}(K)$, then E_f has the properties (1) and (2), whence $E \subseteq E_f$. This means that $\mathrm{C}(K)+E \subseteq E$. In particular, it follows that E_f has properties (1) and (2) even if $f \in E$. Hence $E \subseteq E_f$ for each $f \in E$, whence $E+E \subseteq E$. For $\lambda \in \mathbb{C}$ the set

$$\big\{g \in E : \lambda g \in E\big\}$$

has properties (1) and (2), whence $\lambda E \subseteq E$. This establishes that E is a linear subspace of $\mathrm{BM}(K)$.

In the same way one can show that if $f, g \in E$, then $|f| \in E$ and $fg \in E$. It follows in view of condition (2) that the set

$$\mathcal{F} := \big\{A \in \mathrm{Ba}(K) : \mathbf{1}_A \in E\big\}$$

is a σ-algebra. If $0 \le f \in \mathrm{C}(K)$ then $nf \wedge \mathbf{1} \nearrow \mathbf{1}_{[f>0]}$ pointwise. Hence $[f>0] \in \Sigma$, and therefore $\Sigma = \mathrm{Ba}(K)$. By standard measure theory, see Lemma B.7, $\mathrm{lin}\{\mathbf{1}_A : A \in \mathrm{Ba}(K)\}$ is sup-norm dense in $\mathrm{BM}(K)$, so by (2) we obtain $E = \mathrm{BM}(K)$. □

As a consequence we recover Lemma 5.5.

Corollary D.2 (Uniqueness). *Let K be a compact topological space. If μ, ν are complex Baire measures on K such that*

$$\int_K f \,\mathrm{d}\mu = \int_K f \,\mathrm{d}\nu \qquad \textit{for all } f \in \mathrm{C}(K),$$

then $\mu = \nu$.

Proof. a) Define $E := \{f \in \mathrm{BM}(K) : \int f\mathrm{d}\mu = \int f\mathrm{d}\nu\}$. By dominated convergence, E satisfies (1) and (2) in Theorem D.1, hence $E = \mathrm{BM}(K)$. □

To prove that the mapping d defined in (D.1) is isometric, we need a refined approximation result.

Lemma D.3. *Let $0 \le \mu \in \mathrm{M}(K)$. Then $\mathrm{C}(K)$ is dense in $\mathrm{L}^p(K,\mu)$ for each $1 \le p < \infty$. More precisely, for $f \in \mathrm{L}^\infty(K,\mu)$ there is a sequence of continuous functions $f_n \in \mathrm{C}(K)$ such that $f_n \to f$ μ-a.e. and $|f_n| \le \|f\|_{\mathrm{L}^\infty}$ for all $n \in \mathbb{N}$.*

Proof. Let $1 \le p < \infty$ and let Y be the L^p-closure of $\mathrm{C}(K)$ with respect to μ. Define $E := \mathrm{BM}(K) \cap Y$. Then E satisfies (1) and (2) of Theorem D.1, whence $E = \mathrm{BM}(K)$. But this implies that $\mathrm{L}^\infty(K,\mu) \subseteq Y$, whence $Y = \mathrm{L}^p(K,\mu)$.

For the second assertion, let $0 \ne f \in \mathrm{L}^\infty(K,\mu)$. Then, by the part already proved, there is a sequence $(g_n)_{n\in\mathbb{N}}$ of continuous functions such that $g_n \to f$ in L^1. Passing to a subsequence we may suppose that $g_n \to f$ μ-almost everywhere. Now define

$$f_n := \begin{cases} \|f\|_\infty g_n / |g_n| & \text{on } [\,|g_n| \ge \|f\|_\infty], \\ g_n & \text{on } [\,|g_n| \le \|f\|_\infty]. \end{cases}$$

Then f_n is continuous, $|f_n| \le \|f\|_\infty$, and $f_n \to f$ μ-almost everywhere. □

Now the claimed isometric property of the mapping d is in reach.

Corollary D.4. *The map $\mathrm{d} : \mathrm{M}(K) \to \mathrm{C}(K)'$ is isometric.*

Proof. Fix $\mu \in \mathrm{M}(K)$. The functional $\mathrm{d}\mu$ is bounded by 1 with respect to the norm of $\mathrm{L}^1(K,|\mu|)$:

$$|\langle f,\mu\rangle| \le \langle |f|, |\mu|\rangle = \|f\|_{\mathrm{L}^1(|\mu|)}$$

for $f \in \mathrm{C}(K)$. Hence there is $g \in \mathrm{L}^\infty(K,|\mu|)$ such that

$$\langle f,\mu\rangle = \int_K fg \,\mathrm{d}|\mu| \qquad \text{for all } f \in \mathrm{C}(K).$$

Let $h := \overline{g}/|g|$ on the set where $g \ne 0$, and $h = 0$ elsewhere. By Lemma D.3 from above there is a sequence of continuous functions $f_n \in \mathrm{C}(K)$ with $\|f_n\|_\infty \le 1$ and $f_n \to h$ in $\mathrm{L}^1(K,|\mu|)$. But then

$$\langle f_n,\mu\rangle = \int_K f_n g \,\mathrm{d}|\mu| \to \int_K hg \,\mathrm{d}|\mu| = |\mu|(K) = \|\mu\|_{\mathrm{M}}.$$

This implies that $\|\mu\|_{\mathrm{M}} \le \|\mathrm{d}\mu\|_{\mathrm{C}(K)'}$, whence d is isometric. □

D.2 The Space $\mathrm{C}(K)'$ as a Banach Lattice

In this section we reduce the problem of proving the surjectivity of the mapping d from (D.1), i.e., the representation problem, to positive functionals. We do this by showing that each bounded linear functional on $\mathrm{C}(K)$ can be written as a linear combination of positive functionals. Actually we shall show more: $\mathrm{C}(K)'$ is a *complex*

Banach lattice. (See Definition 7.2 for the terminology and recall from Example 7.3.2 that $\mathrm{C}(K)$ is a complex Banach lattice.)

Let us recall that a linear functional $\lambda : \mathrm{C}(K) \to \mathbb{C}$ is called **positive**, written $\lambda \geq 0$, if

$$\lambda(f) \geq 0 \quad \text{for all } 0 \leq f \in \mathrm{C}(K).$$

This means that λ is a **positive operator** in the sense of Section 7.1.

Lemma D.5. *Each positive linear functional λ on $\mathrm{C}(K)$ is bounded with norm $\|\lambda\| = \lambda(\mathbf{1})$.*

Proof. This is actually a general fact from Banach lattice theory, see Lemma 7.5, but we give a direct proof here. Let $f \in \mathrm{C}(K)$ and determine $c \in \mathbb{T}$ such that $c\lambda(f) = |\lambda(f)|$. Then

$$|\lambda(f)| = \mathrm{Re}(c\lambda(f)) = \mathrm{Re}\,\lambda(cf) = \lambda(\mathrm{Re}\,cf) \leq \lambda\,|f| \leq \lambda(\mathbf{1})\,\|f\|_\infty .$$

Inserting $f = \mathbf{1}$ concludes the proof. □

Our aim is to show that the notion of positivity from the above turns $\mathrm{C}(K)'$ into a Banach lattice. In order to to decompose $\mathrm{C}(K)'$ naturally into a real and an imaginary part, for $\lambda \in \mathrm{C}(K)'$ we define $\lambda^* \in \mathrm{C}(K)$ by

$$\lambda^*(f) = \overline{\lambda(\overline{f})} \qquad (f \in \mathrm{C}(K))$$

and subsequently

$$\mathrm{Re}\,\lambda := \frac{1}{2}(\lambda + \lambda^*) \quad \text{and} \quad \mathrm{Im}\,\lambda := \frac{1}{2\mathrm{i}}(\lambda - \lambda^*).$$

A linear functional $\lambda \in \mathrm{C}(K)'$ is called **real** if $\lambda = \lambda^*$, and

$$\mathrm{C}(K)'_{\mathbb{R}} := \{\lambda \in \mathrm{C}(K)' : \lambda = \lambda^*\}$$

is the $\mathbb{R}$-linear subspace of real functionals. Then we have

$$\lambda \in \mathrm{C}(K)'_{\mathbb{R}} \quad \Leftrightarrow \quad \lambda(f) \in \mathbb{R} \quad \text{for all } 0 \leq f \in \mathrm{C}(K).$$

Moreover, $\lambda = \mathrm{Re}\,\lambda + \mathrm{i}\,\mathrm{Im}\,\lambda$ for each $\lambda \in \mathrm{C}(K)'$, and hence

$$\mathrm{C}(K)' = \mathrm{C}(K)'_{\mathbb{R}} \oplus \mathrm{i}\mathrm{C}(K)'_{\mathbb{R}}$$

is the desired decomposition. Clearly, each positive linear functional is real, and by

$$\lambda \leq \mu \quad \overset{\text{Def.}}{\Leftrightarrow} \quad \lambda(f) \leq \mu(f) \quad \text{for each } 0 \leq f \in \mathrm{C}(K)$$

a partial order is defined in $\mathrm{C}(K)'_{\mathbb{R}}$ turning it into a real ordered vector space.

In the next step we construct the modulus mapping, and for the we need a technical tool.

Theorem D.6 (Partition of Unity). *Let K be a compact space, and let $O_1,\dots,O_n \subseteq K$ be open subsets such that*

$$K = O_1 \cup \cdots \cup O_n.$$

Then there are functions $0 \le f_j \in \mathrm{C}(K)$, $j = 1,\dots,n$, with

$$\sum_{j=1}^n f_j = \mathbf{1} \quad \textit{and} \quad \operatorname{supp}(f_j) \subseteq O_j \quad \textit{for all } j = 1,\dots,n. \tag{D.2}$$

Proof. We first construct compact sets $A_j \subseteq O_j$ for $j = 1,\dots,n$ with

$$K = \bigcup_{j=1}^n A_j. \tag{D.3}$$

By hypothesis, for each $x \in K$ there is $j(x) \in \{1,\dots,n\}$ such that $x \in O_{j(x)}$. Since singletons are closed, Lemma 4.1 yields a closed set B_x and an open set U_x such that

$$x \in U_x \subseteq B_x \subseteq O_{j(x)} \qquad (x \in K).$$

By compactness, there is a finite set $F \subseteq K$ such that $K \subseteq \bigcup_{x\in F} U_x$. Define

$$A_j := \bigcup \big\{ A_x : x \in F,\, j(x) = j \big\} \qquad (j = 1,\dots,n).$$

Then every A_j is closed, $A_j \subseteq O_j$ and (D.3) holds as required.

In the second step we apply Urysohn's Lemma 4.2 to obtain continuous functions $g_j \in \mathrm{C}(K)$ such that

$$\mathbf{1}_{A_j} \le g_j \le \mathbf{1}_{O_j} \quad \text{and} \quad \operatorname{supp}(g_j) \subseteq O_j \qquad (j = 1,\dots,n).$$

Now note that $\sum_{k=1}^n g_k \ge \sum_{k=1}^n \mathbf{1}_{A_k} \ge \mathbf{1}$, whence the well-defined continuous functions

$$f_j := \frac{g_j}{\sum_{k=1}^n g_k} \qquad (j = 1,\dots,n)$$

are positive and satisfy $\sum_{j=1}^n f_j = \mathbf{1}$. Moreover, $\operatorname{supp}(f_j) \subseteq \operatorname{supp}(g_j) \subseteq O_j$ for each $j = 1,\dots,n$, as required. □

The sequence $(f_j)_{j=1}^n$ yielded by the theorem above are called **partition of unity subordinate to the covering** $O_1,\dots,O_n$.

For $0 \le f \in \mathrm{C}(K)$ we consider the "order-ball"

$$B_f := \big\{ g \in \mathrm{C}(K) : |g| \le f \big\}$$

around 0 with "radius" f. Based on the partitions of unity, we can prove the following result.

Lemma D.7. *For a compact space K the following assertions hold.*

a) *Given* $0 \le f \in \mathrm{C}(K)$*, the set of functions* $g = \sum_{j=1}^{n} \alpha_j f_j$ *with*

$$n \in \mathbb{N},\ 0 \le f_j \in \mathrm{C}(K),\ \alpha_j \in \mathbb{T} \quad (j = 1, \dots, n), \quad \sum_{j=1}^{n} f_j \le f$$

is dense in B_f.

b) *Given* $0 \le f_1, f_2 \dots, f_n \in \mathrm{C}(K)$*, the set of functions* $g = \sum_{j=1}^{n} g_j$ *with*

$$g_j \in \mathrm{C}(K),\ |g_j| \le f_j \quad (j = 1, \dots, n)$$

is dense in $B_{f_1 + \cdots + f_n}$.

Proof. a) Let $|g| \le f$ and $\varepsilon > 0$. For each $x \in K$ let $\alpha_x \in \mathbb{T}$ such that $\alpha_x |g(x)| = g(x)$. By compactness of K there are finitely many points $x_1, \dots, x_n \in K$ such that the open sets

$$U_j := \big[\, |g - \alpha_{x_j} |g|| < \varepsilon \,\big] \qquad (j = 1, \dots, n)$$

cover K. Let $(\psi_j)_{j=1}^n$ be a subordinate partition of unity and define $f_j := |g|\, \psi_j \ge 0$. Then

$$\sum_{j=1}^{n} f_j = |g| \le f \quad \text{and} \quad \Big| g - \sum_{j=1}^{n} \alpha_{x_j} f_j \Big| \le \varepsilon.$$

b) Fix $|g| \le f_1 + \cdots + f_n$ and $\varepsilon > 0$. The open sets

$$U_1 := [\, |g| > 0 \,] \quad \text{and} \quad U_2 := [\, |g| < \varepsilon \,]$$

cover K. Take a subordinate partition of unity ψ and $\mathbf{1} - \psi$ and define

$$g_j := \begin{cases} \psi g \frac{f_j}{f_1 + \cdots + f_n} & \text{on} \quad \operatorname{supp}(\psi), \\ 0 & \text{on} \quad [\, \psi = 0 \,] \end{cases}$$

for $j = 1, \dots, n$. Then $g_1 + g_2 + \cdots + g_n = \psi_2 g$, $|g_j| \le f_j$ on K and $g_j \in \mathrm{C}(K)$. The proof is concluded by

$$\Big| g - \sum_{j=1}^{n} g_j \Big| = |(\mathbf{1} - \psi)|\, |g| \le \varepsilon \mathbf{1}. \qquad \square$$

Finally, we can define the **modulus** of a functional $\lambda \in \mathrm{C}(K)'$ by

$$|\lambda|\,(f) := \sup\big\{ |\lambda(g)| : g \in \mathrm{C}(K),\ |g| \le f \big\} = \sup_{g \in B_f} |\lambda(g)| \tag{D.4}$$

for $0 \le f \in \mathrm{C}(K)$. This is a finite number since we have $|\lambda|\,(f) \le \|\lambda\|\, \|f\|_\infty < \infty$. On a moments reflection, we also notice that

$$\|\lambda\| = |\lambda|\,(\mathbf{1}) \qquad (\lambda \in \mathrm{C}(K)'). \tag{D.5}$$

The following is the key step.

Lemma D.8. *For* $\lambda \in \mathrm{C}(K)'$*, the mapping* $|\lambda|$ *defined by* (D.4) *on the positive cone on* $\mathrm{C}(K)$ *extends (uniquely) to a bounded linear functional on* $\mathrm{C}(K)$.

Proof. The positive homogeneity of $|\lambda|$ is obvious from the definition. We turn to prove additivity of $|\lambda|$. Take $f_1, f_2 \geq 0$ and $|g_1| \leq f_1$ and $|g_2| \leq f_2$. Then for certain $c_1, c_2 \in \mathbb{T}$

$$|\lambda(g_1)| + |\lambda(g_2)| = |c_1\lambda(g_1) + c_2\lambda(g_2)| = |\lambda(c_1 g_1 + c_2 g_2)| \leq |\lambda|(f_1 + f_2)$$

since $|c_1 g_1 + c_2 g_2| \leq f_1 + f_2$. By varying g_1, g_2 we obtain

$$|\lambda| f_1 + |\lambda| f_2 \leq |\lambda|(f_1 + f_2).$$

For the converse inequality take $|g| \leq f_1 + f_2$ and $\varepsilon > 0$. By Lemma D.7 we find $|g_1| \leq f_1$ and $|g_2| \leq f_2$ such that $\|g - (g_1 + g_2)\|_\infty \leq \varepsilon$. This yields

$$|\lambda(g)| \leq |\lambda(g_1 + g_2)| + \varepsilon \|\mu\| \leq |\lambda| f_1 + |\lambda| f_2 + \varepsilon \|\lambda\|.$$

Letting $\varepsilon \searrow 0$ and varying g yields $|\lambda|(f_1 + f_2) \leq |\lambda| f_1 + |\lambda| f_2$ as desired.

Finally, we extend $|\lambda|$ first to all of $C(K;\mathbb{R})$ by

$$|\lambda|(f) := |\lambda|(f^+) - |\lambda|(f^-) \qquad (f \in C(K;\mathbb{R})),$$

and then to $C(K)$ by

$$|\lambda|(f) := |\lambda|(\operatorname{Re} f) + \mathrm{i}|\lambda|(\operatorname{Im} f) \qquad (f \in C(K)).$$

To prove that in this way $|\lambda|$ becomes a linear mapping on all of $C(K)$ is tedious but straighforward, and we omit the proof. □

We note that by definition

$$|\lambda(f)| \leq |\lambda||f| \qquad \text{for all } f \in C(K).$$

It follows that if λ is a real functional, then $\lambda \leq |\lambda|$. Hence $\lambda = \lambda - (|\lambda| - \lambda)$ is a difference of positive functionals.

Corollary D.9. *Every bounded real functional on* $C(K)$ *is the difference of two positive linear functionals. The positive linear functionals generate* $C(K)'$ *as a vector space.*

In order to complete the proof that $C(K)'$ is a Banach lattice we note that by (D.5) we have

$$|\lambda| \leq |\mu| \quad \Rightarrow \quad \|\lambda\| \leq \|\mu\|,$$

and hence the norm on $C(K)'$ is a *lattice norm*. The next result connects the modulus and the real lattice structure, characteristic for complex vector lattices, see Definition 7.2 and [Schaefer (1974), Definition II.11.3].

Proposition D.10. *For* $\lambda \in C(K)'$ *one has* $|\lambda| = \sup_{c\in\mathbb{T}} \operatorname{Re}(c\lambda)$ *in the ordering of* $C(K)'_{\mathbb{R}}$.

Proof. Fix $c \in \mathbb{T}$ and $f \geq 0$. Then

$$\begin{aligned}[\mathrm{Re}(c\lambda)]f &= \frac{1}{2}(c\lambda + \overline{c}\lambda^*)(f) = \frac{1}{2}\big(c\lambda(f) + \overline{c}\overline{\lambda(f)}\big) \\ &= \mathrm{Re}[c\lambda(f)] \leq |\lambda(f)| \leq |\lambda| f.\end{aligned}$$

On the other hand, suppose that ν satisfies $\mathrm{Re}(c\lambda) \leq \nu$ for all $c \in \mathbb{T}$. Let $g = \sum_{j=1}^n \alpha_j f_j \in B_f$ be as in Lemma D.7, and let $c_j \in \mathbb{T}$ be such that $c_j \lambda(f_j) = \big|\lambda(f_j)\big|$ for each j. Then

$$\begin{aligned}|\lambda(g)| &\leq \textstyle\sum_{j=1}^n \big|\alpha_j\big|\big|\lambda(f_j)\big| = \sum_{j=1}^n \big|\lambda(f_j)\big| \\ &= \textstyle\sum_{j=1}^n \big(\mathrm{Re}(c_j\lambda)\big)(f_j) \leq \sum_{j=1}^n \nu(f_j) \leq \nu(f).\end{aligned}$$

By Lemma D.7 we obtain that $|\lambda(g)| \leq \nu(f)$ for all $g \in B_f$, and hence $|\lambda| \leq \nu$. This completes the proof. □

D.3 Representation of Positive Functionals

By Corollary D.9, to complete the proof of the Riesz representation theorem it suffices to establish the following result.

Theorem D.11 (Riesz–Markov). *Let λ be a positive linear functional on the Banach space $\mathrm{C}(K)$. Then λ is bounded and there is a unique positive regular Borel measure $\mu \in \mathrm{M}(K)$ such that*

$$\lambda(f) = \int_K f \,\mathrm{d}\mu \qquad \textit{for all } f \in \mathrm{C}(K).$$

Our presentation is a modification of the treatments in [Rudin (1987), Chapter 2] and [Lang (1993), Chapter IX].

Let K be a compact topological space, and let $\lambda : \mathrm{C}(K) \to \mathbb{C}$ be a positive linear functional. We shall construct a positive regular Borel measure $\mu \in \mathrm{M}(K)$ with $\lambda = \langle \cdot, \mathrm{d}\mu \rangle$. Such representing measure is necessarily unique: If μ, ν are both regular Borel measures representing λ, then there Baire restrictions must coincide by Corollary D.2, and then they must be equal by Proposition 5.6.

The program for the construction of the measure μ is now as follows. First, we construct an outer measure μ^*. Next, we show that every open set is μ^*-measurable in the sense of Carathéodory. This yields the measure μ on the Borel sets. Then we show that μ is regular and, finally, that λ is given by integration against μ.

For an open set $O \subseteq K$ we define

$$\mu^*(O) := \sup\big\{\lambda(f) : f \in \mathrm{C}(K),\ 0 \leq f \leq 1,\ \mathrm{supp}(f) \subseteq O\big\}.$$

Then $\mu^*(\emptyset) = 0$, μ^* is monotone, and $\mu^*(\emptyset) \leq \mu^*(K) = \lambda(\mathbf{1})$.

Lemma D.12. *The map μ^* is σ-subadditive on open sets.*

Proof. Suppose that $O = \bigcup_{k\in\mathbb{N}} O_k$, and take $0 \le g \in \mathrm{C}(K)$ such that $0 \le g \le \mathbf{1}$ and $\mathrm{supp}(g) \subseteq O$. Then the compact set $\mathrm{supp}(g)$ is covered by the collection of open sets O_k, and hence there is $n \in \mathbb{N}$ such that

$$\mathrm{supp}(g) \subseteq \bigcup_{j=1}^{n} O_j.$$

Take a partition of unity $f_0, \dots, f_n$ on K subordinate to the cover $O_0 := K \setminus \mathrm{supp}(g)$, $O_1, \dots, O_n$, see Theorem D.6. Then each $g_j := f \cdot f_j$ has support within O_j and

$$g = g \cdot \textstyle\sum_{j=0}^{m} f_j = \sum_{j=1}^{m} g_j.$$

It follows that

$$\lambda(g) = \textstyle\sum_{j=1}^{m} \lambda(g_j) \le \sum_{j=1}^{m} \mu^*(O_j) \le \sum_{j=1}^{\infty} \mu^*(O_j).$$

Taking the supremum on the left-hand side yields $\mu^*(O) \le \sum_{j=1}^{\infty} \mu^*(O_j)$ as claimed. □

We now extend the definition of μ^* to all subsets $A \subseteq K$ by

$$\mu^*(A) := \inf\big\{\mu^*(O) : A \subseteq O \subseteq K,\ O \text{ open}\big\}. \tag{D.6}$$

Recall from Appendix B.3 the notion of an outer measure.

Lemma D.13. *The set function $\mu^* : \mathcal{P}(K) \to [0, \lambda(\mathbf{1})]$ is an outer measure.*

Proof. By Lemma D.12, μ^* coincides with the Hahn-extension to $\mathcal{P}(K)$ of its restriction to open sets, see Appendix B.4. Hence it is an outer measure. □

By Carathéodory's Theorem B.3, the outer measure μ^* is a measure on the σ-algebra

$$\mathcal{M}(\mu^*) := \big\{A \subseteq K : \forall\, H \subseteq K :\ \mu^*(H) \ge \mu^*(H \cap A) + \mu^*(H \setminus A)\big\}$$

of "μ^*-measurable" sets. Our next aim is to show that each open set is μ^*-measurable. We need the following auxiliary result.

Lemma D.14. *If $A, B \subseteq K$ such that $\overline{A} \cap \overline{B} = \emptyset$, then $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$.*

Proof. Fix $\varepsilon > 0$ and take an open set O of K such that

$$\mu^*(O) \le \mu^*(A \cup B) + \varepsilon.$$

Then take, by Lemma 4.1, open sets U, V of K such that

$$\overline{A} \subseteq U, \quad \overline{B} \subseteq V, \quad U \cap V = \emptyset.$$

Now take continuous functions $0 \le f, g \le \mathbf{1}$ such that $\operatorname{supp}(f) \subseteq U \cap O$ and $\operatorname{supp}(g) \subseteq V \cap O$. Then $0 \le f+g \le \mathbf{1}$ and

$$\operatorname{supp}(f+g) \subseteq \operatorname{supp}(f) \cup \operatorname{supp}(g) \subseteq (U \cap O) \cup (V \cap O) \subseteq O.$$

Hence $\lambda(f) + \lambda(g) = \lambda(f+g) \le \mu^*(O) \le \mu^*(A \cup B) + \varepsilon$. By varying f and g we conclude that

$$\mu^*(A) + \mu^*(B) \le \mu^*(U \cap O) + \mu^*(V \cap O) \le \mu^*(A \cup B) + \varepsilon,$$

and this finishes that proof. □

The next result says that μ^* is regular on open sets.

Lemma D.15. *One has $\mu^*(O) = \sup\{\mu^*(L) : L \subseteq O,\ L \text{ compact}\}$ for each open set $O \subseteq K$.*

Proof. Fix $t < \mu^*(O)$ and $f \in \mathrm{C}(K)$ such that $0 \le f \le 1$, $L := \operatorname{supp}(f) \subseteq O$ and $t < \lambda(f)$. Then for any open set $U \supseteq L$ we have $\operatorname{supp}(f) \subseteq U$, and hence $\lambda(f) \le \mu^*(U)$. As U was arbitrary, $\lambda(f) \le \mu^*(L)$, and this concludes the proof. □

We have all tools in our hands to take the next step.

Lemma D.16. *Every open set is μ^*-measurable.*

Proof. Let O be an open set, and let $H \subseteq K$ be an arbitrary subset. We need to show that

$$\mu^*(H) \ge \mu^*(H \cap O) + \mu^*(H \setminus O). \tag{D.7}$$

Fix an open set $U \supseteq H$ and a compact set $L \subseteq U \cap O$. Then $H \setminus O \subseteq O^{\mathrm{c}}$, which is disjoint from L. Hence by Lemma D.14 we obtain

$$\mu^*(U) \ge \mu^*(L \cup (H \setminus O)) = \mu^*(L) + \mu^*(H \setminus O).$$

By Lemma D.15 it follows that

$$\mu^*(U) \ge \mu^*(U \cap O) + \mu^*(H \setminus O) \ge \mu^*(H \cap O) + \mu^*(H \setminus O).$$

By taking the infimum with respect to $U \supseteq H$ we arrive at (D.7). □

At this stage we conclude from Carathéodory's Theorem B.3 that the restriction μ of μ^* to the Borel algebra of K is a (finite) measure of total mass $\mu(K) = \lambda(\mathbf{1})$.

Lemma D.17. *The measure μ is regular.*

Proof. Define

$$\mathcal{D} := \big\{B \in \mathrm{Bo}(K) : B \text{ has the regularity properties as in (5.3)}\big\}.$$

By Lemma D.15, Σ contains all open sets. We shall show that $\mathcal{D}$ is a Dynkin system. Since the open sets form a $\cap$-stable generator of the Borel algebra, it follows that $\mathcal{D} = \mathrm{Bo}(K)$.

Clearly $\emptyset, K \in \mathcal{D}$. If $A, B \in \mathcal{D}$ with $A \subseteq B$ and $\varepsilon > 0$ then there are $L \subseteq A \subseteq O$, $L' \subseteq B \subseteq O'$, L, L' compact and O, O' open, such that $\mu(O \setminus L) + \mu(O' \setminus L') < \varepsilon$. Then

$$L' \setminus O \subseteq B \setminus A \subseteq O' \setminus L, \quad L' \setminus O \text{ compact, } O' \setminus L \text{ open.}$$

Moreover, $(O' \setminus L) \setminus (L' \setminus O) \subseteq (O \setminus L) \cup (O' \setminus L')$ and hence

$$\mu((O' \setminus L) \setminus (L' \setminus O)) \leq \mu(O \setminus L) + \mu(O' \setminus L') < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $B \setminus A \in \mathcal{D}$. Finally, suppose that $A_n \in \mathcal{D}$ and $A_n \nearrow A$. Find $L_n \subseteq A_n \subseteq O_n$, L_n compact and O_n open, such that $\mu(O_n \setminus L_n) < \varepsilon$. Without loss of generality we may suppose that $L_n \nearrow$. Let $O := \bigcup_{n \in \mathbb{N}} O_n$, and find N such that $\mu(O \setminus O_N) < \varepsilon$. Then $L_N \subseteq A \subseteq O$ and

$$\mu(O \setminus L_N) \leq \mu(O \setminus O_N) + \mu(O_N \setminus L_N) \leq 2\varepsilon.$$

It follows that $A \in \mathcal{D}$, and that concludes the proof. □

Finally, we conclude the proof of the Riesz–Markov Theorem D.11 by showing that μ induces the functional λ.

Lemma D.18. *We have $\lambda(f) = \int_K f \, \mathrm{d}\mu$ for all $f \in \mathrm{C}(K)$.*

Proof. By considering real and imaginary part separately we may suppose that f is real-valued. Then there are real numbers a, b such that $K = [a \leq f < b]$. Fix $\varepsilon > 0$ and take points

$$a = t_0 < t_1 < \cdots < t_n = b$$

with $t_j - t_{j-1} < \varepsilon$. The sets

$$A_j := \left[t_{j-1} < f \leq t_j\right] \qquad (j = 1, \ldots, n)$$

form a partition of K into disjoint Baire sets, and the step function

$$g := \sum_{j=1}^{n} t_j \mathbf{1}_{A_j} \quad \text{satisfies} \quad f \leq g \leq f + \varepsilon \mathbf{1}.$$

For each j we choose $s_j < t_j$. Then the open sets $\left[s_{j-1} < f < t_j\right]$, $j = 1, \ldots, n$, cover K. Let $(f_j)_{j=1}^n$ be a partition of unity on K subordinate to this cover. Then $f f_j \leq t_j f_j$ for each j, hence

$$\lambda(f) = \sum_{j=1}^{n} \lambda(f f_j) \leq \sum_{j=1}^{n} t_j \lambda(f_j) \leq \sum_{j=1}^{n} t_j \mu\left[s_{j-1} < f < t_j\right].$$

(The last inequality here comes from the definition (D.6) of μ.) Now we let $s_{j-1} \nearrow t_{j-1}$ for each $j = 1, \ldots, n$ and obtain

$$\lambda(f) \le \sum_{j=1}^{n} t_j \mu\left[t_{j-1} \le f < t_j\right] = \int_K g \,\mathrm{d}\mu \le \int_K f \,\mathrm{d}\mu + \varepsilon\mu(K).$$

Since $\varepsilon > 0$ was arbitrary, this implies that $\lambda(f) \le \int_K f \,\mathrm{d}\mu$. Finally, we replace f by $-f$ and arrive at $\lambda(f) = \int_K f \,\mathrm{d}\mu$. □

With the Riesz theorem we have established an isometric isomorphism between the dual space $\mathrm{C}(K)'$ and the space $\mathrm{M}(K)$ of complex Baire (regular Borel) measures on K, allowing us to *identify* functionals with measures. This may lead to the problem that the symbol $|\mu|$ now has two meanings, depending on whether we interpret μ as a functional (Section D.2 or as a measure (Appendix B.9.) The next result shows that this ambiguity is only virtual. More precisely, it says that the isometric isomorphism

$$\mathrm{d} : \mathrm{M}(K) \to \mathrm{C}(K)'$$

from the Riesz' representation theorem is a *lattice isomorphism*.

Lemma D.19. *For $\mu \in \mathrm{M}(K)$ one has $|\mathrm{d}\mu| = \mathrm{d}\,|\mu|$.*

Proof. For $|g| \le f \in \mathrm{C}(K)$ we have

$$|(\mathrm{d}\mu)(g)| = \left|\int_K g \,\mathrm{d}\mu\right| \le \int_K |g| \,\mathrm{d}|\mu| \le (\mathrm{d}\,|\mu|)(f).$$

Hence, $|\mathrm{d}\mu| \le \mathrm{d}\,|\mu|$. For the converse, let ν be the unique positive regular Borel measure on K such that $\mathrm{d}\nu = |\mathrm{d}\mu|$. Then $|\int_K f \,\mathrm{d}\mu| \le \int_K |f| \,\mathrm{d}\nu$ for all $f \in \mathrm{C}(K)$. By virtue of Theorem D.1 we obtain that

$$\left|\int_K f \,\mathrm{d}\mu\right| \le \int_K |f| \,\mathrm{d}\nu$$

holds even for all $f \in \mathrm{BM}(K)$. By the definition of the measure $|\mu|$ (Appendix B.9) it follows that $|\mu| \le \nu$, and hence $\mathrm{d}\,|\mu| \le \mathrm{d}\nu = |\mathrm{d}\mu|$. □

Appendix E
Theorems of Eberlein, Grothendieck, and Ellis

Let K be a compact space. We denote by $\mathrm{C}_p(K)$ the space $\mathrm{C}(K)$ endowed with the **pointwise topology**, i.e., as a topological subspace of

$$X := \mathbb{C}^K = \prod_{x\in K}\mathbb{C}$$

with the product topology. If $M \subseteq \mathrm{C}_p(K)$ then its closure in X is denoted by $\overline{M}^p$, so that $\overline{M}^p \cap \mathrm{C}(K)$ is its closure in $\mathrm{C}_p(K)$. We shall also speak of p-open/closed/dense/... subsets of $\mathrm{C}(K)$ to refer to these notions in the pointwise topology.

We begin with a central auxiliary result.

Lemma E.1. *Let $M \subseteq \mathrm{C}(K)$, and let $(x_m)_{m\in\mathbb{N}} \subseteq K$ and $x \in K$ such that*

$$f(x_m) \to f(x) \tag{E.1}$$

for every $f \in M$. Then (E.1) *holds for every $f \in \overline{M}^p \cap \mathrm{C}(K)$.*

Proof. To begin with, let $A := \bigcap_{j\in\mathbb{N}} \overline{\{x_m : m \geq j\}}$ be the set of accumulation points of the sequence $(x_n)_n$. By assumption, for each $g \in M$ we have

$$g(A) \subseteq \bigcap\nolimits_{j\in\mathbb{N}} \overline{\{g(x_m) : m \geq j\}} = \{g(x)\},$$

i.e., g is constant with value $g(x)$ on A. Hence, by the definition of the pointwise topology, it follows that $f(A) \subseteq \{f(x)\}$ for every $f \in \overline{M}^p$.

On the other hand, if $f \in \mathrm{C}(K)$ and $s \in \mathbb{C}$ is an accumulation point of $(f(x_n))_n$, then, by compactness of K and the continuity of f, we have $s \in f(A)$. Combining this with the previous observation, we see that for $f \in \overline{M}^p \cap \mathrm{C}(K)$ the point $f(x)$ is the *unique* accumulation point of the sequence $(f(x_n))_n$, and hence its limit. □

E.1 Separability and Metrisability

Every compact metrisable space is separable, but the converse is not true. The next result shows that compact subsets of $\mathrm{C}_p(K)$, however, are special in this respect.

Theorem E.2. *Let $A \subseteq \mathrm{C}_p(K)$ be compact. Then A is separable if and only if A is metrisable.*

Proof. For the nontrivial implication suppose that A is separable, and let $M = \{f_n : n \in \mathbb{N}\} \subseteq A$ be p-dense in A. The function

$$d : K \times K \to \mathbb{R}_+, \qquad d(x,y) := \sum_{n=1}^{\infty} \frac{|f_n(x) - f_n(y)|}{\|f_n\|_\infty 2^n}.$$

is a continuous pseudo-metric on K. By looking at the d-balls $U_x := [d(x,\cdot) < 1/n]$ and using the compactness of K one finds a countable set $\{x_m : m \in \mathbb{N}\} \subseteq K$ that is d-dense in K.

Now define $c_m := 1 + \sup_{f \in A} |f(x_m)|$, which is finite by the compactness of A, and

$$e : A \times A \to \mathbb{R}_+, \qquad e(f,g) := \sum_{m=1}^{\infty} \frac{|f(x_m) - g(x_m)|}{2^m c_m}$$

for $f, g \in A$. We claim that e is a metric inducing the topology of A. Clearly, e is continuous, hence by compactness it suffices to show that e is indeed a metric. The triangle inequality is trivial. Suppose that $e(f,g) = 0$. Then $f(x_m) = g(x_m)$ for all $m \in \mathbb{N}$. Let $x \in K$ be arbitrary. By passing to a subsequence we my suppose that $d(x_m, x) \to 0$, i.e., $f_n(x_m) \to f_n(x)$ for all $n \in \mathbb{N}$. By Lemma E.1, it follows that $f(x_m) \to f(x)$ and $g(x_m) \to g(x)$. Hence $f(x) = g(x)$. □

Recall that a subset A of a topological space X is called **relatively compact** in X if its closure $\overline{A}$ is compact. Note that $A \subseteq \mathrm{C}(K)$ is relatively compact in $\mathrm{C}_p(K)$ if and only if the pointwise closure $\overline{A}^p$ is compact and contained in $\mathrm{C}(K)$.

Theorem E.3 (Eberlein). *Let $M \subseteq \mathrm{C}(K)$ and $f \in \overline{M}^p \cap \mathrm{C}(K)$. Then there is a countable $M_0 \subseteq M$ such that $f \in \overline{M_0}^p$. If in addition M is relatively compact in $\mathrm{C}_p(K)$, then f is the pointwise limit of a sequence in M.*

Proof. For $n, m \in \mathbb{N}$ and $x := (x_1, \ldots, x_n) \in K^n$ take $g_x \in M$ such that $\big|g_x(x_j) - f(x_j)\big| < 1/m$ for $j = 1, \ldots n$. Then

$$x \in U_x := \prod_{j=1}^{n} [\,|g_x - f| < 1/m\,],$$

which is an open subset of K^n. By compactness, there is a finite set $F_{n,m} \subseteq K^n$ such that $\{U_x : x \in F_{n,m}\}$ is a cover of K^n. Then $M_0 := \bigcup_{n,m} \{g_x : x \in F_{n,m}\}$ is countable and its pointwise closure contains f.

To prove the second statement, suppose that M is relatively compact in $\mathrm{C}_p(K)$. Then $\overline{M_0}^p$ is a separable and compact subset of $\mathrm{C}_p(K)$, whence by Theorem E.2 it is metrisable. Consequently, there is a sequence in M_0 converging to f. □

A subset A of a topological space X is called **sequentially closed** in X if it contains the limit of each sequence $(f_n)_n$ that converges in X.

Corollary E.4. *If $A \subseteq \mathrm{C}(K)$ is relatively compact and sequentially closed in $\mathrm{C}_p(K)$, then it is compact.*

E.2 Grothendieck's Theorem

The next result connects the pointwise topology on $\mathrm{C}(K)$ with its weak topology $\mathrm{C}(K)$ as a Banach space. To facilitate argumentation, we shall write (A,p) and (A,w) to denote the pointwise and weak topologies on $A \subseteq \mathrm{C}(K)$, respectively.

Theorem E.5 (Grothendieck). *Let $M \subseteq \mathrm{C}(K)$ be norm-bounded. Then the following statements are equivalent.*

(i) *M is weakly compact.*

(ii) *M is compact in $\mathrm{C}_p(K)$.*

If (i) *and* (ii) *hold, then the two mentioned topologies coincide on M.*

Proof. Note first that the identity mapping

$$\mathrm{id} : (\mathrm{C}(K), w) \to (\mathrm{C}(K), p)$$

is continuous, and both topologies are Hausdorff. Hence (i) implies (ii) and both topologies coincide on M.

Conversely, suppose that M is p-compact in $\mathrm{C}(K)$. It suffices to show that $\mathrm{id} : (M,p) \to (M,w)$ is continuous. To this end, let $A \subseteq M$ be weakly closed in M and $f \in \overline{A}^p$. Then $f \in M \subseteq \mathrm{C}(K)$ by hypothesis. By Eberlein's Theorem E.3 there is a sequence $(f_n)_n \subseteq A$ such that $f_n \to f$ pointwise. By the Riesz Representation Theorem 5.7 and dominated convergence, we obtain $f_n \to f$ weakly. Since A is weakly closed in M, we obtain that $f \in A$. Consequently, A is p-closed. □

Corollary E.6. *For a norm-bounded set $M \subseteq \mathrm{C}(K)$ the following statements are equivalent.*

(i) *M is relatively weakly compact.*

(ii) *M is relatively compact in $\mathrm{C}_p(K)$.*

If (i) *and* (ii) *hold, then $\overline{M}^p = \overline{M}^w$ and that the two mentioned topologies coincide on M.*

Proof. Let $A := \overline{M}^w$ and $B = \overline{M}^p$. Then $A \subseteq B$. If (i) holds, then A is weakly compact, hence by Theorem E.5 it is p-compact. This implies that $A = B$ and that the two topologies coincide on A, hence also on M.

Conversely, if (ii) holds then by Theorem E.5 B is weakly compact. Hence $A \subseteq B$ is also weakly compact, i.e., (i) holds. □

E.3 The Theorem of Kreĭn–Šmulian

Grothendieck's theorem has an enormous range of applications. As an example, we give a quite simple proof of the following result, see Theorem C.11.

Theorem E.7 (Kreĭn–Šmulian). *Let $K \subseteq X$ be a weakly compact subset of a Banach space X. Then $\overline{\operatorname{conv}}(K)$ is weakly compact.*

Proof. Let $B := \{x' \in X' : \|x'\| \leq 1\}$ be the dual unit ball. Then B is weakly* compact, by the Banach–Alaoglu theorem. The map

$$\varphi : (B, w^*) \to \mathrm{C}_p(K) \qquad x' \mapsto x'\big|_K$$

is continuous and its image $C := \varphi(B)$ is norm-bounded. By Grothendieck's theorem, C is weakly compact and the pointwise topology coincides with the weak topology on C. It follows that $\varphi : (B, w^*) \to (\mathrm{C}(K), w)$ is continuous. For $\mu \in \mathrm{M}(K)$ we define $f = \Phi(\mu)$ by

$$f(x') := \int_K \varphi(x')\,\mathrm{d}\mu \qquad (x' \in X').$$

Then $f \in X''$ and $f\big|_B$ is weakly* continuous. By Proposition E.8 below, $f \in X$. Hence $\Phi : \mathrm{M}(K) \to X$ is a well-defined bounded linear mapping, weakly*-to-weak continuous. Since the set $\mathrm{M}^1(K)$ of probability measures is a weakly* compact and convex set, its image $\Phi(\mathrm{M}^1(K))$ is weakly compact and convex. But $\Phi(\delta_x) = x$ for every $x \in K$, whence $K \subseteq \Phi(\mathrm{M}^1(K))$. Consequently

$$\overline{\operatorname{conv}}(K) \subseteq \Phi(\mathrm{M}^1(K))$$

and since the latter is weakly compact, the proof is complete. □

Our proof of the Kreĭn–Šmulian theorem still rests on an auxiliary result from the theory of locally convex vector spaces. For $A \subseteq X$, where X is a locally convex space, the **polar** is defined as

$$A^\circ := \big\{x' \in X' : \big|\langle x', x\rangle\big| \leq 1 \text{ for all } x \in A\big\}.$$

The **bipolar theorem** [Conway (1990), Theorem V.1.8]—a simple consequence of the Hahn–Banach separation theorems [Conway (1990), Cor. VI.3.10]—states that

$$A^{\circ\circ} = \overline{\text{absconv}}(A)$$

where on the X' one takes the weak*, i.e., $\sigma(X,X')$-topology. (The set $\overline{\text{absconv}}(A)$ is the **closed absolutely convex hull** of A, i.e., the the intersection of all closed absolutely convex set containing A.) We can now state the result, interesting in its own right.

Theorem E.8. *Let X be a Banach space and $f \in X''$ be such that its restriction to $B := \{x' \in X' : \|x'\| \le 1\}$ is weakly* continuous. Then $f \in X$.*

Proof. For the proof we regard X as a subset of X'' and consider the dual pair (X'',X') of locally convex spaces, each with the topology induced by the other. Let $\varepsilon > 0$. Then the assumption implies that there are vectors $x_1,\dots,x_n \in X$ and $\delta > 0$ such that

$$\|x'\| \le 1, \quad |\langle x', x_j\rangle| \le \delta \quad (j = 1,\dots,n) \quad \Rightarrow \quad |f(x')| \le \varepsilon. \tag{E.2}$$

By scaling, we may suppose that $\delta = 1$. Let

$$U := \{x' \in X' : |\langle x', x_j\rangle| \le 1 \text{ for } j = 1,\dots,n\} = \{x_1,\dots,x_n\}^\circ = K^\circ,$$

where $K = \overline{\text{absconv}}\{x_1,\dots,x_n\}$. Then (E.2) simply says that $(1/\varepsilon)f \in (U \cap B)^\circ$. Now,

$$(U \cap B)^\circ = (K^\circ \cap B_{X''}^\circ)^\circ = (K \cup B_{X''})^{\circ\circ} = \overline{\text{absconv}}(K \cup B_{X''}) \subseteq \overline{K + B_{X''}}$$

where the closure is in the weak* topology on X''. But K is compact, and hence $K + B_{X''}$ is already weakly* closed. This yields $(1/\varepsilon)f \in K + B_{X''}$, and hence there is $x \in K \subseteq X$ such that $\|f - \varepsilon x\| \le \varepsilon$. As $\varepsilon > 0$ was arbitrary and X is complete, it follows that $f \in X$. □

E.4 Ellis' Theorem and the Haar Measure on a Compact Group

A semigroup S endowed with a topology is called a **semitopological semigroup** if the multiplication mapping

$$S \times S \to S \qquad (a,b) \mapsto ab$$

is *separately continuous*. A group G endowed with a topology is a **topological group** if multiplication is jointly continuous and also inversion is continuous. Equivalently, the mapping

$$G \times G \to G, \quad (a,b) \mapsto ab^{-1}$$

is continuous.

An important theorem of R. Ellis states that a compact semitopological semigroup that is algebraically a group must be a topological group. Another important

theorem, due to A. Haar, states that on a compact topological group there exists a unique invariant probability measure, called the *Haar measure*. In this section we shall prove both results, however, in reverse order. We shall first construct the Haar measure (on a compact semitopological semigroup which is algebraically a group) and afterwards prove Ellis' theorem. The cornerstones for the proofs are Grothendieck's theorem and the Kreĭn–Šmulian theorem.

Construction of the Haar Measure

Let G be a compact semitopological semigroup, which is algebraically a group. We denote by

$$L_a, R_a : \mathrm{C}(G) \to \mathrm{C}(G), \qquad (L_a f)(x) := f(ax), \quad (R_a f)(x) := f(xa)$$

the left and the right rotation by $a \in G$. The operators L_a and R_a are isometric isomorphisms on $\mathrm{C}(G)$. Since the multiplication is separately continuous, for $f \in \mathrm{C}(G)$ the mapping

$$G \to \mathrm{C}_p(G) \qquad a \mapsto L_a f$$

is continuous. Hence, by Grothendieck's Theorem E.5, the orbit

$$\big\{L_a f \,:\, a \in G\big\}$$

is weakly compact. By the Kreĭn–Šmulian Theorem E.7, its closed convex hull

$$K_f := \overline{\mathrm{conv}}\big\{L_a f \,:\, a \in G\big\}$$

is weakly compact, too. (The closure here is the same in the weak and in norm topology, by Mazur's Theorem C.7.) Grothendieck's theorem, again, implies that K_f is p-compact.

The first step now consists in finding a constant function $c_f \in K_f$. To achieve this, let us define the **oscillation**

$$\mathrm{osc}(g) := \sup_{x,y \in G} |g(x) - g(y)|$$

of $g \in \mathrm{C}(G)$. Then g is constant if and only if $\mathrm{osc}(g) = 0$. Note that

$$g \in K_f \quad \text{implies} \quad K_g \subseteq K_f \quad \text{and} \quad \mathrm{osc}(g) \le \mathrm{osc}(f).$$

Now we let $s := \inf_{g \in K_f} \mathrm{osc}(g)$. For every $n \in \mathbb{N}$, the set

$$\big\{g \in K_f \,:\, \mathrm{osc}(g) \le s + \tfrac{1}{n}\big\}$$

is not empty and p-closed, whence also weakly closed, by Grothendieck's theorem. By compactness, the intersection of these sets is not empty, hence there is $g \in K_f$

such that $\operatorname{osc}(g) = s$. The following lemma now shows that in case f is real-valued, $s = 0$, i.e., g is constant.

Lemma E.9. *Let $g \in \mathrm{C}(K;\mathbb{R})$ and $\operatorname{osc}(g) > 0$. Then there exists $h \in K_g$ with $\operatorname{osc}(h) < \operatorname{osc}(g)$.*

Proof. After scaling and shifting we may suppose that $-1 \le g \le 1$ and $\operatorname{osc}(g) = 2$. By compactness, there are $a_1, \dots, a_n \in G$ such that

$$[g \ge 1/2] \subseteq \bigcup\nolimits_{j=1}^{n} a_j^{-1}[g < -1/2].$$

Define $h := \frac{1}{n+1}(g + \sum_{j=1}^{n} L_{a_j} g) \in K_g$. Then clearly $-1 \le h$. If $g(x) \ge 1/2$ then there exists at least on a_j such that $g(a_j x) < -1/2$. Hence

$$h(x) \le \frac{1}{n+1}\Big(1 - \frac{1}{2} + (n-1)\Big) = \frac{n - 1/2}{n+1}.$$

If $g(x) < 1/2$, then

$$h(x) \le \frac{1}{n+1}\Big(\frac{1}{2} + n\Big) = \frac{n + 1/2}{n+1}.$$

It follows that $h \le (n + \frac{1}{2})/(n+1)$ and hence $\operatorname{osc}(h) < 2 = \operatorname{osc}(g)$. □

By decomposing a function f into real and imaginary parts, we conclude that each set K_f contains a constant function c_f. Of course, the same argument is valid for right rotations, and hence for $f \in \mathrm{C}(K)$ there is a constant function

$$c'_f \in K'_f := \overline{\operatorname{conv}}\big\{R_a f : a \in G\big\}.$$

Claim: *We have $c_f = c'_f$.*

Proof. By construction there are sequences $S_n \in \operatorname{conv}\{L_a : a \in G\}$ and $T_n \in \operatorname{conv}\{R_a : a \in G\}$ such that $S_n f \to c_f$ and $T_n f \to c'_f$. Hence $T_m S_n f \to c_f$ as $n \to \infty$ and

$$T_m S_n = S_n T_m f \to c'_f \quad \text{as } m \to \infty.$$

But since these convergences are *in norm* and all the operators are contractions, the claim follows. □

It follows that c_f is the *unique* constant function in K_f as well as the unique constant function in K'_f. We shall write

$$c_f = Pf.$$

It is clear that $P\mathbf{1} = \mathbf{1}$ and $Pf \ge 0$ whenever $f \ge 0$. To show that $PR_a = P$ for $a \in G$, note that since left and right rotations commute, we have

$$Pf = R_a Pf \in R_a(K_f) = K_{R_a f},$$

whence $Pf = PR_a f$, by uniqueness. Analogously, $Pf = PL_a$ for every $a \in G$.

It remains to show that P is linear. Clearly P is $\mathbb{R}$-homogeneous, and $Pf = P(\operatorname{Re} f) + iP(\operatorname{Im} f)$ by uniqueness. It follows that P is $\mathbb{C}$-homogeneous.

Claim: $P(f+g) = Pf + Pg$.

Proof. Let $\varepsilon > 0$. Then there are $S,T \in \operatorname{conv}\{L_a : a \in G\}$ such that $\|Sf - Pf\|, \|TSg - PSg\| \leq \varepsilon$. Note that $TPf = Pf$ and $PSg = Pg$. Hence

$$\begin{aligned}\|TS(f+g) - (Pf+Pg)\| &\leq \|TSf - Pf\| + \|TSg - Pg\| \\ &= \|T(Sf-Pf)\| + \|TSg - PSg\| \leq 2\varepsilon,\end{aligned}$$

since T is a contraction. It follows that $Pf + Pg \in K_{f+g}$, and hence $P(f+g) = Pf + Pg$ by uniqueness. □

Since Pf is a constant function for each $f \in \mathrm{C}(K)$ we may write

$$Pf = \mathrm{m}(f)\mathbf{1} \qquad (f \in \mathrm{C}(K)).$$

Then m is a positive, linear functional, invariant under left and right rotations. We have proved the major part of the following theorem.

Theorem E.10 (Haar Measure). *Let G be a compact semitopological semigroup which is algebraically a group. Then there is a unique probability measure* m *on G that is invariant under all left rotations. This measure has the following additional properties:*

a) m *is strictly positive, i.e.,* $\operatorname{supp}(m) = G$.

b) m *is invariant under right rotations.*

c) m *is invariant under inversion.*

Proof. Existence was proved above, as well as the invariance under right rotations. For uniqueness, suppose that μ is a probability measure on G that is invariant under all left rotations. Given $f \in \mathrm{C}(G)$, it follows that $\langle f, \mu\rangle = \langle g, \mu\rangle$ for all $g \in K_f$, and hence

$$\langle f, \mu\rangle = \langle Pf, \mu\rangle = \mathrm{m}(f)\,\langle \mathbf{1}, \mu\rangle = \mathrm{m}(f).$$

In order to see that m is strictly positive, let $0 \leq f \in \mathrm{C}(G)$ such that $f \neq 0$. Define $U_x := [L_x f > 0]$ for $x \in K$ and note that the sets U_x cover K. By compactness, there are $x_1, \ldots, x_n \in K$ such that $K = \bigcup_{j=1}^n U_{x_j}$. Hence the function $g := \sum_{j=1}^n L_{x_j} f$ is strictly positive, so that there exists $c > 0$ such that $c\mathbf{1} \leq g$. It follows that

$$c = \mathrm{m}(c\mathbf{1}) \leq \mathrm{m}(g) = \textstyle\sum_{j=1}^n \mathrm{m}(L_{x_j} f) = n\,\mathrm{m}(f),$$

which implies that $\mathrm{m}(f) > 0$.

Finally, let $S : \mathrm{C}(G) \to \mathrm{C}(G)$ be the reflection mapping defined by $(Sf)(x) := f(x^{-1})$. Since m is right-invariant, the probability measure $\mu := S'\mathrm{m} \in \mathrm{M}(G)$ is left invariant, and hence by uniqueness $\mu = \mathrm{m}$. This is assertion c). □

The unique probability measure m from Theorem E.10 is called the **Haar measure** on G.

Ellis' Theorem

For a Hilbert space H let

$$\mathrm{Iso}(H) := \{T \in \mathscr{L}(H) : T^*T = \mathrm{I}\} \quad \text{and} \quad \mathrm{U}(H) := \{U \in \mathscr{L}(H) : U^* = U^{-1}\}$$

be the semigroup of isometries and the **unitary group**, respectively.

Lemma E.11. *For a Hilbert space H, the strong and the weak operator topologies coincide on* $\mathrm{Iso}(H)$. *The unitary group* $\mathrm{U}(H)$ *is a topological group with respect to this topology.*

Proof. By an elementary computation one shows that

$$\|Sf - Tf\|^2 = 2\,\mathrm{Re}\,(Tf - Sf \,|\, Tf) \qquad (f \in H)$$

whenever $S, T \in \mathrm{Iso}(H)$. The first claim follows directly from this. By Proposition C.18 the second claim is true. □

We are now in a position to prove Ellis' theorem.

Theorem E.12 (Ellis). *Let G be a compact semitopological semigroup. If G is algebraically a group, it is a compact group.*

Proof. By Theorem E.10 we can employ the Haar measure m on G. Let $H := \mathrm{L}^2(G)$ and let

$$R : G \to \mathscr{L}(H), \quad a \mapsto R_a$$

be the *right regular representation* of G on H. Then R is a homomorphism of G into the unitary group $\mathrm{U}(H)$ on H. Also, R is injective, by Urysohn's lemma and since m is strictly positive.

We claim that R is continuous with respect to the weak operator topology on H. To prove this claim, we have to show that for $f, g \in H$ the map

$$\psi = \psi_{f,g} : G \to \mathbb{C}, \qquad \psi(a) := (R_a f \,|\, g) = \int_G (R_a f) \cdot \overline{g}\,\mathrm{dm}$$

is continuous. By denseness of $\mathrm{C}(K)$ in H we may suppose that $f \in \mathrm{C}(K)$. Then by Grothendieck's Theorem E.5 the mapping $a \mapsto L_a f$ is continuous for the weak topology on $\mathrm{C}(K)$, so the claim is established.

By Lemma E.11, the weak and the strong operator topologies coincide, and $\mathrm{U}(H)$ is a topological group for this latter topology. Since the right regular representation is a homeomorphism onto its image, the theorem is proved. □

E.5 Sequential Compactness and the Eberlein–Šmulian Theorem

A Hausdorff topological space X is called **countably compact** if every sequence $(f_n)_n \subseteq X$ has a cluster point, i.e.,

$$\bigcap\nolimits_n \overline{\{f_k : k \geq n\}} \neq \emptyset.$$

And it is called **sequentially compact** if every sequence has a convergent subsequence. Furthermore, a subset A of a topological space X is called **relatively countably compact** if every sequence in A has a cluster point in X and **relatively sequentially compact** in X if every sequence in A has a subsequence that converges in X.

It is clear that a (relatively) compact or sequentially compact subset is also (relatively) countably compact. However, the converses are false in general. Also, neither notion—(relative) compactness and (relative) sequential compactness—implies the other. Eberlein's Theorem E.2 is the key to the important fact, that for subsets of $\mathrm{C}_p(K)$ these notions all coincide.

Lemma E.13. *Let $A \subseteq \mathrm{C}_p(K)$ be (relatively) compact. Then A is (relatively) sequentially compact.*

Proof. By hypothesis, $B := \overline{A}^p = \overline{A}^p \cap \mathrm{C}(K)$ is compact. Let $(f_n)_n \subseteq A$ be any sequence. Then $Y := \overline{\{f_n : n \in \mathbb{N}\}}^p$ is a separable compact set in $\mathrm{C}_p(K)$. By Theorem E.2, Y is metrisable. Consequently, there is $f \in Y$ and a subsequence $(f_{k_n})_n$ such that $f_{k_n} \to f$ pointwise. □

The next theorem is again due to Grothendieck.

Theorem E.14 (Grothendieck). *A subset $A \subseteq \mathrm{C}_p(K)$ is relatively compact if and only if it is relatively countably compact.*

Proof. If A is relatively compact, it is relatively countably compact. For the converse, suppose that A is relatively countably compact and let $B := \overline{A}^p \subseteq \mathbb{C}^K$. For $x \in K$ the set $\{f(x) : f \in A\}$ is relatively countably compact in $\mathbb{C}$, hence bounded. By Tychonoff's theorem, it follows that B is compact, and it remains to show that $B \subseteq \mathrm{C}(K)$.

Suppose towards a contradiction that there exists $g \in B \setminus \mathrm{C}(K)$. Then there is $y \in K$ at which g fails to be continuous, so there ist $\varepsilon > 0$ such that y is an accumulation point of $Z := [\,|g(y) - g| \geq \varepsilon\,]$. We shall recursively construct points $x_n \in K$, functions f_n and open neighbourhoods U_n of y such that

1) $f_n \in A$;
2) $|f_n(x_m) - g(x_m)| < \varepsilon/2^n$ for $m < n$;
3) $|f_n(y) - g(y)| < \varepsilon/2^n$;
4) $y \in U_n \subseteq \overline{U_n} \subseteq U_{n-1} \cap [\,|f_n - f_n(y)| < \varepsilon/2^n\,]$;
5) $x_n \in U_n \cap Z$.

Suppose that x_j, f_j, and U_j are constructed for $1 \leq j < n$. (This is a vacuous condition for $n = 1$.) Since g is in the p-closure of A, one can find f_n satisfying 1)–3). Since f_n is continuous and by the induction hypothesis, $U_{n-1} \cap [\,|f_n - f_n(y)| < \varepsilon/2^n\,]$ is an open neighbourhood of y, hence one can find another open neighbourhood

U_n of y satisfying 4). Finally, since y is an accumulation point of Z, one can pick $x_n \in U_n \cap Z$, i.e., we have 5).

Now, by relative countable compactness of A in $\mathrm{C}_p(K)$, there is an accumulation point $f \in \mathrm{C}_p(K)$ of $\{f_n : n \geq 1\}$. Since K is compact, there is also an accumulation point x of $\{x_m : m \geq 1\}$.

By 4) and 5), $\{x_m : m \geq n\} \subseteq \overline{U_n}$ and hence $x \in \overline{U_n}$ for each n. It follows that

$$|f_n(x) - f_n(y)| < \varepsilon/2^n \qquad (n \geq 1).$$

Since by 3) also $|f_n(y) - g(y)| < \varepsilon/2^n$ for all $n \in \mathbb{N}$, and by choice of f we obtain $f(x) = g(y)$.

On the other hand, from 2) we obtain $f(x_m) = g(x_m)$ for each $m \in \mathbb{N}$. Since $x_m \in Z$ we also have $|g(x_m) - g(y)| \geq \varepsilon$, and this leads to $|f(x_m) - g(y)| \geq \varepsilon$ for all $m \in \mathbb{N}$. By choice of x we arrive at $|f(x) - g(y)| \geq \varepsilon$, a contradiction. □

Corollary E.15. *For a subset $A \subseteq \mathrm{C}_p(K)$ the following assertions are equivalent.*

(i) *A is relatively compact in $\mathrm{C}_p(K)$.*

(ii) *A is relatively countably compact in $\mathrm{C}_p(K)$.*

(iii) *A is relatively sequentially compact in $\mathrm{C}_p(K)$.*

This is just a combination of Lemma E.13 and Theorem E.14.

Corollary E.16. *For a subset $A \subseteq \mathrm{C}_p(K)$ the following assertions are equivalent.*

(i) *A is compact in $\mathrm{C}_p(K)$.*

(ii) *A is countably compact in $\mathrm{C}_p(K)$.*

(iii) *A is sequentially compact in $\mathrm{C}_p(K)$.*

Proof. By Lemma E.13 (i) implies (iii), and (iii) trivially implies (ii). Suppose that (ii) holds. Then A is relatively compact by Grothendieck's Theorem E.14, and sequentially closed. Hence A is compact by Corollary E.4. □

The Theorem of Eberlein–Šmulian

For (our) convenience, we state and prove only the restricted version of the Eberlein–Šmulian theorem as stated already in Appendix C.6.

Theorem E.17 (Eberlein–Šmulian). *Let X be a Banach space. Then $A \subseteq X$ is (relatively) weakly compact if and only if A is (relatively) weakly sequentially compact.*

In this case, every $f \in \overline{A}^w$ is the weak limit of a sequence in A.

Proof. Let $K := \{x' \in X' : \|x'\| \leq 1\}$ with the weak* topology inherited from X'. Then K is compact by the Banach–Alaoglu Theorem. Let $\Phi : X \to \mathrm{C}(K)$ be the canonical map which maps an element $x \in X$ to $\langle \cdot, x \rangle \big|_K$. Then

$$\Phi : (X, w) \to \mathrm{C}_p(K)$$

is a homeomorphism onto its image. If A is relatively weakly compact, then $\overline{A}$ is weakly compact, and hence $\Phi(\overline{A})$ is compact in $\mathrm{C}_p(K)$. By Corollary E.16, $\Phi(\overline{A})$ is sequentially compact in $\mathrm{C}_p(K)$. Since Φ is a homeomorphism, $\overline{A}$ is seqentially compact. In particular, A is relatively sequentially compact.

Conversely, suppose that A is relatively weakly sequentially compact. Then $\Phi(A)$ is relatively sequentially compact in $\Phi(X)$, and a fortiori in $\mathrm{C}_p(K)$. By Corollary E.15 it follows that $\Phi(A)$ is relatively compact in $\mathrm{C}_p(K)$. By Eberlein's Theorem E.3, every $f \in \overline{\Phi(A)}^p$ is the p-limit of a sequence $(\Phi(a_n))_{n\in\mathbb{N}}$ for some sequence $(a_n)_n \subseteq A$. By hypothesis, there is $a \in X$ such that $a_n \to a$ weakly. Hence $f = \Phi(a) \in \Phi(X)$. This shows that $\overline{\Phi(A)}^p \subseteq X$, hence $\overline{\Phi(A)}^p = \Phi(\overline{A})$. Since Φ is a homeomorphism onto its image, A is relatively compact.

Finally, suppose that A is sequentially weakly compact. Then $\Phi(A)$ is sequentially compact in $\mathrm{C}_p(K)$, whence by Corollary E.16, it is compact. Since Φ is a homeomorphism onto its image, A is weakly compact as well. □

Bibliography

M. Akcoglu and L. Sucheston

[1972] *On operator convergence in Hilbert space and in Lebesgue space*, Period. Math. Hungar. **2** (1972), 235–244. Collection of articles dedicated to the memory of Alfréd Rényi, I.

M. A. Akcoglu

[1975] *A pointwise ergodic theorem in L_p-spaces*, Canad. J. Math. **27** (1975), no. 5, 1075–1082.

A. Baker

[1984] *A Concise Introduction to the Theory of Numbers*, Cambridge University Press, Cambridge, 1984.

H. Bauer

[1981] *Probability Theory and Elements of Measure Theory*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1981. 2nd edition of the translation by R. B. Burckel from the 3rd German edition, Probability and Mathematical Statistics.

[1990] *Maß- und Integrationstheorie*, de Gruyter Lehrbuch, Walter de Gruyter & Co., Berlin, 1990.

V. Bergelson

[2004] *Some historical comments and modern questions around the Ergodic Theorem*, Dynamics of Complex Systems, Kyoto, 2004, pp. 1–11.

P. Billingsley

[1979] *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York-Chichester-Brisbane, 1979.

G. D. Birkhoff

[1912] *Quelques théorèmes sur le mouvement des systèmes dynamiques*, S. M. F. Bull. **40** (1912), 305–323 (French).

[1931] *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. U. S. A. **17** (1931), 656–660 (English).

G. Birkhoff

[1948] *Lattice Theory*, revised ed., American Mathematical Society Colloquium Publications, vol. 25, American Mathematical Society, New York, N. Y., 1948.

J. R. Blum and D. L. Hanson

[1960] *On the mean ergodic theorem for subsequences*, Bull. Amer. Math. Soc. **66** (1960), 308–311.

V. I. Bogachev

[2007] *Measure Theory. Vol. I, II*, Springer-Verlag, Berlin, 2007.

L. Boltzmann

[1885] *Ueber die Eigenschaften monocyklischer und anderer damit verwandter Systeme*, J. Reine Angew. Math. **98** (1885), 68–94 (German).

E. Borel

[1909] *Les probabilités denombrables et leurs applications arithmétiques*, Palermo Rend. **27** (1909), 247–271 (French).

S. G. Brush

[1971] *Milestones in mathematical physics. Proof of the impossibility of ergodic systems: the 1913 papers of Rosenthal and Plancherel*, Transport Theory Statist. Phys. **1** (1971), no. 4, 287–298.

D. L. Burkholder

[1962] *Semi-Gaussian subspaces*, Trans. Amer. Math. Soc. **104** (1962), 123–131.

R. V. Chacon

[1964] *A class of linear transformations*, Proc. Amer. Math. Soc. **15** (1964), 560–564.

[1967] *A geometric construction of measure preserving transformations*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 2, Univ. California Press, Berkeley, Calif., 1967, pp. 335–360.

[1969] *Weakly mixing transformations which are not strongly mixing*, Proc. Amer. Math. Soc. **22** (1969), 559–562.

J. B. Conway

[1990] *A Course in Functional Analysis*, 2nd ed., Graduate Texts in Mathematics, vol. 96, Springer-Verlag, New York, NY, 1990.

M. M. Day

[1957a] *Amenable semigroups*, Ill. J. Math. **1** (1957), 509–544 (English).

[1957b] *Amenable semigroups*, Illinois J. Math. **1** (1957), 509–544.

K. de Leeuw and I. Glicksberg

[1959] *Almost periodic compactifications*, Bull. Amer. Math. Soc. **65** (1959), 134–139.

[1961] *Applications of almost periodic compactifications*, Acta Math. **105** (1961), 63–97.

A. Deitmar and S. Echterhoff

[2009] *Principles of Harmonic Analysis*, Universitext, Springer, New York, 2009.

M. Denker, C. Grillenberger, and K. Sigmund

[1976] *Ergodic Theory on Compact Spaces*, Lecture Notes in Mathematics, vol. 527, Springer-Verlag, Berlin, 1976.

R. Derndinger, R. Nagel, and G. Palm

[1987] *Ergodic Theory in the Perspective of Functional Analysis*, unpublished book manuscript, 1987.

J. Dugundji and Granas

[2003] *Fixed Point Theory*, Springer Monographs in Mathematics, Springer-Verlag, New York, 2003.

N. Dunford and J. T. Schwartz

[1958] *Linear Operators. I. General Theory*, Pure and Applied Mathematics, vol. 7, Interscience Publishers, Inc., New York, 1958. With the assistance of W. G. Bade and R. G. Bartle.

P. Ehrenfest and T. Ehrenfest

[1912] *Begriffliche Grundlagen der statistischen Auffassung in der Mechanik*, Encykl. d. Math. Wissensch. IV 2 II, Heft 6, 1912 (German).

M. Einsiedler and T. Ward

[2011] *Ergodic Theory with a View Towards Number Theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag, London, 2011.

T. Eisner

[2010] *Stability of Operators and Operator Semigroups*, Operator Theory: Advances and Applications, vol. 209, Birkhäuser Verlag, Basel, 2010.

R. Ellis

[1957] *Locally compact transformation groups*, Duke Math. J. **24** (1957), 119–125.

K.-J. Engel and R. Nagel

[2000] *One-Parameter Semigroups for Linear Evolution Equations*, Graduate Texts in Mathematics, vol. 194, Springer-Verlag, Berlin, 2000.

S. N. Ethier and T. G. Kurtz

[1986] *Markov Processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986.

G. B. Folland

[1995] *A Course in Abstract Harmonic Analysis*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1995.

[1999] *Real Analysis*, 2nd ed., Pure and Applied Mathematics (New York), John Wiley & Sons Inc., New York, 1999. A Wiley-Interscience Publication.

H. Furstenberg

[1961] *Strict ergodicity and transformation of the torus*, Amer. J. Math. **83** (1961), 573–601.

[1981] *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, Princeton, N.J., 1981. M. B. Porter Lectures.

[1977] *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

H. Furstenberg and B. Weiss

[1996] *A mean ergodic theorem for* $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, Convergence in ergodic theory and probability (Columbus, OH, 1993), Ohio State Univ. Math. Res. Inst. Publ., vol. 5, de Gruyter, Berlin, 1996, pp. 193–227.

G. Gallavotti

[1975] *Ergodicity, ensembles, irreversibility in Boltzmann and beyond*, J. Statist. Phys. **78** (1975), 1571–1589.

A. M. Garsia

[1965] *A simple proof of E. Hopf's maximal ergodic theorem*, J. Math. Mech. **14** (1965), 381–382.

[1970] *Topics in Almost Everywhere Convergence*, Lectures in Advanced Mathematics, vol. 4, Markham Publishing Co., Chicago, Ill., 1970.

E. Glasner

[2003] *Ergodic Theory via Joinings*, Mathematical Surveys and Monographs, vol. 101, American Mathematical Society, Providence, RI, 2003.

W. H. Gottschalk and G. A. Hedlund

[1955] *Topological Dynamics*, American Mathematical Society Colloquium Publications, Vol. 36, American Mathematical Society, Providence, R. I., 1955.

B. Green

[2008] *Lectures on Ergodic Theory, Part III*, http://www.dpmms.cam.ac.uk/ bjg23/ergodic-theory.html, 2008.

B. Green and T. Tao

[2008] *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547.

P. R. Halmos

[1944] *In general a measure preserving transformation is mixing*, Ann. of Math. (2) **45** (1944), 786–792.

[1950] *Measure Theory*, D. Van Nostrand Company, Inc., 1950.

[1956] *Lectures on Ergodic Theory*, Publications of the Mathematical Society of Japan, no. 3, The Mathematical Society of Japan, 1956.

P. R. Halmos and J. von Neumann

[1942] *Operator methods in classical mechanics. II*, Ann. of Math. (2) **43** (1942), 332–350.

E. Hewitt and K. A. Ross

[1979] *Abstract Harmonic Analysis. Vol. I*, 2nd ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 115, Springer-Verlag, Berlin, 1979.

E. Hewitt and K. Stromberg

[1969] *Real and Abstract Analysis. A Modern Treatment of the Theory of Functions of a Real Variable*, 2nd printing corrected, Springer-Verlag, New York, 1969.

N. Hindman and D. Strauss

[1998] *Algebra in the Stone-Čech Compactification*, de Gruyter Expositions in Mathematics, vol. 27, Walter de Gruyter & Co., Berlin, 1998.

E. Hlawka

[1979] *Theorie der Gleichverteilung*, Bibliographisches Institut, Mannheim, 1979.

K. H. Hofmann and S. A. Morris

[2006] *The Structure of Compact Groups*, augmented ed., de Gruyter Studies in Mathematics, vol. 25, Walter de Gruyter & Co., Berlin, 2006.

K. H. Hofmann and P. S. Mostert

[1966] *Elements of Compact Semigroups*, Charles E. Merrill Books Inc, Columbus, OH, 1966.

E. Hopf

[1954] *The general temporally discrete Markoff process*, J. Rational Mech. Anal. **3** (1954), 13–45.

B. Host and B. Kra

[2005a] *Convergence of polynomial ergodic averages*, Israel J. Math. **149** (2005), 1–19.

[2005b] *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488.

A. Ionescu Tulcea

[1964] *Ergodic properties of isometries in L^p spaces,* $1 < p < \infty$, Bull. Amer. Math. Soc. **70** (1964), 366–371.

[1965] *On the category of certain classes of transformations in ergodic theory*, Trans. Amer. Math. Soc. **114** (1965), 261–279.

K. Jacobs

[1956] *Ergodentheorie und fastperiodische Funktionen auf Halbgruppen*, Math. Z. **64** (1956), 298–338.

R. I. Jewett

[1970] *The prevalence of uniquely ergodic systems*, J. Math. Mech. **19** (1969/1970), 717–729.

L. K. Jones and V. Kuftinec

[1971] *A note on the Blum–Hanson theorem*, Proc. Amer. Math. Soc. **30** (1971), 202–203.

L. K. Jones and M. Lin

[1976] *Ergodic theorems of weak mixing type*, Proc. Amer. Math. Soc. **57** (1976), no. 1, 50–52.

S. Kakutani

[1938] *Iteration of linear operations in complex Banach spaces*, Proc. Imp. Acad. **14** (1938), no. 8, 295–300.

[1943] *Induced measure preserving transformations*, Proc. Imp. Acad. Tokyo **19** (1943), 635–641.

S. Kalikow and R. McCutcheon

[2010] *An Outline of Ergodic Theory*, Cambridge Studies in Advanced Mathematics, vol. 122, Cambridge University Press, Cambridge, 2010.

A. Katok and B. Hasselblatt

[1995] *Introduction to the Modern Theory of Dynamical Systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.

A. S. Kechris

[1995] *Classical Descriptive Set Theory*, Graduate Texts in Mathematics, vol. 156, Springer-Verlag, New York, 1995.

J. L. Kelley

[1975] *General Topology*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1975. Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.].

M. Kern, R. Nagel, and G. Palm

[1977] *Dilations of positive operators: construction and ergodic theory*, Math. Z. **156** (1977), no. 3, 265–277.

A. Kolmogoroff

[1933] *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Math. 2, Nr. 3, IV + 62 S , 1933 (German).

A. N. Kolmogoroff

[1925] *Sur les fonctions harmoniques conjugées et les séries de Fourier*, Fund. Math. **7** (1925), 23–28.

B. Koopman and J. von Neumann

[1932] *Dynamical systems of continuous spectra*, Proc. Natl. Acad. Sci. USA **18** (1932), 255–263 (English).

B. Kra

[2006] *The Green-Tao theorem on arithmetic progressions in the primes: an ergodic point of view*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 1, 3–23 (electronic).

[2007] *Ergodic methods in additive combinatorics*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 103–143.

U. Krengel

[1985] *Ergodic Theorems*, de Gruyter Studies in Mathematics, vol. 6, Walter de Gruyter & Co., Berlin, 1985. With a supplement by Antoine Brunel.

W. Krieger

[1972] *On unique ergodicity*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory (Berkeley, Calif.), Univ. California Press, 1972, pp. 327–346.

L. Kuipers and H. Niederreiter

[1974] *Uniform Distribution of Sequences*, Pure and Applied Mathematics, Wiley-Interscience [John Wiley & Sons], New York, 1974.

K. Kuratowski

[1966] *Topology. Vol. I*, Academic Press Inc., New York, 1966. New edition, revised and augmented. Translated from the French by J. Jaworowski.

S. Lang

[1993] *Real and Functional Analysis*, 3rd ed., Graduate Texts in Mathematics, vol. 142, Springer-Verlag, New York, 1993.

[2005] *Algebra*, 3rd ed., Graduate Text in Mathematics, vol. 211, Springer-Verlag, 2005.

E. R. Lorch

[1939] *Means of iterated transformations in reflexive vector spaces*, Bull. Amer. Math. Soc. **45** (1939), 945–947.

M. Mathieu

[1988] *On the origin of the notion 'ergodic theory'*, Expo. Math. **6** (1988), 373–377.

R. E. Megginson

[1998] *An Introduction to Banach Space Theory*, Graduate Texts in Mathematics, vol. 183, Springer-Verlag, New York, NY, 1998.

V. Müller and Y. Tomilov

[2007] *Quasisimilarity of power bounded operators and Blum–Hanson property*, J. Funct. Anal. **246** (2007), no. 2, 385–399.

R. Nagel and G. Palm

[1982] *Lattice dilations of positive contractions on L^p-spaces*, Canad. Math. Bull. **25** (1982), no. 3, 371–374.

R. J. Nagel

[1973] *Mittelergodische Halbgruppen linearer Operatoren*, Ann. Inst. Fourier (Grenoble) **23** (1973), no. 4, 75–87.

R. J. Nagel and M. Wolff

[1972] *Abstract dynamical systems with an application to operators with discrete spectrum*, Arch. Math. (Basel) **23** (1972), 170–176.

I. Namioka

[1974] *Separate continuity and joint continuity*, Pacific J. Math. **51** (1974), 515–531.

G. Palm

[1976] *Entropie und Erzeuger in dynamischen Verbänden*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **36** (1976), no. 1, 27–45.

A. L. Paterson

[1988] *Amenability*, Mathematical Surveys and Monographs, vol. 29, American Mathematical Society, Providence, RI, 1988.

K. Petersen

[1989] *Ergodic Theory*, Cambridge Studies in Advanced Mathematics, vol. 2, Cambridge University Press, Cambridge, 1989. Corrected reprint of the 1983 original.

R. R. Phelps

[1966] *Lectures on Choquet's Theorem*, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1966.

M. Pollicott and M. Yuri

[1998] *Dynamical Systems and Ergodic Theory*, London Mathematical Society Student Texts, vol. 40, Cambridge University Press, Cambridge, 1998.

R. A. Raimi

[1964] *Minimal sets and ergodic measures in $\beta N - N$*, Bull. Amer. Math. Soc. **70** (1964), 711–712.

I. K. Rana

[2002] *An Introduction to Measure and Integration*, 2nd ed., Graduate Studies in Mathematics, vol. 45, American Mathematical Society, Providence, RI, 2002.

M. Rédei and C. Werndl

[2012] *On the history of the isomorphism problem of dynamical systems with special regard to von Neumann's contribution*, Archive for History of Exact Sciences **66** (2012), no. 1, 71–93.

M. Reed and B. Simon

[1972] *Methods of Modern Mathematical Physics. I: Functional Analysis*, Academic Press Inc., New York, San Francisco, London, 1972.

V. Rohlin

[1948] *A "general" measure-preserving transformation is not mixing*, Doklady Akad. Nauk SSSR (N.S.) **60** (1948), 349–351.

A. Rosenthal

[1988] *Strictly ergodic models for noninvertible transformations*, Israel J. Math. **64** (1988), no. 1, 57–72.

H. L. Royden

[1963] *Real Analysis*, The Macmillan Co., New York, 1963.

W. Rudin

[1987] *Real and Complex Analysis*, 3rd ed., McGraw-Hill Book Co., New York, NY, 1987.

[1990] *Fourier Analysis on Groups*, paperback ed., John Wiley & Sons., New York etc., 1990. Wiley Classics Library; A Wiley-Interscience Publication.

[1991] *Functional Analysis*, 2nd ed., International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., New York, NY, 1991.

C. Ryll-Nardzewski

[1967] *On fixed points of semigroups of endomorphisms of linear spaces*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1, Univ. California Press, Berkeley, Calif., 1967, pp. 55–61.

H. H. Schaefer

[1974] *Banach Lattices and Positive Operators*, Die Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 215, Springer-Verlag, Berlin-Heidelberg-New York, 1974.

[1980] *Topological Vector Spaces*, 4th corrected printing, Graduate Texts in Mathematics, vol. 3, Springer-Verlag, New York-Heidelberg-Berlin, 1980.

C. E. Silva

[2008] *Invitation to Ergodic Theory*, Student Mathematical Library, vol. 42, American Mathematical Society, Providence, RI, 2008.

E. M. Stein

[1961a] *On limits of seqences of operators*, Ann. of Math. (2) **74** (1961), 140–170.

[1961b] *On the maximal ergodic theorem*, Proc. Nat. Acad. Sci. U.S.A. **47** (1961), 1894–1897.

[1993] *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton Mathematical Series, vol. 43, Princeton University Press, Princeton, NJ, 1993. With the assistance of Timothy S. Murphy.

E. Szemerédi

[1975] *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245. Collection of articles in memory of Juriĭ Vladimirovič Linnik.

T. Tao

[2007] *The dichotomy between structure and randomness, arithmetic progressions, and the primes*, International Congress of Mathematicians. Vol. I, Eur. Math. Soc., Zürich, 2007, pp. 581–608.

[2008a] *Topics in Ergodic Theory*, http://terrytao.wordpress.com/category/254a-ergodic-theory/, 2008.

[2008b] *The van der Corput trick, and equidistribution on nilmanifolds*, Topics in Ergodic Theory, 2008. http://terrytao.wordpress.com/2008/06/14/the-van-der-corputs-trick-and-equidistribution-on-nilmanifolds.

S. Todorcevic

[1997] *Topics in Topology*, Lecture Notes in Mathematics, vol. 1652, Springer-Verlag, Berlin, 1997.

B. L. van der Waerden

[1927] *Beweis einer Baudetschen Vermutung*, Nieuw Archief **15** (1927), 212–216 (German).

H. Vogt

[2010] *An Eberlein-Šmulian type result for the weak* topology*, Arch. Math. (Basel) **95** (2010), no. 1, 31–34.

J. von Neumann

[1932a] *Einige Sätze über messbare Abbildungen*, Ann. of Math. (2) **33** (1932), no. 3, 574–586.

[1932b] *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. USA **18** (1932), 70–82 (English).

P. Walters

[1982] *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York, 1982.

H. Weyl

[1916] *Über die Gleichverteilung von Zahlen mod. Eins*, Math. Ann. **77** (1916), no. 3, 313–352.

N. Wiener

[1939] *The ergodic theorem*, Duke Math. J. **5** (1939), no. 1, 1–18.

E. Wigner

[1960] *The unreasonable effectiveness of mathematics in the natural sciences*, Comm. Pure Appl. Math. **13** (1960), 1–14.

S. Willard

[2004] *General Topology*, Dover Publications Inc., Mineola, NY, 2004. Reprint of the 1970 original [Addison-Wesley, Reading, MA].

K. Yosida

[1938] *Mean ergodic theorem in Banach spaces*, Proc. Imp. Acad. **14** (1938), no. 8, 292–294.

K. Yosida and S. Kakutani

[1939] *Birkhoff's ergodic theorem and the maximal ergodic theorem*, Proc. Imp. Acad., Tokyo **15** (1939), 165–168.

N. Young

[1988] *An introduction to Hilbert space*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1988.

Index

absolute value, *see* modulus
accumulation point, 296
algebra, 320
  C^*-algebra, 321
  abstract measure algebra, 179
  Boole algebra, 179
  disc algebra, 57
  homomorphism, 321
  isomorphism, 321
  isomorphism of MDSs, 180
  isomorphism of measure spaces, 180
  measure algebra, 75
  of sets, 306
almost everywhere, 313
almost periodic
  sequence, 49
  strongly, 236
  weakly, 236
almost stable subspace, 134
almost weakly stable, 134, 239
  semigroup, 241
alphabet, 8
amenable semigroup, 215, 238
arithmetic progression, 275
asymptotic density, 132
automorphism
  of a TDS, 12

Baire algebra, 66
Baire measure, 66
Baire space, 302
baker's transformation, 61, 178
Banach algebra, 320
Banach density, 286
Banach lattice, 95
Banach limit, 161
Banach space, 319
Bernoulli shift, 64
bi-dual
  group, 223
  of a normed space, 324
bi-invariant set (in a TDS), 15
bi-Markov operator, 193
block, 16
Bohr-compactification, 224
Borel algebra, 66, 306
Borel measure, 66
bounded gaps, 30
bp-covergence, 334

Cantor set, 8, 301
Cantor system, 8
Čech–Stone compactification, 49
ceiling, 86
Cesàro mean, 108
Cesàro-bounded, 117
character
  group, 219
  of $\mathbb{T}$, 48
  of a group, 257
  of an Abelian group, 219
  orthogonality, 220
character group, 219
characteristic function, 305
clopen set, 301
closed absolutely convex hull, 349
cluster point, 296
  of a net, 304
compact
  countably, 353
  relatively countably compact, 354
  relatively sequentially compact, 354
  relatively weakly compact, 326
  relatively weakly sequentially compact, 326

sequentially, 354
compact group, 9, 217
compact space, 300
compactification
Bohr, 224
one-point, 49, 301
two-point, 18
complement (in a lattice), 179
complete
contraction, 119
isomorphism invariant, 266
lattice, 92
metric space, 294
order complete Banach lattice, 97
relatively complete sub-σ-algebra, 201
conditional expectation, 112, 199
operator, 202
conditional probability, 81
congruence (TDS), 12
conjugate exponent, 125
conjugation
invariant, 40
of elements (in a vector lattice), 95
convergence in density, 132
convolution on $\mathrm{L}^1(G)$, 254
coordinate function, 248, 253
cylinder set, 311, 312

d-torus, 13, 269
decomposition
Jacobs–de Leeuw–Glicksberg, 239
mean ergodic, 109, 239
density
asymptotic, 132
Banach density, 286
convergence in density, 132
lower, 143
upper, 143
Diophantine approximation, 33
Dirac functional, 43
directed set, 303
disconnected
extremally, 302
totally, 301
discrete spectrum
dyadic integers, 270
dynamical system, 262
operator, 261
rotation on torus, 269
disjoint union of TDSs, 16
doubling map, 24, 62, 176
TDS, 24
dual group, *see* character group
dual of a normed space, 323
Dunford–Schwartz operator, 119
dyadic adding machine, 11
dyadic integers, 11, 270
dynamical system, 2
measure-preserving, 61
topological, 7
topological MDS, 183
with discrete spectrum, 262
Dynkin system, 307

eigenvalue
of an operator, 330
peripheral, 47
equidistributed sequence, 157
Ergode, 1
Ergodenhypothese, 107
ergodic
group rotation, 154
hypothesis, 107
Markov shift, 84, 115
MDS, 82
MDS (characterisation), 100
measure, 149
rotation on torus, 101
strongly, 136
uniquely, 150
ergodic hypothesis, 4
ergodic theorem
Akcoglu's, 165
dominated, 174
individual, 163
maximal, 168
Mean ergodic theorem, 110
pointwise, 163, 164
stochastic, 165
von Neumann's mean ergodic theorem, 108
ergodisches System, 1
essential inverse of MDSs, 76
essentially
equal sets, 75
invertible (MDS), 76
excluded block, 17
extremally disconnected, 302
extreme point, 327

factor, 205
bi-invariant, 204
Kronecker, 240
Markov operator, 197
model, 205
of an MDS, 204
finite intersection property, 299
first category, 302
fixed space, 47, 100, 108

forward (topologically) transitive, 18
full subalgebra, 185
functional
  Dirac, 43
  evaluation, 43
  linear, 323
  order continuous, 187
  positive, 336
Furstenberg Correspondence, 278, 284

Gauss map, 63
G_δ-set, 302
Gelfand map, 51
Gelfand space, 49, 50
generating element, 263
generator of a σ-algebra, 306
group, 212
  Abelian, 211
  bi-dual, 223
  character, 219
  compact, 9, 217
  dual, 219
  dyadic integers, 11
  Heisenberg, 10
  inverse element, 212
  left rotation, 9
  locally compact group, 217
  monothetic, 263
  representation, 245
  topological, 9, 213
  unitary, 353
group extension of a TDS, 15
group representation, 245
  finite dimensional, 252
  left regular, 254
  strongly continuous, 245
  unitary, 248
  weakly continuous, 245
group rotation, 9
  ergodic, 154
  MDS, 70
  minimal, 154
  TDS, 9
group translation TDS, 9

Haar measure, 70, 218, 352
Heisenberg group, 10
Hilbert space, 330
homeomorphism, 296
homogeneous system, 13
homomorphism
  of algebras, 321
  of Banach lattices, 96
  of lattices, 93
  of MDSs, 179
  of measure algebras, 179
  of TDSs, 11
  point homomorphism of MDSs, 175

ideal
  in a Banach lattice, 98
  in a semigroup, 212
  in an algebra, 321
  maximal, 321
  principal, 321
  proper, 321
idempotent, 211
induced transformation, 85
infimum in a lattice, 92
infinite product probability space, 311
infinitely recurrent set, 78
integral, 309
intertwining, 182, 252
invariant measure, 60
invariant set
  in a TDS, 15
  in an MDS, 82
inverse
  essential (MDS), 76
  in a Banach algebra, 320
invertible
  MDS, 76
  TDS, 7
involution, 321
irrational rotation, 20
irreducible
  matrix, 21
  positive operator, 99
  row-stochastic matrix, 114
  unitary representation, 248, 252
isolated point, 296
isometric TDS, 29
isomorphism
  $*$-isomorphism, 321
  algebra isomorphism of measure spaces, 180
  Markov, 181
  of algebras, 321
  of lattices, 93
  of MDSs, 180
  of TDSs, 12
  of topological groups, 222
  point isomorphism of MDSs, 176
  point isomorphism of measure spaces, 180
  spectral, 269
isomorphism invariant, 266
isomorphism problem, 269
iterate MDS, 128

Jacobs–de Leeuw–Glicksberg decomp., 239
  reversible part, 239
  stable part, 239
JdLG-decomposition, *see* Jacobs–de Leeuw–Glicksberg decomposition

Koopman operator, 37, 45, 91
Kronecker factor, 240

lattice, 92
  Banach lattice, 95
  distributive, 179
law of large numbers
  strong, 171
  weak, 113
left ideal (in a semigroup), 212
left rotation, 9
left rotation operator, 48
left shift, 8
Lemma
  Fekete, 56
  Hopf, 174
  Kakutani–Rokhlin, 85
  Koopman–von Neumann, 133
  Pythagoras', 331
  Urysohn, 38
  van der Corput, 139
  Weil, 263
letters, 8
limit
  of a net, 303
  projective, 11
linear span, 320
locally compact group, 217
locally compact space, 301
locally convex space, 327

Markov
  embedding, 181, 196
  embedding of MDSs, 182
  isomorphism, 181
  measure, 64
  operator, 149, 161, 193
  shift, 65
Markov operator, 180
  automorphism, 199
  embedding, 196
  factor map, 197
  isomorphism, 198
  projection, 199
  trivial, 194
Markov shift, 65
  transition matrix, 65
matrix
  aperiodic, 130
  irreducible, 21, 114
  primitive, 130
  row-stochastic, 114
  row-substochastic, 8
  strictly positive, 114
  transition, 21, 65
maximal ideal, 321
  in an algebra, 44
maximal inequality, 166
maximal operator, 166
MDS, 61
  baker's transformation, 61
  discrete spectrum, 262
  doubling map, 62
  ergodic, 82
  factor, 204
  faithful topological, 183
  Gauss map, 63
  group rotation, 70
  homogeneous system, 71
  invariant set, 82
  invertible, 76
  iterate, 128
  metric, 183
  product, 65
  right rotation system, 70
  skew product, 65
  spectrally isomorphic, 269
  tent map, 62
  topological, 183
mean ergodic
  operator, 109, 116
  projection, 109, 203
  subalgebra, 189
measurable
  cylinder set, 311
  mapping, 308
  rectangle, 316
  set, 307
measurable space, 60, 306
measure, 307
  Baire, 66
  Borel, 66
  complex, 316
  ergodic, 149
  Haar, 70, 218, 352
  invariant, 60
  inversion invariant, 70
  Markov, 64
  outer, 308
  probability, 307
  product, 311
  push-forward, 60, 310

regular, 66
regular Borel, 66
right-invariant, 70
signed, 316
strictly positive, 69, 151
support, 68
total variation, 316
measure algebra, 49, 75
measure space, 307
measure-preserving dyn. system, *see* MDS
measure-preserving map, 60
metric, 293
discrete, 294
model, 185
space, 293
metrisability
of weak topology, 325
of weak* topology, 324
of weakly compact sets, 326
minimal
group rotation, 154
subsystem, 28
TDS, 28
minimal ideal (in a semigroup), 212
minimal idempotent, 212
mixing
Bernoulli shift, 126
Markov shift, 129
strongly, 126
weakly, 132
weakly of all orders, 138, 142
weakly of order k, 138
model
of a factor, 205
Stone, 186
topological, 183
topological model of a factor, 205
modulus
in a lattice, 95
of a functional, 338
Morse sequence, 36
multiplicative (linear map), 321

net, 303
Neumann series, 52
norm
order continuous, 97
strictly monotone, 97
normal
number, 174
simply normal number, 170
nowhere dense, 302
null set, 313

one-dimensional torus, 9
one-sided shift, 8
operator
adjoint, 324
bi-Markov, 193
contraction, 322
Dunford–Schwartz, 119
intertwining (representations), 252
isometry, 207, 322
Koopman, 37, 45, 91
left rotation, 48
Markov, 149, 161, 180, 193
maximal, 166
mean ergodic, 109, 116
pointwise ergodic, 164
positive, 96, 336
positive semi-definite, 208
power-bounded, 329
projection, 322
with discrete spectrum, 261
operator semigroup
weakly compact, 232, 234
orbit, 17
forward, 17
total, 17
order continuous
functional, 187
norm, 97
order interval, 97
order of a subshift, 17
orthogonal projection, 207
oscillation, 350

partition of unity, 337
peripheral point spectrum, 47
Poincaré sequence, 285
point
almost periodic, 30
generic, 173
infinitely recurrent, 30
periodic , 30
recurrent, 30
point factor map, 175
point spectrum
of an operator, 330
peripheral, 47
pointwise topology (on the dual group), 221
polar (of a set), 348
polarisation identity, 330
positive functional, 336
positive operator, 96
Hölder inequality, 103
irreducible, 99
reducible, 99

positive semi-definite operator, 208
power-bounded, 119
principal ideal, 321
Principle
  Banach, 166
  Furstenberg Correspondence, 278, 284
  of determinism, 2
  of uniform boundedness, 322
  pigeonhole, 36
probability
  measure, 307
probability space, 307
product
  σ-algebra, 310
  measure, 311
  of MDSs, 65
  of TDSs, 13
  of TDSs (infinite), 14
  of TDSs (skew), 14
  topology, 298
projection
  mean ergodic, 109, 203
  orthogonal, 207

quotient
  mapping, 297
  space, 320

random literature, 89
rational rotation, 20
recurrent set, 78
reducible positive operator, 99
reducing subspace, 252
reflexive space, 324
relatively compact, 300
relatively complete sub-σ-algebra, 201
relatively dense, *see* bounded gaps, 143, 282
representation
  Gelfand, 50
  of a group, 245
  Stone, 186
  Stone representation space, 186
return time
  expected, 84
  of almost periodic points, 30
  time of first return, 81
reversible part, 239
right ideal (in a semigroup), 212
Rokhlin tower, 86
rotation, 9
  irrational, 20
  rational, 20
  TDS, 9
  torus, 9, 269
row-stochastic matrix, 114
  irreducible, 114
row-substochastic matrix, 8

σ-additive, 305
σ-algebra, 306
  invariant, 112
second category, 302
semigroup, 211
  JdLG-admissible, 237
  Abelian, 211
  amenable, 238
  compact left-topological, 213
  generated by a single operator, 236
  ideal, 212
  idempotent element, 211
  left ideal, 212
  left-topological, 213
  right ideal, 212
  right-topological, 213
  semitopological, 213, 349
  subsemigroup, 212
  weakly compact, 232
separable
  metric space, 320
  topological space, 295
separately continuous, 349
separating, 299
  subalgebra, 40
  subset, 118
sequence
  almost periodic, 49
  equidistributed, 157
  Morse, 36
  Poincaré, 285
shift, 8
  Bernoulli, 64, 176, 177, 271
  left, 8
  left (operator), 123
  Markov, 65
  one-sided, 8
  right (operator), 123
  TDS, 8
  two-sided, 8
  two-sided Bernoulli, 178
skew product
  cocyle, 15
  of MDSs, 65
  of TDSs, 14
skew rotation, 15, 156
spectral radius
  formula, 54
  in unital algebras, 51
  of an operator, 329

spectrally isomorphic, 269
spectrum
  in unital algebras, 51
  of an operator, 329
stable
  $\wedge$, $\vee$-stable, 92
  $\cap$, $\cup$, $\setminus$-stable, 305
stable set in a TDS, 15
step function, 315
Stone representation, 186
strictly convex Banach space, 238
strictly ergodic TDS, 152
strictly positive measure, 69, 151
strongly
  ergodic, 136
  mixing, 126
strongly compact operator group, 245
subalgebra, 40
  conjugation invariant, 40
  full, 185
  in $\mathrm{C}(K)$, 40
  mean ergodic, 189
sublattice, 92
subsemigroup, 212
subshift, 16
  of finite type, 17
subsystem
  maximal surjective, 16, 162
  minimal, 28
  of a TDS, 15
support of a measure, 68, 151
supremum in a lattice, 92
Sushkevich kernel, 212
symmetric set (in a group), 217
syndetic, *see* bounded gaps, 282

TDS, 7
  automorphism, 12
  bi-invariant set, 15
  conjugation, *see* isomorphism of TDSs
  discrete spectrum, 262
  disjoint union, 16
  doubling map, 24
  dyadic adding machine, 11
  extension, 11
  factor, 11
  factor map, 11
  homogeneous, 13
  homomorphism, 11
  invariant set, 15
  invertible, 7
  isometric, 29
  isomorphism, 12
  minimal, 28
  skew rotation, 15
  strictly ergodic, 152
  subshift, 16
  subsystem, 15
  surjective, 7
  tent map, 24
  uniquely ergodic, 150
tent map, 24, 62, 177
  MDS, 62
  TDS, 24
Theorem
  Akcoglu's ergodic theorem, 165
  Banach's principle, 166
  Banach–Alaoglu, 324
  Banach–Steinhaus, 322
  Bipolar theorem, 348
  Birkhoff's pointwise ergodic theorem, 163
  Birkhoff's recurrence theorem, 32
  Blum–Hanson, 130
  Borel, 170
  Carathéodory, 308
  Dirichlet, 36
  Doinated convergence theorem, 314
  Dominated ergodic theorem, 174
  Dynkin, 307
  Eberlein, 346
  Eberlein–Šmulian, 325, 355
  Ellis (minimal idempotents), 213
  Ellis (topological groups), 215, 353
  Existence of Haar measure, 218, 352
  Fubini, 316
  Furstenberg, 156
  Furstenberg–Sárközy, 283
  Furstenberg–Weiss, 285
  Gelfand–Mazur, 53
  Goldstine, 325
  Green–Tao, 4, 276
  Grothendieck, 347, 354
  Hahn, 308
  Hahn–Banach, 323
  Halmos–von Neumann, 268
  Heine–Borel, 301
  Host–Kra, 279
  Inverse Mapping, 323
  Ionescu Tulcea, 312
  Jewett–Krieger, 190
  Jones–Lin, 134
  Kac, 84
  Kolmogorov's law of large numbers, 171
  Kreĭn–Šmulian, 326, 348
  Kreĭn–Milman, 327
  Kronecker, 20, 226
  Krylov–Bogoljubov, 69, 148
  Markov–Kakutani, 148

Maximal ergodic theorem, 168
Maximal inequality, 168
Mazur, 325
Mean ergodic theorem for squares, 284
Milman, 327
Monotone convergence theorem, 310
Perron, 72, 114
Peter–Weyl, 257
Poincaré recurrence theorem, 78
Pointwise ergodic theorem, 164
Pontryagin duality theorem, 224
Principle of Uniform Boundedness, 322
Riesz' representation theorem, 68
Riesz–Fréchet, 331
Riesz–Markov, 340
Roth, 283
Spectral theorem, 332
Stochastic ergodic theorem, 165
Stone–Weierstraß, 40
Szemerédi, 5, 276
Tietze, 39
Tonelli, 310
Tychonov, 300
Uniqueness theorem for measures, 307
Van der Waerden, 275
von Neumann (isomorphism of MDSs), 180
von Neumann's mean ergodic theorem, 108
Weyl, 158
time mean, 3
topological dynamical system, *see* TDS
backward, 7
forward, 7
topological group, 9, 213
topological MDS, 183
topological model, 183
topological semigroup, 213
topological space, 294
connected, 301
disconnected, 301
metrisabile, 294
normal, 38
Polish, 303
separable, 42, 295
topological vector space, 326
topologically transitive, 18
topology, 294
compact, 300
compact-open (on the dual group), 221
discrete, 294
Hausdorff, 295
inductive, 297
pointwise, 345
pointwise (on the dual group), 221
product, 298
projective, 297
quotient, 297
strong operator, 231, 327
subspace, 297
weak, 325
weak operator, 231, 328
weak*, 324
torus, 9
d-torus, 13, 269
monothetic, 269
one-dimensional, 9
rotation, 269
total orbit, 17
totally disconnected, 271, 301
transition matrix, 21
of a Markov shift, 65
translation, *see* rotation
translation mod 1, 9
translation system, 9
trigonometric polynomial, 221
two-sided shift, 8

uniformly continuous, 217, 296
uniquely ergodic TDS, 150
unit element, 320
unitary
group, 353
representation, 248
system, 248
system for a semigroup, 250
unitary group, 233
unitary representation, 252
equivalent, 257
irreducible, 252
upper Banach density, 286

vector lattice
imaginary part, 95
positive cone, 94, 95
positive element, 94
real, 94
real part, 95

weakly compact
operator group, 245
operator semigroup, 232
weakly mixing, 132
of all orders, 138, 142
of order k, 138
weakly stable, 128
words, 8

zero element (in a semigroup), 211

Symbol Index