CAMBRIDGE STUDIES IN ADVANCED MATHEMATICS 151

Foundations of Ergodic Theory

Rich with examples and applications, this textbook provides a coherent and self-contained introduction to ergodic theory suitable for a variety of one- or two-semester courses. The authors' clear and fluent exposition helps the reader to grasp quickly the most important ideas of the theory, and their use of concrete examples illustrates these ideas and puts the results into perspective.

The book requires few prerequisites, with background material supplied in the appendix. The first four chapters cover elementary material suitable for undergraduate students – invariance, recurrence and ergodicity – as well as some of the main examples. The authors then gradually build up to more sophisticated topics, including correlations, equivalent systems, entropy, the variational principle, and thermodynamic formalism. The 400 exercises increase in difficulty through the text and test the reader's understanding of the whole theory. Hints and solutions are provided at the end of the book.

Marcelo Viana is Professor of Mathematics at Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro and a leading research expert in ergodic theory and dynamical systems. He has served in several academic organizations, such as the International Mathematical Union (Vice-president 2011–2014), the Brazilian Mathematical Society (President 2013–2015), the Latin American Mathematical Union (Scientific Coordinator, 2001–2008) and the newly founded Mathematical Council of the Americas. He is also a member of the academies of science of Brazil, Portugal, Chile and the Developing World and he has received several academic distinctions, including the Grand Croix of Scientific Merit, granted by the President of Brazil, in 2000, and the Ramanujan Prize of ICTP and IMU, in 2005. He was an invited speaker at the international congress of mathematicians ICM 1994, in Zurich, a plenary speaker at the International Congress of Mathematical Physics ICMP 1994, and a Plenary Speaker at the ICM 1998, in Berlin. To date, he has supervised 32 doctoral theses. Currently, he leads the organization of the ICM 2018 in Rio de Janeiro and is also involved in initiatives to improve mathematical education in his country.

Krerley Oliveira is Associate Professor at the Universidade Federal de Alagoas (UFAL), Brazil, where he founded the graduate program in mathematics and the State of Alagoas Math Olympiad program, where he promotes young talents in mathematics. He was a medalist at the Brazilian Mathematical Olympiad (1996) and twice at the Iberoamerican Mathematical Olympiad for university students (1999 and 2000). He was elected an Affiliate Member of the Brazilian Academy of Sciences (2007–2012) and his research is focused on dynamical systems and ergodic theory.

CAMBRIDGE STUDIES IN ADVANCED MATHEMATICS

All the titles listed below can be obtained from good booksellers or from Cambridge University Press.
For a complete series listing visit www.cambridge.org/mathematics.

Already published

112 D. W. Stroock *Partial differential equations for probabilists*
113 A. Kirillov, Jr *An introduction to Lie groups and Lie algebras*
114 F. Gesztesy *et al. Soliton equations and their algebro-geometric solutions, II*
115 E. de Faria & W. de Melo *Mathematical tools for one-dimensional dynamics*
116 D. Applebaum *Lévy processes and stochastic calculus (2nd edition)*
117 T. Szamuely *Galois groups and fundamental groups*
118 G. W. Anderson, A. Guionnet & O. Zeitouni *An introduction to random matrices*
119 C. Perez-Garcia & W. H. Schikh of *Locally convex spaces over non-Archimedean valued fields*
120 P. K. Friz & N. B. Victoir *Multidimensional stochastic processes as rough paths*
121 T. Ceccherini-Silberstein, F. Scarabotti & F. Tolli *Representation theory of the symmetric groups*
122 S. Kalikow & R. McCutcheon *An outline of ergodic theory*
123 G. F. Lawler & V. Limic *Random walk: A modern introduction*
124 K. Lux & H. Pahlings *Representations of groups*
125 K. S. Kedlaya *p-adic differential equations*
126 R. Beals & R. Wong *Special functions*
127 E. de Faria & W. de Melo *Mathematical aspects of quantum field theory*
128 A. Terras *Zeta functions of graphs*
129 D. Goldfeld & J. Hundley *Automorphic representations and L-functions for the general lineargroup, I*
130 D. Goldfeld & J. Hundley *Automorphic representations and L-functions for the general lineargroup, II*
131 D. A. Craven *The theory of fusion systems*
132 J. Väänänen *Models and games*
133 G. Malle & D. Testerman *Linear algebraic groups and finite groups of Lie type*
134 P. Li *Geometric analysis*
135 F. Maggi *Sets of finite perimeter and geometric variational problems*
136 M. Brodmann & R. Y. Sharp *Local cohomology (2nd Edition)*
137 C. Muscalu & W. Schlag *Classical and multilinear harmonic analysis, I*
138 C. Muscalu & W. Schlag *Classical and multilinear harmonic analysis, II*
139 B. Helffer *Spectral theory and its applications*
140 R. Pemantle & M. C. Wilson *Analytic combinatorics in several variables*
141 B. Branner & N. Fagella *Quasiconformalsurgery in holomorphic dynamics*
142 R. M. Dudley *Uniform central limit theorems (2nd Edition)*
143 T. Leinster *Basic category theory*
144 I. Arzhantsev, U. Derenthal, J. Hausen & A. Laface *Cox rings*
145 M. Viana *Lectures on Lyapunov exponents*
146 J.-H.Evertse & K. Győry *Unit equations in Diophantine number theory*
147 A. Prasad *Representation theory*
148 S. R. Garcia, J. Mashreghi & W. T. Ross *Introduction to model spaces and their operators*
149 C. Godsil & K. Meagher *Erdős–Ko–Rado theorems: Algebraic approaches*
150 P. Mattila *Fourier analysis and Hausdorff dimension*
151 M. Viana & K. Oliveira *Foundations of ergodic theory*

Foundations of Ergodic Theory

MARCELO VIANA
IMPA, Rio de Janeiro

KRERLEY OLIVEIRA
Universidade Federal de Alagoas, Brazil

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107126961

First published 2016

Printed in the United Kingdom by Clays, St Ives plc

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Viana, Marcelo.
Foundations of ergodic theory / Marcelo Viana, IMPA, Rio de Janeiro, Krerley Oliveira, Universidade Federal de Alagoas, Brazil.
pages cm. – (Cambridge studies in advanced mathematics ; 151)
Includes bibliographical references and index.
ISBN 978-1-107-12696-1 (hbk)
1. Ergodic theory–Textbooks. 2. Topology–Textbooks. I. Oliveira, Krerley. II. Title.
QA611.5.V53 2016
515′.48–dc23 2015022583

ISBN 978-1-107-12696-1 Hardback

Contents

Preface

In short terms, ergodic theory is the mathematical discipline that deals with dynamical systems endowed with invariant measures. Let us begin by explaining what we mean by this and why these mathematical objects are so worth studying. Next, we highlight some of the major achievements in this field, whose roots go back to the physics of the late 19th century. Near the end of the preface, we outline the content of this book, its structure and its prerequisites.

What is a dynamical system?

There are several definitions of what a dynamical system is some more general than others. We restrict ourselves to two main models.

The first one, to which we refer most of the time, is a transformation $f : M \to M$ in some space M. Heuristically, we think of M as the space of all possible states of a given system. Then f is the evolution law, associating with each state $x \in M$ the one state $f(x) \in M$ the system will be in a unit of time later. Thus, time is a discrete parameter in this model.

We also consider models of dynamical systems with continuous time, namely flows. Recall that a *flow* in a space M is a family $f^t : M \to M$, $t \in \mathbb{R}$ of transformations satisfying

$$f^0 = \text{identity} \quad \text{and} \quad f^t \circ f^s = f^{t+s} \text{ for all } t, s \in \mathbb{R}. \tag{0.0.1}$$

Flows appear, most notably, in connection with differential equations: take f^t to be the transformation associating with each $x \in M$ the value at time t of the solution of the equation that passes through x at time zero.

We always assume that the dynamical system is measurable, that is, that the space M carries a σ-algebra of *measurable subsets* that is preserved by the dynamics, in the sense that the pre-image of any measurable subset is still a measurable subset. Often, we take M to be a topological space, or even a metric space, endowed with the Borel σ-algebra, that is, the smallest σ-algebra that contains all open sets. Even more, in many of the situations we consider in this book, M is a smooth manifold and the dynamical system is taken to be differentiable.

What is an invariant measure?

A *measure* in M is a non-negative function μ defined on the σ-algebra of M, such that $\mu(\emptyset) = 0$ and

$$\mu\big(\cup_n A_n\big) = \sum_n \mu(A_n)$$

for any countable family $\{A_n\}$ of pairwise disjoint measurable subsets. We call μ a probability measure if $\mu(M) = 1$. In most cases, we deal with finite measures, that is, such that $\mu(M) < \infty$. Then we can easily turn μ into a probability ν: just define

$$\nu(E) = \frac{\mu(E)}{\mu(M)} \quad \text{for every measurable set } E \subset M.$$

In general, we say that a measure μ is invariant under a transformation f if

$$\mu(E) = \mu(f^{-1}(E)) \quad \text{for every measurable set } E \subset M. \tag{0.0.2}$$

Heuristically, this may be read as follows: the probability that a point is in any given measurable set is the same as the probability that its image is in that set. For flows, we replace (0.0.2) by

$$\mu(E) = \mu(f^{-t}(E)) \quad \text{for every measurable set } E \subset M \text{ and } t \in \mathbb{R}. \tag{0.0.3}$$

Notice that (0.0.2)–(0.0.3) do make sense since, by assumption, the pre-image of a measurable set is also a measurable set.

Why study invariant measures?

As in any other branch of mathematics, an important part of the motivation is intrinsic and aesthetical: as we will see, these mathematical structures have deep and surprising properties, which are expressed through beautiful theorems. Equally fascinating, ideas and results from ergodic theory can be applied in many other areas of mathematics, including some that do not seem to have anything to do with probabilistic concepts, such as combinatorics and number theory.

Another key motivation is that many problems in the experimental sciences, including many complicated natural phenomena, can be modelled by dynamical systems that leave some interesting measure invariant. Historically, the most important example came from physics: Hamiltonian systems, which describe the evolution of conservative systems in Newtonian mechanics, are described by certain flows that preserve a natural measure, the so-called Liouville measure. Actually, we will see that very general dynamical systems do possess invariant measures.

Yet another fundamental reason to be interested in invariant measures is that their study may yield important information on the dynamical system's behavior that would be difficult to obtain otherwise. Poincaré's recurrence theorem, one of the first results we analyze in this book, is a great illustration of this: it asserts that, relative to any finite invariant measure, almost every orbit returns arbitrarily close to its initial state.

Brief historic survey

The word *ergodic* is a concatenation of two Greek words, $\epsilon\rho\gamma o\nu$ (*ergon*) = work and $o\delta o\sigma$ (*odos*) = way, and was introduced in the 19th century by the Austrian physicist L. Boltzmann. The systems that interested Boltzmann, J. C. Maxwell and J. C. Gibbs, the founders of the kinetic theory of gases, can be described by a Hamiltonian flow, associated with a differential equation of the form

$$\left(\frac{dq_1}{dt},\dots,\frac{dq_n}{dt},\frac{dp_1}{dt},\dots,\frac{dp_n}{dt}\right)=\left(\frac{\partial H}{\partial p_1},\dots,\frac{\partial H}{\partial p_n},-\frac{\partial H}{\partial q_1},\dots,-\frac{\partial H}{\partial q_n}\right).$$

Boltzmann believed that typical orbits of such a flow fill in the whole energy surface $H^{-1}(c)$ that contains them. Starting from this *ergodic hypothesis*, he deduced that the (time) averages of observable quantities along typical orbits coincide with the (space) averages of such quantities on the energy surface, which was crucial for his formulation of the kinetic theory of gases.

In fact, the way it was formulated originally by Boltzmann, this hypothesis is clearly false. So, the denomination *ergodic hypothesis* was gradually displaced to what would have been a consequence, namely, the claim that time averages and space averages coincide. Systems for which this is true were called *ergodic*. And it is fair to say that a great part of the progress experienced by ergodic theory in the 20th century was motivated by the quest to understand whether most Hamiltonian systems, especially those that appear in connection with the kinetic theory of gases, are ergodic or not.

The foundations were set in the 1930's, when J. von Neumann and G. D. Birkhoff proved that time averages are indeed well defined for almost every orbit. However, in the mid 1950's, the great Russian mathematician A. N. Kolmogorov observed that many Hamiltonian systems are actually *not* ergodic. This spectacular discovery was much expanded by V. Arnold and J. Moser, in what came to be called KAM (Kolmogorov–Arnold–Moser) theory.

On the other hand, still in the 1930's, E. Hopf had given the first important examples of Hamiltonian systems that *are* ergodic, namely, the geodesic flows on surfaces with negative curvature. His result was generalized to geodesic flows on manifolds of any dimension by D. Anosov, in the 1960's. In fact, Anosov proved ergodicity for a much more general class of systems, both with discrete time and in continuous time, which are now called Anosov systems.

An even broader class, called uniformly hyperbolic systems, was introduced by S. Smale and became a major focus for the theory of dynamical systems through the last half a century or so. In the 1970's, Ya. Sinai developed the theory of Gibbs measures for Anosov systems, conservative or dissipative, which D. Ruelle and R. Bowen rapidly extended to uniformly hyperbolic systems. This certainly ranks among the greatest achievements of smooth ergodic theory.

Two other major contributions must also be mentioned in this brief survey. One is the introduction of the notion of *entropy*, by Kolmogorov and Sinai, near the end of the 1950's. Another is the proof that the entropy is a complete invariant for Bernoulli shifts (two Bernoulli shifts are equivalent if and only if they have the same entropy), by D. Ornstein, some ten years later.

By then, the theory of non-uniformly hyperbolic systems was being initiated by V. I. Oseledets, Ya. Pesin and others. But that would take us beyond the scope of the present book.

How this book came to be

This book grew from lecture notes we wrote for the participants of mini-courses we taught at the Department of Mathematics of the Universidade Federal de Pernambuco (Recife, Brazil), in January 2003, and at the meeting *Novos Talentos em Matemática* held by Fundação Calouste Gulbenkian (Lisbon, Portugal), in September 2004.

In both cases, most of the audience consisted of young undergraduates with little previous contact with measure theory, let alone ergodic theory. Thus, it was necessary to provide very friendly material that allowed such students to follow the main ideas to be presented. Still at that stage, our text was used by other colleagues, such as Vanderlei Horita (São José do Rio Preto, Brazil), for teaching mini-courses to audiences with a similar profile.

As the text evolved, we have tried to preserve this elementary character of the early chapters, especially Chapters 1 and 2, so that they can used independently of the rest of the book, with as few prerequisites as possible.

Starting from the mini-course we gave at the 2005 *Colóquio Brasileiro de Matemática* (IMPA, Rio de Janeiro), this project acquired a broader purpose. Gradually, we evolved towards trying to present in a consistent textbook format the material that, in our view, constitutes the core of ergodic theory. Inspired by our own research experience in this area, we endeavored to assemble in a unified presentation the ideas and facts upon which is built the remarkable development this field experienced over the last decades.

A main concern was to try and keep the text as self-contained as possible. Ergodic theory is based on several other mathematical disciplines, especially measure theory, topology and analysis. In the appendix, we have collected the main material from those disciplines that is used throughout the text. As a rule, proofs are omitted, since they can easily be found in many of the excellent references we provide. However, we do assume that the reader is familiar with the main tools of linear algebra, such as the canonical Jordan form.

Structure of the book

The main part of this book consists of 12 chapters, divided into sections and subsections, and one appendix, also divided into section and subsections. A list of exercises is given at the end of every section, appendix included. Statements (theorems, propositions, lemmas, corollaries, etc.), exercises and formulas are numbered by section and chapter: for instance, (2.3.7) is the seventh formula in the third section of the second chapter and Exercise A.5.1 is the first exercise in the fifth section of the appendix. Hints for selected exercises are given in special chapter after the appendix. At the end, we provide a list of references and an index.

Chapters 1 through 12 organized as follows:

- Chapters 1 through 4 constitute a kind of introductory cycle, in which we present the basic notions and facts in ergodic theory—invariance, recurrence and ergodicity—as well as some main examples. Chapter 3 introduces the fundamental results (ergodic theorems) upon which the whole theory is built.
- Chapter 4, where we introduce the key notion of ergodicity, is a turning point in our text. The next two chapters (Chapters 5 and 6) develop a couple of important related topics: decomposition of invariant measures into ergodic measures and systems admitting a unique, necessarily ergodic, invariant measure.
- Chapters 7 through 9 deal with very diverse subjects—loss of memory, the isomorphism problem and entropy—but they also form a coherent structure, built around the idea of considering increasingly chaotic systems: mixing, Lebesgue spectrum, Kolmogorov and Bernoulli systems.
- Chapter 9 is another turning point. As we introduce the fundamental concept of entropy, we take our time to present it to the reader from several different viewpoints. This is naturally articulated with the content of Chapter 10, where we develop the topological version of entropy, including an important generalization called pressure.
- In the two final chapters, 11 and 12, we focus on a specific class of dynamical systems, called expanding transformations, that allows us to exhibit a concrete (and spectacular!) application of many of the general ideas presented the text. This includes Ruelle's theorem and its applications, which we view as a natural climax of the book.

Appendices A.1 through A.2 cover several basic topics of measure and integration. Appendix A.3 deals with the special case of Borel measures in metric spaces. In Appendix A.4 we recall some basic facts from the theory of manifolds and smooth maps. Similarly, Appendices A.5 and A.6 cover some useful basic material about Banach spaces and Hilbert spaces. Finally, Appendix A.7 is devoted to the spectral theorem.

Examples and applications have a key part in any mathematical discipline and, perhaps, even more so in ergodic theory. For this reason, we devote special attention to presenting concrete situations that illustrate and put in perspective the general results. Such examples and constructions are introduced gradually, whenever the context seems better suited to highlight their relevance. They often return later in the text, to illustrate new fundamental concepts as we introduce them.

The exercises at the end of each section have a threefold purpose. There are routine exercises meant to help the reader become acquainted with the concepts and the results presented in the text. Also, we leave as exercises certain arguments and proofs that are not used in the sequel or belong to more elementary related areas, such as topology or measure theory. Finally, more sophisticated exercises test the reader's global understanding of the theory. For the reader's convenience, hints for selected exercises are given in a special chapter following the appendix.

How to use this book?

These comments are meant, primarily, for the reader who plans to use this book to teach a course. Appendices A.1 through A.7 provide quick references to background material. In principle, they are not meant to be presented in class.

The content of Chapters 1 through 12 is suitable for a one-year course, or a sequence of two one-semester courses. In either case, the reader should be able to cover most of the material, possibly reserving some topics for seminars given by the students. The following sections are especially suited for that:

> Section 1.5, Section 2.5, Section 3.4, Section 4.4, Section 6.4, Section 7.3, Section 7.4, Section 8.3, Section 8.4, Section 8.5, Section 9.5, Section 9.7, Section 10.4, Section 10.5, Section 11.1, Section 11.3, Section 12.3 and Section 12.4.

In this format, Ruelle's theorem (Theorem 12.1) and its applications are a natural closure for the course.

In case only one semester is available, some selection of topics will be necessary. The authors' suggestion is to try and cover the following program:

Chapter 1: Sections 1.1, 1.2 and 1.3.
Chapter 2: Sections 2.1 and 2.2.
Chapter 3: Sections 3.1, 3.2 and 3.3.
Chapter 4: Sections 4.1, 4.2 and 4.3.
Chapter 5: Section 5.1 (mention Rokhlin's theorem).

Chapter 6: Sections 6.1, 6.2 and 6.3.
Chapter 7: Sections 7.1 and 7.2.
Chapter 8: Section 8.1 and 8.2 (mention Ornstein's theorem).
Chapter 9: Sections 9.1, 9.2, 9.3 and 9.4.
Chapter 10: Sections 10.1 and 10.2.
Chapter 11: Section 11.1.

In this format, the course could close either with the proof of the variational principle for the entropy (Theorem 10.1) or with the construction of absolutely continuous invariant measures for expanding maps on manifolds (Theorem 11.1.2).

We have designed the text in such a way as to make it feasible for the lecturer to focus on presenting the central ideas, leaving it to the student to study in detail many of the proofs and complementary results. Indeed, we devoted considerable effort to making the explanations as friendly as possible, detailing the arguments and including plenty of cross-references to previous related results as well to the definitions of the relevant notions.

In addition to the regular appearance of examples, we have often chosen to approach the same notion more than once, from different points of view, if that seemed useful for its in-depth understanding. The special chapter containing the hints for selected exercises is also part of that effort to encourage and facilitate the autonomous use of this book by the student.

Acknowledgments

The writing of this book extended for over a decade. During this period we benefitted from constructive criticism from several colleagues and students.

Many colleagues used different preliminary versions of the book to teach courses and shared their experiences with us. Besides Vanderlei Horita (São José do Rio Preto, Brazil), Nivaldo Muniz (São Luis, Brazil) and Meysam Nassiri (Teheran, Iran), we would like to thank Vítor Araújo (Salvador, Brazil) for an extended list of suggestions that influenced significantly the way the text evolved from then on. François Ledrappier (Paris, France) helped us with questions about substitution systems.

We also had the chance to test the material in a number of regular graduate courses at IMPA (Instituto de Matemática Pura e Aplicada) and at UFAL (Universidade Federal de Alagoas). Feedback from graduate students Adriana Sánchez, Aline Gomes Cerqueira, El Hadji Yaya Tall, Ermerson Araujo, Ignacio Atal, Rafael Lucena, Raphaël Cyna and Xiao-Chuan Liu allowed us to correct many of the weaknesses in earlier versions.

The first draft of Appendices A.1–A.2 was written by João Gouveia, Vítor Saraiva and Ricardo Andrade, who acted as assistants for the course in *Novos Talentos em Matemática 2004* mentioned previously. IMPA students Edileno de Almeida Santos, Felippe Soares Guimarães, Fernando Nera Lenarduzzi,

Ítalo Dowell Lira Melo, Marco Vinicius Bahi Aymone and Renan Henrique Finder wrote many of the hints for the exercises in Chapters 1 through 8 and the appendix.

The original Portuguese version of this book, *Fundamentos da Teoria Ergódica* [VO14], was published in 2014 by SBM (Sociedade Brasileira de Matemática). Feedback from colleagues who used that back to teach graduate courses in different places helped eliminate some of the remaining shortcomings. The extended list of remarks by Bernardo Lima (Belo Horizonte, Brazil) and his student Leonardo Guerini was particularly useful in this regard.

Several other changes and corrections were made in the course of the translation to English. We are grateful to the many colleagues and students who agreed to revise different parts of the translated text, especially Cristina Lizana, Elaís C. Malheiro, Fernando Nera Lenarduzzi, Jiagang Yang, Karina Marin, Lucas Backes, Maria João Resende, Mauricio Poletti, Paulo Varandas, Ricardo Turolla, Sina Türelli, Vanessa Ramos, Vítor Araújo and Xiao-Chuan Liu.

Marcelo Viana[1] and Krerley Oliveira[2]
Rio de Janeiro and Maceió, March 31, 2015

[1] IMPA, Estrada D. Castorina 110, 22460-320 Rio de Janeiro, Brasil. viana@impa.br.
[2] Departamento de Matemática, Universidade Federal de Alagoas, Campus A. C. Simões s/n, 57072-090 Maceió, Brasil. krerley@mat.ufal.br.