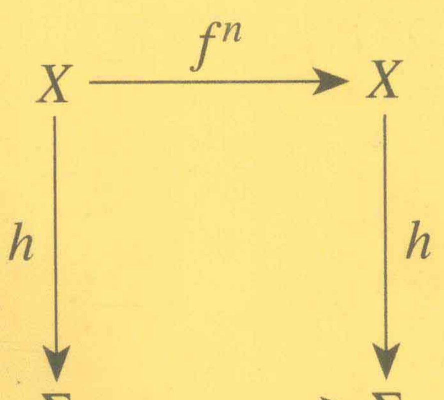
Lecture Notes in Mathematics 1513

L. S. Block W. A. Coppel

Dynamics in One Dimension



Springer-Verlag

L. S. Block W. A. Coppel

Dynamics in One Dimension

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

Authors

Louis Stuart Block
Department of Mathematics
University of Florida
Gainesville, Florida 32611, USA

William Andrew Coppel
Department of Theoretical Physics
Institute of Advanced Studies
Australian National University
GPO Box 4
Canberra 2601, Australia

Mathematics Subject Classification (1991): 26A18, 54H20, 58F08

ISBN 3-540-55309-6 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-55309-6 Springer-Verlag New York Berlin Heidelberg

Printed in Germany

Typesetting: Camera ready by author
Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
46/3140-543210 - Printed on acid-free paper

Editorial Policy

for the publication of monographs

In what follows all references to monographs, are applicable also to multiauthorship volumes such as seminar notes.

§ 1. Lecture Notes aim to report new developments - quickly, informally, and at a high level. Monograph manuscripts should be reasonably self-contained and rounded off. Thus they may, and often will, present not only results of the author but also related work by other people. Furthermore, the manuscripts should provide sufficient motivation, examples and applications. This clearly distinguishes Lecture Notes manuscripts from journal articles which normally are very concise. Articles intended for a journal but too long to be accepted by most journals, usually do not have this "lecture notes" character. For similar reasons it is unusual for Ph. D. theses to be accepted for the Lecture Notes series.

§ 2. Manuscripts or plans for Lecture Notes volumes should be submitted (preferably in duplicate) either to one of the series editors or to Springer- Verlag, Heidelberg . These proposals are then refereed. A final decision concerning publication can only be made on the basis of the complete manuscript, but a preliminary decision can often be based on partial information: a fairly detailed outline describing the planned contents of each chapter, and an indication of the estimated length, a bibliography, and one or two sample chapters - or a first draft of the manuscript. The editors will try to make the preliminary decision as definite as they can on the basis of the available information.

§ 3. Final manuscripts should preferably be in English. They should contain at least 100 pages of scientific text and should include

- a table of contents;
- an informative introduction, perhaps with some historical remarks: it should be accessible to a reader not particularly familiar with the topic treated;
- a subject index: as a rule this is genuinely helpful for the reader.

Further remarks and relevant addresses at the back of this book.

General Remarks

Lecture Notes are printed by photo-offset from the master-copy delivered in camera-ready form by the authors of monographs, resp. editors of proceedings volumes. For this purpose Springer-Verlag provides technical instructions for the preparation of manuscripts. Volume editors are requested to distribute these to all contributing authors of proceedings volumes. Some homogeneity in the presentation of the contributions in a multi-author volume is desirable.

Careful preparation of manuscripts will help keep production time short and ensure a satisfactory appearance of the finished book. The actual production of a Lecture Notes volume normally takes approximately 8 weeks.

Authors of monographs receive 50 free copies of their book. Editors of proceedings volumes similarly receive 50 copies of the book and are responsible for redistributing these to authors etc. at their discretion. No reprints of individual contributions can be supplied. No royalty is paid on Lecture Notes volumes.

Volume authors and editors are entitled to purchase further copies of their book for their personal use at a discount of 33.3 %, other Springer mathematics books at a discount of 20 % directly from Springer-Verlag. Authors contributing to proceedings volumes may purchase the volume in which their article appears at a discount of 20 %.

Commitment to publish is made by letter of intent rather than by signing a formal contract. Springer-Verlag secures the copyright for each volume.

Addresses:

Professor A. Dold
Mathematisches Institut, Universität Heidelberg
Im Neuenheimer Feld 288
W-6900 Heidelberg
Federal Republic of Germany

Professor B. Eckmann
Mathematik, ETH-Zentrum
CH-8092 Zürich, Switzerland

Professor F. Takens
Mathematisch Instituut
Rijksuniversiteit Groningen
Postbus 800
NL-9700 AV Groningen
The Netherlands

Springer-Verlag, Mathematics Editorial
Tiergartenstr. 17
W-6900 Heidelberg
Federal Republic of Germany
Tel.: *49 (6221) 487-410

Lecture Notes in Mathematics 1513

Preface

There has recently been an explosion of interest in one-dimensional dynamics. The extremely complicated — and yet orderly — behaviour exhibited by the logistic map, and by unimodal maps in general, has attracted particular attention. The ease with which such maps can be explored with a personal computer, or even with a pocket calculator, has certainly been a contributing factor. The unimodal case is extensively studied in the book of Collet and Eckmann [49], for example.

It is not so widely known that a substantial theory has by now been built up for *arbitrary* continuous maps of an interval. It is quite remarkable how many strong, general properties can be established, considering that such maps may be either real-analytic or nowhere differentiable. The purpose of the present book is to give a clear, connected account of this subject. Thus it updates and extends the survey article of Nitecki [96]. The two books [112], [113] by Šarkovskii and his collaborators contain material on the same subject. However, they are at present available only in Russian and in general omit proofs. Here complete proofs are given. In many cases these have previously been difficult of access, and in some cases no complete proof has hitherto appeared in print.

Our standpoint is topological. We do not discuss questions of a measure-theoretical nature or connections with ergodic theory. This is not to imply that such matters are without interest, merely that they are outside our scope. [A forthcoming book by de Melo and van Strien discusses these matters, and also the theory of smooth maps.] The material here could indeed form the basis for a course in topological dynamics, with many of the general concepts of that subject appearing in a concrete situation and with much greater effect.

Several of the results included here were first established for piecewise monotone maps. There exist also other results which are valid for piecewise monotone maps, but which do not hold for arbitrary continuous maps. Although we include some results of this nature, we do not attempt to give a full account of the theory of piecewise monotone maps.

The final chapter of the book deals with extensions to maps of a circle of the preceding results for maps of an interval. In contrast to the earlier chapters, the results here are merely stated, with references to the literature for the proofs. [Complete proofs are given in a forthcoming book by Alsedà, Llibre and Misiurewicz, which also discusses the material in our

Chapters 1,7 and 8.] We do not discuss at all some results which have been established for other one-dimensional structures. The pre-eminent importance of the interval and the circle appears to us adequate justification for our title. The list of references at the end of the book, although extensive, has no pretence to completeness.

This book has its origin in a course of lectures which the older author gave at the Australian National University in 1984. The first four chapters are based on the xeroxed notes for that course. However, the older author acknowledges that without the assistance of the younger author the book could never have reached its present greatly expanded form. We accept responsibility equally for the final product.

Our manuscript was originally submitted as a whole volume for the series *Dynamics Reported.* After its submission responsibility for publication of this series passed from Wiley and Teubner to Springer-Verlag. The resulting changes in format would not have presented insurmountable difficulties if the authors had been experts with TEX or LATEX. Since we were not, we decided instead to produce a good camera-ready manuscript, following the instructions to authors provided by Springer-Verlag for its *Lecture Notes in Mathematics* series. We are extremely grateful to the Managing Editors of *Dynamics Reported*, Professors U. Kirchgraber and H.O. Walther, for the time and care they devoted to our manuscript, for obtaining valuable referees' reports, and finally for generously agreeing to its appearance in the *Lecture Notes in Mathematics* rather than in *Dynamics Reported.*

We thank Professor Xiong Jincheng for contributing some unpublished results (Propositions VI.53 and VI.54), and the referees for several useful suggestions. We take this opportunity to thank also the numerous typists who have assisted us over a period of eight years. W.A.C. is grateful to the University of Florida for support during a visit to Gainesville in 1987. L.S.B. would like to thank the Australian National University for its hospitality during visits in 1988 and 1990. These visits considerably accelerated progress on the book. L.S.B. also thanks the University of Göttingen for its hospitality during a visit in 1988, and Zbigniew Nitecki for helpful conversations during that visit. Finally he thanks Ethan Coven for many helpful conversations over the past few years.

We dedicate this book to our families, in gratitude for their support.

Louis Block
Andrew Coppel

Contents

Introduction

This book is primarily concerned with the asymptotic behaviour of sequences (x_n) defined iteratively by $x_{n+1} = f(x_n)$, where f is an arbitrary continuous map of an *interval* into itself. The sequence (x_n) is the *trajectory* of the initial point x_0 under the map f.

An important reason for studying this problem, in addition to its intrinsic interest, comes from higher-dimensional dynamics. The extremely complicated behaviour of some 3-dimensional flows, or 2-dimensional diffeomorphisms, is also observed in non-invertible 1-dimensional maps. We hope to gain a better understanding of this behaviour by studying 1-dimensional maps, since they are much more amenable to mathematical analysis. Many remarkable properties of such maps have been established in recent years.

If in the trajectory (x_n) we have $x_p = x_0$ for some $p > 0$, then $x_{n+p} = x_n$ for every $n > 0$. Thus the trajectory (x_n) is *periodic*. It is said to have *period* p if p is the least positive integer such that $x_p = x_0$. It turns out that if a continuous map f has a periodic trajectory with a given period p, then it necessarily has periodic trajectories with certain other periods. A complete description of all possible sets of periods, for the periodic trajectories of a continuous map of an interval, is given by a theorem of Šarkovskii, which is stated and proved in Chapter I. An interesting feature of this proof is the use of directed graphs.

In Chapter II we begin the study of nonperiodic trajectories. A simple example is a trajectory (x_n) with $x_n = x_2$ for all $n > 2$ and either $x_2 < x_0 < x_1$ or $x_1 < x_0 < x_2$. A map possessing such a trajectory is said to be *turbulent*. It turns out that all trajectories of non-turbulent maps are subject to rather stringent restrictions, whereas turbulent maps possess some trajectories which behave wildly. This wild behaviour is established by using the *shift map* of a symbol space. Thus, even though we are primarily interested in maps of an interval, we are naturally led to consider maps of other spaces. In Chapter II we also study the effects of slightly perturbing the given map f, and we give some results which hold for continuously differentiable or piecewise monotone maps, but not for all continuous maps.

The notions of stable and unstable manifold, of a periodic point, are important in the theory of smooth diffeomorphisms. For continuous maps of an interval, the stable set – or basin of attraction – of a periodic point may not be a manifold or have nice properties. However, as we show in Chapter III, the *unstable manifold* exists and is a well-behaved object. Moreover, one also has *left* and *right* unstable manifolds.

In Chapter III we also study *homoclinic points*. This term was first used by Poincaré, for diffeomorphisms, to describe a point belonging to both the stable and unstable manifolds of a periodic point. In our situation we demand instead that the point hit the periodic point after finitely many iterations, in addition to belonging to its unstable manifold. We show that there is a close relationship between turbulence and the existence of homoclinic points.

Periodicity represents the most precise type of repetitive behaviour. Several other types are studied in *topological dynamics*. In Chapter IV we discuss ordinary *recurrence* (or 'Poisson stability') and *nonwanderingness*. In Chapter V we consider *strong recurrence* (often also called 'recurrence'), *regular recurrence*, and *chain recurrence*. Chain recurrence is the weakest, and the most recently introduced, of these types of repetitive behaviour. Our treatment of it has some novelty, since we adopt a purely topological definition instead of the usual metric one.

Many of the results of Chapters IV and V are valid for continuous maps of any compact metric space. However, there are also results which are specific to maps of an interval. We mention, in particular, a remarkable characterization of *ω-limit points* due to Šarkovskii. These results give a strength to the theory for an interval which is lacking in the general case.

We define a map of an interval into itself to be *chaotic* if some iterate of the map is turbulent or, equivalently, if there exists a periodic point whose period is not a power of 2. It is shown in Chapter VI that there is a marked distinction between the behaviour of chaotic and non-chaotic maps. Exaggerated claims about a new theory of chaos have been appearing in the popular press. In fact there is no generally accepted definition of chaos. It is our view that any definition for more general spaces should agree with ours in the case of an interval. This requirement is not satisfied by some of the definitions used in the literature. The definition given above is strictly 1-dimensional. However, we show that a map is chaotic if and only if some iterate has the shift map as a *factor*, and we propose this as a general definition. Other definitions which meet our requirement are certainly possible, notably that some iterate is *topologically mixing* (as shown in Chapter VI) or that the map has positive *topological entropy* (as shown later in Chapter VIII), but they do not really call for the use of a new word. Ultimately it will probably be necessary to distinguish between different types of chaotic behaviour, in the same way as for recurrence.

To characterize a periodic trajectory we need to know not only its period but also its *type*, i.e. the way in which its points are ordered on the real line. It may be asked if Šarkovskii's theorem on periods can be strengthened to take account of types. That is, if a map has a periodic trajectory of a given type, does it necessarily have periodic trajectories of certain other types? In a sense this question is completely answered by a theorem of Baldwin, which says that a periodic trajectory of type P *forces* a periodic trajectory of type Q if and only if the *linearization* of P has a trajectory of type Q. We prove Baldwin's theorem in Chapter VII, but we do not investigate in detail the rather complicated partial ordering of types which forcing induces.

A periodic trajectory is said to be *primary* if it forces no periodic trajectory with the same period. In Chapter VII we also characterize completely the primary trajectories, and we prove that a map is chaotic if and only if it has a periodic trajectory which is not primary.

Chapter VIII is devoted to the important concept of topological entropy. After establishing the main results which hold for any compact topological space, we devote our attention to results which hold for a compact interval. The most profound of these is a theorem of Misiurewicz, one of whose consequences is the result, already implied, that a map is chaotic if and only if it has positive topological entropy.

Finally, in Chapter IX we summarize, with references only for the proofs, extensions to maps of a *circle* of the foregoing results for maps of an interval. In the literature some results have also been given for 1-dimensional branched manifolds, and in particular for '*Y*', but these lie outside our scope. [See, for example, L. Alsedà, J. Llibre and M. Misiurewicz, *Trans. Amer. Math. Soc.*, **313** (1989), 475-538, and a series of papers by A.M. Blokh in *Teor. Funktsii Funktsional. Anal. i Prilozhen.*]

An introduction, such as this, frequently concludes with some remarks on prerequisites. A most attractive feature of our subject is that the only knowledge demanded of the reader would be contained in a first course on real analysis. For the reader possessing this knowledge we present a variety of interesting and nontrivial results which were unknown thirty years ago! We hope that some readers may be stimulated to make additional contributions of their own, even if it means that our book will become outdated.

I
Periodic Orbits

1 ŠARKOVSKII'S THEOREM

By an *interval* we will always mean, except in Chapter VIII, a connected subset of the real line which contains more than one point. Thus an interval may be open, half-open or closed, but not degenerate, and an endpoint of an interval need not belong to the interval. However, the phrase 'nondegenerate interval' will sometimes be used for emphasis. We will denote by $\langle a,b \rangle$ the closed interval with endpoints a and b, when we do not know (or care) whether $a < b$ or $a > b$.

Let $f: I \to I$ be a continuous map of the interval I into itself. Having performed the map f once we can perform it again, and again, and again. That is, we consider the *iterates* f^n defined inductively by

$$f^1 = f, \quad f^{n+1} = f \circ f^n \qquad (n \geq 1).$$

We also take f^0 to be the identity map, defined by $f^0(x) = x$ for every $x \in I$. Evidently f^n is also a continuous map of I into itself. We are interested in the behaviour of the *trajectory* of x, i.e. the sequence $f^n(x)$ $(n \geq 0)$, for arbitrary $x \in I$. It is convenient to make a distinction between the trajectory of x and the *orbit* of x, which is the set of points $\{f^n(x) : n \geq 0\}$.

There is a very simple graphical procedure for following trajectories. In the (x,y) – plane draw the curve $y = f(x)$ and the straight line $y = x$. To obtain the trajectory with initial point x_0 we go vertically to $y = f(x)$, then horizontally to $y = x$. This gives $x_1 = f(x_0)$, and the process is repeated *ad infinitum* (see Figure 1).

A point $c \in I$ is said to be a *fixed point* of f if $f(c) = c$. Thus the fixed points are given by the intersections of the curve $y = f(x)$ and the straight line $y = x$. If the interval I is compact it necessarily contains at least one fixed point. For if $I = [a,b]$ we have

$$f(a) - a \geq 0 \geq f(b) - b,$$

and so the assertion follows from the intermediate value theorem for continuous functions.

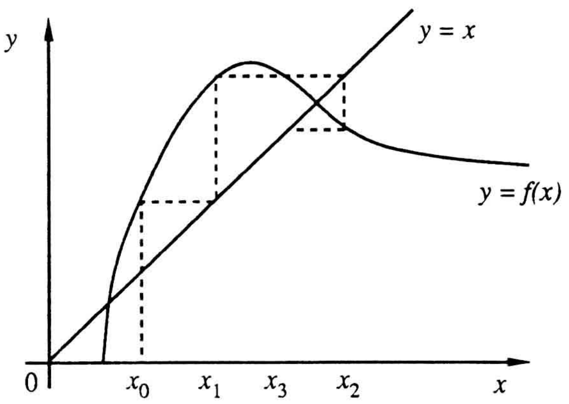


Fig. 1

A point $c \in I$ is said to be a *periodic point* of f with *period* m if $f^m(c) = c, f^k(c) \neq c$ for $1 \leq k < m$. The orbit of c then consists of the m distinct points $c, f(c), ..., f^{m-1}(c)$ and the trajectory of c consists of the same points repeated periodically. By abuse of language the orbit of c will also be said to be periodic. A fixed point is a periodic point of period 1.

If $f: I \to \mathbb{R}$ is a continuous map of an interval into the real line, then all or some of the iterates may be defined on a subinterval of I and we may still talk about periodic points. Throughout this chapter, unless otherwise stated, f will denote an arbitrary continuous map of an arbitrary interval I into the real line.

We are going to study first the periodic orbits of f. Our main objective will be the proof of the following striking theorem, due to Šarkovskii [102].

THEOREM 1 *Let the positive integers be totally ordered in the following way:*

$$3 \prec 5 \prec 7 \prec 9 \prec ... \prec 2.3 \prec 2.5 \prec ... \prec 2^2.3 \prec 2^2.5 \prec ..... \prec 2^3 \prec 2^2 \prec 2 \prec 1.$$

If f has a periodic orbit of period n and if $n \prec m$, then f also has a periodic orbit of period m.

However, on the way we will derive a number of results of independent interest.

LEMMA 2 *If J is a compact subinterval such that $J \subseteq f(J)$, then f has a fixed point in J.*

Proof If $J = [a,b]$ then for some $c,d \in J$ we have $f(c) = a, f(d) = b$. Thus $f(c) \leq c, f(d) \geq d$, and the result follows again from the intermediate value theorem. □

LEMMA 3 *If J, K are compact subintervals such that $K \subseteq f(J)$, then there is a compact subinterval $L \subseteq J$ such that $f(L) = K$.*

Proof Let $K = [a,b]$ and let c be the greatest point in J for which $f(c) = a$. If $f(x) = b$ for some $x \in J$ with $x > c$, let d be the least. Then we can take $L = [c,d]$. Otherwise $f(x) = b$ for some $x \in J$ with $x < c$. Let c' be the greatest and let $d' \leq c$ be the least $x \in J$ with $x > c'$ for which $f(x) = a$. Then we can take $L = [c',d']$. □

LEMMA 4 *If J_0, J_1, ..., J_m are compact subintervals such that $J_k \subseteq f(J_{k-1})$ $(1 \leq k \leq m)$, then there is a compact subinterval $L \subseteq J_0$ such that $f^m(L) = J_m$ and $f^k(L) \subseteq J_k$ $(1 \leq k < m)$.*

If also $J_0 \subseteq J_m$, then there exists a point y such that $f^m(y) = y$ and $f^k(y) \in J_k$ $(0 \leq k < m)$.

Proof The first assertion holds for $m = 1$, by Lemma 3. We assume that $m > 1$ and that it holds for all smaller values of m. Then we can choose $L' \subseteq J_1$ so that $f^{m-1}(L') = J_m$ and $f^k(L') \subseteq J_{k+1}$ $(1 \leq k < m-1)$. We now choose $L \subseteq J_0$ so that $f(L) = L'$.

The second assertion follows from the first, by Lemma 2. □

As a first application of these ideas we prove

PROPOSITION 5 *Between any two points of a periodic orbit of period $n > 1$ there is a point of a periodic orbit of period less than n.*

Proof Let $a < b$ be two adjacent points of the orbit of period n. Since there is one more point of the orbit to the left of b than to the left of a we must have $f^m(a) > a, f^m(b) < b$ for some m such that $1 \leq m < n$. It follows at once that $f^m(c) = c$ for some c such that $a < c < b$, assuming that f^m is defined throughout $[a,b]$. However, the same conclusion can be reached without this assumption. For if $J_k = \langle f^k(a), f^k(b)\rangle$ is the closed interval with endpoints $f^k(a)$ and $f^k(b)$ then $J_k \subseteq f(J_{k-1})$ $(1 \leq k \leq m)$. But $J_0 \subseteq J_m$, since $f^m(a) \geq b, f^m(b) \leq a$. The result now follows from Lemma 4. □

The method of argument used here can be refined. Suppose again that f has a periodic orbit of period $n > 1$. Let $x_1 <...< x_n$ be the distinct points of this orbit and set $I_j = [x_j, x_{j+1}]$ $(1 \leq j < n)$. With the periodic orbit we associate a *directed graph*, or *digraph*, in the following way. The vertices of the directed graph are the subintervals I_1, ..., I_{n-1} and there is an arc $I_j \rightarrow I_k$ if I_k is contained in the closed interval $\langle f(x_j), f(x_{j+1})\rangle$ with endpoints $f(x_j)$ and $f(x_{j+1})$.