# **Introduction**

#### **0.1 About this book**

This book presents the theory of flows, that is, continuous-time dynamical systems from the topological, smooth, and measurable points of view, with an emphasis on the theory of (uniformly) hyperbolic dynamics. It includes both an introduction and an exposition of recent developments in uniformly hyperbolic dynamics, and it can be used as both a textbook and a reference for students and researchers.

Books on dynamics tend to focus on discrete time, largely leaving it to the reader (or unaddressed) to transfer those insights to flows, where the origins of the theory actually lie.<sup>[1](#page-0-0)</sup> It is thus often implicit that "things work analogously for flows," or that "this is different for flows," and aside from geodesic flows, many theorems about flows have had little visibility beyond the research literature. Although much about flows can indeed be found in the research literature, doing so usually involves a combination of diligence and consultation with experts. We fill this gap in the expository literature by giving a deep "flows-first" presentation of dynamical systems and focusing on continuous-time systems, rather than treating these as afterthoughts or exceptions to methods and theory developed for discrete-time systems.

We point to a few additional features of interest and to some new results in this book:

- [Chapter](#page--1-0) 5 is to our knowledge unique in the literature for the extent to which it implements the Anosov–Katok–Bowen program of developing the dynamical features of hyperbolic sets for flows from shadowing alone.[2](#page-0-1)
- [Section](#page--1-1) 5.2 and [Chapter](#page--1-0) 8 provide an exceptional range of examples of (uniformly) hyperbolic flows. To our knowledge, these are more examples than have appeared in any one other source, in no small part because several of them are quite recent discoveries.
- [Chapter](#page--1-0) 5 may be the first account to provide a proper, natural definition of a

<span id="page-0-0"></span> $1$ This might in some part be because there are simpler examples available in discrete time, and longitudinal issues do not obscure the main effects of hyperbolicity—however, these longitudinal effects are quite interesting and indeed define the forefront of some research areas in dynamical systems.

<span id="page-0-1"></span><sup>&</sup>lt;sup>2</sup>Specifically, the Shadowing Lemma and the Shadowing Theorem, which include uniqueness, so in terms of customary usage one should say that shadowing and expansivity produce the insights in [Chapter 5.](#page--1-0)

(uniformly) hyperbolic flow [\(Definition](#page--1-2) 5.3.50) based on the equivalence of the three popular notions [\(Theorem](#page--1-3) 5.3.47 on page [305\)](#page--1-3). Although this equivalence is not new, it does not seem to be broadly known.

- [Section](#page--1-0) 6.6 gives a stronger theorem about existence of Markov sections than anywhere else in the literature.
- Other new results are that discreteness of centralizers is a *topological* fact [\(The](#page--1-4)[orem 9.1.3\)](#page--1-4) and our results on trivial centralizers in [Section 9.1.](#page--1-0)
- In addition to topological and smooth dynamics, we cover the ergodic theory of flows to a considerable extent, and this as well may be singular in the literature—while most of what we present can be found *somewhere* in the (often original research) literature, the ergodic theory of flows is not common textbook material.
- We also call attention to a proof (by Abdenur and Viana) of absolute continuity of the invariant foliations in the generality of partially hyperbolic dynamical systems [\(Section](#page--1-5) B.7). This exceeds what we need, but seemed like a desirable addition to the literature.
- Chapters [8](#page--1-0) and [9](#page--1-0) include a range of advanced topics mainly from the theory of Anosov flows (such as their topology and dynamics, as well as rigidity phenomena); some of these are from recent research and have not previously appeared in any expository literature.
- At the end of the main chapters there are a number of exercises.
- Thorough indexing facilitates the use of this book as a reference.

Chapters [8](#page--1-0) and [9](#page--1-0) are no less accessible than the introductory subjects, but here we take even more opportunities to augment the results we prove with complements whose proofs we do not include, or to outline proofs rather than giving full proofs. This is meant to provide a substantial introduction to these subjects with proofs.

As complements to this book the reader can choose from an abundance of books on ergodic theory and dynamical systems [\[221,](#page--1-6) [204,](#page--1-0) [289,](#page--1-7) [348,](#page--1-0) [317,](#page--1-8) [97,](#page--1-9) [98,](#page--1-10) [274,](#page--1-11) [329,](#page--1-12) [55,](#page--1-13) [70,](#page--1-14) [112,](#page--1-15) [32,](#page--1-16) [283,](#page--1-17) [251,](#page--1-0) [242,](#page--1-18) [263,](#page--1-19) [264,](#page--1-20) [315,](#page--1-21) [301,](#page--1-22) [334,](#page--1-23) [279,](#page--1-24) [268,](#page--1-22) [281,](#page--1-25) [257,](#page--1-26) [113,](#page--1-27) [114,](#page--1-28) [154,](#page--1-29) [176,](#page--1-30) [178,](#page--1-28) [133,](#page--1-31) [71\]](#page--1-31). Here we single out just a brisk introduction in a similar spirit that focuses on discrete time [\[174\]](#page--1-0), the rather larger books [\[213\]](#page--1-32) and [\[314\]](#page--1-33), and the more example-driven text [\[177\]](#page--1-27).

The text is divided into two parts. The first of these develops the general theory of flows (that is, not assuming hyperbolicity, but with a bias toward those aspects of the theory that are most pertinent to hyperbolic flows), both in the topological and measurable realm. The second part is about hyperbolicity and includes an introduction,

advanced material, and a panorama of current topics. The book is self-contained in the technical sense, that is, it includes definitions of all dynamics concepts with which we work, but without any pretense to being comprehensive with introductory material.

We intend this book to be useful for courses, directed study, self-study, and as a reference. For the latter, the broad and deep coverage combined with thorough indexing should be helpful. It has been written in a way that it can be adapted to a course (or independent study) in a number of different ways, depending on the purpose of the course. Starred chapters and sections are optional. They are not necessarily "harder," but the material is not needed for further sections except for an occasional result that can be used as a black box. Much of this material is hard to find in the literature except for original sources.

The core chapters are Chapters [1,](#page--1-0) [5,](#page--1-0) and [6.](#page--1-0) If one wants to emphasize ergodic properties of flows then one could include Chapters [3,](#page--1-0) [4,](#page--1-0) and [7,](#page--1-0) or at least portions of them. For a more topological or geometric course one would instead include [Chapter](#page--1-0) 2, and portions from Chapters [8](#page--1-0) or [9](#page--1-0) (several sections of the latter invoke some ergodic theory, however). A topics course, especially to an audience with some prior knowledge, could more extensively cover those last chapters. The core chapters include exercises.

The appendices contain background material on discrete-time dynamical systems, some of which is invoked on a few occasions in the main text. Those already familiar with it can omit it, refer to it as needed, or review it quickly. For those not familiar with the discrete-time theory, the appendices should provide sufficient background to understand either the material on ergodic theory in [Chapter](#page--1-0) 3 or the material on invariant foliations in [Chapter 6.](#page--1-0)

## **0.2 Continuous and discrete time**

To give our selection of flows versus discrete-time systems some context, we describe a few connections between these. Historically, dynamical systems came about in the form of flows, such as those that arise from differential equations that describe a mechanical system. Poincaré is widely regarded as the founder of the discipline of dynamical systems as we know it, and among the wealth of notions he created is that of a local section, known also as a Poincaré section. This is natural when using periodic orbits (trajectories) of a continuous-time dynamical system as anchors to study other motions in the system. Such a nearby motion will track the periodic motion for possibly considerable amounts of time, and it is often of less interest whether it lags or leads a little than how it moves closer to or further from the periodic orbit. To focus on these transverse phenomena Poincaré considered a small hypersurface

perpendicular to the periodic orbit on which he could track successive "hits" by a nearby motion. This defines a map on this disk, called the Poincaré (first) return map; see [Figure](#page-3-0) 0.2.1. This is an early way in which discrete-time dynamical systems arose.



<span id="page-3-0"></span>Figure 0.2.1. Poincaré section and map. [Reprinted from [\[213\]](#page--1-32) (© Cambridge University Press, all rights reserved) with permission.]

Coming from a different direction, billiard systems illustrate how a similar approach works both naturally and globally. A mathematical billiard system idealizes physical billiards by ignoring the spin and rolling of the balls: a point particle moves along straight lines and is reflected in the boundary with incoming angle equal to outgoing angle. This makes them more like air hockey or a description of light in a mirrored room, and tables of shapes other than rectangular are of considerable interest. These



Figure 9.2.1. A convex billiard Figure 0.2.2. Billiard. [Reprinted from [\[213\]](#page--1-32) (© Cambridge University Press, all rights reserved) with permission.]

are naturally continuous-time systems, but they come with natural discrete moments in time: the moments in which collisions occur. Indeed, all information about the (moderation in a material consistence ∂OOM) meet  $\frac{1}{2}$ 

evolution of such a system is contained in the locations and velocities of all balls at the moment of a collision, because this determines the motion until the next collision and the positions and velocities at that subsequent moment. Therefore, the dynamics can be described as a map on the "collision space" that sends each collision configuration to the next one. Once again, a discrete-time system describes the dynamics of a continuous-time system.[3](#page-4-0)

This latter process can be reversed. Given the discrete-time system, one more piece of information reconstructs the flow entirely: the "return time" from one collision to the next. We call this assembly of a map and a return-time function a suspension if the return time is constant [\(Definition](#page--1-34) 1.2.8), and a special flow or flow under a function otherwise [\(Definition 1.2.11\)](#page--1-35).

We digress to note that some discrete-time dynamical systems arise directly in scientific problems, such as the population biology of species with nonoverlapping generations (cicadas, for instance). While the preceding process could be used to embed this in continuous-time dynamics, this is neither helpful nor meaningful.

There are also aspects of dynamics in which pronounced differences between flows and discrete-time systems are manifested. On one hand, this occurs when "longitudinal" effects matter, that is, when time changes make a difference. In the case of a special flow this amounts to properties that are affected by the choice of "roof" or return-time function versus those that are not. For instance, the existence of a dense orbit is unaffected by the choice of roof function, but whether all periodic orbits are commensurate (their periods are various multiples of one positive number) clearly does depend on return times. Another notable feature of flows is that they permit surgery constructions to construct new flows. Accordingly, such a construction establishes that Anosov flows need not have a dense orbit [\(Section](#page--1-7) 8.3), but it is a long-open and exceedingly difficult problem to decide whether Anosov *diffeomorphisms* always have a dense orbit. In fact, it is not even known whether every Anosov diffeomorphism has a fixed point.

The theory of continuous-time dynamical systems does not directly reduce to that of discrete-time dynamical systems in the most obvious way: few diffeomorphisms arise as time-*t* maps of flows [\(Definition 1.1.1\)](#page--1-36) since (every time-*t* map of) every flow is isotopic to the identity.<sup>[4](#page-4-1)</sup> Also, time-*t* maps of flows have "roots" of all orders, being the *n*th iterate of the time-*t*/*n* map. But one might say that a full continuoustime theory yields a full discrete-time theory because every diffeomorphism can be represented as a Poincaré section for some flow via the suspension/special-flow construction—provided one has a comprehensive understanding of the dynamics of

<span id="page-4-1"></span><span id="page-4-0"></span><sup>&</sup>lt;sup>3</sup>This goes back to Birkhoff; see the discussion leading up to [Theorem 5.2.49.](#page--1-37)

<sup>&</sup>lt;sup>4</sup>One point of view from which flows produce a "sparse" set of maps of a given manifold is related to the mapping-class group. For a manifold M the *mapping-class group* is the set of isotopy-classes of homeomorphisms (or diffeomorphisms) of  $M$ . Flows are contained in the trivial equivalence class of the mapping-class group.

a section in terms of that of the flow. This does not work in reverse because that construction is not unique, and many flows generate a given diffeomorphism, with confounding "longitudinal" effects as above.

More to the point, for the study of hyperbolic flows [\(Chapter 5\)](#page--1-0) it may be useful to know all about hyperbolic maps, but that theory does not apply directly to time-1 maps of flows unless the periodic points for the flow are all hyperbolic equilibria. More specifically, the time-*t* map of a hyperbolic flow satisfies a weaker condition called partial hyperbolicity due to the flow direction, in which neither contraction nor expansion occur. Thus, this "*flows-first*" book complements the existing literature emphasizing discrete-time systems.

Once more, beyond the general theory, our emphasis is on uniformly hyperbolic dynamics. Neither partial nor nonuniform hyperbolicity are themselves subjects in this book. (The sole exception being the proof of absolute continuity of the invariant foliations for partially hyperbolic diffeomorphisms: while it is provided here to be applied to uniformly hyperbolic flows via time-1 maps, the proof covers partially hyperbolic diffeomorphisms in full generality.)

In short, discrete-time dynamics and continuous-time dynamics have closely related toolkits and close interactions, but the discrete-time focus of the existing literature leaves room for an explicit presentation of continuous-time dynamics.<sup>[5](#page-5-0)</sup>

### **0.3 Historical sketch**

We now outline some of the developments that brought about the theory of hyperbolic flows.[6](#page-5-1) There are several intertwined strands of the history of hyperbolic dynamics, including geodesic flows and statistical mechanics on one hand, and hyperbolic phenomena ultimately traceable to some application of dynamical systems. Geodesic flows were studied, for example, by Hadamard, Hedlund, Hopf (primarily either on surfaces or in the case of constant curvature) and Anosov–Sinai (negatively curved surfaces and higher-dimensional manifolds). Other hyperbolic phenomena appear in the work of Poincaré (homoclinic tangles in celestial mechanics [\[295\]](#page--1-38)), Perron (differential equations [\[285\]](#page--1-39)), Cartwright, Littlewood (relaxation oscillations in radio circuits [\[87,](#page--1-40) [88,](#page--1-41) [243\]](#page--1-42)), Levinson (the van der Pol equation [\[241\]](#page--1-43)) and Smale (horseshoes [\[338,](#page--1-44) [337\]](#page--1-45)), to name a few.

**0.3.a Homoclinic tangles and negative curvature.** The advent of complicated dynamics took place in the context of Newtonian mechanics, according to which

<span id="page-5-0"></span><sup>5</sup>To be clear, the *research* literature does not omit the continuous-time theory altogether; it is among *books* that this work occupies a unique place.

<span id="page-5-1"></span><sup>&</sup>lt;sup>6</sup>An expanded version can be found in [\[174\]](#page--1-0).

simple underlying rules governed the evolution of the world in clockwork fashion. The successes of classical and especially celestial mechanics in the 18th and 19th centuries were seemingly unlimited and Pierre Simon de Laplace felt justified in saying (in the opening passage he added to [\[231,](#page--1-46) p. 2]),

Nous devons donc envisager l'état présent de l'univers, comme l'effet de son état antérieur, et comme la cause de celui qui va suivre. Une intelligence qui pour un instant donné, connaîtrait toutes les forces dont la nature est animée, et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule les mouvemens des plus grands corps de l'univers et ceux du plus léger atome: rien ne serait incertain pour elle, et l'avenir comme le passé, serait présent à ses yeux.[7](#page-6-0)

The enthusiasm in this passage is understandable and its forceful description of (theoretical) determinism is a good anchor for an understanding of one of the basic aspects of dynamical systems. Moreover, the titanic life's work of Laplace in celestial mechanics earned him the right to make such bold pronouncements. Another bold pronouncement of his, that the solar system is stable, came under renewed scrutiny later in the 19th century, and Henri Poincaré was expected to win a competition to finally establish this fact. However, Poincaré came upon hyperbolic phenomena in revising his prize memoir [\[295\]](#page--1-38) on the three-body problem. He found that a phenomenon now called homoclinic tangles [\(Figure](#page--1-47) 6.5.1) (which he had initially overlooked) caused great difficulty and necessitated essentially a reversal of the main thrust of that memoir [\[34\]](#page--1-48). He perceived that there is a highly intricate web of invariant curves and that this situation produces dynamics of unprecedented complexity:

Que l'on cherche à se représenter la figure formée par ces deux courbes et leurs intersections en nombre infini dont chacune correspond à une solution doublement asymptotique, ces intersections forment une sorte de treillis, de tissu, de réseau à mailles infiniment serrées; chacune des deux courbes ne doit jamais se recouper elle-même, mais elle doit se replier sur elle-même d'une manière trés complexe pour venir recouper une infinité de fois toutes les mailles du réseau. On sera frappé de la complexité de cette figure, que je ne cherche même pas à tracer $8$ 

<span id="page-6-0"></span> $7$ We ought then to consider the present state of the universe as the effects of its previous state and as the cause of that which is to follow. An intelligence that, at a given instant, could comprehend all the forces by which nature is animated and the respective situation of the beings that make it up, if moreover it were vast enough to submit these data to analysis, would encompass in the same formula the movements of the greatest bodies of the universe and those of the lightest atoms. For such an intelligence nothing would be uncertain, and the future, like the past,



Figure 0.3.1. Homoclinic tangles. [Reprinted from [\[213\]](#page--1-32) (© Cambridge University Press, all rights regard with permission.]<br>reserved) with permission.] reserved) with permission.]

This is often viewed as the moment chaotic dynamics was first noticed. He concluded This is often viewed as the moment chaotic dynamics was first noticed. He concluded<br>that in all likelihood the prize problem could not be solved as posed: which was to that in all likelihood the prize problem could not be solved as posed; which was to<br>find satisfy anonsigns for the motions of the hodies in the solar system that converse find series expansions for the motions of the bodies in the solar system that converge his prize memoir [51] for the Papal A[ca](#page--1-49)demy of Sciences, he noted that and described<br>how this implies complicated dynamics [51] p. 1841 (Theorem 6.5.2) [h](#page--1-49)ow this implies complicated dynamics [51, p. 184] [\(Theorem 6.5.2\)](#page--1-35). uniformly for all time. Indeed, when Birkhoff picked up the study of this situation in

**0.3.b Geodesic flows.** A major class of mathematical examples motivating the devel-**0.3.b Geodesic flows.** A major class of mathematical examples motivating the development of hyperbolic dynamics is that of geodesic flows (that is, free-particle motion) of Riemannian manifolds of negative sectional curva of Riemannian manifolds of negative sectional curvature. Hadamard considered (non-<br>compact) surfaces in  $\mathbb{R}^3$  of negative survature [166] and found, with appear t delight that if the unbounded parts are "large" (do not pinch to arbitrarily small diameter<br>as you go outward along them) then at any point the initial directions of bounded<br>geodesics form a Cantor set. Since only countably many d geodesics form a Cantor set. Since only countably many directions give geodesics<br>that are periodic or asymptotic to a periodic one, this also proves the existence of that are periodic or asymptotic to a periodic one, this also proves the existence of more complicated bounded geodesics. Hadamard was fully aware of the connection to more complicated bounded geodesics. Hadamard was fully aware of the connection to Cantor's work and similar sets discovered by Poincaré, and he appreciated the relation Cantor's work and similar sets discovered by Poincaré, and he appreciated the relation<br>between the complicated dynamics in the two contexts. Hadamard also showed that between the complicated dynamics in the two contexts. Hadamard also showed that each homotopy class (except for the "waists" of cusps) contains a unique geodesic. each homotopy class (except for the "waists" of cusps) contains a unique geodesic. each homotopy class (except for the warsts of cusps) contains a unique geodesic.<br>would be open to its eyes. compact) surfaces in  $\mathbb{R}^3$  of negative curvature [\[166\]](#page--1-50) and found, with apparent delight, that if the unbounded parts are "large" (do not pinch to arbitrarily small diameter geodesics form a Cantor set. Since only countably many directions give geodesics Cantor's work and similar sets discovered by Poincaré, and he appreciated the relation

a full component, we thus have verified (1) of  $D$  of  $D$ 

<span id="page-7-0"></span>would be open to its eyes.

would be open to its eyes.<br><sup>8</sup>If one tries to imagine the figure formed by these two curve<br>corresponding to a doubly asymptotic solution, these intersections In the open to its eyes.<br><sup>8</sup> If one tries to imagine the figure formed by these two curves with an infinite number of intersections, each<br>responding to a doubly asymptotic solution, these intersections form a kind of trall  $\frac{8}{10}$  if one tries to imagine the figure<br>corresponding to a doubly asymptot<br>infinitely tight mesh: each of the two corresponding to a doubly asymptotic solution, these intersections form a kind of trellis, a fabric, a network of<br>infinitely tight mesh; each of the two curves must not cross itself but it must fold on itself in a complic infinitely tight mesh; each of the two curves must not cross itself but it must fold on itself in a complexited way to<br>intersect all of the meshes of the fabric infinitely many times. One will be struck by the complexity o contact the non-term of the method of the method in the method of  $\alpha$  is  $\alpha$  and  $\beta$ , or  $\alpha$  for  $\beta$ ,  $\alpha$  is complexity of this picture, corresponding to a doubly asymptotic solution, these intersections form a kind of trellis, a fabric, a network of infinitely tight mesh; each of the two curves must not cross itself but it must fold on itself in a complicated way to which I will not even attempt to draw



Figure 0.3.2. Negatively curved surface. [Reproduced from Hadamard [\[166\]](#page--1-50) (© 1898 Elsevier Masson SAS, all rights reserved) with permission.]

Duhem [\[123\]](#page--1-51) seized upon this to describe the dynamics of a geodesic flow in terms of what might now be called deterministic chaos—the system is completely determined (no randomness), but one would need infinite precision for long-term predictability.

Several authors trace the introduction of symbolic dynamics to the work of Hadamard on geodesic flows. Birkhoff is among them. Indeed, in his proof of the Birkhoff–Smale Theorem (see [Theorem](#page--1-35) 6.5.2) symbolic sequences appear (as well as a picture that resonates with [Figure](#page--1-47) 6.5.2). It appears, however, that only in 1944 did symbol spaces begin to be seen as dynamical systems, rather than as a coding device [\[99\]](#page--1-52).

**0.3.c Boltzmann's Fundamental Postulate.** Well before Poincaré's work, James Clerk Maxwell (1831–1879) and Ludwig Boltzmann (1844–1906) had aimed to give a rigorous formulation of the kinetic theory of gases and statistical mechanics. A central ingredient was Boltzmann's Fundamental Postulate, which says that the time and space (phase or ensemble) averages of an observable (a function on the phase space) agree. Apparently because of a misstatement by Maxwell,<sup>[9](#page-8-0)</sup> one often ascribes to him the so-called ergodic hypothesis:

*The trajectory of the point representing the state of the system in phase space passes through every point on the constant-energy hypersurface of the phase space.*

Poincaré and many physicists doubted its validity since no example satisfying it had been exhibited [\[296\]](#page--1-53). Accordingly, in 1912 Paul and Tatiana Ehrenfest [\[127\]](#page--1-54) proposed the alternative quasi-ergodic hypothesis:

<span id="page-8-0"></span><sup>&</sup>lt;sup>9</sup>"The system, if left to itself in its actual state of motion, will, sooner or later, pass through every phase."



<span id="page-9-0"></span>Figure 0.3.3. The pseudosphere. [Reprinted from [\[213\]](#page--1-32) (© Cambridge University Press, all rights reserved) with permission.]

# *The trajectory of the point representing the state of the system in phase space is dense on the constant energy hypersurface of the phase space.*

Indeed, within a year proofs (by Rosenthal and Plancherel) appeared that the ergodic hypothesis fails [\[291,](#page--1-55) [316\]](#page--1-0). (This is obvious today because a trajectory has measure 0 in an energy surface.) These difficulties led to the search for *any* mechanical systems with this second property. The motion of a single free particle (that is, the geodesic flow) in a negatively curved space (beginning with the pseudosphere, [Figure](#page-9-0) 0.3.3) emerged as the first and for a long time sole class of examples with this property. Within a decade, the understanding of the problem led to the pertinent contemporary notion, and this turned out to be probabilistic in nature.<sup>[10](#page-9-1)</sup> The 1931 Birkhoff Ergodic Theorem [\(Theorem 3.2.15\)](#page--1-56) ("time averages exist a.e.")<sup>[11](#page-9-2)</sup> laid the foundation for the definition of ergodicity now in use, which is that "any invariant set has zero measure or full measure."[12](#page-9-3) If this is the case, then time averages agree with space averages—Boltzmann's Fundamental Postulate. Furthermore, almost every orbit is dense in the support of the measure.

The 1930s saw a flurry of work in which Artin's 1924 work on the modular surface was duly extended to other manifolds of *constant* negative curvature. For constant curvature, finite volume, and finitely generated fundamental groups the

<span id="page-9-1"></span><sup>&</sup>lt;sup>10</sup>This serves to point out that the earlier quote by Laplace about determinism comes from his "Philosophical essay on probabilities," where he goes on to say that we often do not have sufficiently detailed initial data, and must hence resort to a probabilistic approach. The motion of a molecule of air was a prominent instance he mentioned in that context.

<span id="page-9-2"></span><sup>&</sup>lt;sup>11</sup>This was proved after the von Neumann Ergodic Theorem [\(Theorem 3.2.4\)](#page--1-57) but published earlier [\[358\]](#page--1-58)—and the true foundational paper of ergodic theory is much more likely [\[265\]](#page--1-59).

<span id="page-9-3"></span><sup>&</sup>lt;sup>12</sup>These two combine to give the Strong Law of Large Numbers.

geodesic flow was shown to be topologically transitive [\[225,](#page--1-60) [249\]](#page--1-61), topologically mixing [\[191\]](#page--1-62), ergodic [\[196\]](#page--1-63), and mixing [\[192,](#page--1-64) [197\]](#page--1-65). (In the case of an infinitely generated fundamental group the geodesic flow may be topologically mixing without being ergodic [\[327\]](#page--1-66).) If the curvature is allowed to vary between two negative constants then finite volume implies topological mixing [\[155\]](#page--1-53) (see also [\[158,](#page--1-0) p. 183]). But as Hedlund noted in an address delivered before the New York meeting of the American Mathematical Society on October 27, 1938,

Outstanding problems remain unsolved, a notable one being the problem of metric transitivity [ergodicity] of the geodesic flow on a closed analytic surface of *variable* negative curvature.

It so happens that Eberhard Hopf was just then working on this problem [\[197\]](#page--1-65). He considered compact surfaces of nonconstant (predominantly) negative curvature and was able to show ergodicity of the Liouville measure (phase volume).

From Hopf's work there was no progress in the direction of ergodicity of geodesic flows for almost 30 years. Hopf's argument had shown roughly that Birkhoff averages of a continuous function must be constant on almost every leaf of the horocycle foliation, and, since these foliations are  $C<sup>1</sup>$ , the averages are constant a.e. He realized that much of the argument was independent of the dimension of the manifold (indeed, he carried much of the work out in arbitrary dimension), but could not verify the C<sup>1</sup> condition in higher dimension. Dmitri Anosov [\[10\]](#page--1-67) axiomatized Hopf's instability, defining Anosov flows, and he showed that differentiability may indeed fail in higher dimension, but that the Hopf argument can still be used because the invariant laminations have an absolute continuity property [\[10,](#page--1-67) [12,](#page--1-68) [303,](#page--1-48) [23,](#page--1-69) [70,](#page--1-14) [31\]](#page--1-0). This extension is interesting because, despite the ergodicity paradigm central to statistical mechanics, Boltzmann's Fundamental Postulate, there was a dearth of examples of ergodic Hamiltonian systems. The quintessential model for the Fundamental Postulate, the gas of hard spheres, resisted sustained attempts to prove ergodicity for half a century [\[332,](#page--1-0) [330,](#page--1-70) [331\]](#page--1-25).<sup>[13](#page-10-0)</sup>

The Hopf argument remains a main method for establishing ergodicity of volume in hyperbolic dynamical systems without an algebraic structure (the alternative tool being the theory of equilibrium states; see [\[213,](#page--1-32) Theorem 20.4.1]).

**0.3.d Picking up from Poincaré.** Like Hadamard, several mathematicians had begun to pick up some of Poincaré's work during his lifetime; Birkhoff did so soon after Poincaré's death. He addressed issues that arose from the mathematical development of mechanics and celestial mechanics such as Poincaré's Last Geometric Theorem and the complex dynamics necessitated by homoclinic tangles [\[49,](#page--1-71) Section 9].

<span id="page-10-0"></span><sup>&</sup>lt;sup>13</sup>Half a century because Sinai convinced physicists that he had solved this problem in 1963 [\[232\]](#page--1-24).

He was also important in the development of ergodic theory,<sup>[14](#page-11-0)</sup> notably by proving the Pointwise Ergodic Theorem [\(Theorem 3.2.15\)](#page--1-56).

The work of Cartwright and Littlewood during World War II on relaxation oscillations in radar circuits [\[88,](#page--1-41) [87,](#page--1-40) [243\]](#page--1-42) consciously built on Poincaré's work. Further study of the van der Pol equation by Levinson [\[241\]](#page--1-43) contained the first example of a structurally stable diffeomorphism with infinitely many periodic points. Structural stability had originated in 1937 with Andronov and Pontryagin [\[9\]](#page--1-72) (necessary and sufficient conditions on singularities and periodic orbits for structural stability of vector fields on a disk) but began to flourish only 20 years later—thanks in no small part to Pontryagin's favorite student, Anosov. Inspired by Peixoto's work, which generalized [\[9\]](#page--1-72) to any orientable closed surface [\[284\]](#page--1-73), Smale had been after a program of studying diffeomorphisms with a view to classification [\[339\]](#page--1-74), and he proved that Morse–Smale systems (finitely many periodic points with stable and unstable sets in general position) are structurally stable. The Cartwright–Littlewood example was brought to his attention by Levinson just as he conjectured that Morse–Smale systems are the only structurally stable ones [\[336\]](#page--1-75). He eventually extracted from Levinson's work the horseshoe [\[338,](#page--1-44) [337\]](#page--1-45). Independently, Thom (unpublished) studied hyperbolic toral automorphisms [\(Example 1.5.26\)](#page--1-76) and their structural stability.

Smale in turn was in contact with the Russian school, where Anosov systems (then C- or U-systems) had been shown to be structurally stable, and their ergodic properties were studied by way of further development of the study of geodesic flows in negative curvature.

This book focuses on uniformly hyperbolic flows, and even in this realm there are plenty of new developments. [Section](#page--1-1) 5.2 gives instances of uniformly hyperbolic flows of which several are quite new, and [Chapter](#page--1-0) 8 includes various further constructions of such (notably in Sections [8.2](#page--1-29) and [8.3\)](#page--1-7). Our presentation of these includes results in a range of directions that still await publication.

The initial development of the theory of hyperbolic systems in the 1960s was followed by the founding of the theory of *nonuniformly* hyperbolic dynamical systems in the 1970s, mostly by Pesin [\[273,](#page--1-7) [286,](#page--1-71) [32\]](#page--1-16) (during which time the hyperbolic theory continued its development). One of the high points in the development of smooth dynamics is the proof by Robbin, Robinson, Mañé, and Hayashi [\[189\]](#page--1-77) that structural stability indeed characterizes hyperbolic dynamical systems. For diffeomorphisms this was achieved in the 1980s, for flows in the 1990s. Starting in the 1980s the field of geometric and smooth rigidity came into being and is flourishing now [\(Chapter 9\)](#page--1-0). At the same time topological and stochastic properties of attractors began to be better understood with techniques that nowadays blend ideas from hyperbolic and 1-dimensional dynamics. Meanwhile, the theory of *partially* hyperbolic dynamical

<span id="page-11-0"></span><sup>14</sup>The Poincaré Recurrence Theorem [\(Theorem 3.2.1\)](#page--1-78) is proved in Poincaré's prize memoir [\[295\]](#page--1-38).

systems, which goes back to seminal works of Brin and Pesin in the 1970s, has seen explosive development since the last years of the 20th century [\[288\]](#page--1-79), which in turn has entailed renewed interest in the methods of uniformly hyperbolic dynamical systems and their possible extensions to this new realm.

Of course, insights into complicated dynamics have penetrated well beyond pure mathematics. In the sciences, these ideas have fundamentally changed the appreciation of nonlinear behavior and that complex data may arise from simple models; they have also provided terminology for describing complexity [\[152\]](#page--1-80). Celestial mechanics is the realm where applications have most clearly gone beyond the descriptive; since the 1980s the design of trajectories for space probes has irreversibly moved beyond perturbing the 2-body problem in ways that make entirely new mission designs feasible and economical in astonishing ways [\[38\]](#page--1-81). This can also be said to have added to the very foundation of how evidence is used to build science [\[351\]](#page--1-82).