Christian Bonatti, Lorenzo J. Díaz,
Marcelo Viana

Dynamics Beyond Uniform Hyperbolicity: A Global Geometric and Probabilistic Perspective

May 20, 2004

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Preface

What is Dynamics about ?

In broad terms, the goal of Dynamics is to describe the long term evolution of systems for which an "infinitesimal" evolution rule is known. Examples and applications arise from all branches of science and technology, like physics, chemistry, economics, ecology, communications, biology, computer science, or meteorology, to mention just a few.

These systems have in common the fact that each possible state may be described by a finite (or infinite) number of observable quantities, like position, velocity, temperature, concentration, population density, and the like. Thus, the space of states (*phase space*) is a subset M of an Euclidean space $\mathbb{R}^m$. Usually, there are some constraints between these quantities: for instance, for ideal gases pressure times volume must be proportional to temperature. Then the space M is often a manifold, an n-dimensional surface for some $n < m$.

For *continuous time* systems, the evolution rule may be a differential equation: to each state $x \in M$ one associates the speed and direction in which the system is going to evolve from that state. This corresponds to a vector field $X(x)$ in the phase space. Assuming the vector field is sufficiently regular, for instance continuously differentiable, there exists a unique curve tangent to X at every point and passing through x: we call it the *orbit of* x.

Even when the real phenomenon is supposed to evolve in continuous time, it may be convenient to consider a *discrete time* model, for instance, if observations of the system take place at fixed intervals of time only. In this case the evolution rule is a transformation $f : M \to M$, assigning to the present state $x \in M$ the one $f(x)$ the system will be in after one unit of time. Then the *orbit of* x is the sequence x_n obtained by iteration of the transformation: $x_{n+1} = f(x_n)$ with $x_0 = x$.

In both cases, one main problem is *to describe the behavior as time goes to infinity for the majority of orbits*, for instance, for a full probability set of initial states. Another problem, equally important, is *to understand whether that limit behavior is stable under small changes of the evolution law*, that is, whether it remains essentially the same if the vector field X or the transformation f are slightly modified. It is easy to see why this is such a crucial question, both conceptually and for the practical applications: mathematical models are always simplifications of the real system (a model of a chemical reaction, say, taking into account the whole universe would be obviously unpractical…) and, in the absence of stability, conclusions drawn from the model might be specific to it and not have much to do with the actual phenomenon.

It is tempting to try to address these problems by "solving" the dynamical system, that is, by looking for analytic expressions for the trajectories, and indeed that was the prevailing point of view in differential equations until little more than a century ago. However, that turns out to be impossible in most cases, both theoretically and in practice. Moreover, even when such an analytic expressions can be found, it is usually difficult to deduce from them useful conclusions about the global dynamics.

Then, by the end of the 19th century, Poincaré proposed to bring in methods from other disciplines, such as topology or ergodic theory, *to find qualitative information on the dynamics without actually finding the solutions.* A beautiful example, among many others, is the Poincaré-Birkhoff theorem stating that an area preserving homeomorphism of the annulus which rotates the two boundary circles in opposite directions must have some fixed point. This proposal, which was already present in Poincaré's early works and attained full maturity in his revolutionary contribution to Celestial Mechanics, is usually considered to mark the birth of Dynamics as a mathematical discipline.

Hyperbolicity and stability.

This direction was then pursued by Birkhoff in the thirties. In particular, he was much interested in the phenomenon of transverse *homoclinic points*, that is, points where the stable manifold and the unstable manifold of the same fixed or periodic saddle point intersect transversely. This phenomenon had been discovered in the context of the N-body problem by Poincaré, who immediately recognized it as a major source of dynamical complexity. Birkhoff made this intuition much more precise by proving that any transverse homoclinic orbit is accumulated by periodic points. A definitive understanding of this phenomenon unfolded at the beginning of the sixties, when Smale introduced the *horseshoe*, a simple geometric model whose dynamics can be understood rather completely, and whose presence in the system is equivalent to the existence of transverse homoclinic points.

The horseshoe, and other robust models containing infinitely many periodic orbits, such as Thom's cat map (hyperbolic toral automorphism), were unified by Smale's notion of uniformly *hyperbolic set*: a subset of the phase space invariant under the dynamical system and such that the tangent space at each point splits into two complementary subspaces that are uniformly contracted under, respectively, forward and backward iterations. Then Smale also introduced the notion of *uniformly hyperbolic dynamical system* (Axiom A) which essentially means that the limit set, consisting of all forward or backward accumulation points of orbits, is a hyperbolic set. These ideas much influenced contemporary remarkable work of Anosov where it was shown that the geodesic flow on any manifold with negative curvature is ergodic.

Another major achievement of uniform hyperbolicity was to provide a characterization of structurally stable dynamical systems. The notion of *structural stability*, introduced in the thirties by Andronov, Pontrjagin, means that the whole orbit structure remains the same when the system is slightly modified: there exists a homeomorphism of the ambient manifold mapping orbits of the initial system into orbits of the modified one, and preserving the time arrow. Indeed, uniform hyperbolicity proved to be the key ingredient of structurally stable systems, together with a transversality condition, as conjectured by Palis, Smale.

In the process, a theory of uniformly hyperbolic systems was developed, mostly from the sixties to the mid eighties, whose importance extended much beyond the original objectives. It was part of a revolution in our vision of determinism, strongly driven by observations originating from experimental sciences, which shattered the classical opposition between deterministic evolutions and random evolutions. The uniformly hyperbolic theory provided a mathematical foundation for the fact that deterministic systems, even with a small number of degrees of freedom, often present chaotic behavior in a robust fashion. Thus, it led to the almost paradoxical conclusion that “chaos” may be stable.

On the other hand, structural stability and uniform hyperbolicity were soon realized to be less universal properties than was initially thought: there exist many classes of systems that are robustly unstable and non-hyperbolic and, in fact, that is often the case for specific models coming from concrete applications. The dream of a general paradigm in Dynamics had to be postponed.

Beyond uniform hyperbolicity.

The next years saw the theory being extended in several distinct directions:

- The study of specific classes of systems, such as quadratic maps, Lorenz flows, and Hénon attractors, which introduced a host of new methods and ideas.
- Bifurcation theory including, in particular, the study of the boundary of uniformly hyperbolic systems, and of the local and global mechanisms leading to chaotic behavior, especially homoclinic bifurcations.
- New developments in the ergodic theory of smooth systems and, especially, the theory of non-uniformly hyperbolic systems (Pesin theory).
- Weaker formulations of hyperbolicity, still with a uniform flavor but where one allows for invariant “neutral” directions (partial hyperbolicity, projective hyperbolicity or existence of a dominated splitting).

- The converse implication in the stability conjecture (hyperbolicity is necessary for stability), which led to the introduction of new perturbation lemmas (ergodic closing lemma, connecting lemma).

Building on remarkable progress obtained in these directions, especially in the eighties and early nineties, several ideas have been put forward and a new point of view has emerged recently, which again allow us to dream of a global understanding of "most" dynamical systems. Initiated as a survey paper requested to us by David Ruelle, the present work is an attempt to put such recent developments in a unified perspective, and to point open problems and likely directions of further progress.

Two semi-local mechanisms, very different in nature but certainly not mutually exclusive, have been identified as the main sources of persistently non-hyperbolic dynamics:

- What we call here "critical behavior", corresponding to critical points in one-dimensional dynamics and, more generally, to homoclinic tangencies, and which is at the heart of Hénon-like dynamics. This is now reasonably well understood, in terms of non-uniformly hyperbolic behavior. Moreover, recent results show that this type of behavior is always present in connection to non-hyperbolic dynamics in low dimensions.
- In higher dimensions, dynamical robustness (robust transitivity, stable ergodicity) extends well outside the uniformly hyperbolic domain, roughly speaking associated to coexistence of uniformly hyperbolic behavior with different unstable dimensions. It requires some uniform geometric structure (transverse invariant bundles: partial hyperbolicity, dominated decomposition) that we refer to as "non-critical behavior".

On the other hand, new perturbation lemmas permitted to organize the global dynamics of generic dynamical systems, by breaking it into elementary pieces separated by a filtration. A great challenge is to understand the dynamics on (the neighborhood of) these elementary dynamical pieces, which should involve a deeper analysis of the two mechanisms mentioned previously. Indeed, a good understanding has already been possible in several cases, especially at the statistical level.

What is this book, and what is it not ?

The text is aimed at researchers, both young and senior, willing to get a quick yet broad view of this part of Dynamics. Main ideas, methods, and results are discussed, at variable degrees of depth, with references to the original works for details and complementary information.

We assume the reader is familiar with the fundamental objects of smooth Dynamics, like manifolds or C^r diffeomorphisms and vector fields, as well as with the basic facts in the local theory of dynamical systems close to a hyperbolic periodic point, such as the Hartman-Grobman linearization theorem and the stable manifold theorem. This material is covered by several books, like Bowen [86], Irwin [225], Palis, de Melo [342], Ruelle [394], Katok, Hasselblatt [232], or Robinson [382].

Familiarity with the classical theory of uniformly hyperbolic systems is also desirable, of course. This is also covered by a number of books, including Bowen [86], Shub [411], Mañé [281], Palis, Takens [345], and Katok, Hasselblat [232]. For the reader's convenience, in Chapter 1 we review the main conclusions of the theory that are relevant for our purposes. In that chapter we also give an introductory discussion of robust mechanisms of non-hyperbolicity, and other key issues outside the hyperbolic set-up. This is to be much expanded afterwards, so at that point our presentation is sketchier than elsewhere.

Apart from these pre-requisites, we have tried to keep the text self-contained, giving the precise definitions of all relevant non-elementary notions. Occasionally, this is done in an informal fashion at places where the notion is first needed in a non-crucial way, with the formal definition appearing at some later section where it really is at the heart of the subject. This is especially true about Chapter 1, as explained in the previous paragraph.

Although we have used parts of this book as a basis for graduate courses, it is certainly not designed as a text book that could be used for that purpose all alone. The properties of the main notions are often only stated, and most results are presented with just an outline of the proof.

The book is also not meant to be an exhaustive presentation of the recent results in Dynamics. We are only too conscious of the many fundamental topics we left outside, or touched only briefly. Deciding where to stop could be one of the most difficult and most important problems in this kind of project, and no answer is entirely satisfactory.

How should this book be used and what does it contain ?

The 12 chapters are organized so as to convey a global perspective of dynamical systems. The 5 appendices include several other important results, older and new, which we feel should not be omitted, either because they are used in the text or because they provide complementary views of some aspects of the theory.

Although there is, naturally, a global coherence in the text, we have tried to keep the various chapters rather independent, so that the reader may choose to read one chapter without really needing to go through the previous ones. This means that we often recall main notions and statements introduced elsewhere, or else give precise references to where they can be found. On the other hand, the chapters often rely on ideas and results from the appendices.

The main text may be, loosely, split into the following blocks:

- Chapter 1 contains a brief review of uniformly hyperbolic theory and an introduction to main themes to be developed throughout the text.
- Chapters 2 to 4 are devoted to critical behavior in various aspects: one-dimensional dynamics, homoclinic tangencies, Hénon-like dynamics.

- Chapter 5 shows that, for low dimensional systems, far from critical behavior the dynamical behavior is hyperbolic.
- Chapters 6 to 9 treat non-critical behavior, especially the relation between robustness and existence of invariant splittings. While most of the text focusses on dissipative discrete time systems, Chapter 8 deals with conservative diffeomorphisms and Chapter 9 is devoted to flows.
- In Chapter 10 we try to give a global framework for the dynamics of generic maps, where critical and non-critical behavior could fit together.
- Chapter 11 presents some of the progress attained in describing the dynamics in ergodic terms, both in critical and in non-critical situations (either separate or coexisting). Lyapunov exponents are an important tool in this analysis, and Chapter 12 is devoted to their study and control.

Acknowledgements:

Input from several colleagues greatly helped shape this text and improve our presentation. Besides the referees, we are especially thankful to F. Abdenur, J. Alves, V. Araújo, A. Arbieto, M.-C. Arnaud, A. Avila, M. Benedicks, J. Bochi, S. Crovisier, V. Horita, C. Liverani, S. Luzzatto, L. Macarini, C. Matheus, W. de Melo, C. Morales, C. G. Moreira, K. Oliveira, M. J. Pacifico, J. Palis, E. Pujals, J. Rocha, R. Roussarie, M. Sambarino, M. Shub, A. Tahzibi, M. Tsujii, T. Vivier, A. Wilkinson, J.-C. Yoccoz, and D. Ruelle.

Our collaboration was supported by the Brazil-France Agreement in Mathematics, CNPq-Brazil, CNRS-France, Faperj-Rio de Janeiro, in addition to our own institutions. To all of them we express our warm gratitude.

Contents

1

Hyperbolicity and beyond

Uniformly hyperbolic systems are presently fairly well understood, both from the topological and the ergodic point of view. In Sections 1.1 through1.3 we review some of their main properties (spectral decomposition, stability, physical invariant measures) that one would like to extend to great generality. Several very good references are available for this material, including the books of Shub [411], Palis, Takens [345, Chapter 0] and Katok, Hasselblat [232, Part 4], and Bowen [86] for the ergodic theory of this systems.

Outside the hyperbolic domain, two main phenomena occur: homoclinic tangencies and cycles involving saddles with different indices. These notions are introduced in Section 1.4 and 1.5 and serve as a guiding thread through the chapters that follow, where they will be revisited in much more detail. In Section 1.7, we present a conjecture of Palis pointing at a global description of most dynamical systems. Section 1.6 introduces a few fundamental notions involved in this conjecture and in most of our text.

1.1 Spectral decomposition

Let M be a compact manifold, and $f : M \to M$ be a diffeomorphism.

Definition 1.1. An invariant compact set $\Lambda \subset M$ is a *hyperbolic set* for $f : M \to M$ if the tangent bundle over Λ admits a continuous decomposition

$$T_\Lambda M = E^u \oplus E^s, \tag{1.1}$$

invariant under the derivative and such that $\|Df^{-1} \mid E^u\| \leq \lambda$ and $\|Df \mid E^s\| \leq \lambda$ for some constant $\lambda < 1$ and some choice of a Riemannian metric on the manifold.

A point z is *non-wandering for f* if for every neighborhood U of z there is $n \geq 1$ such that $f^n(U)$ intersects U. The set of non-wandering points is denoted $\Omega(f)$. It contains the set $\mathrm{Per}(f)$ of periodic points, as well as the α-limit set and the ω-limit set of every orbit.

Definition 1.2. The diffeomorphism $f : M \to M$ is *uniformly hyperbolic*, or satisfies the *Axiom A*, if $\Omega(f)$ is a hyperbolic set for f and $\operatorname{Per}(f)$ is dense in $\Omega(f)$.

The definitions for smooth flows $f^t : M \to M$, $t \in \mathbb{R}$, are analogous, except that (unless Λ consists of equilibria) the decomposition (1.1) becomes

$$T_\Lambda M = E^u \oplus E^0 \oplus E^s, \tag{1.2}$$

where E^0 is 1-dimensional and collinear to the flow direction.

The spectral decomposition theorem of Smale [418] asserts that the limit set of a uniformly hyperbolic system splits into a finite number of pairwise disjoint *basic pieces* that are compact, invariant, and dynamically indecomposable. The precise statement follows.

We say that an f-invariant set is *indecomposable*, or *transitive*, if it contains some dense orbit $\{f^n(z) : n \geq 0\}$. An f-invariant set Λ is called *isolated*, or *locally maximal*, if there exists a neighborhood U of Λ such that

$$\Lambda = \bigcap_{n \in \mathbb{Z}} f^{-n}(U). \tag{1.3}$$

That is, Λ coincides with the set of points whose orbits remain in U for all times.

Theorem 1.3. *The non-wandering set $\Omega(f)$ of a uniformly hyperbolic diffeomorphism f decomposes as a finite pairwise disjoint union*

$$\Omega(f) = \Lambda_1 \cup \cdots \cup \Lambda_N$$

of f-invariant transitive sets Λ_i, that are compact and isolated. Moreover, the α-limit set and the ω-limit set of every orbit are contained in some Λ_i.

Here is a sketch of the proof. Consider the equivalence homoclinic relation defined in $\operatorname{Per}(f)$ by $p_1 \sim p_2 \Leftrightarrow$ the stable set of the orbit of each of the points has some transverse intersection with the unstable set of the orbit of the other. The stable manifold theorem implies that there are finitely many equivalence classes, and they are open in $\operatorname{Per}(f)$. The basic pieces Λ_i in the theorem are the closures of the equivalence classes. By construction, they are compact, invariant, and open in $\Omega(f) = \overline{\operatorname{Per}}(f)$. The latter implies that they are isolated sets, because $\overline{\operatorname{Per}}(f)$ is an isolated set if it is hyperbolic. Moreover, the stable and the unstable manifold of any periodic point in some Λ_i are dense in Λ_i. This implies that Λ_i is transitive. Finally, if the α- or ω-limit set of some orbit intersected more than one Λ_i, there would be non-wandering points outside the union of the basic pieces, a contradiction.

Transitive sets and isolated sets of flows $f^t : M \to M$, $t \in \mathbb{R}$, are defined in the same way. Theorem 1.3 remains true for uniformly hyperbolic flows $\{f^t : t \in \mathbb{R}\}$.

Remark 1.4. A question dating from the late sixties asked whether every hyperbolic set is contained in an isolated one. This was recently solved by Crovisier [143], who constructed a transitive diffeomorphism of $M = \mathbb{T}^4$ having a hyperbolic set which is not contained in any isolated hyperbolic set Λ: Crovisier shows that Λ would have to be the whole torus, which is not a hyperbolic set because the diffeomorphism has saddles with different indices. This has been improved by Fisher [183], using different methods: he obtains robust examples in any dimension ≥ 2, and for dimension 4 or higher his examples are also transitive.

1.2 Structural stability

A smooth dynamical system is called ***structurally stable*** [15] if it is equivalent to any other system in a C^1 neighborhood. In the discrete-time case, equivalence means conjugacy by a global homeomorphism. In the case of flows this notion is too restrictive: it forces all periods of closed orbits to be preserved under perturbation. Instead, one asks for the existence of a global homeomorphism sending orbits of one system to orbits of the other, and preserving the direction of time. More generally, replacing C^1 by C^r neighborhoods, any $r \geq 1$, one obtains the notion of C^r structural stability.

The stability conjecture of Palis-Smale [343] proposes a complete characterization of structurally stable systems. Namely, they should coincide with the hyperbolic systems having the property of strong transversality: every stable and unstable manifolds of points in the non-wandering set should be transversal. In fact their conjecture is for C^r structural stability, any $r \geq 1$.

Robbin [377], de Melo [146], and Robinson [379, 380] proved that these are sufficient conditions for structural stability. Strong transversality is also necessary [378]. These results hold in the C^r topology, any $r \geq 1$. The hardest part was to prove that stable systems must be hyperbolic. This was achieved by Mañé [283] in the mid-eighties, for C^1 diffeomorphisms, and extended about ten years later by Hayashi [208] for C^1 flows. Thus

Theorem 1.5. *A C^1 diffeomorphism (or flow) on a compact manifold is structurally stable if and only if it is uniformly hyperbolic and verifies the strong transversality condition.*

A weaker property, called ***Ω-stability*** is defined requiring conjugacy (respectively, equivalence) only restricted to the non-wandering set. The Ω-stability conjecture in [343] proposes a characterization of the Ω-stable systems: they should be the hyperbolic systems having no *cycles*, that is, no basic pieces in their spectral decompositions cyclically related by intersections of the corresponding stable and unstable sets.

The Ω-stability theorem of Smale [418] states that these properties are sufficient. The proof uses the following notion:

Definition 1.6. A *filtration* for a diffeomorphism $f : M \to M$ is a finite family $M_1, M_2, \ldots, M_k$ of submanifolds with boundary and with the same dimension as M, such that

- $M_1 = M$ and M_{i+1} is contained in the interior of M_i for every $1 \le i < k$.
- $f(M_i)$ is contained in the interior of M_i for all $1 \le i \le k$.

The open sets $L_i = \operatorname{int}(M_i \setminus M_{i+1})$ are the *levels of the filtration* (set $M_{k+1} = \emptyset$).

The first step is top show that if f is hyperbolic and has the no-cycles property, then it admits a filtration such that each basic piece coincides with the set of orbits contained in some level: up to reordering,

$$\Lambda_i = \bigcap_{n \in \mathbb{Z}} f^n(L_i) \qquad \text{for all } i.$$

Let g be any diffeomorphism C^r-close to f. Then $M_1, , \ldots, M_k$ is also a filtration for g. Therefore, $\Omega(g)$ is contained in $\Lambda_1(g) \cup \cdots \cup \Lambda_k(g)$, where

$$\Lambda_i(g) = \bigcap_{n \in \mathbb{Z}} g^n(L_i).$$

Stability of hyperbolic sets gives that each $f \mid \Lambda_i$ is conjugate to $g \mid \Lambda_i(g)$. Then, every $\Lambda_i(g)$ is contained in $\Omega(g)$, and $f \mid \Omega(f)$ is conjugate to $g \mid \Omega(g)$. Thus f is Ω-stable, in the C^r sense.

Palis [339] proved that the no-cycles condition is necessary for Ω-stability, in any C^r topology. Necessity of hyperbolicity for Ω-stability was proved by Palis [340], based on Mañé [283], for C^1 diffeomorphisms, and extended to C^1 flows by Hayashi [208].

1.3 Sinai-Ruelle-Bowen theory

A basic piece Λ_i is a *hyperbolic attractor* if the stable set

$$W^s(\Lambda_i) = \{x \in M : \omega(x) \subset \Lambda_i\}$$

contains a neighborhood of Λ_i. In this case we call $W^s(\Lambda_i)$ the *basin* of the attractor Λ_i, and denote it $B(\Lambda_i)$. When the Axiom A system is of class C^2, a basic piece is an attractor if and only if its stable set has positive Lebesgue measure. Thus, the union of the basins of all attractors is a full Lebesgue measure subset of M. This remains true for a residual (dense G_δ) subset of C^1 uniformly hyperbolic diffeomorphisms and flows. See Bowen [86, 87].

The following fundamental result, due to Sinai [416], Ruelle, Bowen [86, 90, 390] says that, no matter how complicated it may be, the behavior of typical orbits in the basin of a hyperbolic attractor is completely well-defined at the statistical level:

Theorem 1.7. *Every attractor Λ of a C^2 uniformly hyperbolic diffeomorphism (or flow) supports a unique invariant probability measure μ such that*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(z)) = \int \varphi\, d\mu \tag{1.4}$$

for every continuous function φ and Lebesgue almost every point $x \in B(\Lambda)$.

One way to construct μ is starting with normalized Lebesgue measure m_D over a compact domain D inside any leaf of the unstable foliation of the attractor. A distortion argument, using the fact that f is C^2, gives that the sequence of iterates $\{f_*^n(m_D) : n \geq 1\}$ has a property of *uniform* absolute continuity along the unstable foliation. Contracting behavior in the transverse direction, together with minimality of the unstable foliation, are used to show that the limit $\mu = \lim_{n\to\infty} f_*^n(m_D)$ exists and is an ergodic measure whose support coincides with Λ. Very important, because of the previous observation μ disintegrates into conditional measures along unstable foliation (see Appendix C) that are equivalent to the Lebesgue measure of each leaf.

Ergodicity gives (1.4) for μ-almost every point in Λ. Using conditional measures we get it for a full Lebesgue measure subset L of some unstable leaf. To prove the much more interesting fact that (1.4) is true for Lebesgue almost every point in the basin of attraction, a whole open set, one uses the observation that time-averages of continuous functions are constant on stable manifolds. The stable manifolds of points in the attractor foliate the whole $B(\Lambda)$. Absolute continuity of this stable foliation (see Appendix C) ensures that the stable manifolds of the points in L cover a full Lebesgue measure of the basin. Theorem 1.7 follows.

Property (1.4) means that the Sinai-Ruelle-Bowen measure μ may be explicitly computed, meaning that the weights of subsets may be found with any degree of precision, as the sojourn-time of any orbit picked "at random" in the basin of attraction:

$$\mu(V) = \text{ fraction of time the orbit of } z \text{ spends in } V$$

for any typical subset V of M (the boundary of V should have zero μ-measure), and for Lebesgue almost any point $z \in B(\Lambda)$. For this reason μ is called a *physical measure.*

There is another sense in which this measure is "physical" and that is that μ is the zero-noise limit of the stationary measures associated to the stochastic processes obtained by adding small random noise to the system. This property is called *stochastic stability*; a formal definition will appear later. For uniformly hyperbolic systems it is due to Sinai [416], Kifer [240, 242], and Young [457]. The model of small stochastic perturbations to represent external influences, too small or too complex to express in deterministic terms, goes back to Kolmogorov and Sinai.

1.4 Heterodimensional cycles

Although uniform hyperbolicity was originally intended to encompass a residual, or at least dense subset of all dynamical systems, it was soon realized that this is not true. There are two main mechanisms that yield robustly non-hyperbolic behavior, that is, whole open sets of non-hyperbolic systems. Not surprisingly, they are at the heart of recent developments that we are going to review in the next sections.

Historically, the first one was the coexistence of periodic points with different Morse indices (dimensions of the unstable manifolds) inside the same transitive set. See Figure 1.1. This is how the first examples of C^1-open subsets of non-hyperbolic diffeomorphisms were obtained by Abraham, Smale [4, 414] on manifolds of dimension $d \geq 3$. It was also the key in the constructions by Shub [410] and Mañé [277] of non-hyperbolic yet robustly transitive diffeomorphisms, that is, such that every diffeomorphism in a C^1 neighborhood has a dense orbit. The examples of Shub and Mañé are outlined in Section 7.1.

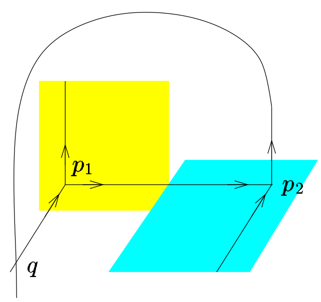

Fig. 1.1. A heterodimensional cycle

For flows this mechanism may assume a novel form, because of the interplay between regular periodic orbits and singularities (equilibrium points). That is, robust non-hyperbolicity may stem from the coexistence of regular and singular orbits in the same transitive set. The first and very striking example was the geometric Lorenz attractor proposed by Afraimovich, Bykov, Shil'nikov [5] and Guckenheimer, Williams [201, 452] to model the behavior of the Lorenz equations [267]. This is a main theme in Chapter 9.

1.5 Homoclinic tangencies

Heterodimensional cycles may exist only in dimension 3 or higher. The first robust examples of non-hyperbolic diffeomorphisms on surfaces were constructed by Newhouse [319], exploiting the second of the mechanisms we mentioned:

homoclinic tangencies, or non-transverse intersections between the stable and the unstable manifold of the same periodic point. See Figure 1.2.

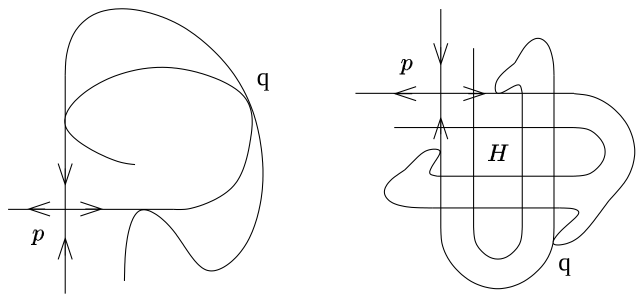

Fig. 1.2. Homoclinic tangencies

It is important to observe that individual homoclinic tangencies are easily destroyed by small perturbations of the invariant manifolds. To construct open examples of surface diffeomorphisms with *some* tangency, Newhouse started from systems where the tangency is associated to a periodic point inside an invariant hyperbolic set with rich geometric structure. See the right hand side of Figure 1.2. His argument requires a very delicate control of distortion, as well as of the dependence of the fractal dimension on the dynamics. Actually, for this reason, his construction is restricted to the C^r topology for $r \geq 2$. By comparison, robust examples of heterodimensional cycles in higher dimensions are obtained by much more elementary transversality and hyperbolicity arguments.

1.6 Attractors and physical measures

Several attempts were made, specially in the seventies, to weaken the definitions of hyperbolicity and structural stability, while keeping their topological flavor, so that they could encompass a residual set of dynamical systems. One such extension was the notion of Ω-stability that we mentioned before.

In parallel, a more probabilistic approach was being proposed, especially by Sinai, Ruelle, Eckmann, where one focus on the statistical behavior of typical orbits, and its stability under perturbations. See [176] for a detailed exposition.

Definition 1.8. A set $\Lambda \subset M$ is an *attractor* for a diffeomorphism (or a flow) on a manifold M if it is invariant and transitive, and the *basin of attraction*,

$$B(\Lambda) = \{x \in M : \omega(x) \subset \Lambda\}$$

has positive Lebesgue measure.

Definition 1.9. A *physical measure*, or *SRB measure*, of a diffeomorphism on a manifold M is an invariant probability measure μ on M, such that the time average of every continuous function $\varphi : M \to \mathbb{R}$ coincides with the corresponding space-average with respect to μ,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(z)) = \int \varphi \, d\mu \tag{1.5}$$

for a set of initial points z with positive Lebesgue measure. We call this set the *basin* of μ, and denote it $B(\mu)$.

For flows (1.5) is replaced by

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \varphi(f^t(z))\, dt = \int \varphi \, d\mu.$$

Problem 1.10. For most dynamical systems, does Lebesgue almost every point have a well-defined time average ? Are there SRB measures whose basins cover almost all M ?

This is the case for C^2 hyperbolic systems, by Theorem 1.7. But the answer can not always be affirmative. A simple counter-example, due to Bowen, is described in Figure 1.3: time averages diverge over any of the spiraling orbits in the region bounded by the saddle connections. However, no robust counter-example is known.

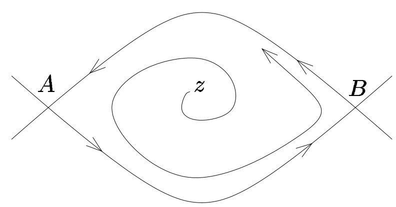

Fig. 1.3. A planar flow with divergent time averages

The following semi-global version of the previous problem also goes back to Sinai and Ruelle:

Problem 1.11. (basin problem) Let Λ be an attractor for a diffeomorphism, or a flow, supporting a unique SRB measure μ. Does

$$B(\mu) = B(\Lambda) \quad \text{up to a zero Lebesgue measure set ?} \tag{1.6}$$

More generally, does $B(\Lambda)$ coincide with the union of the basins of the SRB measures supported in Λ, up to zero Lebesgue measure ?

Again, the answer can not always be positive: [41] shows how Bowen's example can be inserted into an attractor, so that the basin property (1.6) is violated. Note also that general (transitive) attractors may support more than one SRB measure. See Kan [229] and Section 11.1.1. But no robust examples of either of these phenomena have been found to date.

1.7 A conjecture on finitude of attractors

A program towards a global understanding of complex dynamical behavior has been proposed by Palis [341]. Here we quote some of the conjectures embodying his program, others will appear later.

Conjecture 1.12. (finitude of attractors [341]) There is a C^r, $r \geq 1$, dense set subset $\mathcal{D}$ of dynamical systems on any compact manifold that exhibit a finite number of attractors whose basins cover Lebesgue almost all of the manifold.

Conjecture 1.13. (physical measures and stochastic stability [341]) For any element of $\mathcal{D}$ all the attractors support SRB measures and have the basin property. Moreover, any element of $\mathcal{D}$ is stochastically stable on the basin of each of the attractors.

Stochastic stability means that adding small random noise to the system has little effect on its statistical behavior. For discrete-time systems $f : M \to M$, one considers *random orbits* $\{z_j : j \geq 0\}$ where each z_{j+1} is chosen at random in the ε-neighborhood of $f(z_j)$. Then, for ε small, the time averages of continuous functions over almost every random orbit should be close to the corresponding time averages over typical orbits of the original f. There is a corresponding notion for flows, where noise takes place at infinitesimal intervals of time: random orbits are solutions of a stochastic differential equation. Precise definitions are given in Appendix D.

Conjecture 1.14. (metric stability of basins of attraction [341]) Given any element of $\mathcal{D}$ and any of its attractors, then for almost all C^r small perturbations along generic k-parameter families there is a finite set of attractors whose basins cover most (a fraction close to 1 in volume) of the original basin, and these attractors also support physical measures.

A key novelty in the formulation of this conjecture, and in the whole scenario proposed by Palis, is to allow the existence of pathological phenomena, e.g. related to cycles, occupying a small volume in the ambient space. Indeed, cycles have been a main obstruction to the realization of previous global scenarios for Dynamics.

For one-dimensional systems these conjectures take a stronger form: finiteness of attractors, with the nice properties above and with their basins containing Lebesgue almost all points, should correspond to full Lebesgue measure in parameter space, for generic parametrized families. This has been fully verified for C^2 unimodal maps of the interval, as we shall see in Chapter 2. In this case the attractor is unique. There is also substantial partial progress in higher dimensions, some of which is reported throughout the text.

Beyond finiteness ?

Palis' program represents an ambitious attempt to achieve a global description of dynamical systems, and has been inspiring much work in this area. As we have just mentioned, its conclusions have been fully confirmed in the context of unimodal maps of the interval, and it should certainly hold for general smooth maps in dimension 1. At the present stage this program also seems a realistic goal for surface diffeomorphisms. The main remaining difficulty for understanding these systems is given by Newhouse's phenomenon of coexistence of infinitely many attractors or repellers. The previous conjectures are compatible with this phenomenon, except if it occurs robustly, that is, for every diffeomorphism in a whole open set. This is not known at the time, but seems unlikely. Another important related open question is whether this phenomenon may correspond to positive Lebesgue probability in parameter space, on generic parametrized families of diffeomorphisms. In higher dimensions, the situation is presently much less understood, and the previous conjectures remain long term goals.

Even if parts of this program turn out no to be confirmed, investigation of these questions will certainly lead to important further progress. For instance, if coexistence of infinitely many SRB measures does occur for an open subset of systems, but their basins cover a full volume set, and the measures vary continuously with the dynamics, then one will still get a very satisfactory variation of the conclusion in Conjecture 1.14. Also, coexistence of infinitely many independent ergodic behaviors is known to be robust in the conservative setting: by KAM (Kolmogorov, Arnold, Moser) theory many conservative systems exhibit positive volume sets consisting of invariant tori (see for instance [454]). Notwithstanding the existence of a whole continuum of ergodic behaviors, these systems can be understood to some extent. This leaves hope for dissipative systems as well, even if the general finiteness paradigm turns out not to be dense.

Robust non-convergence of time averages for many initial states would, perhaps, be a more disturbing difficulty. We have seen a codimension-2 example in Section 1.6 and it is also known that non-convergence occurs with codimension-1 in the setting of interval maps (see for instance [217]). Different approaches may be envisaged to handling such a difficulty.

One of them, going back to Kolmogorov, is to consider zero-noise limits of stationary measures associated to random perturbations of the system. Indeed, stationary measure exist in great generality (see for instance [18, 19]) and, for small noise levels, they may be considered to provide a certain "physical" perception of the system's behavior. This approached was undertaken in [19] for Bowen's example in Figure 1.3, where the zero-noise limit is an average of the Dirac measures on the two saddle points.

A very different approach is to consider convenient resumations of the time averages, that might be convergent even when the time averages themselves do not. For instance, one may consider higher order averages

$$a_n^{(k)} = \frac{1}{n}\sum_{j=0}^{n-1} a_j^{(k-1)} \qquad a_j^{(0)} = \varphi(f^j(x)).$$

This would yield a stratification of many dynamical systems reflecting, in some sense, their statistical complexity. In this direction we propose the following example, as a test case for both aforementioned approaches:

Problem 1.15. Let X be a vector field on $[0,1]$ which vanishes exactly at the endpoints. Consider the map $f : S^1 \times [0,1] \to S^1 \times [0,1]$ defined by $f(x,y) = (2x \mod \mathbb{Z}, X^{\sin(2\pi x)}(y))$. Observe that the second coordinate of $f^n(x,y)$ is given by the time-T_n map of the vector field X, where $T_n = \sum_{j=0}^{n-1} \sin(2^{n+1}\pi x)$. The Birkhoff time averages $a_n^{(1)}$ of f should diverge almost everywhere. Do second order time averages converge ? Are there zero-noise limits of stationary measures ? Do they provide the same information ?

Notice that it may happen that all these higher order averages $a_k^{(n)}$ diverge; indeed, that seems to be the case in Bowen's example.