DIMENSION PRODUCT STRUCTURE OF HYPERBOLIC SETS

BORIS HASSELBLATT AND JÖRG SCHMELING

(Communicated by Svetlana Katok)

Abstract. We conjecture that the fractal dimension of hyperbolic sets can be computed by adding those of their stable and unstable slices. This would facilitate substantial progress in the calculation or estimation of these dimensions, which are related in deep ways to dynamical properties. We prove the conjecture in a model case of Smale solenoids.

1. Introduction

Since the 1980s pictures of various “fractal” sets have become part of our popular culture, and their fractal dimension has been associated with the smoothness or thickness of these sets. In recent years the rigorous and systematic study of these dimensions (see Section 4 and [F]) and their relation to characteristic properties of the dynamical systems that give rise to these sets has made substantial progress. Much of this progress was initiated by physicists, but concerns of this type have become an important part of mainstream mathematics [P]. While this has had an effect on fields such as number theory, the present work is about invariant sets in hyperbolic dynamical systems.

1.1. Invariant sets in hyperbolic dynamical systems. In this paper a dynamical system is a diffeomorphism or an embedding of a manifold into itself, and the principal interest in the theory is to study the long-term behavior of points under repeated iteration of the map. Often, it is of interest to concentrate on particular invariant sets, and here we concentrate on those that are hyperbolic (see Definition 1), i.e., at each point of the set the tangent space is a direct sum of two subspaces that are contracted and expanded, respectively, by the differential of the map. Simple examples are given by the horseshoe and the “paper clip”, where the map (of the plane or a disk) compresses a rectangle vertically, expands it horizontally, folds it as in Figure 1 and places it over the original rectangle. It is not hard to see that the invariant set is a planar Cantor set; if the map is linear and respects the coordinate directions on the pieces that fall across the original rectangle, then the invariant set is a Cartesian product of two Cantor sets. In this context it has become interesting to obtain (or at least estimate) the fractal dimension (Section 4) of a hyperbolic invariant set from information about the dynamics on that set. This is often difficult to do, and most progress appears to have been made when...
the dynamics on the hyperbolic set is strictly expanding or contracting. The general case combines both contraction and expansion, but topologically, these sets have a (local) product structure, that is, they can locally be represented as products of a representative contracting slice and a representative expanding slice. The techniques that have helped with contracting or expanding hyperbolic sets have been adapted to dealing with these slices, and therefore great progress would be achieved by establishing that the dimension of the hyperbolic set is the sum of those of the contracting (“stable”) and expanding (“unstable”) slices, even though the known product structure is only topological and certainly cannot be achieved with Lipschitz charts [HW]. Charts that are not Lipschitz-continuous can alter fractal dimensions (by a factor given by the Hölder exponent of the chart).

\[ \text{Figure 1. The horseshoe (with second iterate) and the paper clip.} \]

1.2. Dimension product structure. Even though this obstacle is real, we conjecture that the desired dimension product structure still prevails, at least generically, or under mild additional hypotheses. It is clear that establishing this will be difficult, and there are examples that suggest the possible failure of such a result in some situations. But we demonstrate in a model example how the essential difficulties can be addressed (Theorem 2 p. 92 and Theorem 4 p. 93). The central insight is that while [HW] exhibits open sets of symplectic Anosov diffeomorphisms for which the local product charts are non-Lipschitz on a large set, the problem is in an essential way due to the assumption of the Anosov property (that the entire manifold is a hyperbolic set). In this case, of course, little needs to be done in order to determine the fractal dimension of the hyperbolic set, since it is given at the outset. In our work we point to a subtle aspect of the local geometry of these sets that is responsible for a reversal of the situation in the case of fractal hyperbolic sets. For these we show a way of establishing Lipschitz continuity of local product charts off a set that is small enough to be discarded without affecting the dimension of the hyperbolic set or of its stable or unstable slices. This example is different from even nonlinear versions of the maps in Figure 1 because in planar situations we always have a dimension product structure as remarked in Subsection 1.4.

We should point out that a dimension product structure of invariant measures was established some time ago [BPS1] in the course of proving the Eckmann–Ruelle conjecture.

Conjecture. The fractal dimension of a hyperbolic set is (at least generically or under mild hypotheses) the sum of those of its stable and unstable slices, where “fractal” can mean either Hausdorff or upper box dimension.
Here we describe our proof of this conjecture for a class of solenoids [HS]. Even in this situation the dimension decomposition has been an open problem for at least two decades. Both [B] gave conditions of a transversality type under which he could show (assuming estimates on the contraction and expansion rates) that all stable slices have the same dimension, and mild conditions on the contraction and expansion rates guarantee $C^1$-genericity of this transversality condition (see Remark 3).

1.3. Hyperbolic sets. We now introduce hyperbolic sets, their stable and unstable slices as well as the holonomies between them. We begin with the invariant laminations.

Definition 1 ([KH]). Let $M$ be a manifold, $U \subset M$ open, $f: U \to M$ a smooth embedding, $\Lambda \subset M$ a compact $f$-invariant set. Then $\Lambda$ is said to be a hyperbolic set if the tangent bundle $T_\Lambda M$ over $\Lambda$ splits as $T_\Lambda M = E^u \oplus E^s$ in such a way that there are constants $C > 0$, $0 < \lambda < 1 < \eta$ such that $\|Df^{-n}|_{E^u}\| \leq C\eta^{-n}$ and $\|Df^n|_{E^s}\| \leq C\lambda^n$ for $n \in \mathbb{N}$.

The subbundles are then invariant and continuous and have smooth integral manifolds $W^u$ and $W^s$ that are coherent in that $q \in W^u(p) \Rightarrow W^u(q) = W^u(p)$. If $\Lambda = M$ then $W^u$ and $W^s$ define laminations (continuous foliations with smooth leaves) and $f$ is said to be an Anosov diffeomorphism. Even if $\Lambda \neq M$ (the case on which we concentrate) we use the term laminations.

If $A, B$ are two nearby embedded disks transverse to $W^u$ then there is a holonomy map defined on a (possibly empty) subset of $A \cap \Lambda$ by $p \mapsto W^u_{\text{loc}}(p) \cap B$ whenever this makes sense. Here $W^u_{\text{loc}}(p)$ is a small disk in $W^u(p)$ that contains $p$. In other words, we move from $A$ to $B$ along unstable leaves. If $W^u_{\text{loc}}(p) \cap B \neq \emptyset$ then the holonomy is defined on the intersection of a neighborhood of $p$ with $A \cap \Lambda$. These holonomies are always continuous (this is easy to see), and, in fact, Hölder continuous (i.e., $d(f(x), f(y)) \leq Cd(x, y)^\alpha$ for some $\alpha > 0$). There would be no need for the present paper if it were known that these holonomies are always Lipschitz continuous (that is, Hölder continuous with exponent 1). Indeed, that this Lipschitz property is not obviously true might be said to have held back the study of the smooth ergodic theory of geodesic flows and Anosov systems by a quarter century (between the work of Hopf [Ho] and that of Anosov–Sinai [AS]). This is but one motivation for studying the regularity of holonomies and local product charts.

1.4. Regularity. The search for ergodic mechanical systems led to the study of geodesic flows of negatively curved manifolds in the 1920s and ’30s. Eberhard Hopf [Ho] established ergodicity of the geodesic flow of surfaces of negative (and not necessarily constant) curvature and in doing so provided the first and still only argument to establish ergodicity in wide classes of hyperbolic dynamical systems. He used in a critical way that geodesic flows on surfaces have $C^1$ holonomies. There was no progress in higher dimension until the 1960s, and Anosov brought out the technical issue at hand: The holonomies need not be $C^1$ in that case, and they may indeed fail almost everywhere to be Lipschitz continuous [A]. Anosov and Sinai overcame this difficulty by establishing that the holonomies are nevertheless absolutely continuous (with respect to the Riemannian measure on leaves they send null sets to null sets) [AS] [BS], which suffices for the Hopf argument and has become important also in situations where volume is not preserved [AS] [Y].
Subtle issues of regularity of the invariant foliations have been studied since, including in the study of smooth and geometric rigidity (see, e.g., [H2]), where rather high degrees of smoothness play a role. We now outline some of the results of interest that are pertinent to our desire for a Lipschitz property of the holonomies. A careful discussion of the relations between regularity of holonomies and foliation charts is given in [PSW].

If one makes assumptions about particular relations between the various expansion and contraction rates of the diffeomorphism then one can guarantee particular Hölder conditions. Sufficiently stringent such conditions even give $C^1$ holonomies with Hölder derivative, where the Hölder exponent of the derivative can be arbitrarily close to 1. For the holonomies between stable or unstable slices in hyperbolic systems Schmeling and Siegmund-Schultze [SS] found optimal estimates of the Hölder exponent in terms of Lyapunov exponents, while Hasselblatt [H1] produced analogous estimates for the Hölder exponent of the subbundles in terms of more uniform rates. It is not necessary for us to describe the needed conditions, but here are some examples where they follow from assumptions that are simpler to state. If one of the stable and unstable laminations is 1-dimensional then the other one is $C^1$; in particular, in 2-dimensional systems both laminations are $C^1$, which is the context within which Hopf worked (and the reason planar horseshoes are not a useful example for our agenda here). Also, if one foliation is 1-dimensional and the dynamical system preserves volume, then both laminations are $C^1$ (of course, in this case there are no attractors). Thus, there are open sets of hyperbolic dynamical systems where the dimension product structure is known, but one would like to know it in complete generality. We now note that there are significant obstacles to achieving this.

1.5. Failure of Lipschitz continuity. One of our difficulties is that the regularity results mentioned above have been shown to be sharp in a strong sense, beginning with the aforementioned work of Anosov. In at least 3 dimensions the set of maps where the Hölder exponent obtained by these results is less than 1 is open, and we established that within this open set those maps that do not have holonomy maps with a larger Hölder exponent form a residual set. Indeed, Hasselblatt and Wilkinson found open sets of symplectic Anosov maps with the property that on a residual set of full measure (with respect to any invariant measure) the subbundles are not Lipschitz, and the holonomies are non-Lipschitz a.e. with respect to Lebesgue measure [HW]. Earlier, Schmeling found that solenoids often lack regular holonomies but the set of non-Lipschitz points seemed to be rather small in the measure sense [S].

This means that one should expect (or at least allow for the possibility) that in a typical hyperbolic dynamical system the invariant laminations fail to be Lipschitz continuous on a residual set that is conull with respect to any (interesting) invariant measure.

1.6. Controlling the set of non-Lipschitz points. Despite these serious obstacles we endeavor to control the size of the set of non-Lipschitz points in a manner that is sufficient for purposes of dimension calculations. We consider solenoids that carry an invariant smooth hyperbolic coordinate system on the basic set. This implies that local fast stable manifolds depend smoothly on the points. We show for our class of maps that deletion of the set NL of non-Lipschitz points does not diminish the Hausdorff or (upper) box dimension of the hyperbolic set $\Lambda$. More to
the point, we show that all stable slices have the same fractal dimension, and the
sum of this and the unstable dimension (which is 1 here) gives dim \( \Lambda \).

1.7. The solenoid. The situation we consider is the Smale solenoid or “derived
from expanding” (DE-) map, a smooth realization of the natural extension of the
map \( x \mapsto \eta x \) on \( S^1 \) for \( \eta \in \mathbb{N} \times \mathbb{N} \).

Specifically, let \( M := S^1 \times \mathbb{D} \) be the solid torus, where \( \mathbb{D} = \{ v \in \mathbb{R}^2 \mid |v| < 1 \} \)
carries the product distance \( d = d_1 \times d_2 \), and suppose \( f: M \to M, \ (x, y, z) \mapsto (\eta x, \lambda y + u(x), \mu z + v(x)) \) is a \( C^2 \) map, where \( \mu, \lambda, \eta: M \to \mathbb{R}^+ \), \( \mu < \lambda < 1 < \eta \).

Let \( \Lambda := \bigcap_{n \in \mathbb{N}} f^n(M) \) be the attractor and denote by NL the set of non-Lipschitz
points of the unstable lamination, i.e., the set of points \( p \in M \) for which there is a
\( q \in W_{\text{loc}}^u(p) \) such that the unstable holonomy from \( W^s(p) \) to \( W^s(q) \) is not Lipschitz
at \( p \).

2. Results

The natural projection \( \pi_x: (x, y, z) \mapsto x \) allows us to define stable slices: For any
set \( A \subset M \) we write \( A_x := (\pi_x|_A)^{-1}(\{x\}) \). Thus, \( W^s(p) = \Lambda_{\pi_x(p)} \). We also often
project along fast-stable leaves using \( \pi_{(x,y)}: (x, y, z) \mapsto (x, y) \).

**Theorem 2.** Consider a \( C^2 \) map \( f: M \to M \) as above, and to fix ideas assume that \( \lambda, \mu, \eta \) are constant and \( \eta = 2 \) as well as

(i) The slow contraction \( \lambda \) is less than 1/2,

(ii) The unstable foliation is transverse as illustrated in Figure 3. If \( p \in \Lambda \), \( p' \in W^s(p) \), and \( n \in \mathbb{N} \) is such that \( \pi_x(f^{-i}(p)) = \pi_x(f^{-i}(p')) \) for \( 0 \leq i \leq n \),
but not for \( i = n + 1 \), then \( \pi_{(x,y)}(W^u_{\text{loc}}(f^{-n}(p))) \) and \( \pi_{(x,y)}(W^u_{\text{loc}}(f^{-n}(p'))) \)
are transverse (necessarily uniformly so).

Then \( \dim \Lambda_x = \dim \Lambda_{x'} \) for all \( x, x' \in S^1 \), and \( \dim \Lambda = 1 + \dim \Lambda_x \), where \( \dim \) is
either Hausdorff or upper box dimension.

**Remark 3.** This situation was first considered by Bothe [13], who obtained results
on the dimension of \( \Lambda \) and \( \Lambda_x \). In fact, the analysis in [13] implies Theorem 2 for the
case \( \lambda \leq 1/8 \), and Bothe proved that transversality is generic in \( C^1 \) when \( \lambda < 1/4 \).
Contrary to Theorem 2, Bothe does not take the contraction and expansion rates
to be constant. (We do so only in order to keep the notation light.)
Our second (and more difficult) result does not require the transversality assumption at all, although we assume analyticity in this case:

**Theorem 4.** Let \( M := \mathbb{S}^1 \times \mathbb{D} \), \( \mu, \lambda \in \mathbb{R} \), \( 0 < \mu < \lambda < 1/2 \), and let \( f: M \to M \), \( (x, y, z) \mapsto (2x, \lambda y + u(x), \mu z + v(x)) \) be analytic. Then \( \dim \Lambda_x = \dim \Lambda_y \) for all \( x, y \in \mathbb{S}^1 \), and \( \dim \Lambda = 1 + \dim \Lambda_x \), where \( \dim \) is either Hausdorff or upper box dimension and \( \Lambda \) is the attractor.

While Theorem 2 has more robust hypotheses, which is interesting with respect to the prevalence of a dimension product structure, those of Theorem 4 are easily checked in concrete examples.

We expect that dropping the assumption of parallel (in particular, \( C^1 \)) fast directions immediately leads to technical problems, although these may not be fundamental.

### 3. Methods

In this section we give a description of the phenomena that may give rise to non-Lipschitz points. The discussion here requires only the assumptions of Subsection 1.7, *not* the full strength of the hypotheses of Theorem 2 or Theorem 4.

Suppose \( p \in \Lambda \), \( q \in W^u_{loc}(p) \) are such that the holonomy from \( W^u(p) \) to \( W^u(q) \) is not Lipschitz continuous at \( p \), i.e., for all \( C > 2 \) there is a \( p' \in W^u(p) \) arbitrarily near \( p \) such that \( \Delta q := d(q, q') > Cd(p, p') = C \Delta p \), where \( q' \) is the image of \( p' \) under the holonomy. Let \( d_{pq} := d(\pi_x(p), \pi_x(q)) \).

#### 3.1. Pulling the failure of a Lipschitz condition back to unit scale.

We now show that when the “bend” in unstable manifolds that gives rise to a failure of a Lipschitz property at a small scale is pulled back to unit scale, one observes there a corresponding discrepancy in the slow stable direction.

Take \( n \in \mathbb{N} \) such that \( \pi_x(f^{-i}(p)) = \pi_x(f^{-i}(p')) \) for \( 0 \leq i \leq n \), but not for \( i = n + 1 \) (since \( \Delta p \) may be small, \( n \) can be arbitrarily large). Then there exist \( 0 < m \leq M \) independent of \( p, q, p' \) such that \( m \leq d(f^{-n}(p), f^{-n}(p')) =: \Delta p \leq M \) and \( m \leq d(f^{-n}(q), f^{-n}(q')) =: \Delta q \leq M \).

Since \( \mathbb{D} \) carries the product distance \( d = d_1 \times d_2 \) we can write \( \Delta p = \Delta_1 p + \Delta_2 p \) in a natural way, and likewise \( \Delta p = \Delta_1 p + \Delta_2 p \), \( \Delta p = \Delta_1 p + \Delta_2 p \) and \( \Delta q = \Delta_1 q + \Delta_2 q \).

It is now not hard to verify that if \( C \) is sufficiently large and \( \Delta q > C \Delta p \) then \( \Delta_1 q > C' \Delta_1 p \), where \( C' := \frac{C}{1+C_{1.2}d_{pq}} \) for a moderate constant \( C_1 \).

This means that after pulling back to unit scale the failure of the Lipschitz condition shows up precisely as a corresponding discrepancy in the size of the slow-stable components, which also means that these have to be very small.
3.2. Principal case distinction—crossings. The small size of $\Delta_{1p}$ combined with the substantial angle between the projections of the two leaves that is implied by $\Delta q > C'\Delta_{1p}$ suggests that the fast-stable projections of the local unstable leaves in the scaled-up preimage should cross rather near $f^{-n}(p)$. Indeed, the transversality assumption in Theorem 2 implies this: For a point $p \in NL$, there are infinitely many $n \in \mathbb{N}$ such that $f^{-n}(p)$ is within $d_{pq}/\eta_n(q)$ of a crossing of fast-stable projections of local unstable leaves of $f^{-n}(p)$ and $f^{-n}(p')$ for some $p'$ such that $\pi_x(f^{-i}(p)) = \pi_x(f^{-i}(p'))$ for $i \leq n$ but not for $i = n + 1$.

We show that the set of such $p$ (and hence NL) is benign with respect to Hausdorff dimension. (In higher dimension the sets of points with crossings may not be negligible because of the Ledrappier–Young entropy formula, but that very formula bounds the dimension of fast-stable leaves.)

3.3. Absence of crossings. Points $p$ not of the above kind have the property that if $n$ is sufficiently large then $f^{-n}(p)$ is never within $d_{pq}/\eta_n(q)$ of such a crossing. Absent the transversality condition, this failure together with simultaneous failure of a Lipschitz condition imposes additional constraints on the local geometry, which we can control using analyticity.

3.4. A sample of the dimension arguments. As noted above, transversality implies that every $p \in NL$ has a subsequence of preimages that accumulate exponentially fast to the set of crossings. To suggest why one should expect the “bad” set of points to have smaller Hausdorff dimension than $\Lambda$ we outline the proof of the following result.

**Proposition 5.** Under the assumptions of Theorem 2, NL carries less entropy than $\Lambda$.

**Proof.** The dynamical system in Theorem 2 acts affinely on slices, so a cross-section $\{x = \text{const.}\}$ of $f(M)$ consists of two convex sets in $D$. For some $x$ these must have disjoint projections (this is the only place where $\eta \equiv 2$ is used, and we do not need $\Lambda$ and $\mu$ to be constant). Therefore there is a Markov partition such that the $(x,y)$-projections of some of its elements do not contain any crossings. The subshift without any of the corresponding symbols has strictly smaller entropy than the shift over $f$ obtained from the Markov partition, whose entropy we denote by $H$ for the moment.

For a sequence that projects to a point $p$ in NL and an $N \in \mathbb{N}$ there is an $n \geq N$ such that $f^{-n}(p)$ is within $\eta^{-n}$ of a crossing point. This means that for any $\beta \in (0,1)$ the next $\beta n$ iterates $f^{-i}(p)$ lie in a Markov rectangle whose projection contains crossings, i.e., after some $e^{Hn}$ possible choices for the first $n$ entries there are only some $e^{h\beta n}$ choices for the next $\beta n$ entries. This allows for some $e^{Hn}e^{h\beta n}$ initial strings of length $n + \beta n$. Since

$$\frac{\log(e^{Hn}e^{h\beta n})}{(1 + \beta)n} = \frac{H + \beta h}{1 + \beta},$$

we obtain entropy at most $(H + h)/2 < H$ on NL.

4. Appendix: Fractal dimension

For completeness we define here the box dimension and the Hausdorff dimension of a set.
Definition 6. For a totally bounded metric space $Y$ denote by $N(\varepsilon)$ the minimum number of $\varepsilon$-balls needed to cover $Y$. The upper and lower box dimension of $Y$ are defined by

$$\dim B^U Y := \lim_{\varepsilon \to 0} \log N(\varepsilon) / \log 1/\varepsilon$$
and
$$\dim B^L Y := \liminf_{\varepsilon \to 0} \log N(\varepsilon) / \log 1/\varepsilon .$$

If $\dim B^U Y$ and $\dim B^L Y$ exist and agree, then we define the box dimension of $Y$ by $\dim B Y := \dim B^U Y$.

For a subset $U$ of a metric space, we let $\text{diam}(U)$ denote the diameter of the set $U$.

Definition 7. For a subset $Y$ of a metric space $X$ there is a unique number $\dim H Y$, called the Hausdorff dimension of $Y$, such that

$$\lim \inf_{\varepsilon \to 0} \left\{ \sum_{i=1}^\infty (\text{diam}(U_i))^s \ \mid \ Y \subset \bigcup_{i=1}^\infty U_i \text{ and } \sup_{i=1, \ldots, \infty} \text{diam}(U_i) \leq \varepsilon \right\}$$

$$= \begin{cases} \infty & \text{for } s < \dim H Y , \\ 0 & \text{for } s > \dim H Y . \end{cases}$$

Standard arguments give $\dim H Y \leq \dim B^U Y \leq \dim B^L Y$, and these inequalities may be strict.

The box dimension can also be defined in terms of covering sums, as is the Hausdorff dimension, with the only change being that the covering balls all have the same size. In order to estimate the box dimension, it suffices that the sizes of the covering balls tend to 0 along a geometric sequence.

References


Eberhard Hopf: Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung, Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-Physikalische Klasse 91 (1939), 261–304.


François Ledrappier, Lai-Sang Young: The metric entropy of diffeomorphisms. I. Characterization of measures satisfying Pesin’s entropy formula; II. Relations between entropy, exponents and dimension, Annals of Mathematics (2) 122 (1985), no. 3, 509–539; 540–574. MR (87i:58101a); MR (87i:58101b)


Lai-Sang Young: What are SRB measures, and which dynamical systems have them?, J. Statistical Physics 108 (2002), no. 5/6, 733–754. MR (2003g:37042)

Department of Mathematics, Tufts University, Medford, MA 02155
E-mail address: bhasselb@tufts.edu

Lund Institute of Technology, Lunds Universitet, Box 118, SE-22100 Lund, Sweden
E-mail address: Jorg.Schmeling@math.lth.se