WHAT IS...

a Horseshoe?

Michael Shub

The Smale *horseshoe* is the hallmark of chaos. With striking geometric and analytic clarity it robustly describes the homoclinic dynamics encountered by Poincaré and studied by Birkhoff, Cartwright-Littlewood, and Levinson. We give the example first and the definitions later.

Consider the embedding f of the disc Δ into itself exhibited in the figure. It contracts the semidiscs A, E to the semidiscs $f(A), f(E)$ in A and it sends the rectangles B, D linearly to the rectangles $f(B), f(D)$, stretching them vertically and shrinking them horizontally. In the case of D, it also rotates by 180 degrees. We don't really care what the image $f(C)$ of C is as long as it does not intersect the rectangle $B \cup C \cup D$. In the figure it is placed so that the total image resembles a horseshoe, hence the name.

It is easy to see that f extends to a diffeomorphism of the 2-sphere to itself. We also refer to the extension as f and work out its dynamics in Δ, i.e., its iterates f^n for $n \in \mathbb{Z}$.

Necessarily there are three fixed points p, q, s. The point q is a sink in the sense that all points $z \in A \cup E \cup C$ converge to q under forward iteration, $f^n(z) \to q$ as $n \to \infty$.

The points p, s are saddle points. If x lies on the horizontal through p, then f^n squeezes it to p as $n \to \infty$; while if y lies on the vertical through p, then the inverse iterates of f squeeze it to p. With respect to linear coordinates centered at p, $f(x, y) = (kx, my)$ where $(x, y) \in B$ and $0 < k < 1 < m$. Similarly, $f(x, y) = (-kx, -my)$ with respect to linear coordinates on D at s.

The sets

$$W^s = \{z : f^n(z) \to p \text{ as } n \to \infty\},$$
$$W^u = \{z : f^n(z) \to p \text{ as } n \to -\infty\}$$

are the stable and unstable manifolds of p. They intersect at r, which is what Poincaré called a *homoclinic point*. The figure shows these invariant manifolds only locally. Iteration extends them globally.

The key part of the dynamics of f happens on the *horseshoe*

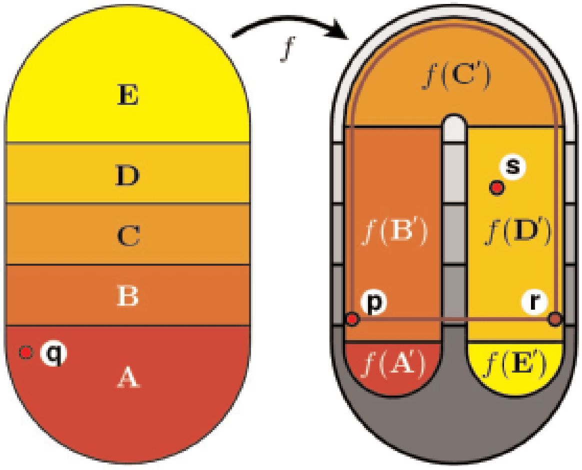
$$\Lambda = \{z : f^n(z) \in B \cup D \text{ for all } n \in \mathbb{Z}\}.$$

Everything there is explained as the "full shift on the space of two symbols". Take two symbols, 0 and 1, and look at the set Σ of all bi-infinite sequences $a = (a_n)$ where $n \in \mathbb{Z}$ and for each n, a_n is 0 or a_n is 1. Thus $\Sigma = \{0, 1\}^{\mathbb{Z}}$ is homeomorphic to the Cantor set. The map $\sigma : \Sigma \to \Sigma$ that sends $a = (a_n)$ to $\sigma(a) = (a_{n+1})$ is a homeomorphism called the shift map. It shifts the decimal point one slot rightward. Every dynamical property of the shift map is possessed equally by $f|_\Lambda$, because there is a homeomorphism $h : \Sigma \to \Lambda$ such that the diagram

---

Michael Shub has been a Research Staff Member at IBM's TJ Watson Research Center for twenty years. He is currently professor of mathematics at the University of Toronto. His email address is `michael.shub@utoronto.ca`.

The author would like to thank Charles Pugh, who suggested changes to an early version of this article and who produced the original version of the figure.



$$\begin{array}{ccc} \Sigma & \xrightarrow{\sigma} & \Sigma \\ h\downarrow & & \downarrow h \\ \Lambda & \xrightarrow{f} & \Lambda \end{array}$$

commutes. The conjugacy h is easy to describe. Given any $a \in \Sigma$, there is a unique $z \in \Lambda$ such that $f^n(z) \in B$ whenever $a_n = 1$, while $f^n(z) \in D$ whenever $a_n = 0$. Thus σ codes the horseshoe dynamics. For instance, $(\cdots 11.111 \cdots)$ codes the point p, $(\cdots 00.000 \cdots)$ codes s, while $(\cdots 111.0111 \cdots)$ codes r.

σ has 2^n periodic orbits of period n, and so must $f|_\Lambda$. The set of periodic orbits of σ is dense in Σ and so must be the set of periodic orbits of $f|_\Lambda$. Small changes of initial conditions in Σ can produce large changes of a σ-orbit, so the same must be true of $f|_\Lambda$. In short, due to conjugacy, the chaos of σ is reproduced exactly in the horseshoe.

The utility of Smale's analysis is this: *every* dynamical system having a transverse homoclinic point, such as r, also has a horseshoe containing r and thus has the shift chaos. Nowadays, this fact is not hard to see, even in higher dimensions. The mere existence of a transverse intersection between the stable and unstable manifolds of a periodic orbit implies a horseshoe. In the case of flows, the corresponding assertion holds for the Poincaré map. To recapitulate,

$$\text{transverse homoclinicity} \Rightarrow \text{horseshoe} \Rightarrow \text{chaos}.$$

Since transversality persists under perturbation, it follows that so does the horseshoe and so does its chaos.

The analytical feature of the horseshoe is *hyperbolicity*, the squeeze/stretch phenomenon expressed via the derivative. The derivative of f stretches tangent vectors that are parallel to the vertical and contracts vectors parallel to the horizontal, not only at the saddle points, but uniformly throughout Λ. In general, hyperbolicity of a compact invariant set such as Λ in any dimension is expressed in terms of expansion and contraction of the derivative on subbundles of the tangent bundle. Smale unified such examples as the horseshoe and the geodesic flow on manifolds of negative curvature, defining what is now called uniformly hyperbolic dynamical systems. The study of these systems has led to many fruitful discoveries in modern dynamical systems theory.

David Ruelle has called Smale's 1967 article [3] "a masterpiece of mathematical literature". It is still worth reading today. Hyperbolic dynamics flourished in the 1960s and 1970s. Anosov proved the stability and ergodicity of the globally hyperbolic systems that now bear his name. Sinai initiated the more general investigation of the ergodic theory of hyperbolic dynamical systems, and in particular showed that the Markov partitions of Adler and Weiss could be constructed for all hyperbolic invariant sets, thus giving a coding similar to the two-shift coding for the horseshoe. This work was carried forward by Ruelle and Bowen. The invariant measures they found, now called Sinai-Ruelle-Bowen (SRB) measures, describe the asymptotic dynamics of most Lebesgue points in the manifold, even for dissipative sytems. Uniformly hyperbolic dynamical systems are remarkable. They exhibit chaotic behaviour. By the work of Anosov, Smale, Palis, and Robbin, they are structurally stable; that is, the dynamics of a perturbation of a uniformly hyperbolic system is topologically conjugate to the original. By the work of Sinai, Ruelle, and Bowen, they are described statistically.

In the early days of the 1960s it was hoped that uniformly hyperbolic dynamical systems might be in some sense typical. While they form a large open set on all manifolds, they are not dense. The search for typical dynamical systems continues to be a great problem. For progress see the survey [1]. Hyperbolic periodic points, their global stable and unstable manifolds, and homoclinic points remain some of the principal features of, and tools for, understanding the dynamics of chaotic systems.

Indeed, transversal homoclinic points are proven to exist in many of the dynamical systems encountered in science and engineering from celestial mechanics, where Poincaré first observed them, to ecology and beyond.

References

[1] Christian Bonatti, Lorenzo J. Diaz, and Marcelo Viana, *Dynamics beyond Uniform Hyperbolicity: A Global Geometric and Probabilistic Approach,* Encyclopedia Math. Sci., Springer, 2004.

[2] Michael Shub, *Global Stability of Dynamical Systems*, Springer, 1986.

[3] Stephen Smale, Differentiable dynamical systems, *Bull. Amer. Math. Soc.* **73** (1967), 747–817.

[4] ____ , Finding a horseshoe on the beaches of Rio, *Math. Intelligencer* **20** (1998), 39–44.