HOMOCLINIC ORBITS AND CHAOTIC BEHAVIOR IN CLASSICAL MECHANICS

PEYAM RYAN TABRIZIAN

DECEMBER 1, 2008

Abstract

This paper is a study of one of the most beautiful phenomena in dynamical systems: homoclinic orbits. We will first define what a homoclinic orbit is, then we will study some of the properties of homoclinic points, using methods from symbolic dynamics. In particular, we will prove that, under certain assumptions, the existence of one homoclinic point implies the existence of a whole Cantor set of homoclinic points. We will also implicitly construct the famous Smale horseshoe. Finally, we will present a topic in physics that is closely related to homoclinic orbits, namely the restricted 3-body problem.

Introduction

Homoclinic orbits have been introduced by $Poincaré$ more than a century ago, and since then, they became a fundamental tool in the study of chaos. In this paper, we are going to study the chaotic dynamics in a neighborhood of a homoclinic orbit. We will see how the analysis of those orbits, which may seem impossible at first, can be reduced to the study of the space of sequences on $\{0, 1, ..., N\}$ (where $N \in \mathbb{N}$ might be equal to ∞), which is pretty straightforward . Our study will closely follow the texts of Ekeland [1] (for the first section) and of Moser [3] (for the remaining sections).

1 Homoclinic orbits

Let ϕ be a C^1-diffeomorphism, and O a fixed point of ϕ. We assume that O is hyperbolic, i.e. $d\phi$ has real eigenvalues λ_1 and λ_2 such that $0 < |\lambda_1| < 1 < |\lambda_2|$. This guarantees the existence of a stable manifold S and an unstable manifold U for ϕ, intersecting in O. Also, we assume that ϕ is an area-preserving transformation, i.e. the determinant of $D\phi$ is 1.

By definition, of S and U, we have:

$$S = \left\{ z \mid \lim_{n \to \infty} \phi^n(z) = O \right\}$$

and

$$U = \left\{ z \mid \lim_{n \to -\infty} \phi^n(z) = O \right\}$$

Basically, S is the set of points that tend to O (under iteration of ϕ) and U is the set of points that emanate from O.

We assume that U and S further intersect in a point H, called *homoclinic point* (see Figure 1). More generally, a homoclinic point is defined as follows:

Definition. *If $r \neq p$ and $\phi^k(r)$ approaches one periodic orbit $\{p, \phi(p), \phi^2(p), ..., \phi^m(p) = p\}$, as $k \to \pm\infty$, then r is called* ***homoclinic point***

Note. *There is also what is called a* ***heteroclinic point*** *in case there are 2 fixed points, p and q, but we won't deal with that situation in this paper.*

$H \in S$, hence $H_1 = \phi(H), H_2 = \phi^2(H)$, etc. all tend to O. We're about to prove the following remarkable fact:

Claim. *$\forall n$, a branch of U goes through H_n (see Figure 2)*

Proof. Because $H \in U$, we have $H_1 = \phi(H) \in \phi(U)$. Hence, it suffices to prove that $\phi(U) \subseteq U$. Let $M \in \phi(U)$. By continuity of ϕ, if M is close to H_1, we have that $M = \phi(P)$, where P is a point in U close to H_1. And so $M^{-1} = P$ (where, in general, $M^{-k} = \phi^{-k}(M)$). Similarly, M^{-1} and P have the same preimages by ϕ,

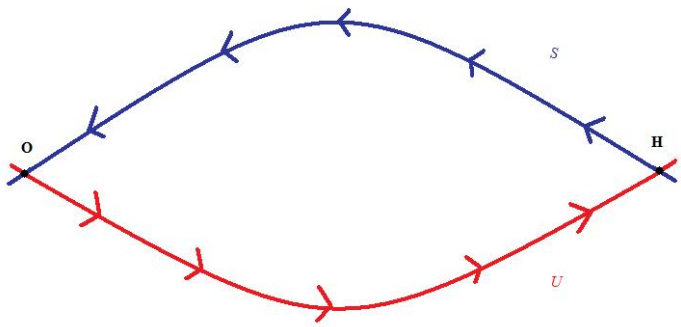

Figure 1: The stable and unstable curves, intersecting in H

which gives $M^{-2} = P^{-1}$ (with $P^{-k} = \phi^{-k}(P)$). In the same way, we conclude that $M^{-3} = P^{-2}$, $M^{-4} = P^{-3}$, and so on. But because $P \in U$, we have, by definition of U,

$$\lim_{n\to\infty} P_{-n} = O \text{ and } \lim_{n\to\infty} M_{-n} = O$$

But this implies, by definition of U, that $M \in U$. Because M was arbitrary, this proves our claim. □

Corollary. *There is a doubly infinite sequence of homoclinic points*

Proof. Just take the sequence $U_k = H_k, k \in \mathbb{Z}$, with $H_k = \phi^k(H)$! □

Moreover, because $D\phi$ has determinant 1, ϕ also preserves orientations, hence, one can prove using the intermediate value theorem that, for all n, between H_n and H_{n+1} there is another point of intersection between U and S, let's call those points K_n (with $K_0 = K$). Obviously, the K_n are homoclinic points as well, so the sequence $V_k = K_k$ is another sequence of new homoclinic points.

So we started with one homoclinic point, and we arrived at two doubly-infinite sequences of homoclinic points! There surely must be something chaotic going on here!

Let's try connect the H_i. We have to keep the following in mind:

1. **U cannot intersect itself**. If it did, then by reverse iteration of ϕ, this would mean that U would intersect itself in a neighborhood of O, which is not possible by a theorem on stable/unstable manifolds.

2. **U and S cannot intersect themselves tangentially**. We would get a similar contradiction.

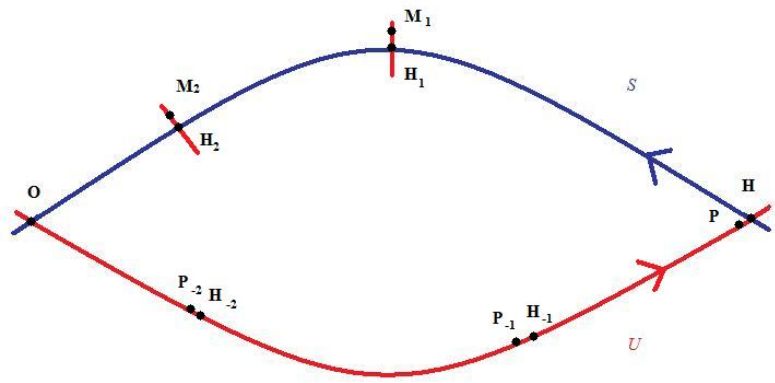

Figure 2: $\forall n$, a branch of U goes through H_n

3. **The areas of each region exterior to S must be equal. Same for the regions exterior to U**. This is because we assume that ϕ is an area-preserving transformation.

4. **U and S must be smooth curves**

5. $\forall n, H_n H_{n+1} < H_{n-1} H_n$

See Figure 3 (at the end of this section) for a picture illustrating those ideas.

The most natural way of connecting the H_i is in such a way that the connected path always remains on one side of the picture, namely on the side of S (see Figure 4). But this is wrong because, by definition of U and S, any point N very close to S but not on S will approach O under iteration of ϕ, but, since $N \notin S$, we have that, eventually, N will leave O under iteration of ϕ. Hence, there must be at least part of U that lands back on S, close to H.

Also note that, under those assumptions, U becomes finer and longer under iteration of ϕ. This implies that U will intersect S in infinitely many new points.

And this is not all of it! Namely, the new points of intersection between U and S are new homoclinic points, and we can repeat this procedure for all points of intersection. The result is a pretty chaotic picture, which means that the study of homoclinic orbits is definitely not as straightforward as people in the past century thought it would be.

And this is still not it! Because so far, we only dealt with one side of the picture. There is still one branch of U and one branch of S we didn't connect. If we assume that those two branches intersect in a new (homoclinic) point H', we get the same picture as before. So we get, again, a bunch of new homoclinic points (see Figure 6). However,

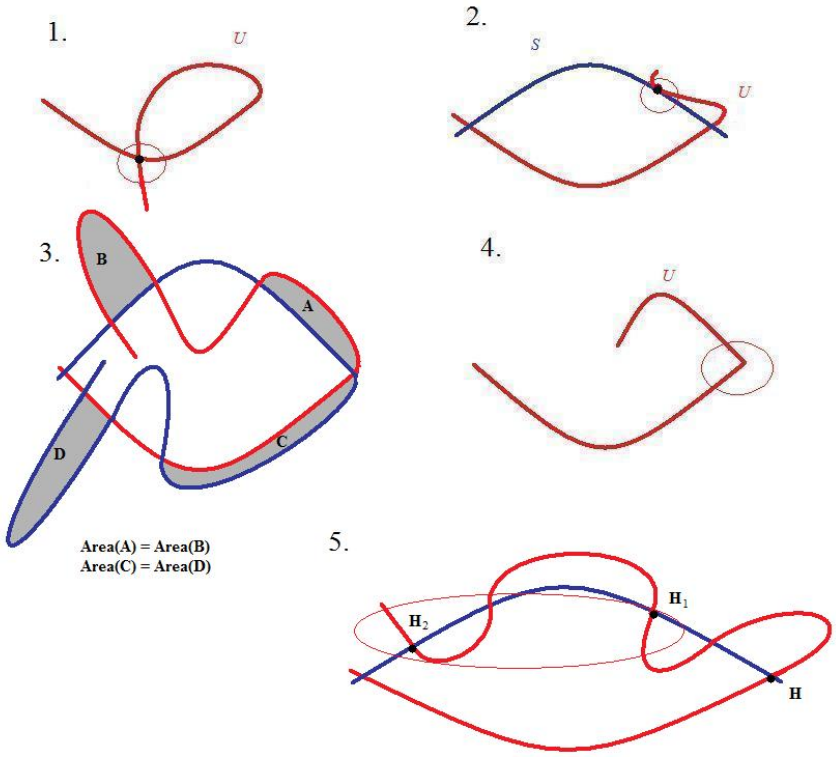

Figure 3: Things U and S cannot do (errors circled in red)

we will not deal with this kind of situation in this paper, because this problem is easily reduced to the problem we are considering. Also, it is physically more realistic to assume that the two branches we didn't connect blow off to infinity (which illustrates the impossibility of going backwards in time).

2 Deeper study of the problem

From now on, for more legibility (in particular in order to not confuse points with sets), H will be denoted by r. U, S, O remain the same as in the previous section.

Consider a quadrilateral R, centered at r, such that one of its sides concides with part of U and another one with part of S, and the other sides are parallel to the tangents of U resp. S at r (see Figure 5). We wish to study the dynamics of ϕ in R. For this, we must consider the set of points which don't eventually escape R. Thus, for $q \in R$ Let $k = k(q)$ be the smallest positive integer for which $\phi^k(q) \in R$, if it exists. Let $D(\widetilde{\phi})$ be the set of $q \in R$ for which such a $k > 0$ exists (this guarantees that we are studying

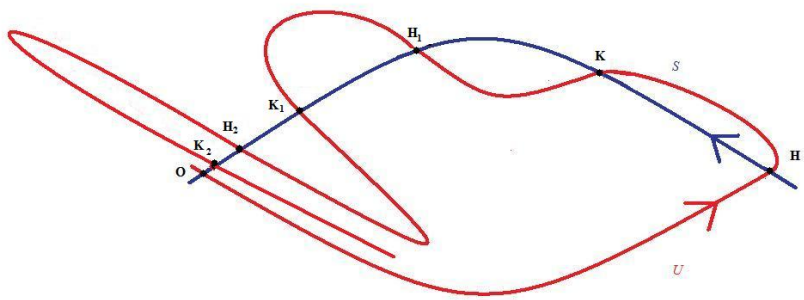

Figure 4: First attempt to connect the H_i

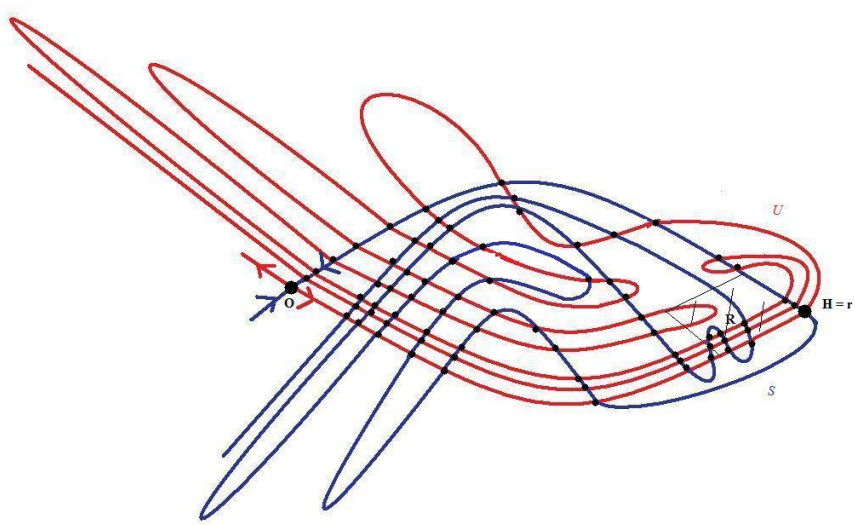

Figure 5: Chaotic behavior of homoclinic orbits (Part I)

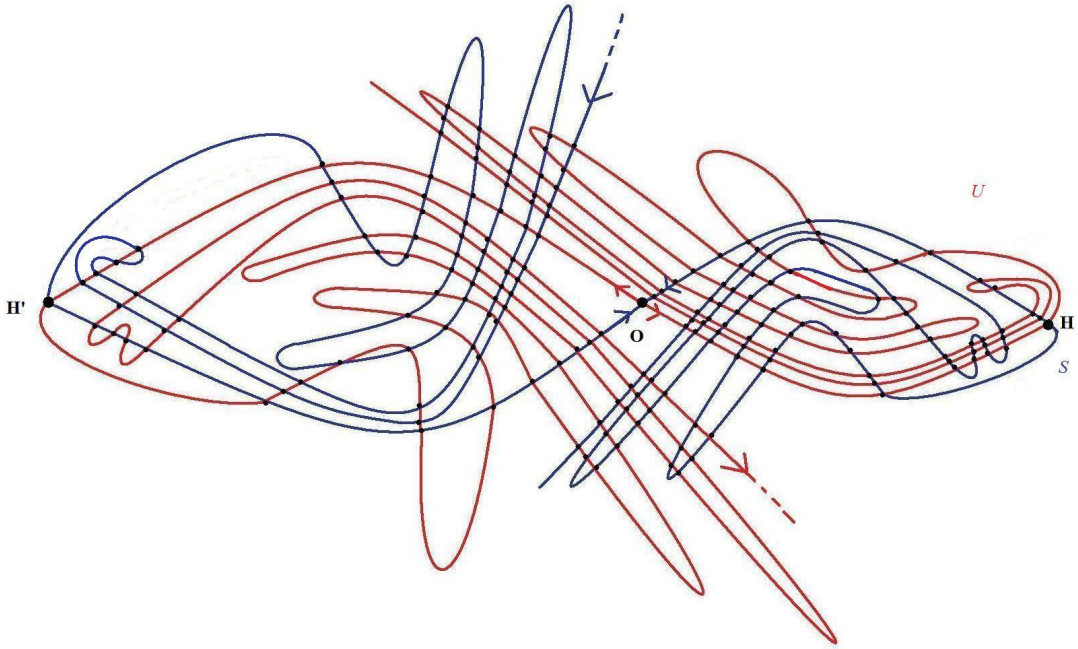

Figure 6: Chaotic behavior of homoclinic orbits (Part II)

iterates that don't blow off to infinity) and set:

$$\widetilde{\phi}(q) = \phi^k(q), q \in D(\widetilde{\phi})$$

$\widetilde{\phi}$ is called the **transversal (or return) map of ϕ for R.** We assume that $D(\widetilde{\phi})$ is not empty.

3 Three crucial properties

$\widetilde{\phi}$ satisfies three crucial properties on R, which will be crucial in our study of homoclinic orbits.

Property 1. *Let $A = \{1, 2, 3, ..., N\}$ if $N \leq \infty$, or $\mathbb{N}$ if $N = \infty$. Then, $\forall a \in A$, there exist U_a and V_a, which are disjoint horizontal resp. vertical strips in $D(\widetilde{\phi})$, and $\widetilde{\phi}$ maps V_a homeomorphically to U_a, i.e.*

$$\widetilde{\phi}(V_a) = U_a$$

Furthermore, the vertical boundaries of V_a are mapped onto the vertical boundaries of U_a, and similarly for the horizontal boundaries.

Definition. *Given μ such that $0 < \mu < 1$, a curve $y = u(x)$ is a **horizontal curve** if $0 \leq u(x) \leq 1$ for $0 \leq x \leq 1$, and*

$$|u(x) - u(x')| \leq \mu \, |x - x'| \, , 0 \leq x \leq x' \leq 1$$

Definition. *If $u_1(x), u_2(x)$ are two horizontal curves, and if*

$$0 \leq u_1(x) < u_2(x) \leq 1$$

then the set

$$U = \{(x, y) | 0 \leq x \leq 1, u_1(x) \leq y \leq u_2(x)\}$$

*is called a **horizontal strip***

Definition. *If U is a horizontal strip*

$$d(U) = \max_{0 \leq x \leq 1} (u_2(x) - u_1(x))$$

*is called the **diameter** of U*

Very similar definitions are valid for vertical strips.

Now because $\widetilde{\phi}$ is C^1 (because ϕ is), $\widetilde{\phi}$ enjoys another useful property. To explain this property, first we represent $\widetilde{\phi}$ in coordinates by

$$\begin{cases} x_1 = f(x_0, y_0) \\ y_1 = g(x_0, y_0) \\ (x_1, y_1) = \widetilde{\phi}(x_0, y_0) \end{cases}$$

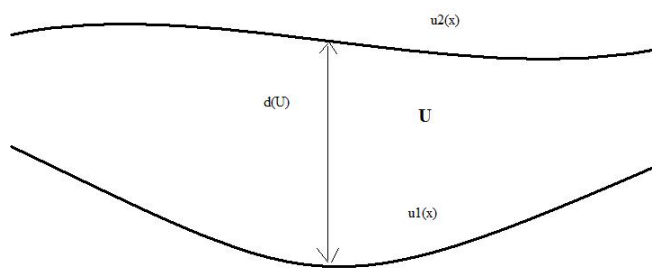

Figure 7: Two horizontal curves, a horizontal strip, and its diameter

Now in particular, $d\widetilde{\phi}$ maps the tangent vector (ξ_0, η_0) at (x_0, y_0) into the tangent vector (ξ_1, η_1) at (x_1, y_1), with:

$$\begin{cases} \xi_1 = f_x \xi_0 + f_y \eta_0 \\ \eta_1 = g_x \xi_0 + g_y \eta_0 \end{cases}$$

Property 2. *There exists μ with $0 < \mu < 1$ such that, if we define the bundle of sectors S^+ and S^- as:*

$$\begin{cases} S^+ := \{(\xi, \eta) | \, |\eta| \leq \mu \, |\xi|\} \\ S^- := \{(\xi, \eta) | \, |\xi| \leq \mu \, |\eta|\} \end{cases}$$

Where S^+ is defined over $\bigcup_{a \in A} V_a$, and S^- over $\bigcup_{a \in A} U_a$, then we have:

$$\begin{cases} d\widetilde{\phi}(S^+) \subseteq S^+ \\ d^{-1}\widetilde{\phi}(S^-) \subseteq S^- \end{cases}$$

Moreover, if $(\xi_1, \eta_1) \in S^-$, and $(\xi_0, \eta_0) = d\widetilde{\phi}^{-1}(\xi_1, \eta_1)$, then we have:

$$|\eta_0| \geq \mu^{-1} \, |\eta_1|$$

Note. *Property 2 basically reflects the fact that $\widetilde{\phi}$ is a very unstable map under iteration, because notice that the horizontal components of a tangent vector get amplified at least by μ^{-n} under $d\widetilde{\phi}^n$ (with $n \geq 1$), and the vertical components by μ^{-n} under $d\widetilde{\phi}^{-n}$.*

It turns out that Properties 1 and 2 imply the following Property 3.

Property 3. *As above, let A be the set $\{1, 2, 3, ..., N\}$ if $N < \infty$, or $\mathbb{N}$ if $N = \infty$. If V is a vertical strip in $\bigcup_{a \in A} V_a$, then, $\forall a \in A$, we have that:*

$$\widetilde{\phi}^{-1}(V) \cap V_a = V'_a$$

is a vertical strip (in particular nonempty), and for some ν with $0 < \nu < 1$:

$$d(V'_a) \leq \nu d(V_a)$$

Similarly, if U is a horizontal strip in $\bigcup_{a \in A} U_a$, then, $\forall a \in A$, we have that:

$$\widetilde{\phi}(U) \cap U_a = U'_a$$

is a horizontal strip, and finally:

$$d(U'_a) \leq \nu d(U_a)$$

Properties 1 and 3 will be crucial later on, when we will prove the main theorem of this paper.

4 Construction of the horizontal/vertical strips

The horizontal and vertical strips in Property 1 are so important for the study of homoclinic orbits that, in this section, we will construct them. This section is independent of the remaining sections. Also note that, in this section, we will be mainly concerned with our original map ϕ, **not** with the transversal map $\widetilde{\phi}$.

First of all, let's introduce again local coordinates, O being the origin of the coordinate system, and U, S the coordinate axes. If $(x_1, y_1) = \phi(x_0, y_0)$, the image of (x_0, y_0) under ϕ, then we have, in the neighborhood of O, that:

$$\begin{cases} x_1 = f(x_0, y_0) \\ y_1 = g(x_0, y_0) \end{cases}$$

with f, g continuously differentiable in this neighborhood, and $f(0, y) = g(x, 0) = 0$, $f_x(0, 0) = \lambda_1$, $g_y(0, 0) = \lambda_2$.
Let's denote by Q the square

$$Q : 0 \leq x \leq a, 0 \leq y \leq a$$

with $a > 0$ small enough. Also, it is natural to denote the kth iterate of (x, y) under ϕ by (x_k, y_k).

Because we can choose the sides of R to be as small as we like, there exist positive integers l, m, such that

$$\begin{cases} \phi^l(r) \in Q \\ \phi^{-m}(r) \in Q \end{cases}$$

Fix l, m, and a, and let δ be such that $0 < \delta < a$, and let b, c such that $0 < b, c < \frac{a}{2}$ and $\phi^l(r)$ has coordinates $(0, 2b)$ and ϕ^{-m} has coordinates $(2c, 0)$. Because we can

choose the sides of R to be as small as we like, and because of continuity of ϕ we can assume that:

$$\begin{cases} \phi^l(R) \in \{0 \leq x \leq \delta, b \leq y \leq a\} \\ \phi^{-m}(R) \in \{c \leq x \leq a, 0 \leq y \leq \delta\} \end{cases}$$

We further assume that:

$$A_0 = \phi^l(R)$$

has one of its sides coinciding with $x = 0$, and one adjacent side intersecting $x = 0$ transversally. And, by choosing the sides of R even smaller if necessary, we may suppose that both sides adjacent to $x = 0$ are of the form $y = h_1(x)$, with h_1 a C^1 function. Similarly, we assume that:

$$A_1 = \phi^{-m}(R)$$

has two sides that can be written in the form $x = h_2(y)$, with h_2 a C^1 function (see Figure 8).

We can now define the transversal map ψ from A_0 to A_1 as follows:
If $q \in A_0$ and $\exists k > 0$ such that $\phi^k(q) \in A_1$, and $\forall 0 < j < k, \phi^j(q) \in Q$, then we say that $q \in D(\psi)$, and we set:

$$\psi(q) = \phi^k(q) \in A_1$$

with k the smallest such $k \geq 0$.

Note. *In $D(\psi) \cap D(\widetilde{\phi})$, we clearly have*

$$\widetilde{\phi} = \phi^m \psi \phi^l$$

Now here comes a remarkable fact:

Claim. *For ψ, defined in $D(\psi) \in A_0$, $\psi(\psi(D)) = \psi^2(D)$ intersects A_1 in infinitely many strips U_k (k = 1,2,...). These strips connect the opposite sides of A_1*

Proof. For sufficiently small a, we have in Q, that $x_1 \geq \sqrt{\lambda_1}x, y_1 \leq \sqrt{\lambda_2}y$, so that the images under ϕ^k of the two curves $y = h_1(x)$ intersect the domain $\{c \leq x \leq a, 0 \leq y \leq \delta\}$ in a curve connecting $x = 0$ and $x = a$. Thus, for large k, the domains $\phi^k(A_0) \cap A_1$ will consist of infinitely many quadrilaterals $\widetilde{U_k}$ connecting opposite sides of A_1 (see Figure 9). Similarly, $\widetilde{V_k} = \psi^{-1}(\widetilde{U_k})$ are vertical strips in A_0, and by definition, we have:

$$\psi(\widetilde{V_k}) = \widetilde{U_k}$$

Now set:

$$\begin{cases} U_k = \phi^m(\widetilde{U_k}) \\ V_k = \phi^{-l}(\widetilde{V_k}) \end{cases}$$

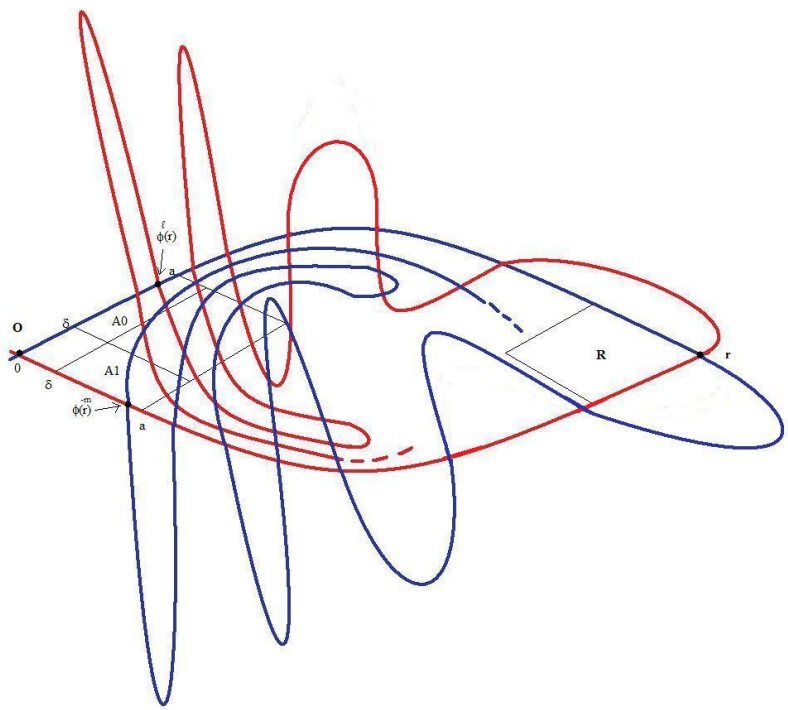

Figure 8: Construction of A_0, A_1

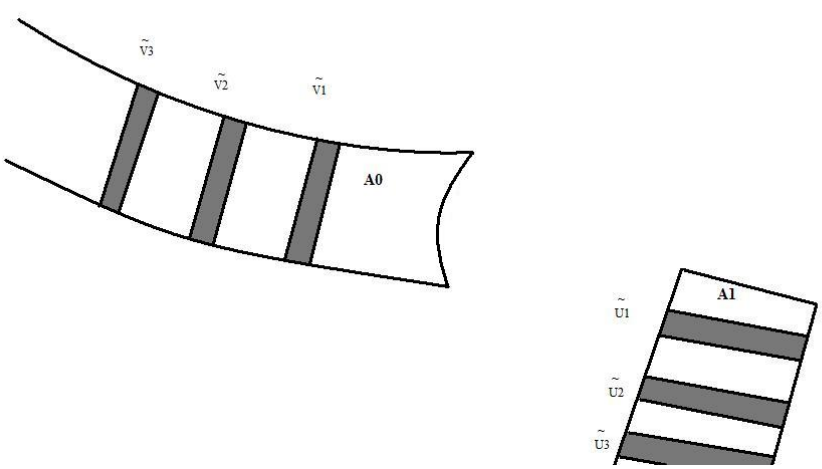

Figure 9: Construction of $\tilde{U}_i, \tilde{V}_i$

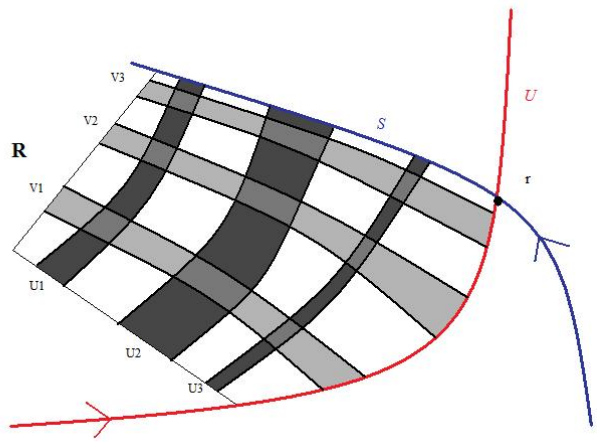

Figure 10: Construction of U_i, V_i

and we obtain similar strips in R (see Figure 10), which satisfy:

$$\widetilde{\phi}(V_k) = \phi^m(\psi(\widetilde{V_k})) = U_k$$

and we get our desired horizontal/vertical strips.

□

5 Main theorem

Here is the main theorem concerning homoclinic orbits.

Main theorem. *There exists an invariant set I in R and a homeomorphism τ of S (the set of sequences on N symbols) into I such that*

$$\widetilde{\phi}\tau = \tau\sigma$$

Where $\sigma : S \to S$ is defined by $(\sigma(s))_n = s_{n-1}$

Proof. We will prove the case where $N < \infty$. The case $N = \infty$ will be dealt with in the next section.

First, define inductively for $n \geq 1$,

$$V_{s_0 s_{-1} \ldots s_{-n}} = V_{s_0} \cap \widetilde{\phi}^{-1}(V_{s_{-1} \ldots s_{-n}})$$

Because of Property 3, and by induction, one can show that these are vertical strips. Also, one can easily see that:

$$d(V_{s_0 s_{-1} \ldots s_{-n}}) \leq \nu d(V_{s_{-1} \ldots s_{-n}}) \leq \nu^n d(V_{s_{-n}}) \leq \nu^n$$

Hence

$$\lim_{n\to\infty} d(V_{s_0 s_{-1}\ldots s_{-n}}) = 0$$

By the definition of $V_{s_0 s_{-1}\ldots s_{-n}}$, we have:

$$\begin{aligned} V_{s_0 s_{-1}\ldots s_{-n}} &= \left\{p \in D(\widetilde{\phi}) \mid \widetilde{\phi}^k(p) \in V_{s_{-k}} (k = 0, 1, ..., n)\right\} \\ &= \left\{p \in D(\widetilde{\phi}) \mid \widetilde{\phi}^{-k}(p) \in V_{s_k} (k = 0, -1, ..., -n)\right\} \end{aligned}$$

Hence:

$$V_{s_0 s_{-1}\ldots s_{-n}} \subset V_{s_0 s_{-1}\ldots s_{-n} s_{-n+1}}$$

We need a little lemma (which follows immediately from the compacity of the family of vertical curves):

Lemma. *If $V^1 \supset V^2 \supset V^3 \supset$ ... is a sequence of nested vertical strips, and if $d(V^k) \to 0$ as $k \to \infty$, then*

$$\bigcap_{k=1}^{\infty} V^k$$

defines a vertical curve

From this lemma, it follows that:

$$V(s) = \bigcap_{n=0}^{\infty} V_{s_0 s_{-1}\ldots s_{-n}} = \left\{p \in D(\widetilde{\phi}) \mid \widetilde{\phi}^{-k}(p) \in V_{s_k} (k = 0, -1, ...)\right\}$$

defines a vertical curve, depending on the left half of the sequence s.

Similarly, define inductively, for $n \geq 2$,

$$U_{s_1 s_2 \ldots s_n} = U_{s_1} \cap \widetilde{\phi}(U_{s_2 \ldots s_n})$$

which, again, are nested horizontal strips whose diameters tend to 0 as $k \to \infty$

Hence, since $\widetilde{\phi}(V_{s_k}) = U_{s_k}$, we see that:

$$\begin{aligned} U(s) = \textstyle\bigcap_{n=1}^{\infty} U_{s_0 s_1 \ldots s_n} &= \left\{p \in D(\widetilde{\phi}) \mid \widetilde{\phi}^{-k+1}(p) \in U_{s_k} (k = 0, 1, ...)\right\} \\ &= \left\{p \in D(\widetilde{\phi}) \mid \widetilde{\phi}^{-k}(p) \in V_{s_k} (k = 0, 1, ...)\right\} \end{aligned}$$

defines a horizontal curve, depending on the right half of s.

Here's another lemma (whose proof, which uses the intermediate value theorem, is left to the reader):

Lemma. *A horizontal and a vertical curve intersect precisely in one point*

Hence, by the above lemma, the intersection

$$V(s) \cap U(s) = \left\{ p \in D(\widetilde{\phi}) \mid \widetilde{\phi}^{-k}(p) \in V_{s_k} (k = 0, \pm 1, \pm 2...) \right\}$$

defines precisely one point, say w, in I (the set of points of intersections of those horizontal and vertical curves, which is clearly invariant under $\widetilde{\phi}$)

Finally, define $\tau : S \to I$ by $\tau(s) = w$, where $s = (...s_1 s_0 s_1 ...) \in S$

From construction, it follows that if $\tau(s) = w$, then $\tau(\sigma(s)) = \widetilde{\phi}(w)$

Hence

$$\tau\sigma = \widetilde{\phi}\tau$$

And it is not very hard to prove that τ is a homeomorphism.

□

Note. *If $N < \infty$, then it can be shown that $\tau(S)$ is a Cantor set. It is therefore uncountable, but has Lebesgue measure zero.*

Note. *We can introduce a topology on S by taking as neighborhood basis of*

$$s = (..., s_{-1}, s_0, s_1, ...)$$

the sets

$$U_j = \{t \in S \mid t_k = s_k, |k| < j\}$$

6 An extension

To deal with the case $N = \infty$, we now discuss an extension of S.

We assume that $A = \mathbb{N}$ and that V_a is ordered according to increasing x-coordinates with increasing a (this ordering is possible, because the V_a are disjoint). Then V_a tend to a vertical curve V_∞, which agrees with $x = 1$. Similarly for $y = 1$.

Now for α, β integers satisfying $\alpha \leq 0, \beta \geq 1$, let

$$s = (\infty, s_{\alpha+1}, ..., s_{\beta-1}, \infty), s_k \in A$$

Note. *The case $\alpha = 0, \beta = 1$ corresponds to the symbolic element (∞, ∞)*

Now if $\alpha = -\infty$, $\beta = \infty$, we get elements in S. Elements with $\alpha = -\infty, \beta < \infty$ or $\alpha > -\infty, \beta = \infty$, correspond to half-infinite sequences. These sequences form the elements of the space $\overline{S} \supset S$.

Now extend σ to $\overline{\sigma}$, in the domain:

$$D(\overline{\sigma}) = \left\{ s \in \overline{S}, s_0 \neq \infty \right\}$$

The range $Im(\overline{\sigma})$ of $\overline{\sigma}$ is given by:

$$R(\overline{\sigma}) = \left\{ s \in \overline{S}, s_1 \neq \infty \right\}$$

Note. *We can introduce a topology on $\overline{S}$ by taking as a neighborhood basis of $s = (..., s_{-1}, s_0, ..., s_{\beta-1}, \infty), \alpha = \infty, \beta < \infty$ the sets*

$$U_K = \left\{ t \in \overline{S} \mid t_k = s_k, -K \leq k < \beta, t_\beta \geq K \right\}$$

The following theorem is a variation of our main theorem:

Theorem. *τ can be extended to $\overline{\tau} : \overline{S} \to \overline{I}$ such that*

$$\widetilde{\phi}\overline{\tau} = \overline{\tau}\overline{\sigma}$$

if both sides are restricted to $D(\overline{\sigma})$

Note. *$\overline{\tau}\overline{S}$ is compact with respect to a certain topology, but not anymore invariant under $\widetilde{\phi}$. The new sequences correspond to solutions which 'escape' the homoclinic orbit for positive or negative time.*

7 Consequences

Now witness the power of the previous theorem! There are numerous corollaries, which are very useful in the understanding of the problem. In this section, we'll list and prove some of them.

Corollary. *Homoclinic points are dense in R*

Proof. Just take sequences that end with ∞ (they correspond to homoclinic points). Since those sequences are dense in $\overline{S}$, we are done. □

Corollary. *There exist infinitely many points whose orbits are dense in R*

Proof. Just take a sequence which contains all blocks of sequences. Since there are infinitely many such sequences, we get infinitely many points whose orbits are dense in R. □

Corollary. *Periodic orbits are dense in R*

Proof. Given a sequence $s = (..., s_{-1}, s_0, s_1, ...) \in S$, consider, for any n, the periodic sequence $s' = (s_n, s_{-n}, s_{-n+1}, ..., s_0, ..., s_n, s_{-n}, s_{-n+1}, ...) \in S$, which is clearly periodic, which gives a periodic sequence that is arbitrarily close to s. □

Corollary. *Points in R enjoy the property of **topological entropy** , that is, in any neighborhood of $r \in R$, there exists a point $r' \in R$ such that, under iteration of $\widetilde{\phi}$, those two points do not lie in the same neighborhood any more*

Proof. Given a sequence $s = (..., s_{-1}, s_0, s_1, ...) \in S$ that corresponds to a point r, just take r' to correspond to the sequence $s' = (..., s_{-1}, s_0, s_{-1}, ..., s_n, s'_{n+1}, s'_{n+2}, ...)$, for some appropriate n, and such that, for all $k > 0$, $s'_{n+k} \neq s_{n+k}$, which geometrically means that r and r' first lie in the same horizontal/vertical strips, but that eventually, under iteration of $\widetilde{\phi}$, they end up in completely different horizontal/vertical strips. □

Corollary. *We have **transitivity** : Given two neighborhoods U_1, U_2 in R, there is a point $r \in U_1$ such that r eventually lies in U_2 under iteration of $\widetilde{\phi}$*

Proof. WLOG, assume that those neighborhoods correspond to the horizontal/vertical strips, then the sequence $s = (..., s_{-1}, s_0, s_1, s_1, s_1, ...) \in S$, with $s_1 \neq s_0$ corresponds to a point that first lies in the strip V_0, but then lies in the strip U_1, and stays in that strip for an indefinite amount of time. □

Corollary. *Taking the definition of Smale [2], $\widetilde{\phi}$ is chaotic on R.*

Proof. Follows directly from the fact that periodic orbits are dense, from topological entropy, and from transitivity.

□

Corollary. *ϕ does not possess a real analytic integral in R*

Definition. *Given a diffeomorphism ϕ in a domain $D \subset \mathbb{R}^2$, a real-valued function f in D is an **integral** of ϕ if f is not constant, and $f(\phi(p)) = f(p)$*

Proof. Suppose f is a continuous function on $D(\widetilde{\phi})$ such that $f(\phi(p)) = f(p)$. Hence $\forall k > 0, f(\phi^k(p)) = f(p)$. Choosing $\phi^k(p)$ to be a dense orbit (this is possible by one of the above corollaries), we get, by continuity, that $f(q) = f(p)$, for all $q \in I$.

Now assume that $f \in C^1(R)$, such that $f(\phi(p)) = f(p)$ on I. We'll show that $f_x = f_y = 0$ on I.

In fact, if $p^* = \tau(s^*)$ is an arbitrary point in I, then by the first of the claims in the next section, p^* lies on a horizontal curve $U(s^*)$ and vertical curve $V(s^*)$. By modifying the left tail of the sequence s^*, we can choose a sequence of points $q_\nu \in U(s^*) \cap I, q \neq p^*$, converging to p^*. Thus, if we parametrize $U(s^*)$ by x, the directional derivative $D_x f$ of f along the tangent of $U(s^*)$ at p^* can be expressed limit of a difference quotient of f at p^*.

Since $f(p) = f(q_\nu)$, it follows that $D_x(f) = 0$ at p^*, and by the same argument, the directional derivative $D_y(f)$ along the tangent of $V(s^*)$ vanishes.

Since $D_x(f)$, $D_y(f)$ are linearly independent, we conclude that $f_x = f_y = 0$ on I.

And if $f \in C^\infty(Q)$, then repeat this argument, and conclude that all derivatives of f vanish on I.

Thus, if f is real analytic, it is the zero function. □

8 Further properties

The invariant set $I = \tau(S)$ enjoys further geometric properties, which we will state, but not prove. For this, we need a definition:

Definition. *A set P is said to be **hyperbolic** if one can associate with every point $p \in P$ two linearly independent lines L_p^+, L_p^- in the tangent space at p, such that L_p^+, L_p^- vary continuously with $p \in P$, and, furthermore*

$$d\phi L_p^{\pm} = L_{\phi(p)}^{\pm}$$

and that, with some constant $\lambda > 1$*, and the norm* $|\zeta|$ = *max* $(|\xi|, |\eta|)$*, we have:*

$$\begin{cases} |d\phi(\zeta)| \geq \lambda|\zeta| & for \zeta \in L_p^+ \\ |d\phi^{-1}(\zeta)| \geq \lambda|\zeta| & for \zeta \in L_p^- \end{cases}$$

Let Δ be the Jacobian determinant of $d\phi$ at a point p. Also recall the definition of μ in Property 3.

Claim. *If* $\Delta, \Delta^{-1} \leq \frac{1}{2}\mu^{-2}$*, then I is a hyperbolic set*

Even more is true:

Claim. *If* $0 \leq \mu \leq \frac{1}{2}$ *min*$(|\Delta|^{\frac{1}{2}}, |\Delta|^{-\frac{1}{2}})$*, then U(s), V(s) are continuously differntiable curves whose tangents at points of I agree with the lines of the hyperbolic structure.*

9 The restricted 3-body problem

The same study we've done in this paper can be modified to analyze to a very beautiful physical phenomenon: the restricted 3-body problem. In this section as well, we will state the main results, but not prove them. The setting is as follows:

Suppose we have 2 mass points of equal mass $m_1 = m_2 = \frac{1}{2}$, moving under Newton's law of attraction in the elliptic orbits while the center of mass is at rest. Now consider a third mass point, of mass $m_3 = 0$, moving on the line L perpendicular to the plane of motion s of the first two mass points, and going through the center of mass. Since $m_3 = 0$, it is clear that the third mass point will remain on L. The problem is to describe the motion of the third mass point (see Figure 11). To this end, let z be the coordinate describing the third mass point on L, and t the time parameter. Then we have the following differential equation:

$$\frac{d^2z}{dt^2} = -\frac{z}{(z^2 + r^2(t))^{\frac{3}{2}}}$$

with

$$r(t) = \frac{1}{2}(1 - \epsilon cos(t)) + O(\epsilon^2)$$

and ϵ denotes the eccentricity.

Theorem. *There exists a critical escape velocity* $\tilde{\dot{z}}(0)$ *such that, if* $\dot{z}(0) > \tilde{\dot{z}}(0)$*, the corresponding solutions will tend to infinity*

In this section, we will study the solutions near the critical escape velocity $\tilde{\dot{z}}(0)$.

Consider a solution $z(t)$ with infinitely many zeroes t_k $(k = 0, \pm1, \pm2, ...)$, such that $t_k < t_{k+1}$, then let:

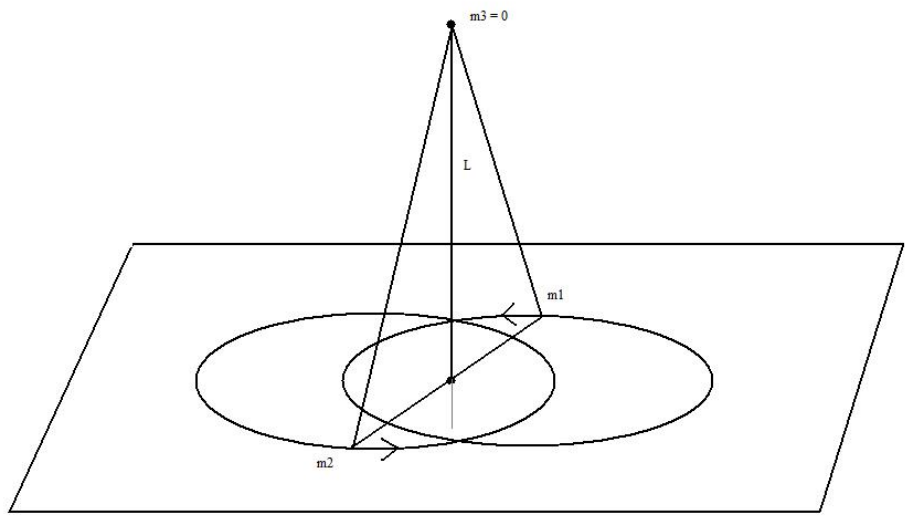

Figure 11: The restricted 3-body problem

$$s_k = \left[\frac{t_{k+1} - t_k}{2\pi}\right]$$

Basically, s_k measures the number of complete revolutions of the primaries between two zeroes of $z(t)$. Hence, we can associate to every such solution a doubly infinite sequence of integers.

The surprising fact is that the converse holds too:

Theorem. *For all but finitely many eccentricites $0 < \epsilon < 1$, there exists an integer $m = m(\epsilon)$ such that any sequence s with $s_k \geq m$ corresponds to a solution of the above differential equation.*

Any solution $z(t) \neq 0$ is determined by giving its velocity $\dot{z_0}$ and time t_0 when $z(t_0) = 0$. Such a zero always exists, otherwise, if $z(t) > 0$ for all t, hence $\ddot{z_0} < 0$, so z(t) is concave for all t, and positive, which is impossible.
Hence, we can describe an arbitrary orbit by giving $t_0[2\pi]$ and $\dot{z_0}$. Since the differential equation is invariant under the reflection $z \mapsto -z$, we may set $v_0 = |\dot{z}(t_0)|$. Now we can define a map ϕ, which maps (v_0, t_0) to (v_1, t_1), where t_1 is the next zero (if it exists), and $v_1 = |\dot{z}(t_1)|$.

Note. *The point $(0, \infty)$ can be viewed in this case as a hyperbolic fixed point of ϕ, thus the importance of studying solutions that blow off to infinity.*

We have the following very important lemmas:

Lemma. *There exists a real analytic simple closed curve in $\mathbb{R}^2$ in whose interior D_0 the mapping ϕ is defined. For (v_0, t_0) outside D_0, the corresponding solutions escape.*

Lemma. *ϕ maps D_0 onto a domain $D_1 = \phi(D_0)$, which agrees with the reflected domain $D_1 = \rho(D_0)$, where $\rho(v,t) = (v,-t)$ (see Figure 12). Moreover, ϕ preserves the area element vdvdt, and:*

$$\phi^{-1} = \rho^{-1}\phi\rho$$

Lemma. *If $\epsilon > 0$ is small enough, then $D_0 \neq D_1$, and the boundary curves $\partial D_0, \partial D_1$ intersect on the symmetry line nontangentially at a point P (see Figure 12).*

Note. *P plays the same role as a homoclinic point*

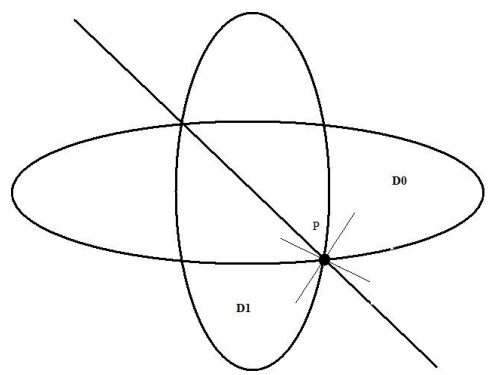

Figure 12: Construction of D_0, D_1, P

Here we come with the main theorem concerning the restricted 3-body-problem:

Theorem. *There exists a homeomorphism τ of S into D_0 such that*

$$\phi\tau = \tau\bar{\sigma}$$

Moreover, there exists a hyperbolic invariant set I homeomorphic to S on which ϕ is topologically equivalent to σ

The idea of the proof is the similar as for the main theorem: Just consider a rectangle R around P, and verify Properties 1 and 3. In this case, the horizontal strips are just the elements of $\phi(R) \cap R$, and the vertical strips are the reflections of those horizontal strips.

We conclude that the movement of the third mass point is very chaotic.

Conclusion

In order to study the dynamics near a homoclinic point H, we introduced a transversal map $\widetilde{\phi}$. This map satisfies 3 properties, which led us to the main theorem. This theorem reduced our study to a simple analysis of sequences, which helped us to prove

corollaries that seem too good to be true. For example, among those corollaries, we have that homoclinic orbits form a chaotic system (according to Smale), and that they do not possess an analytic integral, which completely rules out any regular/nonchaotic behavior. Thus, Moser's paper opened some windows in the understanding of chaos. But nonetheless, this field of study has not yet been completely understood, and will probably one of the hot topics of mathematics in the 21st century.

References

[1] Ivar Ekeland. *Le Calcul, l'Imprévu : les figures du temps de Kepler á Thom*. Paris: Editions du Seuil, 1984.

[2] Stephen Smale et.al. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Elsevier Academic Press, 2004.

[3] Jurgen Moser. *Stable and Random Motions in Dynamical Systems*. Princeton University Press, 2001.