The Baker Transformation and a Mapping Associated to the Restricted Three Body Problem

Robert L. Devaney[1]*

Department of Mathematics, University of Maryland, College Park, MD 20742, USA

Abstract. The nonlinear mapping of the plane

$$x_1 = x_0 + 1/y_0$$
$$y_1 = y_0 - x_0 - 1/y_0$$

was recently introduced by Hénon as an asymptotic form of the equations of motion of the restricted three body problem. This is an area preserving diffeomorphism, except along the x-axis where the mapping is singular. We show that this mapping exhibits a type of stochastic behavior known as topological transitivity, by showing that it is topologically conjugate to the well known baker transformation. Consequently, periodic points are dense in the plane and there is also a dense orbit. We note that the baker transformation also preserves Lebesgue measure and is ergodic, so this raises interesting open questions about the ergodic properties of the non-linear mapping.

Consider the nonlinear mapping of the plane $(x_1, y_1) = F(x_0, y_0)$ given by

$$x_1 = x_0 + \frac{1}{y_0}$$
$$y_1 = y_0 - x_0 - \frac{1}{y_0}.$$

This mapping was recently encountered by Hénon [2] in his studies of the restricted three body problem of classical mechanics. Roughly speaking, this mapping is an asymptotic form of the equations of motion which is related to

---

1   Current address: Department of Mathematics, Boston University, Boston, MA 02215, USA

the ultimate behavior of families of periodic solutions as the mass-ratio tends to zero.

Hénon's numerical work indicates that this mapping has highly unstable and chaotic behavior, and he has conjectured that it is equivalent to the well-known baker transformation. Our goal in this paper is to verify this conjecture by constructing a topological conjugacy between F and the baker transformation.

The conjugacy is given in fact by a natural mapping: to any point $p \in \mathbb{R}^2$, we associate a doubly infinite sequence $s(p)$ of $+1$'s and -1's according to whether successive iterates of p lie above or below the x-axis. The mapping is not defined when $y = 0$, so certain orbits terminate when they meet this line. We can extend the conjugacy to these points by assigning a terminating sequence of $+1$'s and -1's to them.

There is a natural identification of these sequences with the open square $|u|, |v| < 1$ in the plane. Under this identification, the mapping F goes over to the baker transformation. Thus the coding s gives a complete description of the associated dynamical system, since we may take the baker transformation to be completely understood.

Theorem A. *The mapping F of Hénon is topologically conjugate to the baker transformation of the open square $0 \leq |u|, |v| < 1$ in the plane.*

The dynamics of the baker transformation thus yield immediately the following corollaries.

Corollary B. *Periodic points for F are dense in the plane. There are exactly $2^n - 2$ fixed points for F^n for each $n \geq 1$. Moreover, each periodic point is hyperbolic.*

Corollary C. *There is a point $p \in \mathbb{R}^2$ whose F-orbit is dense in $\mathbb{R}^2$. Equivalently, F is topologically transitive.*

The x-axis is a singular set for F since orbits which meet this set cannot be continued. Although the set of all such singular orbits is of measure zero in the plane, it nevertheless is dense.

Corollary D. *Let S denote the set of points in $\mathbb{R}^2$ whose orbits terminate at the singular set $y = 0$. Then S is dense in $\mathbb{R}^2$.*

The proofs of each of these corollaries follow immediately from the existence of the topological conjugacy given by Theorem A and the corresponding statements for the baker transformation.

The mapping F preserves Lebesgue measure in the plane. This raises several intriguing questions about the ergodic properties of this mapping. Is F ergodic? It is well-known that the baker transformation preserves Lebesgue measure in the square and that it is ergodic. So our conjugacy gives a different (infinite) measure on the square which arises naturally in the context of mechanics. It would be interesting to know exactly what this measure on the square is.

1. The Baker Transformation

F is a real analytic, area preserving mapping defined on $\mathbb{R}^2 - (y=0)$. Its inverse is also real analytic and is given by $(x_{-1}, y_{-1}) = F^{-1}(x_0, y_0)$ where

$$x_{-1} = x_0 - \frac{1}{x_0 + y_0}$$

$$y_{-1} = x_0 + y_0.$$

Hence F^{-1} is not defined on the line $y_0 = -x_0$.

We will find it convenient to denote the point $F^j(x_0, y_0)$ by (x_j, y_j) for $j \in \mathbb{Z}$.

Let $p \in \mathbb{R}^2$. If $F^k(p)$ is defined for all $k \in \mathbb{Z}$, we may assign a doubly infinite sequence $s(p)$ of $+1$'s and -1's to p via the rule

$$s(p) = (\ldots s_{-2}, s_{-1}, s_0; s_1, s_2, \ldots),$$

where

$$s_j = s_j(p) = \begin{cases} +1 & \text{if } y_j(p) > 0 \\ -1 & \text{if } y_j(p) < 0. \end{cases}$$

Some F-orbits terminate when $y_j(p) = 0$ under forward iteration, or when $y_j(p) = -x_j(p)$ under backward iteration. To these orbits we assign a terminating sequence of the form

$$(\ldots s_{-2}, s_{-1}, s_0; s_1, \ldots, s_k, 0],$$

if $y_{-k}(p) = -x_{-k}(p)$, or a sequence of the form

$$[0, s_{-j}, \ldots, s_{-1} s_0; s_1, s_2, \ldots),$$

if $y_{j+1}(p) = 0$ with $j \geq 1$. We also allow finite sequences of the form

$$[0, s_{-j}, \ldots, s_{-1}, s_0; s_1, \ldots, s_k, 0].$$

So for each $p \in \mathbb{R}^2$, there is assigned a sequence fo the form

$$[0, s_{-j}, \ldots, s_0; s_1, \ldots, s_k, 0],$$

where j and/or k may be infinite. We assign the sequence $[0; 0]$ to the origin.

Let Σ denote the set of all possible sequences of the above type, excluding those which terminate (to the right and/or left) with an infinite string of $+1$'s or -1's. That is, Σ consists of all sequences except those of the form

$$(\ldots, \alpha, \alpha, \alpha, s_{-j}, \ldots, s_0; s_1, \ldots, s_k, 0]$$

$$[0, s_{-j}, \ldots, s_{-1}, s_0; s_1, \ldots, s_k, \alpha, \alpha, \alpha, \ldots)$$

with an infinite string of α's. Σ may be mapped onto the open square $0 \leq |u|$, $|v| < 1$ in the plane via the transformation

$$u = \sum_{i=1}^{k} \frac{s_i}{2^i} \qquad v = \sum_{i=0}^{-k} \frac{s_i}{2^{1-i}}.$$

Note that sequences of the form $[0; s_1, \ldots, s_k, 0]$ are mapped to $v = 0$, while sequences of the form $[0, s_{-j}, \ldots, s_0; 0]$ are mapped to $u = 0$. The boundaries $|u| = 1$ and $|v| = 1$ are excluded since sequences of the forms

$$(\ldots, \alpha, \alpha, \alpha; s_0, \ldots, s_k, 0]$$

$$[0, s_{-j}, \ldots, s_0; \alpha, \alpha, \alpha, \ldots)$$

are excluded.

Also, one should think of the disallowed sequences of the forms

$$[0, s_{-j}, \ldots, s_{-1}, s_0; \ldots, s_k, +1, -1, -1, -1, \ldots)$$

and

$$[0, s_{-j}, \ldots, s_{-1}, s_0; \ldots, s_k, -1, +1, +1, +1, \ldots)$$

as being identified with the allowed sequence

$$[0, s_{-j}, \ldots, s_0; s_1, \ldots, s_k, 0],$$

this being the usual identification which yields the dyadic rationals. There is a similar identification for negative indices.

On Σ we may define the usual shift automorphism by

$$\sigma([0, s_{-j}, \ldots, s_{-1}, s_0; s_1, \ldots, s_k, 0]) = [0, s_{-j}, \ldots, s_{-1}; s_0, s_1, \ldots, s_k, 0]$$

provided $s_0 \neq 0$. On the square, σ is represented by the well-known baker transformation (see, for example, [4, p. 63] or [1]). If we impose the topology of the square on Σ, then σ is a homeomorphism as long as $s_0 \neq 0$, where σ is undefined. The range of σ includes all sequences except those with $s_1 = 0$.

In the next sections we will prove that s gives a topological conjugacy between F and σ, i.e., s is a homeomorphism which makes the following diagram commute

$$
\begin{array}{ccc}
\mathbb{R}^2 - (y = 0) & \xrightarrow{\ F\ } & \mathbb{R}^2 - (y = -x) \\
\Big\downarrow{\scriptstyle s} & & \Big\downarrow{\scriptstyle s} \\
\Sigma - (s_0 = 0) & \xrightarrow{\ \sigma\ } & \Sigma - (s_1 = 0).
\end{array}
$$

2. Hyperbolicity

In this section we prove that there exists at least one point in the plane which corresponds to any allowable sequence. Our main tool is the (non-uniform) hyperbolicity of F.

Let (ξ_0, η_0) be a tangent vector to $\mathbb{R}^2$. We denote the forward image $dF(\xi_0, \eta_0)$ by (ξ_1, η_1) and the inverse image $dF^{-1}(\xi_0, \eta_0)$ by (ξ_{-1}, η_{-1}). In each tangent space, define the sectors

$$S^u_{(x,y)} = \{(\xi_0, \eta_0) \in T_{(x,y)} \mathbb{R}^2 \mid \xi_0 \eta_0 \leq 0\}$$

$$S^s_{(x,y)} = \{(\xi_0, \eta_0) \in T_{(x,y)} \mathbb{R}^2 \mid \xi_0 \eta_0 \geq 0\}$$

and let

$$S^u = \bigcup_{(x,y)} S^u_{(x,y)} \qquad S^s = \bigcup_{(x,y)} S^s.$$

So the unstable sectors S^u consist of the second and fourth quadrants in each tangent space, while S^s consists of the other quadrants. Let $\|(\xi_0, \eta_0)\|$ denote the sup norm on each tangent space, i.e., $\|(\xi_0, \eta_0)\| = \sup(|\xi_0| + |\eta_0|)$.

The proof of the following proposition is straightforward.

Proposition 1.

    i) $dF(S^u) \subset S^u$, $dF^{-1}(S^s) \subset S^s$.

    ii) *For* $(\xi_0, \eta_0) \in S^u_{(x,y)}$,

$$\|dF(\xi_0, \eta_0)\| \geqq \min\left(2, 1 + \frac{2}{y^2}\right) \|(\xi_0, \eta_0)\|.$$

    iii) *For* $(\xi_0, \eta_0) \in S^s(x, y)$,

$$\|dF^{-1}(\xi_0, \eta_0)\| \geqq \left(1 + \frac{1}{(x+y)^2}\right) \|(\xi_0, \eta_0)\|.$$

Remark. This proposition implies that F is non-uniformly hyperbolic; the rate of expansion in the unstable sector tends to 1 as $|y| \to \infty$. The existence of invariant manifolds almost everywhere for such mappings is a difficult problem studied by Pesin [6].

Definition. A smooth curve $\gamma(t)$ is called an *unstable curve* if $\gamma'(t)$ lies in the interior of $S^u_{\gamma(t)}$ for all t. An unstable curve of the form $(t, f(t))$ for $-\infty < t < \infty$ is called an *unstable separatrix* if $\lim\limits_{t \to \pm\infty} f(t) = \mp\infty$. Similarly, $\gamma(t)$ is a *stable curve* if $\gamma'(t)$ lies in the interior of $S^s_{\gamma(t)}$ for all t. And a stable curve of the form $(t, f(t))$ is a *stable separatrix* if either

    i) $\lim\limits_{t \to \infty} f(t) = \infty$ and $\lim\limits_{t \to -\infty} f(t) = 0$, or

    ii) $\lim\limits_{t \to \infty} f(t) = 0$ and $\lim\limits_{t \to -\infty} f(t) = -\infty$.

See Fig. 1.

Proposition 2. *If* $\gamma(t)$ *is an unstable (respectively stable) curve, then so is* $F(\gamma(t))$ *(respectively,* $F^{-1}(\gamma(t))$*). Moreover, if* $\gamma'(t)$ *is contained in a particular quadrant of* $S^u_{\gamma(t)}$ *(respectively,* $S^s_{\gamma(t)}$*), then* $dF(\gamma'(t))$ *(respectively,* $dF^{-1}(\gamma'(t))$*) is contained in the same quadrant.*

Proof. One checks easily that ξ_1 and ξ_{-1} have the same sign as ξ_0, and that η_1 or η_{-1} have the same sign as η_0, provided (ξ_0, η_0) lies in the appropriate sector.

Proposition 3. i) *Let* $\gamma(t)$ *be an unstable separatix. Then* $F(\gamma(t))$ *is a pair of unstable separatrices, one on each side of the line* $y = -x$.

    ii) *Let* $\mu(t)$ *be a stable separatrix. Then* $F^{-1}(\mu(t))$ *is a pair of stable separatrices, one of each of the above types (i.e., one on each side of the x-axis).*

Proof. We prove i); case ii) is similar. Let $\gamma(t) = (t, f(t))$ for $t \in \mathbb{R}$ and suppose t^* is the unique point for which $f(t^*) = 0$. Let $\gamma_1(t) = (t, f(t))$ with $t < t^*$ and $\gamma_2(t)$
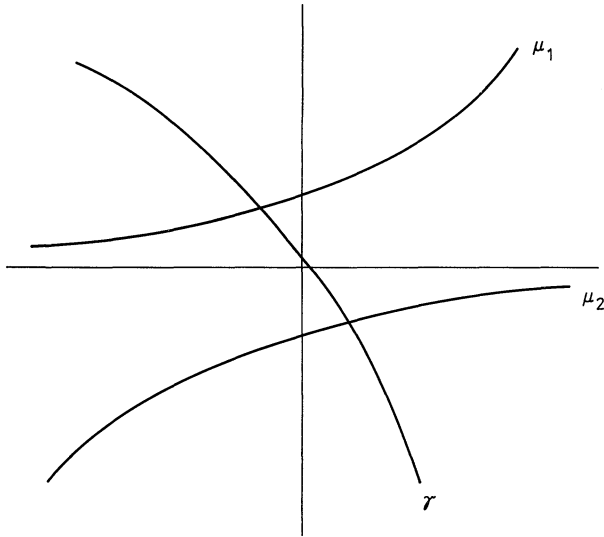
Fig. 1. γ is an unstable separatrix, while μ_1 and μ_2 are stable separatrices

$=(t, f(t))$ with $t > t^*$. Then $F(\gamma_1(t))$ and $F(\gamma_2(t))$ satisfy the conclusion of the proposition.

Definition. An *unstable strip* is a closed region in the plane bounded by a pair of non-intersecting unstable separatrices, or else the closed region to the left or right of a single unstable separatrix. *Stable strips* are defined analogously.

Corollary 4. *Let V be an unstable strip. Then $F(V)$ is a pair of unstable strips, one on each side of $y = -x$. If H is a stable strip, then $F^{-1}(H)$ is a pair of stable strips, one on each side of $y = 0$.*

Note that stable and unstable separatrices must meet at a unique point. Consequently, the intersection of stable and unstable strips is non-empty and bounded by stable and unstable curves.

We now prove that there exists at least one point in $\mathbb{R}^2$ corresponding to any sequence of the form $(\dots s_{-2}, s_{-1}, s_0; s_1, s_2, \dots)$ as long as $s_i \neq s_j$ for some pair $i, j \geq 0$ and $s_k \neq s_\ell$ for another pair $k, \ell < 0$. The proof is analogous to the proof of the conjugacy of the well-known Smale "horseshoe" diffeomorphism with the binary shift, so we merely sketch the proof. Further details can be found in Smale's original paper [7], or in the books of Moser [4] or Nitecki [5].

Let Q_1 (respectively, Q_{-1}) denote the upper (respectively, lower) half plane. The closures of Q_1 and Q_{-1} are stable strips, while the closures of $F(Q_1)$ and $F(Q_{-1})$ are unstable strips bounded by the line $y = -x$.

Define inductively

$$H_{s_0 s_{-1} \dots s_{-n}} = \text{closure}(Q_{s_0} \cap F^{-1}(Q_{s_{-1}}) \cap \dots \cap F^{-n}(Q_{s_{-n}}))$$

$$= \text{closure}(Q_{s_0} \cap F^{-1}(H_{s_{-1} \dots s_{-n}})).$$

One checks immediately that the $H_{s_0 s_{-1}...s_{-n}}$ form a nested sequence of stable strips. See, for example, Moser [4, p. 172]. Moreover, any point on the boundary of $H_{s_0 s_{-1}...s_{-n}}$ is mapped to $y=0$ by some iterate of F.

Similarly, define

$$V_{s_1...s_n} = \text{closure}(F(Q_{s_1}) \cap ... \cap F^n(Q_{s_n}))$$
$$= \text{closure}(F(Q_{s_1}) \cap F(V_{s_2...s_n})).$$

These are nested unstable strips whose boundaries consist of unstable separatrices mapped to $y = -x$ by some iterate of F^{-1}.

Lemma 5. *If some $s_i \neq s_j$, then $V_{s_1...s_n}$ with $n \geq i, j \geq 1$ is bounded by two unstable separatrices.*

Proof. For fixed n, the set of unstable strips $V_{s_0...s_n}$ fills the plane, and any two strips meet only along at most one unstable separatrix in their boundaries. One checks easily that $V_{s_0...s_n}$ is bounded by only one unstable separatrix when $s_j = 1$ or $s_j = -1$ for all j, $1 \leq j \leq n$. Hence all other strips lie between these two, and hence they are bounded by two unstable separatrices.

Now suppose $(s) = (... s_{-2}, s_{-1}, s_0; s_1, s_2 ...)$ is any sequence satisfying $s_i \neq s_j$ for some $i, j > 0$ and $s_k \neq s_\ell$ for some $k, \ell \leq 0$. Consider the sets $V_{s_1...s_n} \cap H_{s_0 s_{-1}...s_{-n}}$. By the lemma, these form a nested sequence of compact sets. The intersection

$$\bigcap_{n=1}^{\infty} (V_{s_1...s_n} \cap H_{s_0 s_{-1}...s_{-n}})$$

then contains at least one point. This point must then have associated sequence (s), as the kth iterate of this point lies in $Q_{s_{-k}}$. This proves the existence of a point corresponding to each doubly infinite sequence.

For the terminating sequences, we argue as follows. To find a point corresponding to the one sided terminating sequence $(... s_{-2}, s_{-1}, s_0; s_1, ..., s_n, 0]$ we first note that $V_{s_1...s_n}$ is bounded by one or two unstable separatrices. Only one of these curves is mapped to the line $y = -x$ by F^{1-n}. Call this curve γ. Then any point in

$$\left(\bigcap_{n=0}^{\infty} H_{s_0 s_{-1}...s_{-n}} \right) \cap \gamma$$

is associated to the sequence

$$(... s_{-2}, s_{-1}, s_0; s_1, s_2, ..., s_n, 0].$$

The other types of terminating sequences are handled similarly.

We conclude this section by proving that s gives a continuous mapping from $\mathbb{R}^2$ to Σ. If $s(p)$ is a non-terminating sequence, then our construction of the stable and unstable strips shows that s is continuous at p. So we confine our attention to terminating sequences.

First suppose $p^* = (x_0^*, 0)$ is on the x-axis. We will prove that s is continuous at p^*; all other one- or two-sided terminating sequences are handled similarly.

Lemma 6. *Let $M > 1$. Suppose $p = (x_0, y_0)$. Then there exists $\varepsilon > 0$ such that*

    i) *If $0 < y_0 < \varepsilon$ and $|x_0^* - x_0| < \varepsilon$, then $x_1 > M$ and $y_1 < -M$.*
    ii) *If $-\varepsilon < y_0 < 0$ and $|x_0^* - x_0| < \varepsilon$, then $x_1 < -M$ and $y_1 > M$.*

Proof. For part i) we choose ε so that $x_0^* + \dfrac{1}{\varepsilon} - 2\varepsilon > M$. Then we have

$$x_1 = x_0 + \frac{1}{y_0} > x_0^* - \varepsilon + \frac{1}{\varepsilon} > M + \varepsilon > M$$

$$y_1 = y_0 - x_1 < \varepsilon - M - \varepsilon.$$

Part ii) follows similarly.

Remark. This lemma shows that points near the x-axis are mapped arbitrarily far away by F, with y_1 approaching either $+\infty$ or $-\infty$, depending upon the sign of y_0.

Now suppose $s(p^*) = [0; s_1^*, s_2^*, \ldots)$ and $k \in \mathbb{Z}^+$. We claim that there is a neighborhood W of p^* such that if $p = (x_0, y_0) \in W$ and $y_0 \neq 0$, then $s(p) = (\ldots s_{-1}, s_0; s_1 \ldots)$ satisfies

    i) $s_j = s_j^*$ for $1 \leq j \leq k$.
    ii) if $y_0 > 0$, then $s_0 = +1$ while $s_{-i} = -1$ for $1 \leq i \leq k$.
    iii) if $y_0 < 0$, then $s_0 = -1$ while $s_{-i} = +1$ for $1 \leq i \leq k$.

In the topology of the square, this means that $s(p)$ is close to $s(p^*)$, proving continuity.

Choose $M > k$, so that $2M - k > 0$. If $y_0 > 0$, then by the lemma, there exists $\varepsilon > 0$ such that if $|x_0^* - x_0| < \varepsilon$ and $y_0 < \varepsilon$, then $x_1 > M$ and $y_1 < -M$. We also have

$$x_2 = x_1 + \frac{1}{y_1} > M - \frac{1}{M} > M - 1$$

$$y_2 = y_1 - x_2 < -M - M + 1 = -2M + 1 < 0.$$

Continuing inductively, it follows that for $2 \leq i \leq k$, we have

$$x_{i+1} > M - i > 0$$

$$y_{i+1} < -2M + i < 0.$$

Hence $s_{-i}(p) = -1$ for $1 \leq i \leq k$ and $s_0(p) = +1$ as required.

If $y_0 < 0$, the argument is similar.

3. Proof of Theorem A

In this section we complete the proof of Theorem A by showing that s is $1-1$. Suppose $p = (x_0, y_0)$. We show first that the orbit of p either changes sign infinitely often or else meets $y = 0$. This follows from several lemmas.

Lemma 1. *Suppose $x_0 \geq 0$, $y_0 > 0$. Then there exists $n > 0$ such that $y_n \leq 0$. If $x_0 \leq 0$, $y_0 < 0$, there exists $n' > 0$ such that $y_{n'} \geq 0$.*

Proof. We prove the first statement; the proof of the second is similar. Suppose $y_n > 0$ for all n. Then we have

$$x_n = x_{n-1} + \frac{1}{y_{n-1}} > x_{n-1} > \ldots > x_0 \geqq 0.$$

We also have

$$y_1 = y_0 - x_0 - \frac{1}{y_0} < y_0 - x_0$$

and by induction

$$y_n < y_0 - nx_0.$$

Therefore, if $x_0 > 0$, we have a contradiction. If $x_0 = 0$, then $x_1 > 0$ and we may apply the above argument using x_1 instead, again providing a contradiction.

Lemma 2. *Suppose* $x_0 < 0$ *and* $y_0 > 0$. *Then there exists* $n > 0$ *such that either* $y_n \leqq 0$ *or else* $x_n \geqq 0$. *If* $x_0 > 0$ *and* $y_0 < 0$, *then there exists* $n' > 0$ *such that either* $y_{n'} \geqq 0$ *or else* $x_{n'} \leqq 0$.

Proof. Again we prove only the first statement. Assume $y_n > 0$ for all n. As in Lemma 1 we have $x_n > x_0$ and $0 < y_n < y_0 < nx_0$. Now

$$x_2 = x_1 + \frac{1}{y_1} > x_0 + \frac{1}{y_0} + \frac{1}{y_0 - x_0}$$

and by induction,

$$x_{n+1} > x_0 + \frac{1}{y_0} + \frac{1}{y_0 - x_0} + \ldots + \frac{1}{y_0 - nx_0}.$$

Now the series $\sum_{n=0}^{\infty} \frac{1}{y_0 - nx_0}$ diverges, so that $x_n \to \infty$, which gives a contradiction.

Combining Lemmas 1 and 2, we find that the orbit of p must continually cross the x-axis, or else meet it after a finite number of iterates. It is for this reason that sequences which end with an infinite sequence of $+1$'s or -1's are disallowed.

For the remainder of this section we let $p = (x_0, y_0)$ and $p' = (x'_0, y'_0)$. Define $\Delta x_n = x'_n - x_n$ and $\Delta y_n = y'_n - y_n$ for each $n \in \mathbb{Z}$.

Lemma 3. *Suppose* $s(p) = s(p')$.

i) *If* $\Delta x_0 < 0$ *and* $\Delta y_0 > 0$, *then* $\Delta x_n < \Delta x_{n-1}$ *and* $\Delta y_n > \Delta y_{n-1}$ *for all* $n > 0$. *Furthermore,* $\lim_{n \to \infty} \Delta y_n = \infty$.

ii) *If* $\Delta x_0 > 0$, $\Delta y_0 > 0$, *then* $\Delta x_{-n} > \Delta x_0$ *and* $\Delta y_{-n} > \Delta y_{-n+1}$ *for all* n. *Furthermore,* $\lim_{n \to \infty} \Delta y_{-n} = \infty$.

Remark. In the first case we are assuming that the straight line connecting p to p' is an unstable curve, while in the second case, it is a stable curve. If p and p' lie on the same horizontal or vertical line, the lemma also follows by applying it to $F(p)$ and $F(p')$ or $F^{-1}(p)$ and $F^{-1}(p')$.

Proof. Again we prove only part i). Since p and p' have the same sequences, it follows that y_n' and y_n have the same signs. Hence

$$\Delta x_1 = \Delta x_0 + \frac{1}{y_0'} - \frac{1}{y_0} < \Delta x_0 < 0$$

$$\Delta y_1 = \Delta y_0 - \Delta x_1 > \Delta y_0 - \Delta x_0 > \Delta y_0.$$

By induction, if $\Delta y_{n-1} > 0$, then

$$\frac{1}{y_{n-1}'} - \frac{1}{y_{n-1}} < 0$$

so that $\Delta x_n < \Delta x_{n-1} < \Delta x_0$. Hence

$$\Delta y_n = \Delta y_{n-1} - \Delta x_n > \Delta y_0 - (n-1) \Delta x_0 - \Delta x_0.$$

It follows that $\Delta y_n \to \infty$.

Remark. It follows from Lemma 3 that if $s(p) = s(p')$ and the straight line joining p to p' is an unstable (respectively, stable) curve, then so is the straight line joining $F^n(p)$ to $F^n(p')$ for all $n > 0$ (respectively, $n < 0$).

Lemma 4. *Suppose p and p' have the same two-sided terminating sequence. Then $p = p'$.*

Proof. The result is clear if $p = 0$. If $y_k = y_k' = 0$ for some $k > 0$, then the straight line joining $F^{k-1}(p)$ to $F^{k-1}(p')$ is a stable curve since y_{k-1} and y_{k-1}' have the same sign. But then the remark above implies that the straight line joining $F^{k-\alpha}(p)$ to $F^{k-\alpha}(p')$ for all $\alpha \geq 1$ is a stable curve. Hence this line can never have slope -1, and so there does not exist $\alpha > 0$ such that $y_{k-\alpha} = -x_{k-\alpha}$ and $y_{k-\alpha}' = -x_{-k-\alpha}'$ contradicting the assumption that $s(p)$ is a two-sided terminating sequence.

Lemma 5. *Suppose $s(p) = s(p')$ and $\Delta x_0 < 0$, $\Delta y_0 > 0$. Then for any $M > 0$ there exists $j, k \geq 0$ such that $y_j' > M$ and $y_{-k} < -M$.*

Proof. Assume $y_i' < M$ for all $i \geq 0$. By Lemma 3, Δy_n increases to ∞ as $n \to \infty$. Hence there exists $N \geq 0$ such that for $n \geq N$, $\Delta y_n > M$. By Lemmas 1 and 2, there exists $k \geq N$ such that $y_k' > 0$. But then $y_k = y_k' - \Delta y_k < 0$, contradicting the fact that $s(p) = s(p')$.

Lemma 6. *Suppose $s(p) = s(p')$ and $\Delta x_0 < 0$, $\Delta y_0 > 0$. Then there exists $k > 0$ such that*

    i) $x_k', x_{k+1}' \leq 0$.
    ii) $x_k, x_{k+1} > 0$.
    iii) $y_k, y_{k+1} > 0$.

Proof. By Lemma 5, there exists $j > 0$ such that

$$y_j' > \frac{3}{\Delta x_0}$$

and $x'_j < 0$. Now $y'_{j+1} > y'_j$ if $x'_{j+1} < 0$, so there exists $\alpha \geqq 0$ such that $x'_{j+\alpha} \leqq 0$ but $x'_{j+\alpha+1} > 0$. Then

$$0 < x'_{j+\alpha+1} = x'_{j+\alpha} + \frac{1}{y'_{j+\alpha}}$$

$$\leqq x'_{j+\alpha} + \frac{1}{y'_j}$$

$$\leqq x'_{j+\alpha} + \frac{\Delta x}{3}.$$

Hence $x'_{j+\alpha} > -\dfrac{\Delta x}{3}$. Also $x'_{j+\alpha-1} > -\frac{2}{3}\Delta x$. Let $k = j + \alpha - 1$. Then we have $x'_k < x'_{k+1} \leqq 0$ and $y'_{k+1} > y'_k > 0$. Furthermore,

$$x_{k+1} > x_k > x'_k + \Delta x_0 > \frac{\Delta x}{3} > 0.$$

This completes the proof.

Lemma 7. *Suppose $s(p) = s(p')$ and $\Delta x_0, \Delta y_0 > 0$. Then there exists $k < 0$ such that*

    i) $x'_k, x'_{k-1} \leqq 0$.
    ii) $x_k, x_{k-1} > 0$.
    iii) $y_k, y_{k-1} > 0$.

Proof. The proof is similar to the preceding lemma and hence is omitted.

Proof of uniqueness. Suppose $s(p) = s(p')$ and $\Delta x_0 < 0$, $\Delta y_0 > 0$. The other case is handled similarly, so the proof is omitted.

    Choose $k \geqq 0$ as in Lemma 6, and let ζ be the straight line connecting $F^k(p)$ to $F^{k+1}(p')$. See Fig. 2. We assume that the tangent vectors ζ' lie in the second quadrant, so that $dF^\alpha(\zeta')$ also lies in the second quadrant for $\alpha > 0$. In the terminology of Sect. 1, ζ is an unstable curve. By Lemma 1, there exists $n > 0$ such that $y_n(F^{k+1}(p)) < 0$ but $y_n(F^k(p)) = y_{n-1}(F^{k+1}(p)) > 0$. Hence $y_n(F^{k+1}(p')) < 0$. Then $F^n(\zeta)$ is an unstable curve connecting $F^{n+k}(p)$ to $F^{n+k+1}(p')$. Hence tangent vectors to this curve cannot lie in the second quadrant. This contradiction establishes the result.
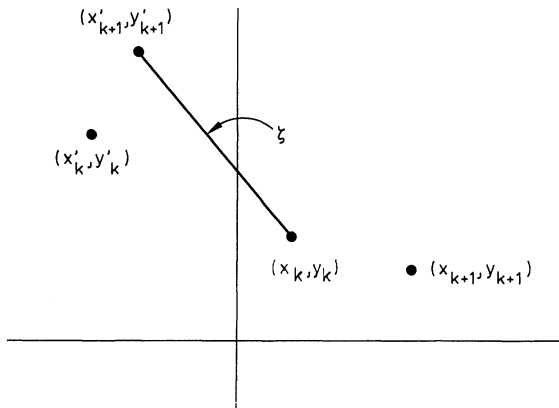

Fig. 2

References

1. Arnold, V.I., Avez, A.: Ergodic Problems of Classical Mechanics. New York: Benjamin 1968
2. Hénon, M.: Notes on the restricted three body problem. Mimeographed
3. Katok, A., Strelcyn, J.-M.: Invariant manifolds for smooth maps with singularities. (To appear)
4. Moser, J.: Stable and random motions in dynamical systems. Annals of Mathematics Studies, No. 77. Princeton, New Jersey: Princeton University Press 1973
5. Nitecki, Z.: Differentiable Dynamics. Cambridge, Massachusetts: MIT Press 1971
6. Pesin, Ja.B.: Math. USSR Izvest. **10**, 1261–1305 (1976)
7. Smale, S.: Diffeomorphisms with many periodic points. In: Differential and Combinatorial Topology, 63–80. Princeton, New Jersey: Princeton University Press 1965