7

Physica D 46 (1990) 149–162
North-Holland

ON AUBRY–MATHER SETS

J.J.P. VEERMAN
Institute for the Mathematical Sciences, SUNY at Stony Brook, Stony Brook, NY 11794, USA

and

F.M. TANGERMAN
Mathematics Department, SUNY at Stony Brook, Stony Brook, NY 11794, USA

Received 17 April 1990
Accepted 6 June 1990
Communicated by R.S. MacKay

Let f be a two-dimensional area preserving twist map. For each irrational rotation number in a certain (nontrivial) interval, there is an f-invariant minimal set which preserves order with respect to that rotation number. For large nonlinearity these sets are, typically, Cantor sets and they are referred to as Aubry–Mather sets. We prove that under some assumptions these sets are ordered vertically according to ascending rotation number ("monotonicity"). Furthermore, if f satisfies certain conditions, the right hand points of the gaps in an irrational Cantor set lie on a single orbit ("single gap") and diffusion through these Aubry–Mather sets can be understood as a limit of resonance overlaps ("convergence of turnstiles"). These conditions essentially establish the existence of a hyperbolic structure and limit the number of homoclinic minimizing orbits. Some other results along similar lines are given, such as the continuity at irrational rotation numbers of the Lyapunov exponent on Aubry–Mather sets.

1. Introduction

Let f be an area preserving monotone twist map on the cylinder $A = S^1 \times \mathbb{R}$. For each number α in the rotation interval I of f, Aubry [1] and Mather [12] have constructed f-invariant sets M_α that have the given number as rotation number. These invariant sets are constructed as the global minima of a certain action functional. The topologically minimal sets E_α with irrational rotation number are precisely the Aubry–Mather sets.

These sets are well-defined [12], lie on Lipschitz graphs over S^1 and on them the dynamics preserves the circular ordering [4]. They can be smooth invariant, homotopically non-trivial curves, so-called KAM curves. For large enough non-linearity, though, one expects them to be broken up into Cantor sets [3]. In fact, the parameter value at which a set breaks up depends to a large extent on the number-theoretical properties of the rotation number in question [10, 13]. These Cantor sets are then the "remnants" of the invariant KAM curves of the nearly integrable case.

Aubry–Mather sets play an important role in the global dynamics of the map, especially in stability questions. As invariant curves, they confine the dynamics of all orbits to narrow regions. However, numerical experiments indicate, that even as Cantor sets, these sets continue to restrict vertical motion [11]. These attempts to understand this led to a geometrical construction called "turnstiles". The idea was to construct the stable and unstable manifolds in the gaps of the Cantor sets, thus capturing the area per iterate that diffuses across the set.

In this article, we prove a number of theorems. One of these (the monotonicity theorem) states that under certain conditions the E_ρ admit a vertical ordering in the cylinder similar to ref. [16]. Another has been conjectured before on the basis of numerical evidence: this is the theorem that asserts that the diffusion through an Aubry–Mather set can be considered as a limit of resonance overlaps. Finally, the single gap theorem, which says that under a geometric assumption Aubry–Mather sets have only one gap orbit in them, has not appeared in the literature, as far as we are aware.

It is often convenient to consider the rotation number as being an element of the extended rotation interval I^+ defined as follows. Replace each rational number p/q with the set $\{p/q-, p/q, p/q+\}$ with the natural ordering between them. With this ordering, the ordering on I induces an ordering on I^+. The topology on I^+ is the order topology. Notice that p/q is an isolated point. We will often use I and I^+ interchangeably.

If ρ is rational, say p/q, then $E_{p/q}$ will denote a minimizing q-periodic orbit with rotation number p/q. Katok [5] proved the existence of minimizing orbits that are homoclinic to $E_{p/q}$, one advancing, $E_{p/q+}$, and one receding, $E_{p/q-}$. In this work, the only C^k generic, $k \geq 1$, properties of f that we use, are the following. First of all, $E_{p/q}$ consists of a single hyperbolic periodic orbit. Second, $M_{p/q}$ consists of $E_{p/q}$ plus a single advancing orbit, $E_{p/q+}$, homoclinic to $E_{p/q}$, and a single receding orbit, $E_{p/q-}$, also homoclinic to $E_{p/q}$ (in the sense that qth iterates of points move between successive points in $E_{p/q}$). The proof of this is standard and an outline is given in the appendix. (In a forthcoming work [21] we prove uniqueness of these orbits in the case of the standard map with large enough non-linearity parameter.)

Hausdorff limits (Hlim) of these sets are well-defined [12]:

$$\operatorname*{Hlim}_{\alpha \downarrow p/q+} E_\alpha = E_{p/q} \cup E_{p/q+} = \operatorname{clos}\{E_{p/q+}\},$$

$$\operatorname*{Hlim}_{\alpha \uparrow p/q-} E_\alpha = E_{p/q} \cup E_{p/q-} = \operatorname{clos}\{E_{p/q-}\}. \tag{1.1a}$$

For ω irrational

$$E_\omega \subseteq \operatorname*{Hlim}_{\alpha \to \omega} E_\alpha. \tag{1.1b}$$

In order to avoid notational complications, results will be stated and proved, where possible, in the universal covering space of the cylinder without further comment. For the lift of f, the notation F will be used.

2. Monotonicity

In this section, the monotonicity result will be proven. This result restricts the region that Aubry–Mather sets with irrational rotation number can inhabit.

Let f be a C^k $(k \geq 1)$ area preserving monotone twist map on the cylinder $A = S^1 \times \mathbb{R}$. Fix a lift $F: \mathbb{R}^2 \to \mathbb{R}^2$ of f. When necessary, we will use coordinates (x, y) on $\mathbb{R}^2$. Note that F commutes with the unit translation in the x direction. Each point $p = (x_0, y_0)$ defines an orbit $\{F^i(p)\}$ and the projections $\pi(x_i, y_i)$ on the x-coordinate are called x_i. Denote by M_α the set of minimizing points of rotation number α, i.e., $\lim x_i/i = \alpha$.

We define the local stable and unstable manifolds $W_\epsilon^{s/u}(x)$ and their inverse forward images respectively as the stable and unstable manifolds in the usual way (see for example ref. [6]).

Let V denote the foliation of $A = \mathbb{R}^2$ by vertical lines, so that $F(V)$ and $F^{-1}(V)$ are the corresponding images of V under F. At a point p in A, we can now define the open cone $C_p = C_p^- \cup C_p^+$ bounded by $F(V)_p$ and $F^{-1}(V)_p$ and containing V_p (see fig. 1). Here V_p denotes the leaf of the foliation V through p and $-$ or $+$ indicates the downward, respectively, the upward component. Define

$$C^{+/-}(p) = \bigcup_{i=-\infty}^{i=+\infty} C^{+/-}_{F^i(p)}.$$

Similarly, define tangent cone $TC_p = TC_p^- \cup TC_p^+$ as the cone in the tangent space to p whose boundary is formed by the tangent lines to C_p.

The fundamental wisdom that underlines this section, is that two minimizing points, s and p, with different rotation numbers in I^+, satisfy a

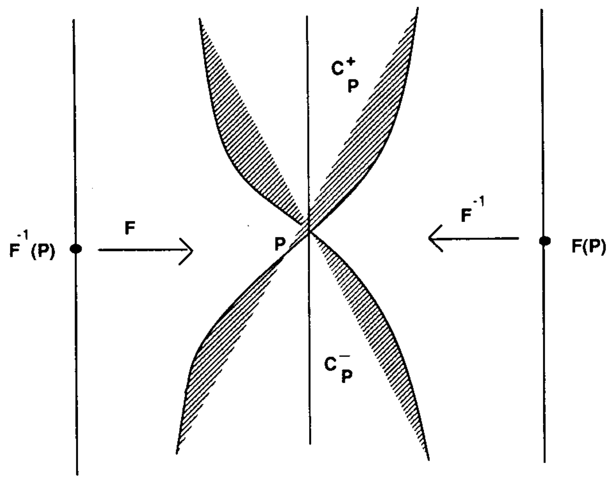

Fig. 1. The cone C_p.

geometrical inequality. Aubry's fundamental lemma [1] implies that s cannot lie in C_p^- if its rotation number in I is greater than or equal to that of p. But then, of course, s cannot lie in iterates of C_p^- either, and the same is true for iterates of s.

Lemma 2.1. Let p be a point of $E_{p/q+}$ or $E_{p/q-}$; the tangent to the stable and unstable manifolds $W^s(p)$ and $W^u(p)$ at p is given by $\lim_{n\to\infty} t_n^s/|t_n^s|$ and $\lim_{n\to\infty} t_n^u/|t_n^u|$ respectively, where

$$t_n^s = DF^{-nq}(F^{nq}(p)) \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

$$t_n^u = DF^{nq}(F^{-nq}(p)) \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Proof. This follows directly from the definition of the local stable and unstable manifolds plus the fact that for n large enough the vector $(0, -1)$ does not lie on $W_\epsilon^u(F^{nq}(p))$ nor on $W_\epsilon^s(F^{-nq}(p))$. The latter claim is proved as follows.

The points $f^{nq}(p)$ and $f^{-nq}(p)$ are close to two points x and y in $E_{p/q}$. By general hyperbolic theory (see also section 3) tangents to their (local) invariant manifolds are nearly parallel to those of x and y. But the tangents at x and y cannot be contained in TC_x, because if they were, Aubry's fundamental lemma would be violated since $E_{p/q-}$ and $E_{p/q+}$ accumulate onto x and y along those tangents. □

The following lemma shows that the invariant manifolds emanating from a point $p \in E_{p/q}$, $E_{p/q-}$, $E_{p/q+}$ can never be very close to vertical, and that the upper right branch is an unstable manifold. For clarity, we number the branches clockwise, starting from the vertical (see fig. 2).

Lemma 2.2. If $p \in E_{p/q}$, $E_{p/q-}$, or $E_{p/q+}$, then the first clockwise branch, W_1, is an unstable one. Moreover, corresponding branches on the orbit of p map into each other.

Proof. The region

$$TC_1^- = DF(F^{-1}(p))(C_{F^{-1}(p)}^-) \cup TC_p^- \cup DF^{-1}(F(p))(C_{F(p)}^-)$$

forms a new local cone (each original cone has V_p^- in its closure). It is then easy to see that TC_n^-, defined in the obvious way, contains TC_{n-1}^-, for all n. By lemma 2.1, the boundary of TC_{nq}^- (q fixed) must accumulate on a stable and an unstable direction. Observe that TC_{nq}^- cannot contain any stable or unstable directions and further that the interior of TC_{nq}^- is connected. Therefore, its boundaries can only accumulate onto W_2 and W_3.

The second statement follows from the fact that $C^-(p)$ is mapped to $C^-(f(p))$. Thus the tangents to the boundaries at p and $f(p)$ are also mapped to one another. □

For each $p/q+$ or $p/q-$, separating curves $\gamma(p/q+)$ or $\gamma(p/q-)$ on the cylinder can be defined as follows. For each pair of neighboring points p_1 and p_2 in $E_{p/q}$, pick one point s in $E_{p/q+}$ (or $E_{p/q-}$) between them. Connect s to the points p_1 and p_2 along their invariant manifolds. We define $\gamma(p/q+)$ (or $\gamma(p/q-)$) as the closed curve obtained as the union of the segments (see fig. 3). An orientation along the curve can be defined in such a way that the

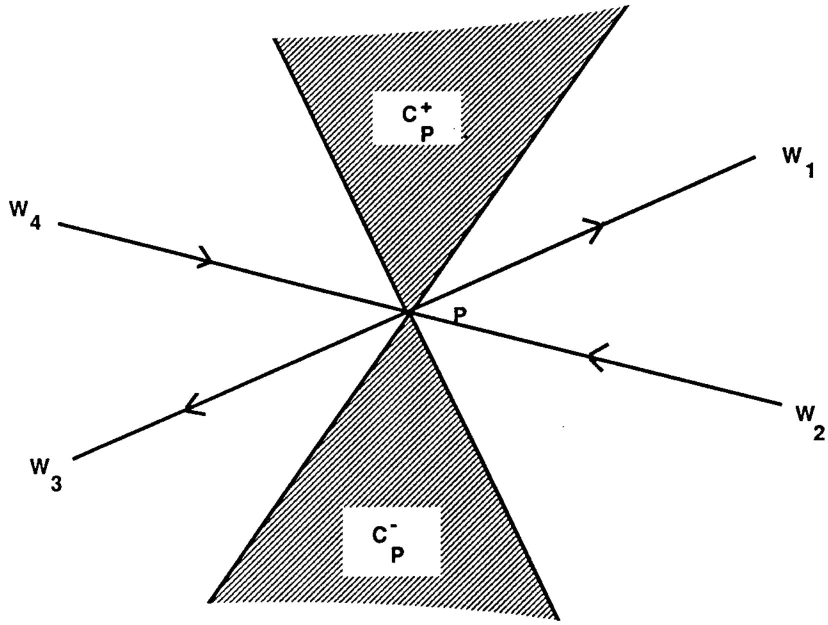

Fig. 2. The clockwise enumeration of the invariant manifolds.

orientation has a component to the right on the points of $E_{p/q}$ for the construction of $\gamma(p/q+)$ and to the left for the construction of $\gamma(p/q-)$. With this orientation, there is now a slight extension of lemma 2.2.

Corollary 2.3. At a hyperbolic minimizing point in $E_{p/q+}$ the first clockwise branch is unstable and oriented with a component to the right, the second is stable and oriented to the right, and so on. The orientation is opposite for $E_{p/q-}$.

The curves γ are not necessarily Jordan. Clearly, $\gamma(p/q+)$ separates A in an upper component containing $+\infty$, a lower component containing $-\infty$ and finitely many other components. Note, that these curves and components depend on the choice of q points $s_1, \ldots, s_q$ in $E_{p/q+}$, respectively, $E_{p/q-}$. In the following we will make a simplifying assumption on the character of these curves (which will be proved to hold in the case of the standard map with large enough k in a forthcoming work [21].

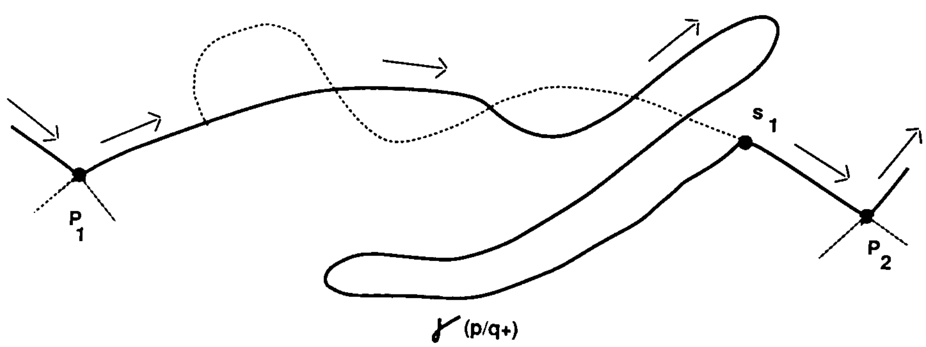

Fig. 3. Construction of the curve $\gamma(p/q+)$.

Let G be a pair of two neighboring points a and b of $E_{p/q+}$ such that in the plane $\pi(a) < \pi(b)$. The segment $\gamma(p/q+)$ that connects them is called $W^u(G)$ if it is an unstable segment, or $W^s(G)$ if it is a stable one. Let ℓ_a and ℓ_b be the left boundary of C_a^- and C_b^- respectively and r_a and r_b the right sides.

Condition 2.4. For each neighboring pair G of points a and b in $E_{p/q+}$(a to left of b):

$W^u(G) \cap (\ell_a \cup \ell_b) = \varnothing$ if the connecting segment is unstable,

$W^s(G) \cap (r_a \cup r_b) = \varnothing$ if the connecting segment is stable.

Similarly for $\gamma(p/q-)$.

We remark that some types of intersection are a priori excluded. Suppose for example that we are interested in an unstable connecting segment of $\gamma(p/q+)$ which we parametrize, starting at the point a, by $\gamma(t)$. Let $\psi(\gamma(t))$ be the angle of the tangent to $\gamma(t)$ with the positive vertical. Counting clockwise as positive, define $\phi(\gamma(t))$ as

$$\phi(\gamma(t)) = \int_0^t \frac{\mathrm{d}}{\mathrm{d}t} \psi(\gamma(t))\, \mathrm{d}t.$$

If i_a is in $W^u(G) \cap \ell_a$ and i_b in $W^u(G) \cap \ell_b$, the remark is that $\phi(i_a) > 0$ and $\phi(i_b) > 0$.

One proves this by showing that if for example $\phi(i_a) < 0$, then the inverse image under f of $W^u(G)$ has an intersection point with the same property. But we know that inverse images of $W^u(G)$ eventually land in the local unstable manifolds to $E_{p/q}$ which do not have this property.

Suppose we choose points $s_1, \ldots, s_q$ in the construction of $\gamma(p/q+)$ such that condition 2.4 holds. The figure consisting of $W^u(G)$, ℓ_a, and ℓ_b (see fig. 4) then separates the plane in two components (similarly for $W^s(G)$, r_a, and r_b). Only one of these components contains $+\infty$. The other one is called "below ab". We now define "below $\gamma(p/q+)$" as follows: a point x is "below $\gamma(p/q+)$" if

$x \in C_a^-$ for some $a \in E_{p/q+}$

or

$x \in$ "below ab"

for some neighboring pair a and b in $E_{p/q+}$.

"Above $\gamma(p/q+)$" is the complement of $\{\gamma \cup$ "below $\gamma(p/q+)$"$\}$. One gives a similar definition for "above $\gamma(p/q-)$" and "below $\gamma(p/q-)$". We will use the symbols $<$ and $>$ for "below" and "above", respectively.

Remark. Note that the definition of above and below $\gamma(p/q+)$ is not symmetric. The same

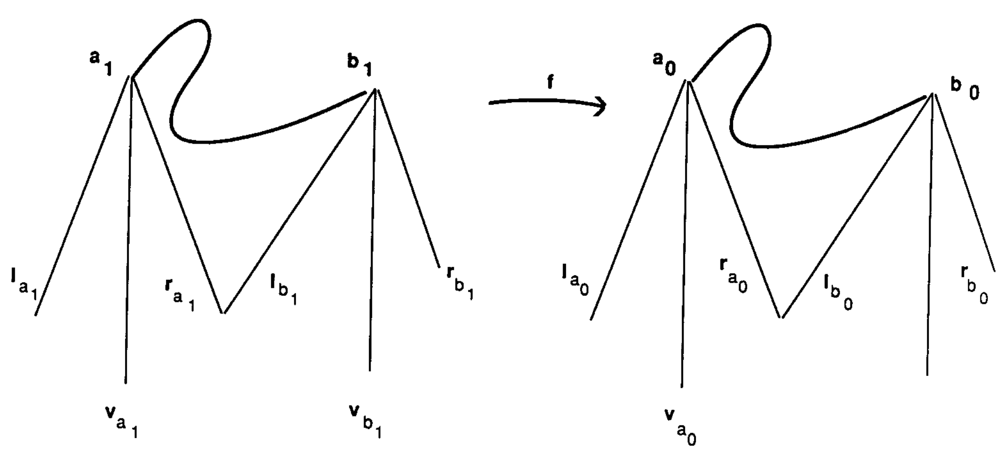

Fig. 4. The separating sets in the plane.

holds for $\gamma(p/q-)$. Note further that in the case that γ is a Jordon curve, this notion coincides with the standard interpretation of "above" and "below".

Theorem 2.5. For any $\gamma(p/q+), \gamma(p/q-)$ for which condition 2.4 holds, we have:

(i) If $\alpha > p/q$, then $M_\alpha > \gamma(p/q+)$ and $M_\alpha > \gamma(p/q-)$, and

(ii) if $\alpha < p/q$, then $M_\alpha < \gamma(p/q-)$ and $M_\alpha < \gamma(p/q+)$.

Proof. We only prove the first half of the first statement. The statement will follow from a contradiction by supposing that there is a point $x \in M_a$ such that x lies on the "wrong" side of $\gamma(p/q+)$.

Suppose without loss of generality that the segment that connects the neighboring points a_0 and b_0 of a "gap" G_0 is unstable. Let $x \in M_\alpha$ with $\alpha > p/q$. As noted before, x cannot lie in $C^-_{a_0}$ or $C^-_{b_0}$. It remains to be proved that x is not contained in the region S_0 (possibly consisting of more than one component) bounded by $W^u(G_0)$, r_{a_0}, and ℓ_{b_0}.

Iterate by f^{-1}. Then S_0 is mapped into the region bounded by $W^u(G_1)$, r_{a_1}, and ℓ_{b_1}. By Aubry's fundamental lemma, the point $x \in S_0$ cannot be mapped to a cone. Therefore it must land in S_1, which is the region bounded by $W^u(G_1)$, r_{a_1}, and ℓ_{b_1}. We can continue this, inductively defining S_n containing $f^{-n}(x)$, until, for some n, S_n lies in an ϵ-neighborhood of a hyperbolic periodic point.

But this neighborhood can be chosen so small that f^q restricted to it is very nearly linear. By lemma 2.2, we know the orientations of the local invariant manifolds (see fig 5). Orbits of points in S_n under f^q lie on hyperbolae. Any order preserving orbit in S_n with rotation number greater than p/q must satisfy

$$\pi(f^q(y)) > \pi(y).$$

These requirements are incompatible and thus x cannot map to S_n. □

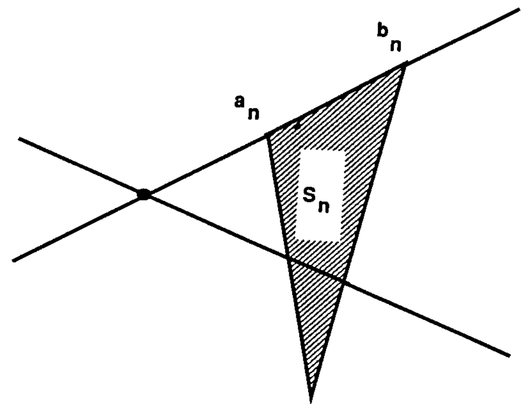

Fig. 5. The region S_n in a small neighborhood of a hyperbolic periodic point.

Notice, that we can compare two irrational sets as well, since there are always pairs of $\gamma(p/q+)$ and $\gamma(p/q-)$ that separate them. So, theorem 2.6 follows immediately.

Theorem 2.6. "Monotonicity". If for all $p/q+$ and $p/q-$ in I, $\gamma(p/q+)$ and $\gamma(p/q-)$ can be constructed that satisfy condition 2.4, then $\alpha > \beta$ implies E_α lies above E_β.

3. Hyperbolicity

Here, we prove that the invariant minimizing sets close enough to rationals are hyperbolic (and thus for irrational rotation numbers Cantor sets). One expects these sets to be hyperbolic as soon they break up (see ref. [8]). The rest of the section is devoted to a corollary stating that the Lyapunov exponent of such sets depends continuously on the rotation number.

We start with some generalities concerning the hyperbolic sets that we are interested in. Again, we assume f to be generic, so that eqs. (1.1) hold. The set $H_N = \bigcup_{\rho \in N} E_\rho$ is compact if N is a closed interval in I^+. According to Lanford [6], a compact invariant set H is a (uniformly) hyperbolic set, if the tangent space of each point x of the set is spanned by stable and unstable spaces and if the following holds. The tangent vectors in the stable space must be contracted exponentially (as μ^n, $\mu < 1$ uniformly on the set) under $Df^n(x)$, and the same holds for vectors in the unstable spaces under $Df^{-n}(x)$. These requirements imply that the local stable and unstable manifolds $W^s_\epsilon(x)$ and W^u_ϵ, tangent to the stable and unstable spaces, are continuous as functions of $x \in H$, and that their diameter is uniformly bounded away from zero [6]. That, in turn, implies, that there is a $\delta_0 > 0$, so that for any pair x and y in H whose distance is less than δ_0, $W^s_\epsilon(x)$ and $W^u_\epsilon(y)$ have a unique intersection point [6].

In the following, we will establish the genericity of hyperbolicity. To do that, we use a cone field criterion as also described in ref. [7].

Theorem 3.1. "Hyperbolicity". Let h be a hyperbolic set for f, then there exists a compact neighborhood H of h so that $\bigcap_{i=-\infty}^{\infty} f^i(H)$ is also a hyperbolic set for f.

Proof. Since h is compact and hyperbolic, one can construct a cone field $\{C_x\}_{x\in h}$ which is mapped strictly into itself by Df. One does this by constructing a norm on the tangent bundle restricted to h such that Df is expanding on the unstable bundle (choose unit vector $e_u(x)$) and contracting on the stable bundle (choose unit vector $e_s(x)$), see lemma 2.1 of ref. [15]. Then choose the cone field C_x as follows: a vector $v = ae_u(x) + be_s(x)$ is in C_x if $|a| \geq |b|$.

By continuity of Df, we can extend this cone field C to a cone field $\{C'_x\}_{x\in H}$ defined on a sufficiently small neighborhood H of h such that Df maps the cone field C' on $f^{-1}(H) \cap H$ strictly into C' on H. Consequently, any invariant compact set in $\bigcap_{i=-\infty}^{\infty} f^i(H)$ is also a hyperbolic set. □

Corollary 3.2. (See ref. [7].) For generic f, there exists an open neighborhood U of the rational rotation numbers such that the collection of minimizing sets with rotation number in U forms a hyperbolic set.

Proof. Take $H = E_{p/q} \cup E_{p/q+} \cup E_{p/q-}$ and pick H as above. For generic f, the set H is hyperbolic. □

We let $\lambda(\rho)$ denote the Lyapunov coefficient ≥ 1 for an order preserving minimal set E_ρ. Since these sets are uniquely ergodic with invariant probability measure $\mu(\rho)$ (see ref. [12]), $\lambda(\rho)$ is well defined and constant μ almost everywhere.

Proposition 3.3. Let M_α be a hyperbolic minimizing set with irrational rotation number, then $\lambda(\rho)$ is continuous at $\rho = \alpha$.

Proof. Let h be a hyperbolic set for f with one-dimensional unstable bundle E^u. Assume E^u is orientable. Choose a continuous nowhere zero section v of E^u and consider the function

$$D(x) = \left|\frac{Df \cdot v(x)}{v(f(x))}\right|.$$

One observes that D is continuous on h. For an ergodic probability measure μ on h, its Lyapunov coefficient $\lambda(\mu)$ equals

$$\lambda(\mu) = \exp \int \ln D \, d\mu.$$

Since M_α is hyperbolic, we have by theorem 3.1 that $\bigcap_{i=-\infty}^{\infty} f^i(H)$ is also hyperbolic where H is a sufficiently small neighborhood of M_ρ. One knows that [12], for α irrational, $\bigcap_{i=-\infty}^{\infty} f^i(H)$ contains nearby minimizing sets M_ρ with invariant probability measures $\mu(\rho)$ and that $\lim_{\rho\to\alpha} \mu(\rho) = \mu(\alpha)$, in the weak topology. Consequently $\lim_{\rho\to\alpha} \lambda(\rho) = \lambda(\alpha)$. □

Remark. The function $D(x)$ is not canonical. If one chooses a different section $v'(x) = \phi(x) v(x)$, then

$$D'(x) = \frac{\phi(x)\, D(x)}{\phi(f(x))}.$$

However, the Lyapunov coefficient is insensitive to this: one easily checks that

$$\int \ln D' \,\mathrm{d}\mu = \int \ln D \,\mathrm{d}\mu .$$

Finally, a simple result that follows from hyperbolicity. Let x be a point in E_α where both forward and backward images of the gap G accumulate. In a neighborhood of x one can connect these tiny gaps by local stable and unstable manifolds which are almost straight segments that make a positive angle with each other. From the accumulation of these different gaps, one concludes the following.

Remark. The set E_α cannot be imbedded in a C^1 curve.

4. Single gap

Consider the projections of the Aubry–Mather sets on the x-axis. By a "gap" G in E_ρ, we mean [4, 5] a pair of points in E_ρ whose projections bound an interval that contains no point of the projection of E_ρ. The length $|G|$ of the gap is simply the length of that interval. The meaning of $F^i(G)$ is then also clear. The main result of this section is that, under certain assumptions, E_α has only one gap orbit.

Before we embark on the general course, we first formulate the single intersection hypothesis, which will be needed in theorem 4.3. Denote the finite pieces of invariant manifolds to E_ρ that connect the endpoints of a gap J in E_ρ by $W^s(J)$ and $W^u(J)$. We will say that f satisfies the single intersection hypothesis if all $E_{p/q}$, $E_{p/q-}$, and $E_{p/q+}$ are unique (true for generic f, see section 1), and if, for a gap J in $E_{p/q+}$ or $E_{p/q-}$, $W^u(J) \cap W^s(J)$ contains a single point (which then has to be the minimax), see fig. 6.

Lemma 4.1. If E_α is hyperbolic, then it has at most finitely many gap orbits.

Proof. From the generalities mentioned in section 2, one can deduce that for a hyperbolic set, there is a $\delta_0 > 0$ with the property that if x and y are points in E_α, then

$$\text{if} \quad d\big(F^i(x), F^i(y)\big) < \delta \quad \text{for all } i \in \mathbb{Z},$$
$$\text{then} \quad x = y.$$

So, each gap orbit must have a gap of length greater than δ. □

Remark. A different proof of this fact was given by MacKay [9].

Proposition 4.2. For f generic, there is a neighborhood N of p/q, such that if $\alpha \in N$, then E_α has one gap orbit.

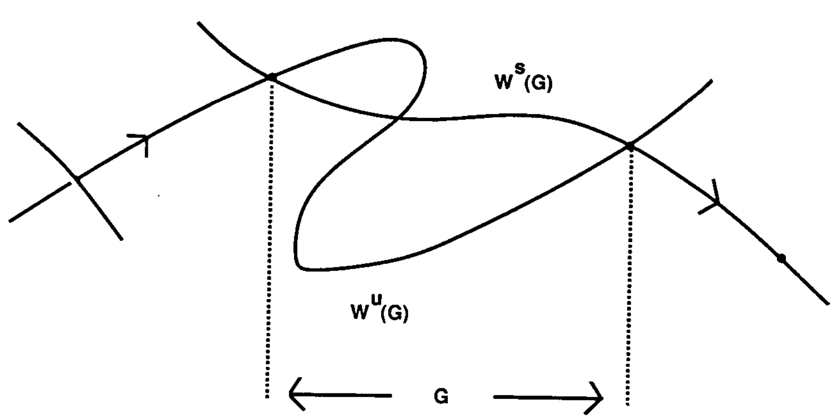

Fig. 6. The single intersection hypothesis.

Proof. If f is generic, then $E_{p/q+}$ and $E_{p/q-}$ consist of a single homoclinic orbit (see section 1), and thus have one single gap orbit. According to proposition 1.3, for α close enough to p/q (without loss of generality $\alpha > p/q$, E_α is hyperbolic. Pick any gap in $E_{p/q+}$. There is an $m > 0$, such that all gaps in $E_{p/q+}$ with length greater than $\delta/3$ are contained in $\{F^i(G)\}_{i=-m}^{i=+m}$. By eq. 1.1, we can pick α so close to p/q, that $d(E_\alpha, E_{p/q+}) < \delta/3$. Then, by uniform continuity of $\{f^i\}_{i=-m}^{i=+m}$, every gap with length greater than δ is shadowed by a gap in $E_{p/q}$ and vice versa. □

This result is somewhat unsatisfactory, since one would like to have a statement for α fixed. In studying the global stability of these systems, numerical work indicates that the curve with rotation number $(1+\sqrt{5})/2$ (or a related diophantine number, see ref. [10]) is the last one to break up. As a consequence, one is especially interested in the gap structure of this set, being, as it were, a "bottleneck" for the dynamics of f.

We are now in a position to prove the main result.

Theorem 4.3. "Single gap". Let α be irrational. If
(a) f satisfies single intersection,
(b) $\mathrm{Hlim}_{\rho\to\alpha} E_\rho = E_\alpha$, and
(c) E_α is hyperbolic,
then E_α has only one gap orbit.

Remark. To prove the result for a single α, it is enough to require that f satisfy a local variant of the single intersection hypothesis.

Remark. Numerical work suggests that (a) holds for the standard map. The more general case is commented upon after the proof. As stated before, (c) has been proved only in a restricted setting [3], but appears to hold more generally. One suspects that (b) is generically true, see ref. [2].

Proof. We will assume from now on that there are two independent gaps in E_α and eventually deduce from that a contradiction with the single intersection hypothesis.

By assumption (c), we can choose an interval K^+ in I^+ of rotation numbers ρ such that the set $E = \bigcup E_\rho$ (union over K^+) is uniformly hyperbolic. By assumption (b), we can choose the interval K^+ so that in addition we have:

$$\mathrm{Hdist}(E_\rho, E_\alpha) < \delta \ll \delta_0, \tag{4.1}$$

where δ_0 is a lower bound for the diameter of the local invariant manifolds (see section 3) to E_ρ with $\rho \in K^+$. Thus each $E_{p/q+}, E_{p/q-}$ in E is contained in the *local* stable and unstable manifolds to $E_{p/q}$. Also each gap orbit in E must have a gap which is larger than δ_0. Denote the two "big" independent gaps in E_α by G_α and H_α. According to eq. (4.1), we can uniquely define by taking G_ρ and H_ρ to be approximating gaps in E_ρ. If ρ is rational, rational+, or rational−, then, of course, we have that there exists an $m(\rho)$ with

$$f^{m(\rho)}(G_\rho) = H_\rho. \tag{4.2}$$

By the continuity of f, it is clear that $m(\rho)$ has to become unbounded as $f \to \alpha$.

The question we address now, is: How does $m(\rho)$ change as a function of ρ in K^+? From relation (1.1b) and the continuity of f, it transpires that if $\omega \in K^+$ is irrational, then either $m(\omega) = \infty$ or $m(\rho)$ is continuous at ω.

As a consequence, $m(\rho)$ can make finite jumps only at rational values of ρ. This situation is depicted in fig. 7, where $m(p/q-)$ is not equal to $m(p/q+)$. Because $m(p/q)$ is unique, we have

$$m(p/q+) = m(p/q-) + kq. \tag{4.3}$$

Without loss of generality, we take

$$m(p/q-) > 0.$$

We will now argue that k is negative. Suppose, then, that k is positive. Relations (4.1) and (4.2) together with the well ordered character of E_ρ,

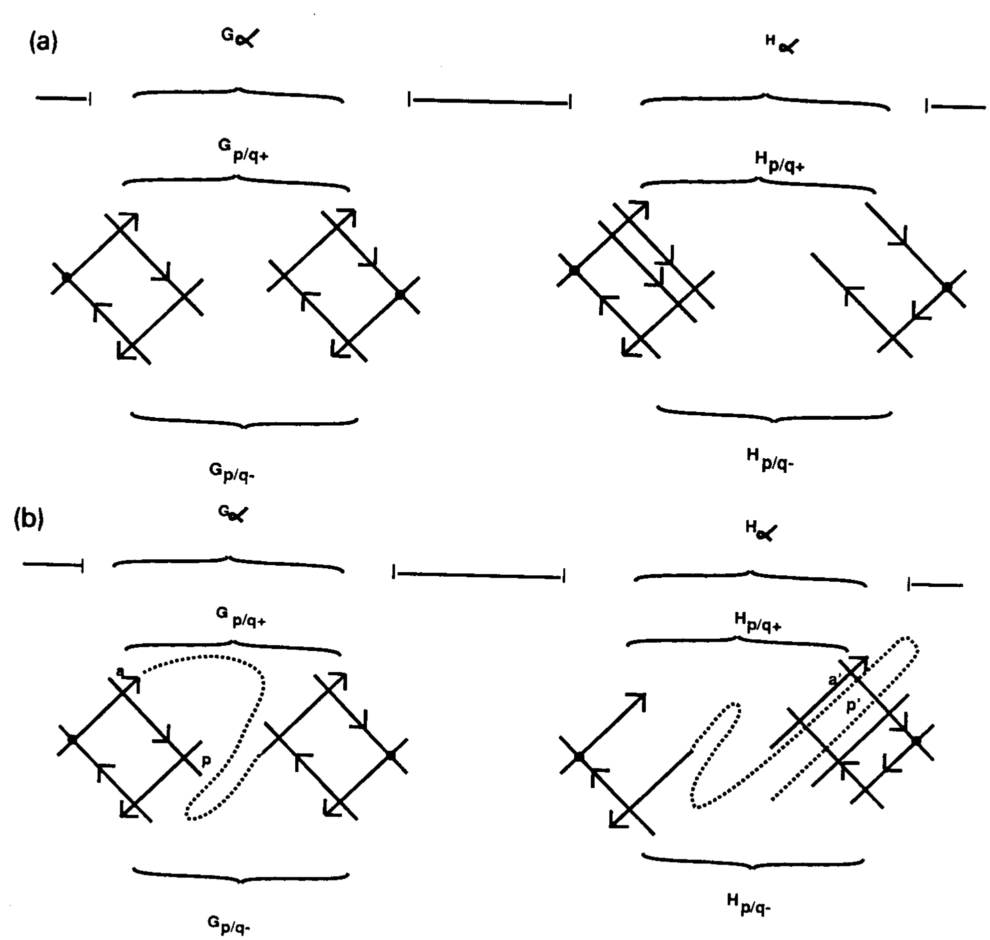

Fig. 7. The local invariant manifolds with rational rotation number of the gap G and their $m(p/q-)$th image, drawn if (a) $m(p/q+)>0$ and (b) $m(p/q+)<0$.

imply that $f^{m(p/q-)}(\mathrm{G}_{p/q+})$ lies in a local unstable manifold. Upon iterating this $m(p/q-)$ times back to the original gap, as in fig. 7, one encounters a contradiction (namely, that points have left the local unstable manifold under the application of f^{-1}). So k is negative.

If $m(p/q+)>0$, then (recall that $m(p/q-)$ is positive) (4.3) implies that $|m(p/q+)|<|m(p/q-)|$. So, in order to allow $m(\rho)$ to become unbounded, we need

$$m(p/q-)>0 \quad \text{and} \quad m(p/q+)<0.$$

This is the situation sketched in fig. 7b, and it is here that the final contradiction with single intersection arises. It can be seen as follows that in this case $\mathrm{W}^{\mathrm{s}}(\mathrm{H}_{p/q-})$ and $\mathrm{W}^{\mathrm{u}}(\mathrm{H}_{p/q+})$ intersect at least two times. Under forward iterates, the number of intersections involving the local stable manifolds along with the relative orientations (use corollary 2.3 to determine the orientations) is conserved. One concludes that the point marked p in the figure is mapped under $f^{m(p/q-)}$ (with $m(p/q-)>0$) to the point labelled p', with the orientation as indicated. So

$$\begin{aligned} p' \in\ & \mathrm{W}^{\mathrm{u}}\big(f^{m(p/q-)}(\mathrm{G}_{p/q-})\big) \\ & \cap \mathrm{W}^{\mathrm{s}}\big(f^{m(p/q-)}(\mathrm{G}_{p/q+})\big) \\ =\ & \mathrm{W}^{\mathrm{u}}(\mathrm{H}_{p/q-}) \cap \mathrm{W}^{\mathrm{s}}\big(f^{-kq}(\mathrm{H}_{p/q+})\big), \end{aligned}$$

where k is negative. Let a' be the image under $f^{m(p/q-)}$. Since $W^u(G_{p/q+})$ does not intersect the a–p (by single intersection) its image under $f^{m(p/q-)}$ does not intersect a'–p'. Therefore, if $W^u(H_{p/q-})$ intersects p', it is caught in an unstable "lobe" of $f^{m(p/q-)}(H_{p/q+})$ and must intersect the local stable manifold to $E_{p/q}$ again in order to leave the "lobe". By uniform hyperbolicity $W^u(H_{p/q-})$ then must intersect $f^{m(p/q-)}(H_{p/q+})$ another time. The intersections with $W^s(H_{p/q-})$ follow by the hyperbolicity of the two thus constructed points (and hence the existence of their local unstable manifolds). □

From the last paragraph of this proof, it is clear that it is sufficient to replace condition (a) in the theorem by the requirement

(a*) $W^s(H_\rho) \cap W^u(H_\rho)$ is bounded away (uniformly in K^+) from the endpoints of H_ρ.

This requirement appears to be borne out [17] by extensive numerical experiments for the standard map. (If this were not so, one would have even longer and thinner lobes formed by $W^u(H_{p/q-})$ as the $H_{p/q-}$ accumulate on the gap H_α.) We suspect that single gap is a persistent property for an open neighborhood of maps around the standard map but can only prove that for large values of the non-linearity parameter [21].

5. Turnstiles

Let α be an irrational rotation number in I and assume that E_α is hyperbolic. This section is dedicated to proving that the leakage of orbits through an Aubry–Mather set can be understood in terms of overlap criteria. It is somewhat speculative in nature, since we have to assume all the conditions that are required for theorem 4.3 to be true. Nevertheless, as stated, there is good reason to believe that the result holds for an open set of maps containing the standard map. This confidence is partly based on numerical results by various authors, especially MacKay, Meiss and Percival [11]. We will thus proceed to elaborate on some of the consequences of the theorem.

We define turnstiles as follows [11]. Let G denote a gap in E_α. Connect the endpoint of $f^n(G)$ with a straight line segment λ^s_n. Similarly, connect the endpoints of $f^{-n}(G)$ with a straight line segment λ^u_n. Clearly, $f^{+n}(\lambda^u_n)$ and $f^{-n}(\lambda^s_n)$ connects the endpoints of the gap G.

Note, that by hyperbolicity and endpoints of every gap G are connected by a branch of stable manifold, and by a branch of the unstable manifold (the future and past iterates of G collapse the gap). We will denote finite branches that connect a gap J by $W^s(J)$ and $W^u(J)$.

Proposition 5.1.

(i) $\operatorname{Hlim}_{i\to\infty} f^{-i}(\lambda^s_i) = W^s(G)$,

(ii) $\operatorname{Hlim}_{i\to\infty} f^{+i}(\lambda^u_i) = W^u(G)$.

Proof. It suffices to prove (i) only. Because the gaps $f^i(G)$ do not overlap, the sum of their lengths is less or equal to one. So there is an N such that the length $|f^{N+i}(G)| < \epsilon$ for all $i \geq 0$. Then, by uniform hyperbolicity, for each i the left and the right endpoint of $f^{N+i}(G)$ have to lie on the same local stable manifold. But then we also have $\operatorname{Hlim} f^{-i}(\lambda^s_{N+i}) = W^s(f^N(G))$, because λ^s_{N+i} is transversal to the local *stable manifold of* f^{-1}. □

As in section 3, we can define a hyperbolic set H_N that contains E_ρ with ρ in N, a neighborhood of α. We now take $\{p_i/q_i\}$ to be a sequence in N with α as its limit. Define $G_i^{+/-}$ as the gaps in $E_{p_iq_i-}$ and E_{p_i/q_i+} that have limit G (as $i \to \infty$).

The main result of this section is

Theorem 5.2. "Convergence of turnstiles". Let f satisfy the same conditions as in theorem 4.3, then

(i) $\operatorname{Hlim}_{i\to\infty} W^s(G_i^+) = W^s(G)$,

(ii) $\operatorname{Hlim}_{i\to\infty} W^s(G_i^-) = W^s(G)$,

(iii) $\operatorname{Hlim}_{i\to\infty} W^u(G_i^+) = W^u(G)$,

(iv) $\operatorname{Hlim}_{i\to\infty} W^u(G_i^-) = W^u(G)$.

Proof. It suffices to prove the first statement only. Since we are dealing with gaps G_i^+ and H_i^s only, we will drop the unnecessary superscripts + and s on them.

By the single gap theorem (4.3), we can choose an integer N such that

$$\sum_{j=-N}^{j=+N} |f^j(G)| > 1 - \epsilon/2.$$

By eqs. (1.1) and the continuity of f it follows that, for i sufficiently large,

$$\sum_{j=-N}^{J=+N} |f^j(G_i)| > 1 - \epsilon.$$

So, if $H_i = f^{+2N}(G_i)$, its length and that of its forward images are smaller than ϵ. Thus, by uniform hyperbolicity, its endpoints lie on the same local stable manifold. Let $H = f^{+2N}(G)$. Then, by taking $2N$ (N fixed) inverse iterates, one concludes that (i) is true, if and only if $W^s(H_i)$ converges to $W^s(H)$, but that follows directly from the fact that uniform hyperbolicity implies that local stable and unstable manifolds vary continuously as function of their base-point. So the theorem is proved. □

This theorem immediately implies that the diffusion through an Aubry–Mather set can be understood as a limit of "resonance overlaps". The way to see this is to construct curves $\gamma(p_i/q_i+)$, with $p_i/q_i \uparrow \alpha$, as in section 2, except that now we take s_i to be the left endpoint of the gaps G_i, defined as in the proof of the previous theo-theorem. The other points needed in the construction are taken to be the $q_i - 1$ images of s_i (see fig. 8). Iterate the area B_i below $\gamma(p_i/q_i+)$ once, and it is clear that one can define a region I_i^+ with

$$f^{-1}(I_i^+) > \gamma(p_i/q_i+) \quad \text{and} \quad I_i^+ < \gamma(p_i/q_i+).$$

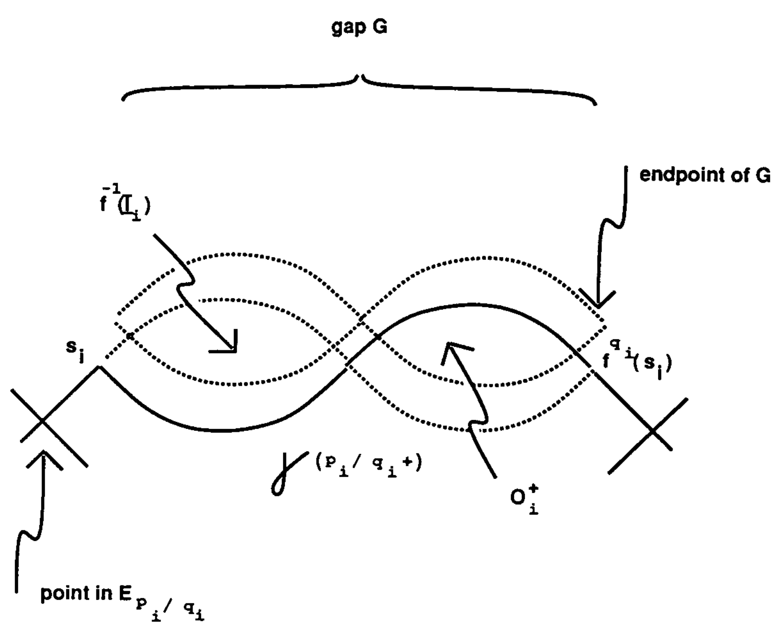

Fig. 8. Resonance overlaps.

Similarly, a region O_i^+ can be defined by

$$O_i^+ < \gamma(p_i/q_i+) \quad \text{and} \quad f^{+1}(O_i^+) > \gamma(p_i/q_i+).$$

Theorem 5.2 immediately implies that I_i^+ converges in the Hausdorff limit to I and O_i^+ to O, and I and O are the areas enclosed by $W^s(G)$ and $W^u(G)$. The corresponding statement holds for $\gamma(r_i/s_i-)$, if we define O_i^- and I_i^- in a similar vein and if s_i now accumulates to the right endpoint of the gap G. By convergence of turnstiles, it follows, that the "resonance overlaps" $O_i^+ \cap O_i^-$ and $I_i^+ \cap I_i^-$ limit on O and I, respectively. We summarize this loosely with the following corollary:

Corollary 5.3. Under the conditions of theorem 4.3, the diffusion through an Aubry–Mather set is a limit of resonance overlaps, if E_α is hyperbolic.

If one assumes that E_α is hyperbolic whenever it is a Cantor set, it is also clear that the above proves the following (geometric) criterion for the non-existence of an invariant circle with rotation number α (compare ref. [12], whose result is more general but less geometric).

Corollary 5.4. If E_α is hyperbolic whenever it is a Cantor set, then, under the conditions of theorem 4.3, $\text{Hlim}\, O_i^+ \cap O_i^-$ converges and has positive area if and only if E_α is not (contained in) an invariant circle.

6. Concluding remarks

We have argued that one of our main results, the single gap theorem, holds for f satisfying a number of conditions (see theorem 4.3). It is not hard to find a counter-example. Suppose f is a twist map for which the single gap theorem holds and let [14] $g = f^2$, then we have the equality $E_\alpha(f) = E_{2\alpha}(g)$ (as sets). If G and f(G) are gaps in $E_\alpha(f)$, they can never be mapped into each other by g. Moreover, if $E_\alpha(f)$ is hyperbolic, then so is $E_{2\alpha}(g)$, and under small perturbations of g' or g, $E_{2\alpha}(g')$ retains the same number of gap orbits. (Note that $g+\epsilon$, for ϵ small enough, does not satisfy requirement a* at the end of section 4.) One can conclude, therefore, that theorem 4.3 certainly will not hold generically.

This counter-example is also of interest in connection with the monotonicity theorem. Note that in our definition of "above" and "below", we have assumed that $E_{p/q}$, $E_{p/q-}$, and $E_{p/q+}$ are unique for all $p/q \in I$ (true for generic f). This does not hold for the map g. As we let ϵ run through zero, we can witness that the minimizing sets jump. And so, our definition 2.4 also yields curves $\gamma(p/q+)$ and $\gamma(p/q-)$ that jump. A similar comment is valid for the convergence of turnstiles theorem.

The hyperbolicity of the Aubry–Mather sets in a Birkhoff zone is a very powerful tool (see section 3). We expect it to hold in a general situation. It would be useful to have a proof for a more general case than the one discussed in ref. [3].

Appendix

The lift F of the twist map has a generating function h satisfying

$$F(x,y) = (x,y') \quad \text{iff} \quad y = -\partial_1 h(x,x'), \quad y' = \partial_2 h(x,x').$$

Proposition A.1. For C^k-generic $(k>1)$ h, the global minima $E_{p/q}$, $E_{p/q-}$, $E_{p/q+}$ are unique.

Proof. The Kupka–Smale theorem for area preserving maps (see ref. [19]) implies in this context that generically there are a finite number of periodic orbits of given type. Moreover, if they are hyperbolic, two fundamental domains of a stable and an unstable invariant manifold associated with these these orbits intersect finitely many times, and the intersections are transversal. This implies that there are finitely many orbits of type $E_{p/q-}$ or $E_{p/q+}$.

For p/q fixed, let $\{\xi_i\}$ be a (globally) minimizing orbit of type $E_{p/q+}$. Consider the collection C of sequences $\{x_i\}_{i=-\infty}^{i=\infty}$ of type $E_{p/q+}$. If $\{\eta_i\}$ in C is an orbit, then each interval (ξ_i, ξ_{+q}) close enough to a point in $E_{p/q}$ contains exactly one of the η_j and none of the ξ_j. We have

$$\sum [h(\eta_i, \eta_{i+1}) - h(\xi_i, \xi_{i+1})] \geq 0.$$

We change

$$h \to h'(x, x') = h(x, x') + \phi(x),$$

where $\phi(x)$ is a "bump" function which is positive on the interval (ξ_i, ξ_{i+q}) for some fixed i but has vanishing first and second derivatives on ξ_i and ξ_{i+q}. Now $\{\xi_i\}$ is a unique minimizing orbit. This proves that the property of having a unique global minimum of type $E_{p/q+}$ is dense.

For p/q fixed, suppose that fundamental domains of stable and unstable manifolds to $E_{p/q}$ intersect finitely many times (open and dense). Then there are finitely many orbits of type $E_{p/q+}$. Since the intersections are transversal, their number is conserved under small perturbations. The value of the above sum then also changes continuously. Therefore, uniqueness is an open property. □

Acknowledgements

One of the authors (J.J.P.V) has the pleasure to thank the Institute for Theoretical Physics of the City University of Amsterdam and the Mathematics Department of Warwick University for their hospitality. He acknowledges useful discussions with Hans Capel, Ko van der Weele, Robert MacKay, and Maciej Wojtkowski. His research was supported by a postdoctoral fellowship at the Autonomous University of Madrid and by DOE grant DE-AC-02-83-ER-13044 at the Rockefeller University. The other (F.M.T.) carried out his research at Queens College in Flushing (NY). This article was written in 1988.

Both of us thank John Mather for a very illuminating conversation.

References

[1] S. Aubry, in: Structures et Instabilitiés ed. C. Godreche (Les Editions de Physique, Paris 1986).
[2] V. Bangert, Mather sets for twist maps and geodesics on tori, preprint (1986).
[3] D. Goroff, Erg. Th. Dyn. Syst. 5 (1985) 337–339.
[4] A. Katok, Erg. Th. Dyn. Syst. 2 (1982) 185–194.
[5] A. Katok, More about Birkhoff periodic and Mather sets for twist maps, Maryland preprint (1982).
[6] O. Lanford, in Regular and Chaotic Motions in Dynamic Systems, ed. G. Velo and A.S. Wightman, NATO ASI series B: Physics, vol. 18 (Plenum, New York, 1985).
[7] P. Le Calvez, Les ensembles de L'Aubry–Mather d'un diffeomorphisme conservatif de l'anneau deviant la verticale sont en general hyperboliques, preprint (1987).
[8] W. Li and M. Bak, Phys. Rev. Lett. 57 (1986) 655–658.
[9] R.S. MacKay, Hyperbolic Cantori ..., Warwick preprint (1987).
[10] R.S. MacKay, Introduction to dynamics of area preserving maps, Warwick preprint (1986).
[11] R.S. MacKay, J.D. Meiss and I.C. Percival, Phys. Rev. Lett. 52 (1984) 697–700.
[12] J.N. Mather, Publ. Math. I.H.E.S. 63 (1986) 153–204.
[13] J.N. Mather, Destruction of invariant circles, ETH Zurich preprint (1987).
[14] R.S. MacKay, private communication.
[15] Z. Nitecki, Differentiable Dynamics (MIT Press, Cambridge 1971).
[16] A. Olvera and C. Simo, Physica D 26 (1987) 181–192.
[17] J. Palis and W. de Melo, Geometric Theory of Dynamical Systems (Springer, Berlin, 1982).
[18] I.C. Percival, private communication.
[19] R.C. Robinson, Am. J. Math. 92 (1970) 562–603, 897–906.
[20] J.J.P. Veerman and F.M. Tangerman, J. Stat. Mech. 56 (1989) 83–98.
[21] J.J.P. Veerman and F.M. Tangerman, Intersection properties of invariant manifolds in certain twist maps, preprint (1989).