PUBLICATIONS MATHÉMATIQUES DE L'I.H.É.S.

JOHN N. MATHER

A criterion for the non-existence of invariant circles

Publications mathématiques de l'I.H.É.S., tome 63 (1986), p. 153-204
<http://www.numdam.org/item?id=PMIHES_1986__63__153_0>

A CRITERION FOR THE NON-EXISTENCE OF INVARIANT CIRCLES

by JOHN MATHER

TABLE OF CONTENTS

1. Introduction

This paper is based on joint work with A. Katok. Results similar to those obtained here have been obtained by Aubry, Le Daeron and André [7]. Our method is inspired by some remarks of G. D. Birkhoff on the billiard ball problem. We quote Birkhoff's remarks in § 2. The minimax principle, described by Birkhoff, plays a fundamental role in this paper. We discuss the minimax principle in § 3. Birkhoff's remarks generalize in a straightforward way to area-preserving monotone twist mappings of the annulus. This generalization is described in § 4. Our new results are described in §§ 5-10. The rest of the paper consists of proofs of the results announced in §§ 5-10.

Our main result is a criterion for the existence of invariant circles for a certain class of area preserving diffeomorphisms of the annulus, which we call " monotone twist maps ". Our motivation for studying this question comes from the version of celestial mechanics studied by Poincaré, Levi-Civita, G. D. Birkhoff, Kolmogoroff, Arnold, Moser, Sternberg, and others who are essentially pure mathematicians. The studies of these mathematicians led to the consideration of invariant circles in area preserving mappings. It is seen from these studies that it is of fundamental importance to understand when invariant circles do and do not exist.

More recently, Percival used a variational principle to study the question of the existence of invariant circles numerically [27], [28]. This reminded me of Birkhoff's proof [9] of the existence of periodic orbits in the billiard ball problem. Of course, one cannot prove the existence of invariant circles by such a method, because very frequently invariant circles do not exist. The orbits $\{(x_n, y_n)\}_{n\in\mathbf{Z}}$ which lie in the invariant circles are called *quasi-periodic*, because they have expansions in Fourier series

$$(x_n, y_n) = \sum_k C_k e^{i\omega k n},$$

where C_k denotes a vector in $\mathbf{R}^2$. Around March 1981, I realized that a simple modification of Birkhoff's argument [9], using Percival's Lagrangian, provides a proof of the existence of quasi-periodic orbits in this sense. Since there may be no invariant circle of frequency ω, these quasi-periodic orbits do not necessarily lie on an invariant circle. However, under the hypothesis which I considered (monotone twist hypothesis), they necessarily exist and lie on a minimal set. The restriction of the original transformation to this minimal set has dynamical properties very similar to an irrational rotation of a circle. In fact, it is semi-conjugate to such a rotation by a continuous mapping which is $1-1$ except on a countable set. Moreover, if there is an invariant circle, this minimal set is necessarily in the circle.

I proved these results by maximizing Percival's Lagrangian. Another proof was later given by Katok, who showed that if ω is an irrational number, and $\frac{p_n}{q_n}$ is a sequence of rational numbers converging to ω, then the Birkhoff periodic orbits of type (p_n, q_n) converge in the Hausdorff metric to an invariant set. This invariant set contains a minimal set, which is precisely the minimal set (associated to the frequency ω) which I constructed. Still another proof has been given by Aubry, Le Daeron, and André [7]. It seems that this proof is the result of ideas Aubry had developed over several years ([1], [2], [3], [4], [5], [6], [8]). The proof of Aubry et al. is quite different from either Katok's proof or my proof.

In March 1982, after long discussions with Katok, I discovered the results which

I will describe in this paper. In the second half of my graduate course in the spring of 1982, I lectured on these results. The proofs I give here follow the method I described in my course, although I have made considerable efforts to improve the exposition. I would like to thank the students in my course, T. Folk, D. Goroff, R. Llave, R. MacKay, D. Nance and T. Pignataro, whose questions, comments, enthusiasm and willingness to listen to some very obscure lectures greatly aided me in writing up this article.

The principle result of this paper gives a numerical invariant, for each irrational number ω, which is always non-negative, and which vanishes if and only if there is an invariant circle of frequency ω. One way of defining this number is the following. Let p/q be a rational number, expressed in lowest terms. In the case of the billiard ball problem, Birkhoff [9] showed the existence of at least two periodic orbits of type (p, q). One of these is obtained by maximizing the perimeter of a polygon; the other is obtained by a minimax principle. Birkhoff's arguments generalize to the case of monotone twist mappings of the annulus. In this more general setting, the perimeter of the polygon is replaced by the " action " of a sequence of points in $\mathbf{R}$. But, we still get a max orbit of type (p, q), with action $W_{p,q,\max}$, and a minimax orbit of type (p, q) with action $W_{p,q,\text{minimax}}$. We set $\Delta W_{p,q} = W_{p,q,\max} - W_{p,q,\text{minimax}}$. For an irrational number ω, we will prove $\Delta W_{p,q}$ converges to a limit ΔW_ω as the rational number p/q (in lowest terms) tends to ω. Our principle result states that $\Delta W_\omega = 0$ if and only if there exists an invariant circle of frequency ω.

While I was preparing this text, I became aware (in May 1982) of Aubry's work ([1], [2], [3], [4], [5], [6], [7], [8]). Starting from a question in solid state physics which is completely different from the question we started with, he has arrived at results which are similar to ours. In particular, he defines a number which he calls the Peierls energy barrier. A principal result in his paper [7] is that the Peierls energy barrier vanishes if and only if there is an invariant circle (for the given frequency). This is closely related to our result. The Peierls energy barrier is a lower bound for ΔW_ω. We will discuss the Peierls energy barrier in our terminology in § 25.

The methods of the paper of Aubry, Le Daeron, and André are very interesting and quite different from our methods. They also have a number of results in their preprint which we have not proved, and this paper contains a number of results which they have not proved. Specifically, we have the following results: continuity results for ΔW_ω and semi-continuity results for the minimal sets whose existence was proved in [21]. Moreover, we show the connection with the classical results described by Birkhoff in [9].

Our setting is somewhat different from Aubry's. We consider diffeomorphisms of a bounded annulus $\overline{A} = (\mathbf{R}/\mathbf{Z}) \times [0, 1]$; Aubry considers diffeomorphisms of an infinite annulus $(\mathbf{R}/\mathbf{Z}) \times \mathbf{R}$. The difference is that our annulus has finite area and a boundary; Aubry's has infinite area and no boundary; moreover, in Aubry's set-up the diffeomorphism twists arbitrarily much in the negative direction near the lower end, and arbitrarily much in the positive direction near the upper end. However, these are only technical differences; they do not affect the main ideas. Nevertheless, treating the case where there is a boundary does require a lengthy discussion of technicalities at several points.

One of my principle aims in developing this theory was to find a rigorous criterion for existence or non-existence of invariant circles which would be possible to implement numerically. I have had several very useful conversations with J. Greene, J. Percival and R. MacKay, who have been in the forefront of numerical studies of invariant circles. In particular, I learned the whole approach of studying invariant circles through maximization-minimization in Percival's formulation) from Percival. The influence of Greene may be seen in the use of approximating periodic orbits to study quasi-periodic orbits, which follows an idea Greene has used for numerical purposes.

This idea of using approximating periodic orbits to study quasi-periodic orbits also derives from Katok's ideas in his paper [17] and in many conversations I have had with him. In [21], I used a method analogous to Birkhoff's to construct quasi-periodic orbits. In [17], Katok showed that the existence theorem for quasi-periodic orbits could be obtained from Birkhoff's existence theorem for periodic orbits and a limiting procedure. He also suggested to me (in conversation) that there should be something which corresponds to the Birkhoff minimax orbits for the quasi-periodic (as opposed to periodic) case. His original idea was a second Cantor set, in the case the invariant set which I constructed was a Cantor set and not a circle. This idea turned out to be very valuable; following it, I found not a second Cantor set, but an orbit homoclinic to the first Cantor set. This, in turn, led to the orbits I describe in this paper.

I believe it should be possible to develop numerical methods to compute ΔW_ω to an arbitrary degree of precision, together with a rigorous estimate for the error. If that were done, the main result of this paper would give a means of proving the non-existence of invariant circles when they, in fact, do not exist. (Hence, the title of this paper.) On the other hand, computing ΔW_ω to arbitrary precision will never tell whether it is zero or not, so the result in this paper does not provide a means of proving the existence of invariant circles, when they do, in fact, exist.

Newman and Percival [26] also have a criterion (different from the above) for proving non-existence of invariant circles. A rigorous proof of their criterion follows easily from the work of Aubry, Le Daeron and André [7].

2. The Periodic Orbits of Birkhoff in the Billiard Ball Problem

Since I got some of my basic ideas for this paper and some of the previous papers I have written on this subject ([21], [22], [23], [24]) from considering Birkhoff's description of the billiard ball problem, I will quote what he says at length [9, § 2]:

" In order to see how the theorem of Poincaré and its generalization can be applied to dynamical systems with two degrees of freedom, I propose to draw attention to a special but highly typical system of this sort, namely that afforded by the motion of a billiard ball upon a convex billiard table (Fig. 1). This example is very illuminating for the following reason: Any dynamical system with two degrees of freedom is isomorphic with the motion of a particle on a smooth surface rotating uniformly about a fixed axis and carrying a conservative field of force with it. In particular if the surface is not rotating and if the field of force is lacking, the paths of particles will be geodesics. If the surface is conceived of as convex to begin with and then gradually flattened to the form of a plane convex curve C, the ' billiard ball ' problem results. But in this problem, the formal side, usually so formidable in dynamics, almost completely disappears, and only the interesting qualitative questions need to be considered. If C happens to be an ellipse an integrable system results, namely as a limiting case of the geodesics on an ellipsoid treated by Jacobi.

" In this problem one can arrive at the existence of certain periodic motions by direct maximum-minimum methods. As of interest in itself I wish to show how this can be done. Results which are being obtained by Morse (but not yet published) indicated that the scope of these methods, already developed to some extent by Hadamard, Poincaré, Whittaker, and myself, can further be extended. Thus the power of such maximum-minimum considerations in the billiard ball problem is likely to prove typical of the general case.

" Any longest chord of the curve C (or boundary of the billiard table) when traversed in both directions evidently yields one of the simplest periodic motions. The billiard ball moving along this chord strikes the curved boundary at right angles and recoils along it in the opposite direction. If we seek to vary this chord continuously, while diminishing its length as little as possible, so as to finally to interchange its two ends, there will be an intermediate position at least length which will be the chord C where C is of least breadth. Detailed computation of the slightly perturbed motions indicates that the first of these two periodic motions is unstable, while the second is stable, i.e. with formal trigonometric series for the perturbations.

" Next we ask for the triangle of maximum length inscribed in C. Evidently at least one such triangle will exist, and can have no degenerate side of zero length. At each of its vertices the tangent will, of course, make equal angles with the two sides passing through the vertex. Hence a *harmonic* triangle is obtained which will correspond to two distinct motions, one for each of the two possible senses of description.

" Moreover if we seek to vary this triangle continuously, without changing the order of the vertices and diminishing the perimeter as little as possible, so as finally to advance the vertices cyclically, we discover a second harmonic triangle, also corresponding to two periodic motions.

" In this way the existence of two harmonic n sided polygons which make k circuits of the curve C (k less than $n/2$ and prime to n) can be proved. The [motion] corresponding to the polygon of maximum type will be unstable, while the other of minimax type may be stable or unstable. "

3. The Minimax Principle

Here, we give a version of the minimax principle which is suited to our needs.

We begin by considering a C^1 function H on a smooth, connected, compact manifold $\mathscr{X}$. Since $\mathscr{X}$ is compact, H has a maximum on $\mathscr{X}$. We will suppose that H takes its maximum value at at least two distinct points x^0 and x^1. Intuitively, one expects to find a " pass " between two " peaks " represented by x^0 and x^1 (Fig. 2). If one imagines a traveler traveling between two peaks who wishes to stay at as high an altitude as possible, i.e. keep H as large as possible, it appears that he must travel through a pass. This leads to the following definitions, which make sense when $\mathscr{X}$ is a compact, Hausdorff topological space, which is connected and locally pathwise connected, H is a continuous real valued function on $\mathscr{X}$, and H takes its maximum value at at least two points x^0 and x^1.

Definition. — A *path* connecting two points x^0 and x^1 is a continuous mapping $\gamma : [0, 1] \to \mathscr{X}$ such that $\gamma(0) = x^0$ and $\gamma(1) = x^1$.

Definition. — The *minimax value* of H associated to the two points x^0 and x^1 is $\sup_\gamma \min_t H(\gamma(t))$, where γ ranges over all paths connecting x^0 and x^1, and t ranges over the unit interval $[0, 1]$.

For any real number a, let $\{H \geq a\} = \{x \in \mathscr{X} : H(x) \geq a\}$.

Proposition **(3.1)**. — *The minimax value of* H *associated to the two points* x^0 *and* x^1 *is* $\max\{a : x^0$ *and* x^1 *are in the same connected component of* $\{H \geq a\}\}$.

The proof follows from our assumption that $\mathscr{X}$ is locally pathwise connected, by means of elementary topological arguments. Note that the maximum of $\{a : x^0$ and x^1 are in the same connected component of $\{H \geq a\}\}$ is actually achieved, as may also be seen by an elementary topological argument based on the fact that $\mathscr{X}$ is compact. We omit these arguments. □

Note that there may be no path γ connecting x^0 and x^1 such that

$$\min_t H(\gamma(t)) = \sup_\gamma \min_t H(\gamma(t)),$$

even if H is a C^∞ function on a C^∞ manifold, although there is such a path if H is an analytic function on an analytic manifold, by the Bruhat-Cartan Selection Lemma ([14], [19], [20]).

Let H_{minimax} denote the minimax value of H associated to the two points x^0 and x^1.

Definition. — A point $y \in \{H = H_{\text{minimax}}\}$ will be said to be *free* (*with respect to* $(H, \mathscr{X}, x^0, x^1)$) if there exists of continuous mapping F of $\{H \geq H_{\text{minimax}}\}$ into itself with the following properties:

1) $H \circ F \geq H$,
2) $H \circ F(y) > H(y)$,
3) $H \circ F(x) = H(x) \Rightarrow F(x) = x$, for all $x \in \{H \geq H_{\text{minimax}}\}$.

Other points in $\{H = H_{\text{minimax}}\}$ will be said to be *bound* (*with respect to* $(H, \mathscr{X}, x^0, x^1)$).

Proposition (3.2). — *There exist bound points in* $\{H = H_{\text{minimax}}\}$.

Proof. — Suppose the contrary. Then for each $y \in \{H = H_{\text{minimax}}\}$, there exists $F_y = F$ with the properties listed above. Let

$$U_y = \{y' \in \{H = H_{\text{minimax}}\} : HF_y(y') > H(y')\}.$$

Since U_y is open and $y \in U_y$, the family of sets $\{U_y\}$ forms an open cover of $\{H = H_{\text{minimax}}\}$. Since this set is compact, the open cover $\{U_y\}$ of $\{H = H_{\text{minimax}}\}$ has a finite subcover $U_{y(1)}, \ldots, U_{y(n)}$. Let

$$G = F_{y(1)} \circ \ldots \circ F_{y(n)}.$$

Then $H \circ G(y) > H(y)$, for all $y \in \{H = H_{\text{minimax}}\}$. Moreover, since x^0 and x^1 maximize H, it follows from 1) that $H \circ F_y(x^0) = H(x^0) = H(x^1) = H \circ F_y(x^1)$, and from 3) that $F_y(x^0) = x^0$, $F_y(x^1) = x^1$, for all $y \in \{H = H_{\text{minimax}}\}$. Then $G(x^0) = x^0$ and $G(x^1) = x^1$.

We have proved that $G\{H \geq H_{\text{minimax}}\} \subset \{H > H_{\text{minimax}}\}$, G is continuous, and $G(x^0) = x^0$, $G(x^1) = x^1$. But this is impossible because x^0 and x^1 are in the same connected component of $\{H \geq H_{\text{minimax}}\}$, but in different connected components of $\{H > H_{\text{minimax}}\}$. This contradiction proves the proposition. □

Proposition (3.3). — *If $\mathscr{X}$ is a smooth manifold without boundary (in addition to being compact, connected, and Hausdorff) and* H *is a* C^1 *function on* $\mathscr{X}$, *then any bound point in* $\{H = H_{\text{minimax}}\}$ *is a critical point of* H, *i.e.* dH *vanishes there.*

Proof. — Let $y \in \{H = H_{\text{minimax}}\}$. If $dH(y) \neq 0$, then there is a C^1 vector field ξ supported in a small neighborhood of y such that $\xi . H(y) > 0$, $\xi . H \geq 0$, and $\xi . H = 0$ only where $\xi = 0$. Let $\exp \xi$ denote the exponential of ξ (i.e. the time one map of the flow generated by ξ). Then $F = \exp \xi$ has the properties listed in the definition of a free point, so y is free. □

Proposition (3.4). — *If $\mathscr{X}$ is a smooth manifold with boundary,* H *is a* C^1 *function on* $\mathscr{X}$, *and for each* $y \in \partial\mathscr{X}$, *there exists a tangent vector* ξ_y *pointing into the interior of* $\mathscr{X}$ *such that* $\xi_y . H > 0$, *then any bound point in* $\{H = H_{\text{minimax}}\}$ *is in the interior of* $\mathscr{X}$ *and is a critical point of* H.

Proof. — Same as for Proposition (3.3). □

To summarize, the minimax principal allows us, given two points which maximize a C^1 function, to find a third critical point, a sort of " pass " between the two " peaks ".

All this has been well known for a century or so (as indicated by Birkhoff in our quote from his article). Nonetheless, because it forms the basis of our reasoning in what follows, we felt it would be helpful to give an exposition of it.

4. Generalization of Birkhoff's Results to Area-Preserving Twist Mappings of the Annulus

In [9, § 3], G. D. Birkhoff points out that the billiard ball problem is equivalent to a dynamical system which may be described as follows: Let $\bar{A} = \{(x, v) : x \in C$ and v is a unit vector at x directed towards the interior of C or tangent to $\partial C\}$ (Fig. 3). Let (x', v') be the pair, where x' is the point in C where the ray starting at x and directed in the direction of v intersects C, and v' is the unit vector obtained by reflecting this ray in the tangent to C at x'. Topologically $\bar{A}$ is an annulus, and we have a homeomorphism $\bar{f}$ of this annulus into itself, defined by $\bar{f}(x, v) = (x', v')$. Periodic orbits in the dynamical system generated by f correspond to harmonic polygons in Birkhoff's sense (§ 2). Therefore, Birkhoff's argument proves the existence of periodic orbits in the billiard ball problem.

Birkhoff's argument generalizes without any difficulty to a class of mappings of the annulus which we call area-preserving monotone twist homeomorphisms. To describe the condition which we impose, we consider not the annulus $\bar{A}$, but its universal cover A, which we represent as $\{(x,y) \in \mathbf{R}^2 : 0 \leq y \leq 1\}$. We consider a homeomorphism $f : A \to A$. We suppose that the representation of A as the universal cover of $\bar{A}$ is chosen so that the translation by the unit, $T(x,y) = (x + 1, y)$, is a generator of the group of Deck transformations. In other words, $\bar{A} = A/T$. We suppose that f is the lifting of a homeomorphism of $\bar{A}$, i.e. $fT = Tf$. We suppose that f is area-preserving (for the usual area in the plane), orientation preserving, maps each boundary component of A into itself, and satisfies the following *monotone twist condition*:

$$\pi_1 f(x,y) > \pi_1 f(x, z), \quad \text{when } y > z,$$

where

$$\pi_1 : A = \mathbf{R} \times [0, 1] \to \mathbf{R}$$

denotes the projection on the first factor.

For $x \in \mathbf{R}$, we let $f_0(x) = \pi_1 f(x, 0)$, $f_1(x) = \pi_1 f(x, 1)$. We let

$$B = \{(x, x') \in \mathbf{R}^2 : f_0(x) \leq x' \leq f_1(x)\}.$$

We may associate to f a real-valued continuous function h, defined on B, such that h is C^1 in the interior of B, and

$$(4.1) \qquad f(x,y) = (x',y') \Leftrightarrow \begin{cases} y = \dfrac{\partial h(x, x')}{\partial x} \\ y' = -\dfrac{\partial h(x, x')}{\partial x'} \end{cases}$$

For a convenience, we will assume throughout this paper that f is C^1 and

$$\text{(4.2)} \qquad \frac{\partial(\pi_1 f(x,y))}{\partial y} > 0.$$

Then h is C^2 and

$$\text{(4.3)} \qquad h_{12} > 0$$

on the interior of B, where

$$h_{12}(x, x') = \frac{\partial^2 h}{\partial x\, \partial x'}(x, x').$$

Physicists call h a *generating function* for f. Obviously, h uniquely determines f. Conversely, f determines h up to an additive constant. The class of area preserving monotone twist mappings is precisely the class of mappings which can be defined by a generating function in this way. Birkhoff's reasoning, quoted in § 2, generalizes without any difficulty to this class of mappings, as we will explain below.

In the case of the billiard ball problem, f is area preserving for the area form $ds\, du$, where ds is the element of arc-length on C and $u = \cos\varphi$, where φ is the angle which the inward pointing vector v makes with the tangent to C. In this case, f is obviously a monotone twist map. Moreover $h(x, x')$ is the Euclidean planar distance between x and x', for all $x, x' \in C$.

Next, we outline how Birkhoff's arguments generalize to monotone twist mappings. Here, and in the sequel, we let p and q be relatively prime integers, with $q > 0$. We will also suppose that p/q lies between the rotation numbers of f_0 and f_1, i.e.,

$$\rho(f_0) < p/q < \rho(f_1).$$

We let $\mathscr{X}_{p,q}$ denote the set of all bi-infinite sequences $x = (\ldots, x_i, \ldots)$ such that $(x_i, x_{i+1}) \in B$, $x_{i+q} = x_i + p$, for all $i \in \mathbf{Z}$, and

$$\text{(4.4)} \qquad x_i + j \leq x_{i'} + j' \Leftarrow pi + qj \leq pi' + qj',$$

for all $i, j, i', j' \in \mathbf{Z}$.

We provide $\mathbf{R}^\infty$ with the product topology, and its subset $\mathscr{X}_{p,q}$ with the induced topology. The mapping $(\ldots, x_i, \ldots) \mapsto (x_0, \ldots, x_{q-1})$ embeds $\mathscr{X}_{p,q}$ as a subspace of $\mathbf{R}^q$.

We let $T^\infty : \mathbf{R}^\infty \to \mathbf{R}^\infty$ be defined by $T^\infty(\ldots, x_i, \ldots) = (\ldots, x_i + 1, \ldots)$. Then $\mathscr{X}_{p,q}$ is invariant under T^∞. The quotient space $\mathscr{X}_{p,q}/T^\infty$ is compact, in view of the fact that $f_0^i(x_0) \leq x_i \leq f_1^i(x_0)$, for $x = (\ldots, x_i, \ldots) \in \mathscr{X}_{p,q}$, and $i > 0$.

For $x \in \mathscr{X}_{p,q}$, we define

$$W(x) = \sum_{i=0}^{q-1} h(x_i, x_{i+1}).$$

This is the analogue, for the discrete dynamical system generated by f, of what physicists call the " action " of a periodic orbit of a continuous dynamical system. Accordingly, we call $W(x)$ the *action* of x. In the case of the billiard ball problem, $W(x)$ is the perimeter of the polygon which corresponds to x.

Using W, it is possible to define Birkhoff max and minimax orbits of f just as we did for the billiard ball problem. To do this it is convenient to introduce the following notation. We let $\mathscr{X}$ denote the set of bi-infinite sequences $(\ldots, x_i, \ldots) \in \mathbf{R}^\infty$ such that $(x_i, x_{i+1}) \in B$. An element $x = (\ldots, x_i, \ldots) \in \mathscr{X}$ will be said to be an *equilibrium sequence* if

$$h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) = 0, \tag{4.5}$$

for all $i \in \mathbf{Z}$. Here, and in the sequel, we use the notation

$$h_1(x, x') = \frac{\partial h(x, x')}{\partial x}, \quad h_2(x, x') = \frac{\partial h(x, x')}{\partial x'}.$$

Using the fact that h is a generating function for f, we may set up a $1-1$ correspondence between equilibrium sequences and orbits of f, as follows: Let $(\ldots, x_i, \ldots)$ be an equilibrium sequence. Set

$$y_i = h_1(x_i, x_{i+1}).$$

From (4.1) and (4.5), it follows easily that

$$f(x_i, y_i) = (x_{i+1}, y_{i+1}),$$

so $(\ldots, (x_i, y_i), \ldots)$ is an orbit of f. Conversely, if $(\ldots, (x_i, y_i), \ldots)$ is an orbit of f, it follows easily from (4.1) that $(\ldots, x_i, \ldots)$ is an equilibrium sequence.

Clearly $W \circ T^\infty = W$. Hence W induces a function on $\mathscr{X}_{p,q}/T^\infty$, which we continue to denote by the same letter. Obviously, W is continuous (whether considered as a function on $\mathscr{X}_{p,q}$ or on $\mathscr{X}_{p,q}/T^\infty$). Since W is a continuous function on the compact space $\mathscr{X}_{p,q}/T^\infty$, it has a maximum value. Therefore, it also has a maximum value on $\mathscr{X}_{p,q}$.

Proposition **(4.1)**. — *Any point in* $\mathscr{X}_{p,q}$ *where* W *takes its maximum is an equilibrium sequence.*

This will be proved in § 11.

Let $x = (\ldots, x_i, \ldots) \in \mathscr{X}_{p,q}$ maximize W. Since it is an equilibrium sequence, there is a corresponding orbit $(\ldots, (x_i, y_i), \ldots)$ of f. We will call this orbit a *Birkhoff max orbit of type* (p, q). Its projection on $\bar{A}$ is obviously a periodic orbit of period q.

We will give a somewhat involved topological description of the Birkhoff minimax orbit(s), following the ideas discussed in §§ 2 and 3. First, we choose $x = x_{p,q,\max} \in \mathscr{X}_{p,q}$ which maximize W. We define $x_+ \in \mathscr{X}_{p,q}$ by setting

$$(x_+)_i = x_{i+i_0} + j_0,$$

where $pi_0 + qj_0$ is the minimum positive element of the set $\{pi + qj : i, j \in \mathbf{Z}\}$. We set

$$\mathscr{X}_{p,q}(x) = \{x' \in \mathscr{X}_{p,q} : x_i \leq x'_i \leq (x_+)_i, \text{ all } i \in \mathbf{Z}\}.$$

We let $W_{p,q,x,\text{minimax}}$ denote the minimax value of $W \mid \mathscr{X}_{p,q}(x)$ associated to the points x and x_+.

Clearly, $\mathscr{X}_{p,q}(x)$ is a compact, Hausdorff topological space, which is connected and locally pathwise connected, W is a continuous real valued function on $\mathscr{X}_{p,q}(x)$, and W takes its maximum value on $\mathscr{X}_{p,q}(x)$ at x and x_+, which are distinct points. So, the definition of bound points (§ 3) applies. By Proposition (3.2), there exist bound points in

$$\mathscr{X}_{p,q}(x) \cap \{W = W_{p,q,x,\text{minimax}}\}.$$

Proposition **(4.2)**. — *Any bound point in* $\mathscr{X}_{p,q}(x) \cap \{W = W_{p,q,x,\text{minimax}}\}$ *is an equilibrium sequence.*

Here, we mean bound with respect to $(W, \mathscr{X}_{p,q}(x), x, x_+)$. Proposition (4.2) will be proved in § 12.

Let $x' = (\ldots, x'_i, \ldots)$ be a bound point in $\mathscr{X}_{p,q}(x) \cap \{W = W_{p,q,x,\text{minimax}}\}$. Since it is an equilibrium sequence, there is a corresponding orbit $(\ldots, (x'_i, y'_i), \ldots)$ of f. We will call this orbit a Birkhoff minimax orbit of type (p, q). Its projection on $\overline{A}$ is obviously a periodic orbit of period q.

Proposition **(4.3)**. — *If* $x, x_1 \in \mathscr{X}_{p,q}$ *both maximize* W *over* $\mathscr{X}_{p,q}$, *then*

$$W_{p,q,x,\text{minimax}} = W_{p,q,x(1),\text{minimax}}.$$

This will be proved in § 19.

We will write $W_{p,q,\text{minimax}}$ for $W_{p,q,x,\text{minimax}}$, where x is any element of $\mathscr{X}_{p,q}$ which maximizes W. The above proposition says this is well defined. We will also write $W_{p,q,\max}$ for the maximum of W over $\mathscr{X}_{p,q}$. We set

$$\Delta W_{p,q} = W_{p,q,\max} - W_{p,q,\text{minimax}}.$$

5. The Necessary and Sufficient Conditions for the Existence of Invariant Circles

In the last section, we defined $\Delta W_{p,q}$ to be the difference of the action of a Birkhoff max orbit of type (p, q) and a Birkhoff minimax orbit of type (p, q).

Proposition **(5.1)**. — *Consider an irrational number* ω *satisfying* $\rho(f_0) < \omega < \rho(f_1)$. *Then* $\Delta W_{p/q}$ *converges to a limit as* p/q *tends to* ω.

We denote the limit by ΔW_ω. In other words,

$$\Delta W_\omega = \lim_{p/q \to \omega} \Delta W_{p/q}.$$

Proposition (5.1) will be proved in § 24.

Proposition **(5.2)**. — *Consider an irrational number* ω *satisfying* $\rho(f_0) < \omega < \rho(f_1)$. *There exists an* $\bar{f}$ *invariant circle of rotation number (or frequency)* ω *if and only if* $\Delta W_\omega = 0$.

This will be proved in § 20, for the direct definition of ΔW_ω given in § 6.

Note that nothing is said about whether $\bar{f}$ is topologically transitive on this circle.

For our proof, we will give a direct definition of ΔW_ω in § 6. Our direct definition is vaguely similar to the definition given in [7]. However, what Aubry calls the Peierls energy barrier is not the same as our ΔW_ω. It is a lower bound for ΔW_ω. In [7] Aubry, Le Daeron, and André prove a version of our Proposition (5.2), where ΔW_ω is replaced by the Peierls energy barrier. They do not have the analogue of our Proposition (5.1) in [7].

Propositions (5.1) and (5.2) were suggested to me by lengthy conversations I had with A. Katok, when I was in Maryland in January 1982, and previously. My existence theorem for quasi-periodic orbits was suggested by Birkhoff's method of finding periodic orbits in the billiard ball problem. But, I had forgotten that Birhoff had constructed periodic orbits also by a minimax principle. Katok reminded me of this and suggested finding a second invariant Cantor set by means of a minimax principle. We played around with this idea quite a bit. A couple of months after I came home from Maryland, I realized that the ideas that Katok had suggested lead, not to a second invariant Cantor set, but to an orbit which is homoclinic to the original Cantor set. This orbit will be described in § 7.

6. Direct Definition of ΔW_ω

To give the direct definition of ΔW_ω, we have to recall some features of the proof of the existence of quasi-periodic orbits which we gave in [21].

Let ω be an irrational number, $\rho(f_0) < \omega < \rho(f_1)$. In [21], we set

$$Y_\omega = \{\varphi : \mathbf{R} \to \mathbf{R} : \varphi \text{ is weakly order preserving, } \varphi(t+1) = \varphi(t) + 1, \\ (\varphi(t), \varphi(t+\omega)) \in B, \text{ for all } t \in \mathbf{R}, \text{ and } \varphi \text{ is left continuous}\}.$$

It will be convenient to also introduce here the following subset of Y_ω:

$$X_\omega = \{\varphi \in Y_\omega : \sup\{s : \varphi(s) < 0\} = 0\}.$$

WARNING: This differs from the set which we called X_ω in [21].

The advantage of introducing this set is that for each $\varphi \in Y_\omega$, there is a unique $a = a(\varphi)$ such that $\varphi T_a \in X_\omega$. Explicitly,

$$a(\varphi) = \sup\{s : \varphi(s) < 0\}.$$

This provides us with a representation of Y_ω as a product:

(6.1) $$Y_\omega = X_\omega \times \mathbf{R},$$

where we set up the identification

$$\text{(6.2)} \qquad \varphi = (\varphi T_a, a), \quad \text{where } a = a(\varphi).$$

Here $T_a : \mathbf{R} \to \mathbf{R}$ is translation by a. In other words, $T_a(t) = t + a$. For the topology on Y_ω which we introduced in [21], (6.1) cannot be a topological product; the projection on $\mathbf{R}$ associated to (6.1) is not continuous, and $X_\omega \times 0$ is not closed in Y_ω. Nevertheless, it will be useful to give X_ω the quotient topology coming from the topology on Y_ω defined in [21] *via* the projection of Y_ω on X_ω associated to (6.1).

In [22], we defined an isometry $I : Y_\omega \to Y_\omega^-$; we have $X_\omega^- = I(X_\omega)$, and the topology we have introduced here on X_ω corresponds *via* I to the topology we introduced in [22] on X_ω^-. Thus, the discussion of [22] applies; in particular, we see that X_ω is compact. More directly, the set which we called X_ω in [21] maps surjectively and continuously onto the set we call X_ω here. Since the former is compact, so is the latter.

In [21], we defined

$$F_\omega(\varphi) = \int_0^1 h(\varphi(t), \varphi(t + \omega))\, dt.$$

This is Percival's Lagrangian. In [21], we showed that $F_\omega \circ T_a = F_\omega$ and F_ω is continuous on Y_ω. It follows that F_ω is continuous on X_ω. Since X_ω is compact, there is an element of it which maximizes F_ω. We denote this element by φ_ω. In [21], we proved that there is only one element which maximizes $F_\omega \mid X_\omega$.

For $\varphi \in Y_\omega$ and $t \in \mathbf{R}$, we define $\hat{\varphi}(t-), \hat{\varphi}(t+) \in \mathscr{X}$ by

$$\hat{\varphi}(t-)_i = \varphi(t + \omega i -), \qquad \hat{\varphi}(t+)_i = \hat{\varphi}(t + \omega i +).$$

Note that $\hat{\varphi}(t-) = \hat{\varphi}(t)$ (with the obvious meaning of the latter), since φ is left-continuous. In [21], we proved that $\hat{\varphi}_\omega(t-)$ and $\hat{\varphi}_\omega(t+)$ are equilibrium sequences.

Since $\hat{\varphi}_\omega(t-)$ and $\hat{\varphi}_\omega(t+)$ are equilibrium sequences, there are orbits of f associated to them. Explicitly, we let

$$\hat{\eta}_\omega(t\pm)_i = h_1(\hat{\varphi}_\omega(t\pm)_i, \hat{\varphi}_\omega(t\pm)_{i+1}).$$

Then $(\ldots, (\hat{\varphi}_\omega(t\pm)_i, \hat{\eta}_\omega(t\pm)_i), \ldots)$ is an orbit of f. We will denote the union of all such orbits by M_ω. (This is the same as the set which we denoted M_φ in [21].) Then M_ω is a closed set, invariant for both f and T. The image of M_ω in $\bar{A}$ under the covering mappings $A \to \bar{A}$ is the minimal set for $\bar{f}$ mentioned in the introduction ("... under the hypothesis I considered (monotone twist hypothesis), [the quasi-periodic orbits whose existence I proved] necessarily exist and lie on a minimal set").

We showed [21] that φ_ω is strictly order preserving and $\pi_1 : M_\omega \to \mathbf{R}$ is injective. In Aubry's terminology, $\pi_1 M_\omega$ is the union of all ground states of mean atomic distance ω.

For $t \in \mathbf{R}$, we let $\mathscr{X}_{\omega,t}$ denote the set of all $x = (\ldots, x_i, \ldots) \in \mathscr{X}$ which satisfy

$$\hat{\varphi}_\omega(t-)_i \leq x_i \leq \hat{\varphi}_\omega(t+)_i.$$

We topologize $\mathscr{X}_{\omega,t}$ with the topology induced by the product topology on $\mathbf{R}^\infty$. Note that $\mathscr{X}_{\omega,t}$ is compact, by the Tychonoff product theorem. We set $x_i^0 = x^0(t)_i = \varphi_\omega(t-)_i$, $x_i^1 = x^1(t)_i = \varphi_\omega(t+)_i$. For $x = (\ldots, x_i, \ldots)$ $x = (\ldots, x_i, \ldots) \in \mathscr{X}_{\omega,t}$, we set

$$G_\omega(x) = \sum_{i=-\infty}^{\infty} [h(x_i, x_{i+1}) - h(x_i^0, x_{i+1}^0)].$$

Lemma **(6.1)**. — *The sum on the right side of the above equation is uniformly (over $\mathscr{X}_{\omega,t}$) absolutely convergent.*

This will be proved in § 13.

Consequently, G_ω is continuous on $\mathscr{X}_{\omega,t}$.

Lemma **(6.2)**. — *For all* $x \in \mathscr{X}_{\omega,t}$, $G_\omega(x) \leq 0$. *Moreover,* $G_\omega(x^0) = G_\omega(x^1) = 0$.

The statement that $G_\omega(x^0) = 0$ is obvious from the definitions. The other statements will be proved in § 14.

Now we are in the situation considered in § 3, if t is a point of discontinuity of φ_ω. In other words, $\mathscr{X}_{\omega,t}$ is a compact, Hausdorff space, which is connected and locally pathwise connected, G_ω is a continuous function on $\mathscr{X}_{\omega,t}$, and G_ω takes its maximum value at x^0 and x^1. Clearly, $x^0 \neq x^1$, when t is a point of discontinuity of φ_ω. (When t is a point of continuity of φ_ω, then $\mathscr{X}_{\omega,t}$ consists of just one point, in view of the fact, which we proved in [21], that if t is a point of continuity of φ_ω then so is $t + \omega i$, for all $i \in \mathbf{Z}$. Obviously $x^0 = x^1$, in this case.)

Definition. — When t is a point of discontinuity of φ_ω, we let $\Delta W_{\omega,t}$ denote the negative of the minimax value of $G_\omega : \mathscr{X}_{\omega,t} \to \mathbf{R}$ associated to x^0 and x^1. When t is a point of continuity of φ_ω, we let $\Delta W_{\omega,t} = 0$.

Obviously, $\Delta W_{\omega,t} = \Delta W_{\omega,t'}$ when $t' = t + \omega i + j$, where $i, j \in \mathbf{Z}$.

Definition. — We will say that two points $t, t' \in \mathbf{R}$ are *ω-independent* if

$$t' \neq t + \omega i + j,$$

for all $i, j \in \mathbf{Z}$.

Lemma **(6.3)**. — *Let* I *be a maximal ω-independent set of points of discontinuity of* φ_ω. *We have*

$$\sum_{t \in I} \Delta W_{\omega,t} < \infty.$$

This will be proved in § 17.

We set

$$\Delta W_\omega = \max_{t \in \mathbf{R}} \Delta W_{\omega,t}.$$

By the preceding discussion, this is finite and non-negative. The statement that this agrees with the previous definition amounts to the assertion:

Proposition (**6.4**). — $\Delta W_\omega = \lim_{p/q \to \infty} \Delta W_{p,q}$.

This will be proved in § 24. It obviously implies Proposition (5.1).

Note that this implies that ΔW_ω is continuous as a function of ω at all irrational points ω. On the other hand, it is not necessarily continuous at rational points.

Example. — In the billiard ball problem in the ellipse, $\Delta W_\omega > 0$, when $\omega = \frac{1}{2}$, $\Delta W_\omega = 0$, when $\omega \neq \frac{1}{2}$. This can be seen from the discussion of the billiard ball problem in the ellipse given in [29, p. 86].

Consider an invariant circle for $\bar{f}$. We will say that such an invariant circle is *transitive* if the restriction of $\bar{f}$ to it is topologically transitive; otherwise, we will say it is *intransitive*. For area preserving monotone twist mappings $\bar{f}$ of high differentiability class, it is an unsolved problem as to whether intransitive circles of irrational rotation number which wind once around the annulus exist. On the other hand, M. R. Herman has constructed [16] mappings $\bar{f}$ of class $C^{3-\varepsilon}$ which are monotone twist diffeomorphisms in our sense, and have intransitive circles of irrational rotation number which wind once around the annulus.

The necessary and sufficient condition for there to be a transitive circle of frequency ω is that φ_ω be continuous. For, if φ_ω is continuous, M_ω is homeomorphic to $\mathbf{R}$, and its image in the annulus $\bar{A}$ under the covering mapping is a circle. Conversely, if there is an invariant circle, Birkhoff's theorem [10], [11] implies that it is the graph of a function. Its lifting to the universal cover A of the annulus is a curve which separates A into two parts. From the monotone twist hypothesis, it follows that the points in the lower part advance slower than the rate ω, and points in the upper part advance faster. Using this idea, one proves easily that M_ω cannot be in the lower or upper part, so it must be in the curve itself. Since the circle is assumed to be transitive, M_ω must be the whole curve, i.e. φ must be continuous.

7. The Minimax Orbits of Irrational Frequencies

Let $\rho(f_0) < \omega < \rho(f_1)$, suppose ω is irrational, and suppose t is a point of discontinuity of φ_ω.

Proposition (**7.1**). — *Let* $x \in \mathscr{X}_{\omega,t}$ *and suppose* $G_\omega(x) = -\Delta W_{\omega,t}$. *If* x *is bound with respect to* $(G_\omega, \mathscr{X}_{\omega,t}, x^0, x^1)$, *then* x *is an equilibrium sequence.*

According to Proposition (3.2) such bound points exist.

Since $x = (\ldots, x_i, \ldots)$ is an equilibrium sequence, there is associated an orbit $(\ldots, (x_i, y_i), \ldots)$ of f. We will call this a *minimax orbit of frequency* ω associated to a point t of discontinuity of φ_ω. It is easily seen that the projection of this orbit on the annulus is homoclinic to the Cantor set M_ω/T, i.e. it approaches the Cantor set under both forward and backward iteration.

Proposition (7.1) was suggested to me by Katok's idea that there should be an analogue for irrational frequencies of the Birkhoff minimax orbits of type (p, q). Originally, he thought one might find a second Cantor set, but I observed that his idea leads to an orbit homoclinic to the Cantor set. Proposition (7.1) led to the other results in this paper.

Proposition (7.1) will be proved in § 18.

8. Limit Infinum and Limit Supremum of Subsets

In the next section, we will define an f and T-invariant set M'_ω which contains M_ω and whose projection onto $\mathbf{R}$ is injective. The principal results which we state in that section can be stated briefly, as follows: 1) M'_ω is an upper semi-continuous function of f and ω, at irrational ω. 2) M_ω if a lower semi-continuous function of f and ω, at irrational ω.

Some care in formulating these results is required, because there are various inequivalent notions of semi-continuity for set-valued functions. These notions are surveyed in [13], [25] and the references there.

For definitions of convergence, the notion due to H. Cartan [15] of *filter* is convenient (cf. [12]). A filter on a set X is a non-empty collection $\mathscr{A}$ of subsets of X such that: 1) The intersection of any finite family of members of $\mathscr{A}$ is a member of $\mathscr{A}$, 2) any subset of X which contains a member of $\mathscr{A}$ is itself a member of $\mathscr{A}$, and 3) the empty set is not a member of $\mathscr{A}$.

Examples. — If X is a topological space and $x_0 \in X$, the collection of neighborhoods of x_0 in X is a filter on X, called the *filter of neighborhoods* of x_0 in X. If x_0 is not isolated in X, then the collection of sets $N \backslash x_0$, where N is a neighborhood of x_0 in X, is a filter on $X \backslash x_0$, called the *filter of punctured neighborhoods* of x_0.

Definition. — Let X be a set with a filter $\mathscr{A}$ and Y a topological space. Let $f: X \to Y$ be a mapping. We say *f converges to y over $\mathscr{A}$* if for every neighborhood of y, there is an element of $\mathscr{A}$ whose image under f lies in the given neighborhood. We will express this notion also by writing

$$\lim_{\mathscr{A}} f = y \quad \text{or} \quad \lim_{x, \mathscr{A}} f(x) = y.$$

In the two examples above, the expression $\lim_{\mathscr{A}} f = y$ is equivalent to the usual expressions $\lim_{x \to x_0} f(x) = y$ and $\lim_{x \to x_0,\, x \neq x_0} f(x) = y$, respectively.

Definition. — Let $\mathscr{A}$ be a filter on a set X. Let Y be a topological space. For each $x \in X$, let Z_x be a subset of Y. We define

$$\limsup_{\mathscr{A}} Z = \limsup_{x,\mathscr{A}} Z_x = \{y \in Y : \text{for any neighborhood N of } y$$
$$\text{and any } U \in \mathscr{A}, \text{ there exists } x \in U \text{ such that } Z_x \cap N \neq \emptyset\},$$
$$\liminf_{\mathscr{A}} Z = \liminf_{x,\mathscr{A}} Z_x = \{y \in Y : \text{for any neighborhood N of } y \text{ in Y},$$
$$\text{there exists } U \in \mathscr{A}, \text{ such that for any } x \in U, \text{ we have } Z_x \cap N \neq \emptyset\}.$$

Obviously, $\limsup_{\mathscr{A}} Z$ and $\liminf_{\mathscr{A}} Z$ are both closed subsets of Y.

For us, the following are the convenient definitions of upper and lower semi-continuity.

Definition. — Let X be a topological space. Let $x \to Z_x$ be a mapping of X into the set of subsets of a second topological space. This mapping will be said to be *upper semi-continuous* at $x_0 \in X$ if

$$\limsup_{x \to x_0} Z_x = Z_{x(0)}.$$

Note that we necessarily have the inclusion $\supset$ and that upper semi-continuity at x_0 implies that $Z_{x(0)}$ is closed.

Definition. — The mapping $x \to Z_x$ will be said to be *lower semi-continuous* at $x_0 \in X$ if

$$\liminf_{x \to x_0} Z_x = Z_{x(0)}.$$

Note that we necessarily have the inclusion $\liminf_{x \to x_0} Z_x \subset \bar{Z}_{x(0)}$ and lower semi-continuity at x_0 implies $Z_{x(0)}$ is closed.

9. Semi-Continuity Properties

Suppose ω is irrational and $\rho(f_0) < \omega < \rho(f_1)$. Let t be a point of discontinuity of φ_ω.

Proposition (**9.1**). — *Suppose* $x = (\ldots, x_i, \ldots) \in \mathscr{X}_{\omega,t}$ *and* $G_\omega(x) = 0$. *Then* x *is an equilibrium sequence.*

This will be proved in § 15.

We define $M'_{\omega,t}$ to be the set of all (x_i, y_i), where x ranges over $G^{-1}(0) \cap \mathscr{X}_{\omega,t}$, i ranges over all integers, and $y_i = h_1(x_i, x_{i+1})$. We let

$$M'_\omega = M_\omega \cup \bigcup_t M'_{\omega,t},$$

where t ranges over all points of discontinuity of φ_ω.

By Proposition (9.1), M'_ω is invariant under f. Obviously, it is invariant under T.

Now suppose ω is rational, say $\omega = p/q$ in lowest terms. In this case, we define M'_ω to be the union of all Birkhoff max orbits of type (p, q).

In this section, we will want to state some semi-continuity results with respect to variation of f as well as ω. For this purpose, we need to introduce the set $\mathscr{F}^1$ consisting of all area, orientation, and boundary component preserving C^1 diffeomorphisms $f: A \to A$ such that $fT = Tf$ and $\partial(\pi_1 f(x,y))/\partial y > 0$, provided with the C^1 topology. We call $\mathscr{F}^1$ *the space of area-preserving monotone twist diffeomorphisms.*

We will indicate the dependence on f of various quantities which we introduced above by means of a subscript, e.g. $M_{f,\omega}$, $M'_{f,\omega}$, $Y_{f,\omega}$, $X_{f,\omega}$, $F_{f,\omega}$, h_f, etc.

The domain $\mathscr{D}$ of the mapping $(f, \omega) \to M'_{f,\omega}$ is the set of pairs (f, ω) such that $f \in \mathscr{F}^1$ and $\rho(f_0) < \omega < \rho(f_1)$. We provide $\mathscr{D}$ with the topology induced from the topology on $\mathscr{F}^1 \times \mathbf{R}$.

Proposition (**9.2**). — *The mapping $(f, \omega) \to M'_{f,\omega}$ is upper semi-continuous at (f, ω) if ω is irrational.*

This will be proved in § 23.

We will denote by (f, T) the action of $\mathbf{Z}^2$ on A generated by f and T. By an (f, T)-orbit, we mean an orbit of this action. If $\mathcal{O}$ is an orbit under f, obviously $\bigcup_i T^i \mathcal{O}$ is an (f, T)-orbit. If $\mathcal{O}$ is a max (resp. minimax) Birkhoff orbit of type (p, q), we will call $\bigcup_i T^i \mathcal{O}$ a *max* (resp. *minimax*) (f, T)-*Birkhoff orbit of type* (p, q). For generic $f \in \mathscr{F}^1$, there is only one max (resp. minimax) (f, T)-Birkhoff orbit of type (p, q), although in general there may be more than one.

Proposition (**9.3**). — *For each $f \in \mathscr{F}^1$, and each rational number p/q satisfying $\rho(f_0) < p/q < \rho(f_1)$, let $M_{f,p/q}$ denote one of the max (f, T)-Birkhoff orbits of type (p, q). Then the function $(f, \omega) \to M_{f,\omega}$ is defined on all of $\mathscr{D}$. It is lower semi-continuous at all points (f, ω) where there is no choice in the definition of $M_{f,\omega}$, i.e. whenever ω is irrational or $\omega = p/q$ and there is only one max Birkhoff orbit of type (p, q).*

This will be proved in § 21, where we will also discuss some questions concerning convergence in measure.

Proposition (**9.4**). — *$\Delta W_{f,\omega}$ is continuous as a function of $(f, \omega) \in \mathscr{D}$, at any point $(f, \omega) \in \mathscr{D}$ where ω is irrational.*

This will be proved in § 24. It obviously implies Propositions (6.4) and (5.1).

10. Order Properties

Let $(f, \omega) \in \mathscr{D}$. Let M be a closed subset of A which is invariant under f and T. We will say that M is f-monotone (for projection on the first factor) if $\pi_1 : M \to \mathbf{R}$ is injective, and for $(x,y) \in M$, $(x',y') \in M$, we have

$$x < x' \Rightarrow \pi_1 f(x,y) < \pi_1 f(x',y') \quad \text{and} \quad \pi_1 f^{-1}(x,y) < \pi_1 f^{-1}(x',y').$$

(Recall that $\pi_1 : A = \mathbf{R} \times [0, 1] \to \mathbf{R}$ is the projection on the first factor.)

Proposition (**10.1**). — $M'_{f,\omega}$ *is f-monotone (for projection on the first factor).*

This will be proved in § 16.

In Aubry's terminology, $\pi_1 M'_{f,\omega}$ is the union of all m.e. states of atomic mean distance ω. This assertion follows from Aubry's results (but not from ours). Aubry's Fundamental Lemma is the basic step in proving such a result in his approach. Proposition (10.1) can also be easily proved from Aubry's results. Indeed, his Fundamental Lemma goes further in one direction than anything which we have done. He does not assume any order property on m.e. states, but proves they have an order property which corresponds to Proposition (10.1), as a consequence of his Fundamental Lemma. In contrast, the fact that M_ω is monotone is an immediate consequence of our definition of it.

11. Proof of Proposition (4.1)

For the proof, it is convenient to introduce the following definition.

Definition. — By the *boundary* $\partial\mathscr{X}_{p,q}$ of $\mathscr{X}_{p,q}$, we mean $\{x \in \mathscr{X}_{p,q} : (x_i, x_{i+1}) \in \partial B$ for some i, or $(x_+)_i = x_i$, for some $i\}$.

It is not difficult to show that $\mathscr{X}_{p,q}$ is a q-dimensional topological manifold with boundary, and $\partial\mathscr{X}_{p,q}$ is its boundary in this sense. Indeed for $x \in \mathscr{X}_{p,q}$, let $\varphi_x \in Y_\omega$ (where $\omega = p/q$) be the step function associated to x, and let $\psi_x = I(\varphi_x) \in Y^-_\omega$ (in the notation of [22]). The family $\{\psi_x : x \in \mathscr{X}_{p,q}\}$ is a family of step functions, which forms a convex, q-dimensional set. The fact that $\partial\mathscr{X}_{p,q}$ is the boundary of $\mathscr{X}_{p,q}$ in the sense of manifold with boundary follows easily from this representation of it. However, we will not need to use this fact.

Likewise, it is not difficult to show that $\partial\mathscr{X}_{p,q}$ is the point-set boundary of $\mathscr{X}_{p,q}$, when the latter is considered as the subspace of the affine space of all sequences $x = (\ldots, x_i, \ldots)$ such that $x_{i+q} = x_i + p$.

Now suppose that x maximizes W over $\mathscr{X}_{p,q}$.

Step 1. — $x \notin \partial\mathscr{X}_{p,q}$. If $x \in \partial\mathscr{X}_{p,q}$, we have one of two possibilities: $(x_i, x_{i+1}) \in \partial B$ or $(x_+)_i = x_i$, for some i. In the case that $(x_i, x_{i+1}) \in \partial B$, we have $x_{i+1} = f_0(x_i)$ or $x_{i+1} = f_1(x_i)$. Suppose $x_{i+1} = f_0(x_i)$, for some i.

We cannot have $x_{i+1} = f_0(x_i)$, for all large i, since this would imply

$$\lim_{n\to\pm\infty} \frac{x_n}{n} = \rho(f_0),$$

whereas $$\lim_{n\to\pm\infty} \frac{x_n}{n} = \frac{p}{q},$$

by the hypothesis that $x \in \mathscr{X}_{p,q}$. (Recall that we assumed $\rho(f_0) < \frac{p}{q} < \rho(f_1)$.) Hence there exists i such that $x_i = f_0(x_{i-1})$, but $x_{i+1} > f_0(x_i)$. Then $\frac{\partial}{\partial x_i} h(x_{i-1}, x_i) = 0$ and $\frac{\partial}{\partial x_i} h(x_i, x_{i+1}) > 0$, so we have

$$\frac{\partial}{\partial x_i}[h(x_{i-1}, x_i) + h(x_i, x_{i+1})] > 0. \tag{11.1}$$

In the case $(x_+)_i > x_i$, we may replace x with $x' \in \mathscr{X}_{p,q}$ for which $W(x') > W(x)$, by increasing x_i slightly.

In the case $(x_+)_i = x_i$, we have $f_0((x_+)_{i-1}) \geq f_0(x_{i-1}) = x_i = (x_+)_i$, so $f_0((x_+)_{i-1}) = (x_+)_i$ (since $((x_+)_{i-1}, (x_+)_i) \in B$, i.e. $f_0((x_+)_{i-1}) \leq (x_+)_i \leq f_1((x_+)_{i-1})$). Likewise, $f_0((x_+)_i) = f_0(x_i) < x_{i+1} \leq (x_+)_{i+1}$. So, just as before,

$$\frac{\partial}{\partial (x_+)_i}[h((x_+)_{i-1}, (x_+)_i) + h((x_+)_i, (x_+)_{i+1})] > 0.$$

It follows that, in the case $(x_{++})_i > (x_+)_i$, we may replace x with $x' \in \mathscr{X}_{p,q}$ for which $W(x') > W(x)$, by increasing both x_i and $(x_+)_i$ slightly.

Likewise, if $(x_{++})_i = (x_+)_i = x_i$, but $(x_{+++})_i > (x_{++})_i$, we may replace x with $x' \in \mathscr{X}_{p,q}$ for which $W(x') > W(x)$, by increasing $(x_{++})_i$, $(x_+)_i$, and x_i slightly, and so on, for ever larger number of $+$'s. For a large enough number of $+$'s, we must have $(x_{+\dots+})_i > x_i$, so in any case we obtain the existence of x' for which $W(x') > W(x)$.

Since x was assumed to maximize W, this is a contradiction. This contradiction shows that we cannot have $x_{i+1} = f_0(x_i)$, for any i. A similar argument shows that we cannot have $x_{i+1} = f_1(x_i)$, for any i.

Suppose $(x_+)_i = x_i$, for some i. We cannot have $(x_+)_i = x_i$ for all large i, for otherwise, we would have $x_i + j = x_{i'} + j'$, for all i, j, i', j', which is clearly impossible. Hence $(x_+)_i = x_i$ and $(x_+)_{i+1} > x_{i+1}$, for some i.

Recall formula (4.3), $h_{12} > 0$. This implies $h_1(x_i, x_{i+1}) < h_1((x_+)_i, (x_+)_{i+1})$. Since $(x_+)_{i-1} \geq x_{i-1}$, we also have $h_2(x_{i-1}, x_i) \leq h_2((x_+)_{i-1}, (x_+)_i)$. We must therefore have

$$h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) < h_2((x_+)_{i-1}, (x_+)_i) + h_1((x_+)_i, (x_+)_{i+1}).$$

Now suppose $(x_{++})_i > (x_+)_i$ and $(x_-)_i < x_i$ (where $x_- \in \mathscr{X}_{p,q}$ is defined by $(x_-)_+ = x = (x_+)_-$). From the above inequality, we have

$$h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) < 0,$$

or

$$h_2((x_+)_{i-1}, (x_+)_i) + h_1((x_+)_i, (x_+)_{i+1}) > 0$$

(or both). In the first case, we may replace x with $x' \in \mathscr{X}_{p,q}$, for which $W(x') > W(x)$, by decreasing x_i slightly. In the second case, we may replace x with $x' \in \mathscr{X}_{p,q}$ for which $W(x') > W(x)$, by increasing $(x_+)_i$ slightly. In either case, we obtain a contradiction to the assumption that x maximizes W.

In the general case, we write $x_{(1)} = x_+$, $x_{(2)} = x_{++}$, etc., and $x_{(-1)} = x_-$, $x_{(-2)} = x_{--}$, etc. Obviously, we may choose integers $\alpha \leq 0$, $\beta > 0$, such that

$$(x_{(\alpha-1)})_i < (x_{(\alpha)})_i = (x_{(\beta)})_i < (x_{\beta+1})_i.$$

The argument we gave above shows that

$$h_2((x_{(\gamma)})_{i-1}, (x_{(\gamma)})_i) + h_1((x_{(\gamma)})_i, (x_{(\gamma)})_{i+1}) \leq h_2((x_{(\gamma+1)})_{i-1}, (x_{(\gamma+1)})_i) + h_1((x_{(\gamma+1)})_i, (x_{(\gamma+1)})_{i+1})$$

for $\alpha \leq \gamma < \beta$, with strict inequality for $0 \leq \gamma$. Consequently, slightly decreasing $(x_{(\gamma)})_i$ for which

$$h_2((x_{(\gamma)})_{i-1}, (x_{(\gamma)})_i) + h_1((x_{(\gamma)})_i, (x_{(\gamma)})_{i+1}) < 0$$

and slightly increasing $(x_{(\gamma)})_i$ for which

$$h_2((x_{(\gamma)})_{i-1}, (x_{(\gamma)})_i) + h_1((x_{(\gamma)})_i, (x_{(\gamma)})_{i+1}) > 0,$$

produces an x' for which $W(x') > W(x)$.

Hence $x \notin \partial\mathscr{X}_{p,q}$.

Step 2. — From the fact that x maximizes W and $x \notin \partial\mathscr{X}_{p,q}$, it follows immediately that

$$h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) = \frac{\partial W(x)}{\partial x_i} = 0,$$

so, x is an equilibrium sequence. □

12. Proof of Proposition (4.2)

The idea is to follow the proof of Proposition (3.4), with $\mathscr{X}_{p,q}(x)$ in place of $\mathscr{X}$. Proposition (3.4) doesn't strictly apply, because the boundary of $\mathscr{X}_{p,q}(x)$ may have corners and not be very smooth. However, this doesn't affect the proof any. The main point is that there is an appropriate modification of the condition (in Prop. (3.4)) that for each $y \in \partial\mathscr{X}$, there exist a tangent vector ξ_y pointing into the interior of $\mathscr{X}$ such that $\xi_y.H > 0$.

We define $\partial\mathscr{X}_{p,q}(x) = \{x' \in \mathscr{X}_{p,q}(x) : x'_i = x_i$ or $x'_i = (x_+)_i$ or $f_0(x'_i) = x'_{i+1}$ or $f_1(x'_i) = x'_{i+1}$, for some $i\}$. The space $\mathscr{X}_{p,q}(x)$ is a topological manifold with boundary $\partial\mathscr{X}_{p,q}(x)$. However, we will not need to use this fact. The important thing for us is that $\partial\mathscr{X}_{p,q}(x)$ is the point-set boundary of $\mathscr{X}_{p,q}(x)$, when the latter is considered as a subspace of the affine space of all sequences $x = (\ldots, x_i, \ldots)$ such that $x_{i+q} = x_i + p$.

What replaces the condition concerning ξ_y is the following result. For any real number a, we let $\mathscr{X}_{p,q}(x)\ (\geq a) = \{x' \in \mathscr{X}_{p,q}(x) : W(x') \geq a\}$.

Lemma. — *Let* $x' \in \partial\mathcal{X}_{p,q}(x)$ *and suppose* $x' \neq x$ *or* x_+. *Let* $a = W(x')$. *Then there is a continuous mapping* F *of* $\mathcal{X}_{p,q}(x)$ $(\geq a)$ *into itself which satisfies*

1) $W \circ F \geq W$,

2) $W \circ F(x') > W(x')$,

3) $W \circ F(x'') = W(x'') \Rightarrow F(x'') = x''$, *for all* $x'' \in \mathcal{X}_{p,q}(x)$ $(\geq a)$.

Proof. — This follows the main lines of step 1 in § 11. For $y \in \mathcal{X}_{p,q}(x)$ $(\geq a)$ near x', we construct a point $F(y)$ by means of the construction we used there. In other words, $F(y)$ is obtained by modifying y in the same way as x' was obtained by modifying x in step 1 in § 11. Recall that the modification which we used there depended on which case held, i.e. whether $x_{i+1} = f_0(x_i)$, $x_{i+1} = f_1(x_i)$, or $(x_+)_i = x_i$, and then which subcase held. Here the cases are $x'_{i+1} = f_0(x'_i)$, $x'_{i+1} = f_1(x'_i)$, $x'_i = x_i$, and $x'_i = (x_+)_i$. We do the appropriate modification of all y in a neighborhood of x' depending on which case x' belongs to (not which case y belongs to, because that would lead to a discontinuous F). The construction in step 1 in § 11 was such that we can make the modification of y as small as we like. If we make the modification small enough, we will have $F(y) \in \mathcal{X}_{p,q}(x)$ $(\geq a)$. (The proof of this will be discussed below.) Moreover, it is easily seen from the construction in step 1 in § 11 that we can make the modification so that $F(y)$ depends continuously on y and $F(y) = y$ outside a small neighborhood of x', but $F(x') \neq x'$. It is clear from the construction used in step 1 in § 11 that $W \circ F(y) > W(y)$ when $F(y) \neq y$. In this way, we see that all the conditions on the mapping F listed in the lemma will be satisfied.

For example, in the case $x'_i = f_0(x'_{i-1})$, $x'_{i+1} > f_0(x'_i)$, we modify y by increasing y_i slightly. The only way a small enough modification of y of this type could fail to be in $\mathcal{X}_{p,q}(x)$ is if $x'_i = (x_+)_i$. But this cannot be:

$$x'_i = f_0(x'_{i-1}) \leq f_0((x_+)_{i-1}) < (x_+)_i,$$

where the strict inequality is a consequence of the fact that $x \notin \partial\mathcal{X}_{p,q}$.

The case $x'_i = f_1(x'_{i-1})$, $x'_{i+1} > f_1(x'_i)$ may be treated similarly.

In the case $x_i = x'_i$, we have that for some i, either $x_{i-1} < x'_{i-1}$, $x_i = x'_i$, and $x_{i+1} \leq x'_{i+1}$, or $x_{i-1} \leq x'_{i-1}$, $x_i = x'_i$, and $x_{i+1} < x'_{i+1}$ (or both), since $x \neq x'$, by assumption. In either case, we have

$$\frac{\partial}{\partial x'_i}[h(x'_{i-1}, x'_i) + h(x'_i, x'_{i+1})] > 0,$$

since $h_{12} > 0$ and

$$\frac{\partial}{\partial x_i}[h(x_{i-1}, x_i) + h(x_i, x_{i+1})] = 0,$$

since x was assumed to maximize W, and therefore is an equilibrium sequence by Proposition (4.1). So, by slightly increasing y_i, we get $F(y)$ having the required properties.

The case $(x_+)_i = x'_i$ may be treated similarly. □

End of the Proof of Proposition (4.2). — Let x' be a bound point in $\mathcal{X}_{p,q}(x) \cap \{W = W_{p,q,\,\text{minimax}}\}$. If $x' = x$ or x_+ then it is an equilibrium sequence. Otherwise, $x' \notin \partial\mathcal{X}_{p,q}(x)$, according to the lemma and the definition of bound point. Since this is so, it follows, just as in the proof of Proposition (3.3), that

$$h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) = \frac{\partial W(x)}{\partial x_i} = 0,$$

so, x is an equilibrium sequence. □

13. Proof of Lemma (6.1)

Let $\rho : \mathbf{R} \to \mathbf{R}/\mathbf{Z}$ denote the projection. The intervals $\rho[x_i^0, x_i^1]$, $i \in \mathbf{Z}$ are mutually disjoint subsets of $\mathbf{R}/\mathbf{Z}$, by the fact that φ_ω is strictly order-preserving, the fact that $\varphi_\omega(t+1) = \varphi_\omega(t) + 1$, and the fact that ω is irrational. Hence,

$$(13.1) \qquad \sum_{i=-\infty}^{\infty} (x_i^1 - x_i^0) \leq 1.$$

Since $x_i^0 \leq x_i \leq x_i^1$, we obtain

$$(13.2) \qquad \sum_{i=-\infty}^{\infty} |h(x_i, x_{i+1}) - h(x_i^0, x_{i+1}^0)| \leq 2C,$$

where

$$(13.3) \qquad C = \max\{\sup_B h_1, \sup_B h_2\}.$$

Note that C is finite, since $h(x+1, x'+1) = h(x, x')$, and h_1 and h_2 are continuous. □

14. Proof of Lemma (6.2)

Let $x = (\ldots, x_i, \ldots) \in \mathcal{X}_{\omega,t}$ and suppose $G_\omega(x) > 0$. We will construct $\varphi' \in Y_\omega$ such that $F_\omega(\varphi') > F_\omega(\varphi_\omega)$. This will contradict our assumption that φ_ω maximizes F_ω over Y_ω.

Choose a large positive integer N and a small positive number δ. Define $\varphi' \in Y_\omega$ by

$$\varphi'(s) = x_i \qquad \text{if } t + \omega i - \delta \leq s \leq t + \omega i, \quad |i| \leq N,$$
$$\varphi'(s) = \varphi_\omega(s), \qquad \text{for all other } s.$$

For fixed N and small enough δ (depending on N), φ' is well-defined by the above specifications.

For N large enough and δ small enough, we have $F_\omega(\varphi') > F_\omega(\varphi_\omega)$. Indeed,

$$\begin{aligned} F_\omega(\varphi') - F_\omega(\varphi_\omega) = \sum_{i=-N}^{N-1} \int_{t+\omega i-\delta}^{t+\omega i} [h(x_i, x_{i+1}) - h(\varphi_\omega(s), \varphi_\omega(s+\omega))]\, ds \\ + \int_{t+\omega N-\delta}^{t+\omega N} [h(x_N, \varphi_\omega(s+\omega)) - h(\varphi_\omega(s), \varphi_\omega(s+\omega))]\, ds \\ + \int_{t-\omega(N+1)-\delta}^{t-\omega(N+1)} [h(\varphi_\omega(s), x_{-N}) - h(\varphi_\omega(s), \varphi_\omega(s+\omega))]\, ds. \end{aligned}$$

It follows that if we fix N, then

$$\lim_{\delta\to 0}\frac{F_\omega(\varphi') - F_\omega(\varphi_\omega)}{\delta} = \sum_{i=-N}^{N-1}(h(x_i, x_{i+1}) - h(x_i^0, x_{i+1}^0))$$

$$+ [h(x_N, x_{N+1}^0) - h(x_N^0, x_{N+1}^0)] + [h(x_{-N-1}^0, x_N) - h(x_{-N-1}^0), x_{-N}^0)].$$

Therefore, we will have $F_\omega(\varphi') - F_\omega(\varphi_\omega) > 0$ for small enough δ, if the right side above is positive. But, as $N \to \infty$, the right side converges to $G_\omega(x)$, which is positive by assumption. Hence, the right side is positive for all sufficiently large N, and we obtain the existence of φ' for which $F_\omega(\varphi') > F_\omega(\varphi_\omega)$, which is the desired contradiction.

We have shown $G_\omega(x) \leq 0$, for all $x \in \mathscr{X}_{\omega,t}$. In particular $G_\omega(x^1) \leq 0$.

Define $G'_\omega(x)$ in the same way as $G_\omega(x)$, but with x^0 replaced by x^1, i.e.

$$G'_\omega(x) = \sum_{i=-\infty}^{\infty}(h(x_i, x_{i+1}) - h(x_i^1, x_{i+1}^1)).$$

We may show that $G'_\omega(x) \leq 0$, for all $x \in \mathscr{X}_{\omega,t}$, by an argument just like that which we have just given. Clearly $G'_\omega(x^0) = - G_\omega(x^1)$, so from $G'_\omega(x^0) \leq 0$, we obtain $G_\omega(x^1) \geq 0$. We have shown both $G_\omega(x^1) \leq 0$ and $G_\omega(x^1) \geq 0$, so we have $G_\omega(x^1) = 0$. □

15. Proof of Proposition (9.1)

We set $x^0 = \hat{\varphi}_\omega(t-)$ and $x^1 = \hat{\varphi}_\omega(t+)$, following the notation we introduced in § 6. We define $\partial\mathscr{X}_{\omega,t} = \{x \in \mathscr{X}_{\omega,t} :$ for some i, we have $x_i = x_i^0$ or $x_i = x_i^1$ or $f_0(x_i) = x_{i+1}$ or $f_1(x_i) = x_{i+1}\}$. In the proof we are about to give, $\partial\mathscr{X}_{\omega,t}$ plays a role analogous to that played by $\partial\mathscr{X}_{p,q}$ in the proof of Proposition (4.1) (§ 11) and that played by $\partial\mathscr{X}_{p,q}(x)$ in the proof of Proposition (4.2) (§ 12). However, in this case $\mathscr{X}_{\omega,t}$ is infinite dimensional (in fact, a Hilbert cube manifold), and $\partial\mathscr{X}_{\omega,t}$ has no intrinsic topological meaning. But this doesn't affect the proof any.

Suppose $x = (\ldots, x_i, \ldots) \in \mathscr{X}_{\omega,t}$ and $G_\omega(x) = 0$.

Step 1. — We will show that if $x \in \partial\mathscr{X}_{\omega,t}$, then $x = x^0$ or x^1. There are several cases to be considered:

Case 1. — $x_i = x_i^0$ for some i. If $x \neq x^0$, then either there exists i such that $x_{i-1} > x_{i-1}^0$ and $x_i = x_i^0$ or there exists an i such that $x_i = x_i^0$ and $x_{i+1} > x_{i+1}^0$. Consider the first subcase. Since $h_{12} > 0$ and $x_{i+1} \geq x_{i+1}^0$,

$$h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) > h_2(x_{i-1}^0, x_i^0) + h_1(x_i^0, x_{i+1}^0) = 0,$$

so if we replace x by x' by slightly increasing x_i, we find $G_\omega(x') > G_\omega(x) = 0$. But since $x' \in \mathscr{X}_{\omega,t}$, this contradicts Lemma (6.2).

The subcase when $x_i = x_i^0$ and $x_{i+1} > x_{i+1}^0$ may be treated similarly.

Case 2. — $x_i = x_i^1$ for some i. Similar to Case 1.

Case 3. — $x_i = f_0(x_{i-1})$, for some i. If we had $x_i = f_0(x_{i-1})$ for all large i, it would follow that

$$\rho(f_0) = \lim_{i \to +\infty} \frac{f_0^i(x_0)}{i} = \lim_{i \to +\infty} \frac{x_i}{i} = \omega,$$

which contradicts our assumption that $\rho(f_0) < \omega$. Hence, there exists i for which $x_i = f_0(x_{i-1})$ and $x_{i+1} > f_0(x_i)$. Then

$$h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) = h_1(x_i, x_{i+1}) > 0,$$

so, slightly increasing x_i produces $x' \in \mathscr{X}_{\omega, t}$ for which $G_\omega(x') > G_\omega(x) = 0$. This contradicts Lemma (6.2).

Case 4. — $x_i = f_1(x_{i-1})$, for some i. Similar to Case 3.

Step 2. — If $x = x^0$ or x^1, then it is an equilibrium point. Otherwise, it is not in $\partial\mathscr{X}_{\omega, t}$, by Step 1.

In the latter case, if we had $h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) \neq 0$ for some i, then by slightly increasing or decreasing x_i, we would get $x' \in \mathscr{X}_{\omega, t}$ for which $G_\omega(x') > G_\omega(x) = 0$. This contradicts Lemma (6.2). Therefore, x is an equilibrium point. □

16. Proof of Proposition (10.1)

We begin with the case when ω is irrational. Suppose that the conclusion of Proposition (10.1) does not hold. One possibility is that there exist two orbits (x_i, y_i) and (x'_i, y'_i) in M'_ω ($= M'_{f,\omega}$) such that $x_i < x'_i$, $x_j < x'_j$, and $x_k \geq x'_k$ for $i < k < j$, where i and j are some positive integers such that $j > i + 1$. In this case, $x = (\ldots, x_k, \ldots) \in \mathscr{X}_{\omega, t}$ and $x' = (\ldots, x'_k, \ldots) \in \mathscr{X}_{\omega, t}$ for the same value of t. We define $x'', x''' \in \mathscr{X}_{\omega, t}$ by

$$\begin{aligned} x''_k &= x_k, \quad k \leq i \text{ or } k \geq j, \\ &= x'_k, \quad i < k < j, \\ x'''_k &= x'_k, \quad k \leq i \text{ or } k \geq j \\ &= x_k \quad i < k < j. \end{aligned}$$

We have

$$\begin{aligned} &G_\omega(x'') + G_\omega(x''') - G_\omega(x) - G_\omega(x') \\ &\quad = h(x_i, x'_{i+1}) + h(x'_i, x_{i+1}) - h(x_i, x_{i+1}) - h(x'_i, x'_{i+1}) + h(x_{j-1}, x'_j) \\ &\qquad + h(x'_{j-1}, x_j) - h(x_{j-1}, x_j) - h(x'_j, x'_{j+1}). \\ &\quad = \left\{ \int_{x_i}^{x'_i} \int_{x'_{i+1}}^{x_{i+1}} + \int_{x'_{j-1}}^{x_{j-1}} \int_{x_j}^{x'_j} \right\} h_{12}(s, t)\, ds\, dt > 0. \end{aligned}$$

Since $G_\omega(x) = G_\omega(x') = 0$, we obtain

$$G_\omega(x'') + G_\omega(x''') > 0.$$

This contradicts Lemma (6.2).

Another possibility is that there exist two orbits (x_i, y_i) and (x'_i, y'_i) in M'_ω such that $x_i < x'_i$ and $x_j \geq x'_j$, for all $j > i$. Again $x = (\ldots, x_i, \ldots)$ and $x' = (\ldots, x'_i, \ldots)$ are both in $\mathscr{X}_{\omega, t}$, for some t. In this case we define

$$\begin{aligned} x''_k &= x_k, \quad k \leq i, \\ &= x'_k, \quad i < k, \\ x'''_k &= x'_k, \quad k \leq i, \\ &= x_k, \quad i < k. \end{aligned}$$

We have

$$\begin{aligned} G_\omega(x'') &+ G_\omega(x''') - G_\omega(x) - G_\omega(x') \\ &= h(x_i, x'_{i+1}) + h(x'_i, x_{i+1}) - h(x_i, x_{i+1}) - h(x'_i, x'_{i+1}) \\ &= \int_{x_i}^{x'_i} \int_{x'_{i+1}}^{x_{i+1}} h_{12}(s, t)\, ds\, dt > 0. \end{aligned}$$

So, again we get a contradiction.

The remaining possibility leads to a contradiction in the same way. For every possibility of the conclusion of Proposition (10.1) being false when ω is irrational, we have obtained a contradiction. So, we have proved Proposition (10.1) when ω is irrational.

Now, suppose $\omega = p/q$, where p and q are relatively prime integers, $q > 0$. Let x and $x' \in \mathscr{X}_{p,q}$ both maximize W over $\mathscr{X}_{p,q}$. We set

$$\begin{aligned} x''_i &= \min(x_i, x'_i), \\ x'''_i &= \max(x_i, x'_i). \end{aligned}$$

It is easily verified that $x'', x''' \in \mathscr{X}_{p,q}$. Moreover, the argument we have just given shows that $W(x'') + W(x''') - W(x) - W(x') > 0$, unless $x_i < x'_i$, for all i, or $x'_i < x_i$, for all i, or $x_i = x'_i$, for all i.

For, we can evaluate $W(x'') + W(x''') - W(x) - W(x')$ in the same way as we evaluated $G_\omega(x'') + G_\omega(x''') - G_\omega(x) - G_\omega(x')$ above. The resulting sum has a positive summand for every i for which $x_i < x'_i$ and $x_{i+1} \geq x'_{i+1}$, or $x'_i < x_i$ and $x'_{i+1} \geq x_{i+1}$, or $x_i \leq x'_i$ and $x_{i+1} > x'_{i+1}$, or $x'_i \leq x_i$ and $x'_{i+1} > x_{i+1}$. These are all the summands, so $W(x'') + W(x''') - W(x) - W(x') > 0$.

But, since x and x' maximize W over $\mathscr{X}_{p,q}$ and $x'', x''' \in \mathscr{X}_{p,q}$, this inequality gives a contradiction. Hence, we have $x_i < x'_i$, for all i, or $x'_i < x_i$, for all i, or $x_i = x'_i$ for all i. □

17. Proof of Lemma (6.3)

We use the notation of § 13. The reasoning used in § 13 shows that the intervals $\rho[x^0(t)_i, x^1(t)_i] \subset \mathbf{R}/\mathbf{Z}$ are mutually disjoint, where t ranges over the maximal independent set I of points of discontinuity of φ_ω and i ranges over $\mathbf{Z}$. It follows that

$$\sum_{t \in I} \sum_{i=-\infty}^{\infty} (x^1(t)_i - x^0(t)_i) \leq 1.$$

From this and the fact that $x_i^0 \leq x_i \leq x_i^1$ for $x \in \mathcal{X}_{\omega,t}$, we obtain

$$\sum_{t \in I} \sup_{x \in \mathcal{X}_{\omega,t}} \sum_{i=-\infty}^{\infty} |h(x_i, x_{i+1}) - h(x^0(t)_i, x^0(t)_{i+1})| \leq 2C,$$

where C is given by (13.3).

Obviously, $G_\omega \geq - \sup_{x \in \mathcal{X}_{\omega,t}} \sum_{i=-\infty}^{\infty} |h(x_i, x_{i+1}) - h(x^0(t)_i, x^0(t)_{i+1})|$, on $\mathcal{X}_{\omega,t}$.

Hence

$$\Delta W_{\omega,t} \leq \sup_{x \in \mathcal{X}_{\omega,t}} \sum_{i=-\infty}^{\infty} |h(x_i, x_{i+1}) - h(x^0(t)_i, x^0(t)_{i+1})|,$$

so we get

$$\sum_{t \in I} \Delta W_{\omega,t} \leq 2C. \quad \square$$

18. More on Minimax Orbits

Let $f \in \mathcal{F}^1$ and let $\rho(f_0) < \omega < \rho(f_1)$. We consider a closed non-void subset M of $M'_{f,\omega}$ which is invariant under f and T. In view of Proposition (10.1) (proved in § 16), such a set is f-monotone. In the case ω is irrational, we have $M_{f,\omega} \subset M$, since for any $(x,y) \in M'_{f,\omega}$, the closure of $\{f^i T^j(x,y) : i,j \in \mathbf{Z}\}$ contains $M_{f,\omega}$.

By a *complementary interval* of $\pi_1 M$, we will mean the closure of a component of the complement of $\pi_1 M$. Let J be a complementary interval of $\pi_1 M$. We will denote its endpoints by $x_0^0 = x^0(J)_0$ and $x_0^1 = x^1(J)_0$, where $x_0^0 < x_0^1$. Let $y_0^0, y_0^1 \in [0, 1]$ be numbers defined by $(x_0^0, y_0^0) \in M$, $(x_0^1, y_0^1) \in M$. Set $f^i(x_0^j, y_0^j) = (x_i^j, y_i^j)$, for $j = 0, 1$, $i \in \mathbf{Z}$. Set $x^0 = x^0(J) = (\ldots, x_i^0, \ldots)$ and $x^1 = x^1(J) = (\ldots, x_i^1, \ldots)$.

In the case ω is irrational, we let $\mathcal{X}_J = \{x = (\ldots, x_i, \ldots) \in \mathcal{X} : x_i^0 \leq x_i \leq x_i^j,$ for $i \in \mathbf{Z}\}$. For $x \in \mathcal{X}_J$, we have

$$G_\omega(x) = \sum_{i=-\infty}^{\infty} (h(x_i, x_{i+1}) - h(x_i^0, x_{i+1}^0)) = \sum_{i=-\infty}^{\infty} (h(x_i, x_{i+1}) - h(x_i^1, x_{i+1}^1)).$$

In the case ω is rational, say $\omega = p/q$ in lowest terms, we let

$$\mathcal{X}_J = \{x = (\ldots, x_i, \ldots) \in \mathcal{X} : x_i^0 \leq x_i \leq x_i^1 \text{ and } x_{i+q} = x_i + p, \text{ for } i \in \mathbf{Z}\}.$$

We let

$$G_\omega(x) = \sum_{i=0}^{q-1} (h(x_i, x_{i+1}) - h(x_i^0, x_{i+1}^0)) = W(x) - W(x^0).$$

Since x^0 maximizes W over $\mathcal{X}_{p,q}$, we have $G_\omega(x) \leq 0$ everywhere. Since x^1 also maximizes W over $\mathcal{X}_{p,q}$, we have $G_\omega(x^1) = G_\omega(x^0) = 0$.

Now we are again in the situation considered § 3. In other words, $\mathcal{X}_J$ is a compact, Hausdorff space, which is connected and locally pathwise connected, G_ω is a continuous function on $\mathcal{X}_J$, and G_ω takes its maximum value at x^0 and x^1. Clearly $x^0 \neq x^1$.

Definition. — We let ΔW_J denote the negative of the minimax value of $G_\omega : \mathscr{X}_J \to \mathbf{R}$ associated to x^0 and x^1.

This generalizes two definitions which we have made previously. In the case that ω is irrational, $M = M_\omega$, t is a point of discontinuity of φ_ω, and $J = [\varphi(t-), \varphi(t+)]$, we have $\Delta W_J = \Delta W_{\omega,t}$. (The latter was defined in § 6.) In the case that $\omega = p/q$, we consider a max (f, T)-Birkhoff orbit M of type (p, q). The obvious action of $\mathbf{Z}^2$ (generated by f and T) on the set of complementary intervals of $\pi_1 M$ is transitive. It follows that ΔW_J is independent of the choice of the interval J, complementary to $\pi_1 M$, in this case. Moreover, if $(\dots(x_i, y_i), \dots)$ is an f-orbit in M, then

(18.1) $$\Delta W_J = W_{p,q,\max} - W_{p,q,x,\text{minimax}},$$

where $x = (\dots, x_i, \dots)$. This is what we denoted $\Delta W_{p/q}$ in § 4, although $\Delta W_{p/q}$ will not be proved to be independent of the choice of $x \in \mathscr{X}_{p/q}$ maximizing W (or, equivalently, of the choice of max (f, T)-Birkhoff orbit M) until § 19.

Proposition **(18.1)**. — *Let* $x \in \mathscr{X}_J$ *and suppose* $G_\omega(x) = -\Delta W_J$. *If* x *is bound with respect to* $(G_\omega, \mathscr{X}_J, x^0, x^1)$, *then* x *is an equilibrium sequence.*

According to Proposition (3.2) such bound points exist.

Since $x = (\dots, x_i, \dots)$ is an equilibrium sequence, there is associated an orbit $(\dots, (x_i, y_i), \dots)$ of f. We will call this a *minimax orbit of frequency* ω *associated to* J.

Proposition (18.1) generalizes both Proposition (4.2) and Proposition (7.1).

We set $\partial\mathscr{X}_J = \{x \in \mathscr{X}_J$: for some i, we have $x_i = x_i^0$ or $x_i = x_i^1$ or $f_0(x_i) = x_{i+1}$ or $f_1(x_i) = x_{i+1}\}$. If $\omega = p/q$, then $\mathscr{X}_J$ is a q-dimensional manifold with boundary $\partial\mathscr{X}_J$. On the other hand, if ω is irrational, then $\mathscr{X}_J$ is infinite dimensional (in fact, a Hilbert cube manifold), and $\partial\mathscr{X}_J$ has no intrinsic topological meaning. However, this is of no importance in what we do.

We have the following generalization of the lemma in § 12. For $a \in \mathbf{R}$, let $\mathscr{X}_J(\geq a) = \{x' \in \mathscr{X}_J : G_\omega(x') \geq a\}$.

Lemma. — *Let* $x' \in \partial\mathscr{X}_J$ *and suppose* $x' \neq x^0$ *or* x^1. *Let* $a = G_\omega(x')$. *Then there is a continuous mapping* F *of* $\mathscr{X}_J(\geq a)$ *into itself which satisfies:*

1) $G_\omega \circ F \geq G_\omega$,
2) $G_\omega \circ F(x') > G_\omega(x')$,
3) $G_\omega \circ F(x'') = G_\omega(x'') \Rightarrow F(x'') = x''$,

for all $x'' \in \mathscr{X}_J(\geq a)$.

If we take $\omega = p/q$, and let M be a max (f, T)-Birkhoff orbit of type (p, q), this lemma becomes a restatement of the lemma in § 12, since $\mathscr{X}_J = \mathscr{X}_{p,q}(x)$, where $(\dots, (x_i, y_i), \dots)$ is an appropriate f-orbit in M and $x = (\dots, x_i, \dots)$, and G_ω differs from W by a constant.

Proof. — Same as the proof of the lemma in § 12, with appropriate notational changes. Note that since $F(y)$ is a modification of y obtained by increasing or decreasing a single coordinate y_i of y, the fact that G_ω is defined by an infinite sum causes no difficulty. □

Proof of Proposition (18.1). — If $x = x^0$ or x^1, then it is an equilibrium sequence. Otherwise, it is not in $\partial\mathscr{X}_J$, by the lemma. Suppose

$$h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) \neq 0.$$

Depending on the sign, we may construct $F(y)$ by slightly increasing or decreasing the i-th coordinate of y, for all y in a small neighborhood of x and letting $F(y) = y$ outside such a small neighborhood. Such an F satisfies the conditions we gave in the definition of a free point. Therefore, x is free. But, we assumed x was bound. This contradiction shows that

$$h_2(x_{i-1}, x_i) + h_1(x_i, x_{i+1}) = 0,$$

so x is an equilibrium sequence. □

19. Comparison of Minimax Orbits

Let $(f, \omega) \in \mathscr{D}$. We consider two closed non-void subsets M, M′ of M'_ω, both of which are invariant under f and T. We suppose $M \subset M'$.

Proposition **(19.1)**. — *Let* J *be a complementary interval of* π_1 M. *Then*

$$\Delta W_J = \max\{\Delta W_{J'} : J' \text{ is a complementary interval of } \pi_1 M' \text{ and } J' \subset J\}.$$

We let $J'_i = [x^0(J')_i, x^1(J')_i]$, in the notation of § 18. Obviously,

$$\Delta W_{J'(i)+j} = \Delta W_{J'},$$

for any integers i and j. We will say that two complementary intervals J′ and J″ of π_1 M are *ω-independent* if $J'' \neq J'_i + j$, for all $i, j \in \mathbf{Z}$.

Lemma **(19.2)**. — $\sum_{J'} \Delta W_{J'} < \infty$, *where the sum is taken over a maximal ω-independent set of complementary intervals of* π_1 M′.

Proof. — Same as the proof of Lemma (6.3) (§ 17). □

From Lemma (19.2), it follows that $\{\Delta W_{J'} : J'$ is a complementary interval of $\pi_1 M'$ and $J' \subset J\}$ has a maximum.

Notation: For $x \in \pi_1 M'_\omega$, we define $\hat{x} = (\ldots, \hat{x}_i, \ldots)$ by choosing y so that $(x, y) \in M'_\omega$, and setting $\hat{x}_i = \pi_1 f^i(x, y)$.

Proof of Proposition (19.1). — For each complementary interval J' of $\pi_1 M$, let $C_{J'}$ be the connected component of $\{x \in \mathscr{X}_{J'} : G_\omega(x) \geq -\Delta W_{J'}\}$ which contains $x^0(J')$. In view of the definition of $\Delta W_{J'}$, we have that $C_{J'}$ also contains $x^1(J')$. We let

$$C = \cup\{C_{J'} : J' \subset J\} \cup \{\hat{x} : x \in \pi_1 M' \cap J\}.$$

It is easily seen that C is a connected subset of $\mathscr{X}_J$ which contains $x^0(J)$ and $x^1(J)$. Hence

$$\Delta W_J \leq \max\{\Delta W_{J'}\},$$

where J' ranges over all complementary intervals of $\pi_1 M'$ such that $J' \subset J$.

To prove the opposite inequality, we define a retraction

$$\pi_{J,J'} : \mathscr{X}_J \to \mathscr{X}_{J'},$$

for each complementary interval J' of $\pi_1 M'$ in J. We set

$$\begin{aligned} \pi_{J,J'}(x)_i &= x^0(J')_i, && \text{if } x_i \leq x^0(J')_i \\ &= x_i, && \text{if } x^0(J')_i \leq x_i \leq x^1(J')_i \\ &= x^1(J')_i, && \text{if } x^1(J')_i \leq x_i. \end{aligned}$$

We will show that $G_\omega \circ \pi_{J,J'}(x) \geq G_\omega(x)$ with equality only if $\pi_{J,J'}(x) = x$.

The algebra is a little easier if we do this in two steps. Let $x^0 = x^0(J')$, $x'_i = \max(x^0_i, x_i)$, $x''_i = \min(x^0_i, x_i)$. The argument used in § 16 shows that

$$G_\omega(x') + G_\omega(x'') \geq G_\omega(x) + G_\omega(x^0),$$

with equality if and only if $x = x'$. Since $M' \subset M'_\omega$, it follows from the defining property of M'_ω that $G_\omega(x^0) = 0$. Moreover, $G_\omega(x'') \leq 0$, since this is true for all elements of $\mathscr{X}_J$. Hence

$$G_\omega(x') \geq G_\omega(x),$$

with equality only if $x = x'$.

Now set $x^1 = x^1(J')$, $x'''_i = \min(x^1_i, x'_i)$, $x^{(iv)}_i = \max(x^1_i, x'_i)$. Again, we find

$$G_\omega(x''') + G_\omega(x^{(iv)}) \geq G_\omega(x) + G_\omega(x^1),$$

and $G_\omega(x^1) = 0$, $G_\omega(x^{(iv)}) \leq 0$, so we obtain

$$G_\omega(x''') \geq G_\omega(x').$$

Since $x''' = \pi_{J,J'}(x)$, we have shown

$$G_\omega \circ \pi_{J,J'}(x) \geq G_\omega(x).$$

Let C be the connected component of $\{x \in \mathscr{X}_J : G_\omega(x) \geq -\Delta W_J\}$ which contains $x^0(J)$ and $x^1(J)$. Then $\pi_{J,J'}(C)$ is a connected subset of $\mathscr{X}_{J'}$ which contains $x^0(J')$ and $x^1(J')$. Consequently, the inequality above implies

$$\Delta W_{J'} \leq \Delta W_J. \quad \square$$

Proposition (4.3) follows from Proposition (19.1). For, take $\omega = p/q$, $M = \{(x_i + j, y_i) : i, j \in \mathbf{Z}\}$, where $y_i = h_1(x_i, x_{i+1})$, and take $M' = M'_\omega$. If J is a complementary interval of $\pi_1 M$, then (18.1) holds. Hence, Proposition (19.1) and the fact that ΔW_J is independent of the choice of complementary interval J of $\pi_1 M$ (in this case) imply that

$$W_{p,q,x,\text{minimax}} = W_{p,q,\max} - \max_{J'} \Delta W_{J'},$$

where J′ runs over all complementary intervals of M′. Since M′ is independent of x, this shows that $W_{p,q,x,\text{minimax}}$ is independent of x, which is the statement of Proposition (4.3).

20. Proof of Proposition (5.2)

We will prove Proposition (5.2) using the definition of ΔW_ω given in § 6.

First, suppose that there is an f-invariant circle, going once around the annulus, of rotation number ω. Its lifting to A is a curve which separates A into two parts. From the monotone twist hypothesis, it follows that points in the lower part advance slower than the rate ω, and points in the upper part advance faster. Hence, M_ω is not in the lower or upper part; it is in the curve itself.

Lemma **(20.1)**. — *M'_ω is the invariant curve.*

Proof. — By a theorem of G. D. Birkhoff [10, § 44] and [11, § 3], the invariant curve is the graph of a Lipschitz function $\mu : \mathbf{R} \to [0, 1]$. It is enough to show that graph $\mu \subset M'_\omega$, since M'_ω cannot properly contain graph μ, by the monotone property of M'_ω (Proposition (10.1)). If $M'_\omega = \text{graph}\ \mu$, we are done. Otherwise, we consider a complementary interval J of $\pi_1 M_\omega$. Let x_0^0 and x_0^1 denote its endpoints. Let $x_i^\varepsilon = \pi_1 f^i(x_0^\varepsilon, \mu(x_0^\varepsilon))$, $\varepsilon = 0, 1$. Let $x^\varepsilon = (\ldots, x_i^\varepsilon, \ldots)$, $\varepsilon = 0, 1$. For $x_0 \in J$, let $x_i = \pi_1 f^i(x_0, \mu(x_0))$. Let $x = (\ldots, x_i, \ldots)$. It is enough to prove that

$$\textbf{(20.1)} \qquad \sum_{i=-\infty}^{\infty} (h(x_i, x_{i+1}) - h(x_i^0, x_{i+1}^0)) = 0,$$

for, then, we will have graph $\mu \mid J \subset M'_\omega$.

First, we will give a formal argument which indicates why (20.1) should be true. Then we will give the rigorous version of the formal argument.

Formally, we differentiate (20.1) with respect to x_0:

$$\frac{\partial}{\partial x_0} \sum_{i=-\infty}^{\infty} (h(x_i, x_{i+1}) - h(x_i^0, x_{i+1}^0))$$
$$= \sum_{i=-\infty}^{\infty} \left(h_1(x_i, x_{i+1}) \frac{dx_i}{dx_0} + h_2(x_i, x_{i+1}) \frac{dx_{i+1}}{dx_0} \right)$$
$$= \sum_{i=-\infty}^{\infty} (h_1(x_i, x_{i+1}) + h_2(x_{i-1}, x_i)) \frac{dx_i}{dx_0} = 0.$$

We have $h_1(x_i, x_{i+1}) + h_2(x_{i-1}, x_i) = 0$, because graph μ is f-invariant.

The difficulties with this argument are, first, that x_i is not necessarily a differentiable function of x_0. Second, we have to justify differentiating under the summation sign and changing the order of summation in the second sum.

By Birkhoff's theorem, μ is Lipschitz, so x_i is a Lipschitz function of x_0. Obviously, x_i is an increasing function of x_0. Since x_i is a Lipschitz function of x_0, its derivative dx_i/dx_0 exists almost everywhere, and

$$x_i^1 - x_i^0 = \int_{x_0^0}^{x_0^1} \frac{dx_i}{dx_0}\, dx_0 = \int_{x_0^0}^{x_0^1} \left| \frac{dx_i}{dx_0} \right| dx_0,$$

since the Fundamental Theorem of Calculus is true for absolutely continuous functions, and any Lipschitz function is absolutely continuous (cf. [30, 11.7]).

The reasoning which gives (13.1) is valid in this case, too, so we get

$$\sum_{i=-\infty}^{\infty} \int_{x_0^0}^{x_0^1} \left| \frac{dx_i}{dx_0} \right| dx_0 = \sum_{i=-\infty}^{\infty} (x_i^1 - x_i^0) \leq 1.$$

Since h_1 and h_2 are bounded and continuous, the sums which appear in the above calculation are absolutely convergent in the L^1 norm on functions defined on J. So, the changes of order of summation are valid if the various terms appearing are considered as elements of $L^1(J)$. So, we get

$$\begin{aligned} 0 &= \int_{x_0=x_0^0}^{x_0^1} \left[\sum_{i=-\infty}^{\infty} h_1(x_i, x_{i+1}) \frac{dx_i}{dx_0} + h_2(x_i, x_{i+1}) \frac{dx_{i+1}}{dx_0} \right] dx_0 \\ &= \sum_{i=-\infty}^{\infty} \int_{x_0=} \frac{d}{dx_0} h(x_i, x_{i+1})\, dx_0 \\ &= \sum_{i=-\infty}^{\infty} (h(x_i, x_{i+1}) - h(x_i^0, x_{i+1}^0)). \quad \square \end{aligned}$$

From Lemma (20.1), it follows immediately that $\Delta W_\omega = 0$, for the definition of the latter given in § 6. We have in this way shown that the existence of an invariant circle implies $\Delta W_\omega = 0$.

Conversely, suppose $\Delta W_\omega = 0$. From the definition of ΔW_ω given in § 6, it follows that for every point t of discontinuity of φ_ω, we have that $\mathscr{X}_{\omega,t}$ (≥ 0) is connected. It follows that $\pi_1(M'_\omega) = \mathbf{R}$. In view of the monotonicity of M'_ω, it follows easily from this that M'_ω is the graph of a function. The projection of this graph on the annulus is the required circle. $\square$

21. Convergence in Measure

Let $(f, \omega) \in \mathscr{D}$. If ω is irrational, the set M_ω/T supports a unique Borel probability measure μ_ω which is $\bar{f}$-invariant. This is a consequence of the semi-conjugacy of $\bar{f} \mid (M_\omega/T)$ with the rotation of the circle through ω. Note that μ_ω is also the only Borel probability measure on M'_ω/T which is $\bar{f}$-invariant. This is because every $\bar{f}$-orbit in M'_ω/T is either in M_ω/T or is homoclinic to M_ω/T (i.e. approaches arbitrarily close to M_ω/T under indefinite forward and backward iteration).

When ω is rational, M'_ω/T may support more than one Borel probability measure which is $\bar{f}$-invariant. In fact, the $\bar{f}$-ergodic Borel probability measures on M'_ω/T are in $1-1$ correspondence with the $\bar{f}$-orbits in M'_ω/T; each $\bar{f}$-orbit supports one such measure. In particular, M'_ω/T supports only one $\bar{f}$-invariant Borel probability measure if and only if M'_ω/T is reduced to a single $\bar{f}$-orbit. This is the same as saying that M'_ω is reduced to a single (f, T) orbit. Obviously, this is the case for generic f.

We provide the space of Borel probability measures on A/T with the weak topology. A basis of open sets for this topology consists of sets of the form

$$\{\mu : a_i < \mu(f_i) < b_i, \text{ for } i = 1, \ldots, n\},$$

where $a_i < b_i$ are real numbers and f_i is a continuous function on A/T, for $i = 1, \ldots, n$.

We will indicate the dependence of μ_ω on f explicitly, by writing $\mu_{f,\omega}$.

Proposition **(21.1)**. — *When $(f, \omega) \in \mathscr{D}$ and $\omega = p/q$, let $\mu_{f,\omega}$ denote any Borel probability measure on $M'_{f,\omega}/T$ which is f-invariant. (For other $(f, \omega) \in \mathscr{D}$, let $\mu_{f,\omega}$ be as defined above.) The mapping $(f, \omega) \to \mu_{f,\omega}$ is continuous at all points of $\mathscr{D}$ where it is uniquely defined, i.e. where ω is irrational or where $\omega = p/q$ and there is only one max (f, T)-Birkhoff orbit of type (p, q).*

Proof. — Let $\mu_{f,\omega}$ be an element of $X_{f,\omega}$ which maximizes $F_{f,\omega}$. When ω is irrational, $\varphi_{f,\omega}$ is unique by the main result of [22]. When $\omega = p/q$ and there is only one max (f, T)-Birkhoff orbit of type (p, q), it is easily seen that $\varphi_{f,\omega}$ is unique.

To $\varphi_{f,\omega}$ we may associate a measure $\mu_{f,\omega}$, as follows. We define

$$\eta_{f,\omega}(t) = h_1(\varphi_{f,\omega}(t), \varphi_{f,\omega}(t+\omega)).$$

Then $(\varphi_{f,\omega}, \eta_{f,\omega})$ is a mapping of $\mathbf{R}$ into A. Let $\pi_T : A \to A/T$ denote the projection. We set

$$\mu_{f,\omega} = (\pi_T \circ (\varphi_{f,\omega}, \eta_{f,\omega}))_* (\mu \mid [0, 1]),$$

where μ denotes the Lebesgue measure on $[0, 1]$. Then $\mu_{f,\omega}$ is an $\bar{f}$-invariant Borel probability measure on $M'_{f,\omega}/T$. In fact, the correspondence $\varphi_{f,\omega} \leftrightarrow \mu_{f,\omega}$ is a $1-1$ correspondence between elements of $X_{f,\omega}$ which maximize $F_{f,\omega}$ and $\bar{f}$-invariant probability measures on $M'_{f,\omega}/T$. This correspondence is bi-continuous, where we use the weak topology on measures and the topology on the $\varphi_{f,\omega}$ defined by the metric d of [21]. Moreover, because $\varphi_{f,\omega}$ is obtained by maximizing $F_{f,\omega}$ over a compact space $X_{f,\omega}$, and $F_{f,\omega}(\varphi)$ depends continuously on f, ω, and φ, it follows that $\varphi_{f,\omega}$ depends continuously on f and ω, wherever it is uniquely defined. There is a slight technical difficulty in this argument caused by the fact that the space $X_{f,\omega}$ depends on f and ω. But this is overcome by the observation that for each $(f, \omega) \in \mathscr{D}$ there is a compact set $K \subset X_{f,\omega}$ such that

$$\varphi_{f',\omega'} \in K \subset X_{f',\omega'},$$

for all (f', ω') in a sufficiently small neighborhood of (f, ω), where $\varphi_{f',\omega'}$ may be any element of $X_{f',\omega'}$ which maximizes $F_{f',\omega'}$ over $X_{f',\omega'}$. $\square$

Proof of Proposition (9.3). — Immediate from Proposition (21.1) and the fact that $M_{f,\omega} = \operatorname{supp} \mu_{f,\omega}$. $\square$

22. Lemmas Concerning Convergence of Sets

For what we do in the next couple of sections, we will need some elementary results from general topology.

Definition. — By a *basis* $\mathscr{B}$ of a filter $\mathscr{F}$, one means a subcollection $\mathscr{B} \subset \mathscr{F}$ such that for every $U \in \mathscr{F}$, there exists $V \in \mathscr{B}$ such that $V \subset U$.

Definition. — Let $\mathscr{F}$ be a filter on a set X. We say a sequence $x_1, x_2, \ldots$ of elements of X is *$\mathscr{F}$-convergent* if for each $U \in \mathscr{F}$, there exists a positive integer N such that $x_i \in U$, for all $i \geq N$.

Lemma (**22.1**). — *Let $\mathscr{F}$ be a filter on a set* X. *Let* Y *be a topological space. For each $x \in X$, let Z_x be a subset of* Y. *Suppose that there is a countable basis for $\mathscr{F}$ and a countable basis for the topology of* Y.

Let $y \in \limsup_{x,\mathscr{F}} Z_x$. Then there exists an $\mathscr{F}$-convergent sequence $x_1, x_2, \ldots$ of elements of X *such that*

$$y \in \liminf_{i\to\infty} Z_{x(i)} = \limsup_{i\to\infty} Z_{x(i)}.$$

Proof. — Let $U_1, U_2, \ldots$ be a countable basis for $\mathscr{F}$. Since Y has a countable basis for its topology, the system of neighborhoods of y in Y has a countable basis, $V_{1y}, V_{2y}, \ldots$ Since $y \in \limsup_{x,\mathscr{F}} Z_x$, we may choose, for each positive integer i, an element $x_{i1} \in U_i$ such that $Z_{x(i,1)} \cap V_{1y} \neq \emptyset$. Then

$$(\mathbf{22.1}) \qquad y \in \liminf_{i\to\infty} Z_{x(i,1)}.$$

Let $V_1, V_2, \ldots$ be a basis for the topology of Y. We construct, by induction on j, sequences $x_{1j}, x_{2j}, \ldots$ We have just constructed $x_{11}, x_{21}, \ldots$ Suppose $x_{1j}, x_{2j}, \ldots$ has been constructed. If $V_j \cap \limsup_{i\to+\infty} Z_{x(i,j)} = \emptyset$, we let $x_{i,j+1} = x_{ij}$, for all i. Otherwise, we choose $y_j \in V_j \cap \limsup Z_{x(i,j)}$; we let $x_{i,j+1} = x_{ij}$, for $i \leq j$, and let $x_{j+1,j+1}, x_{j+2,j+1}, x_{j+3,j+1}, \ldots$ be a subsequence of $x_{j+1,j}, x_{j+2,j}, x_{j+3,j}, \ldots$ such that

$$(\mathbf{22.2}) \qquad y_j \in \liminf_{i\to\infty} Z_{x(i,j+1)}.$$

The same argument which shows the existence of the sequence $x_{11}, x_{21}, \ldots$ such that (22.1) holds also shows the existence of the subsequence $x_{j+1,j+1}, x_{j+2,j+1}, x_{j+3,j+1}, \ldots$ such that (22.2) holds.

We let $x_i = x_{ii}$. Then we have that

$$\liminf_{i\to\infty} Z_{x(i)} \cap V_j \neq \emptyset \Leftrightarrow \limsup_{i\to\infty} Z_{x(i)} \cap V_j \neq \emptyset.$$

Since $V_1, V_2, \ldots$ is a basis for the topology of Y, and $\liminf_{i\to\infty} Z_{x(i)}$ and $\limsup_{i\to\infty} Z_{x(i)}$ are both closed, it follows that

$$\liminf_{i\to\infty} Z_{x(i)} = \limsup_{i\to\infty} Z_{x(i)}.$$

Since $y \in \liminf_{i\to\infty} Z_{x(i,1)}$, and $x_1, x_2, \ldots$ is a subsequence of $x_{11}, x_{21}, x_{31}, \ldots$, we have that $y \in \liminf_{i\to\infty} Z_{x(i)}$.

Clearly $x_{11}, x_{21}, \ldots$ is $\mathscr{F}$-convergent (since $x_{i1} \in U_i$), so its subsequence $x_1, x_2, \ldots$ is also $\mathscr{F}$-convergent. □

We will adopt the following standard notation. We will say that $\lim_{x,\mathscr{F}} Z_x$ exists and write

(22.3) $$Z_0 = \lim_{x,\mathscr{F}} Z_x$$

to mean that

$$Z_0 = \liminf_{x,\mathscr{F}} Z_x = \limsup_{x,\mathscr{F}} Z_x.$$

In general, there is no convenient topology on the set of subsets of Y such that (22.3) can be interpreted as convergence in that topology [25, p. 179]. However, in a special case there is: We suppose Y is a compact metric space, and the sets Z_x are all *closed* subsets of Y. Let d denote the metric on Y, and use the same symbol to denote the *Hausdorff metric* on the set of closed, non-void subsets of Y. Recall from [18] that this is the metric defined by

(22.4) $$d(Z_1, Z_2) = \max\{\sup_x \inf_y d(x,y), \sup_y \inf_x d(x,y)\},$$

where x ranges of Z_1 and y ranges over Z_2.

We then have the following elementary result: If each Z_x is a closed, non-void subset of the compact metric space Y, then the definition of (22.3) which we gave above is equivalent to convergence with respect to the Hausdorff metric.

In § 24, it will be useful to have a generalization of this result:

Lemma **(22.2)**. — *Let* Y *be a metric space,* T *an isometry of* Y. *Suppose* Y/T *is compact. Define the Hausdorff metric on the set of* T-*invariant, closed, non-void subsets of* Y *by* (22.4). *Suppose each* Z_x *is a* T-*invariant, closed, non-void subset of* Y. *Then the definition of* (22.3) *which we gave above is equivalent to convergence with respect to the Hausdorff metric.*

The proof is elementary. One way to proceed is to observe that the Hausdorff metric on T-orbits is a metric on Y/T. Its underlying topology is the quotient topology, associated to the projection $Y \to Y/T$. Since we have a metric on Y/T, we may speak of the Hausdorff metric on closed, non-void subsets of Y/T. The natural $1-1$ correspondence between T-invariant, closed, non-void subsets of Y and closed, non-void subsets of Y/T is an isometry with respect to the Hausdorff metrics. Since lim inf and lim sup of T-invariant sets commute with passage to the quotient by T, Lemma (22.2) reduces to the special case of it quoted just before the statement of it.

We omit the details. □

23. Upper Semi-Continuity of $M'_{f,\omega}$

Let $(f, \omega) \in \mathscr{D}$ and suppose ω is irrational. Let $(x_0, y_0) \in \limsup_{(f',\omega')\to(f,\omega)} M'_{f',\omega'}$. By Lemma (22.1), there is a sequence $(f_i, \omega_i)_{i=1,2,\dots}$ of elements of $\mathscr{D}$ which converges to (f, ω) as $i \to \infty$, such that $M = \lim_{i\to\infty} M'_{f(i),\omega(i)}$ exists and $(x_0, y_0) \in M$. Since (x_0, y_0) is an arbitrary element of $\limsup_{(f',\omega')\to(f,\omega)} M'_{f',\omega'}$, it is enough to show that $(x_0, y_0) \in M'_{f,\omega}$, in order to prove Proposition (9.2). Since $(x_0, y_0) \in M$, it is then enough to prove $M \subset M'_{f,\omega}$.

In order to prove $M \subset M'_{f,\omega}$, we first develop several properties of M.

By Proposition (9.3) (proved in § 21), we have $M_{f,\omega} \subset M$.

Lemma **(23.1)**. — M *is f-monotone (for the projection on the first factor, in the sense defined in* § 10). *Moreover,* M *satisfies a Lipschitz condition of the form: There exists* $L > 0$ *such that* $(x,y), (x',y') \in M \Rightarrow |y'-y| < L\,|x'-x|$.

In this and the next section, it will be convenient to use the following abbreviations:

Notation. — We set $M^{(i)} = M'_{f(i),\omega(i)}$, $G = G_{f,\omega}$, $G_i = G_{f(i),\omega(i)}$, $h = h(f)$, $h_i = h_{f(i)}$.

Proof of Lemma (23.1). — Since $M^{(i)}$ is (f, T)-invariant and f-monotone, and $M = \lim_{i\to\infty} M^{(i)}$, we obtain that M is *weakly f-monotone*, in the sense that for $(x,y) \in M$, $(x',y') \in M$, we have

$$x \leq x' \Rightarrow \pi_1 f^i(x,y) \leq \pi_1 f^i(x',y'), \quad \text{for all } i \in \mathbf{Z}.$$

Since f commutes with T and A/T is compact, (4.2) implies that there exists $\delta > 0$ such that

$$\frac{\partial(\pi_1 f(x,y))}{\partial y} > \delta \quad \text{and} \quad -\frac{\partial(\pi_1 f^{-1}(x,y))}{\partial y} > \delta,$$

for all $(x,y) \in B$. Likewise, there exists $C > 0$ such that

$$\left|\frac{\partial(\pi_1 f(x,y))}{\partial x}\right| < C \quad \text{and} \quad \left|\frac{\partial(\pi_1 f(x,y))}{\partial x}\right| < C,$$

for all $(x,y) \in \mathrm{B}$. If $\mathrm{L} = \mathrm{C}/\delta$, then $(x,y) \in \mathrm{B}$, $(x',y') \in \mathrm{B}$ and $|y'-y| \geq \mathrm{L}\,|x'-x|$ imply

$$\pi_1 f(x',y') > \pi_1 f(x,y) \quad \text{and} \quad \pi_1 f^{-1}(x',y') < \pi_1 f^{-1}(x,y),$$

or

$$\pi_1 f(x',y') < \pi_1 f(x,y) \quad \text{and} \quad \pi_1 f^{-1}(x',y') > \pi_1 f^{-1}(x,y).$$

Since M is weakly f-monotone, this shows that M satisfies the Lipschitz condition, and is f-monotone. □

Lemma **(23.2)**. — *If* J *is a complementary interval of* $\pi_1 \mathrm{M}$, *then*

$$\mathrm{G}(x^0) = \mathrm{G}(x^1),$$

where $x^0 = x^0(\mathrm{J})$ *and* $x^1 = x^1(\mathrm{J})$.

Note that since $\mathrm{M}_{f,\omega} \subset \mathrm{M}$, we have $\mathrm{J} \subset [\varphi_{f,\omega}(t-), \varphi_{f,\omega}(t+)]$, for some $t \in \mathbf{R}$. (Recall from § 6 that $\varphi_{f,\omega} = \underline{\varphi_\omega}$ denotes the unique element of $\mathrm{X}_{f,\omega}$ which maximizes $\mathrm{F}_{f,\omega}$. We have $\pi_1 \mathrm{M}_{f,\omega} = \overline{\varphi_{f,\omega}(\mathbf{R})}$.) Then $x^0, x^1 \in \mathscr{X}_{f,\omega,t}$, so $\mathrm{G}(x^0)$ and $\mathrm{G}(x^1)$ are defined, and the equation $\mathrm{G}(x^0) = \mathrm{G}(x^1)$ is equivalent to

$$\textbf{(23.1)} \qquad \sum_{j=-\infty}^{\infty} (h(x^1_{j+1}, x^1_j) - h(x^0_{j+1}, x^0_j)) = 0.$$

This sum is absolutely convergent, in view of (13.1), which refers to the endpoints of the interval $[\varphi_{f,\omega}(t-), \varphi_{f,\omega}(t+)]$, but obviously implies the same result for the endpoints of J.

Proof of Lemma (23.2). — Suppose $\mathrm{G}(x^0) > \mathrm{G}(x^1)$. Let

$$x'_j = x^0_j, \quad |j| \leq \mathrm{N},$$
$$\quad = x^1_j, \quad |j| > \mathrm{N},$$

where N is some large positive integer. Then $\mathrm{G}(x') > \mathrm{G}(x^1)$, if N is large enough.

Since $\mathrm{M} = \lim_{i\to\infty} \mathrm{M}^{(i)}$, we have $\pi_1 \mathrm{M} = \lim_{i\to\infty} \pi_1 \mathrm{M}^{(i)}$. According to Lemma (22.2), this can be interpreted as convergence with respect to the Hausdorff metric. (Take $\mathrm{Y} = \mathbf{R}$, $\mathrm{T}(t) = t + 1$, to apply Lemma (22.2).) It follows that we can choose, for each positive integer i, a complementary interval $\mathrm{J}^{(i)}$ of $\mathrm{M}^{(i)}$ such that

$$\mathrm{J} = \lim_{i\to\infty} \mathrm{J}^{(i)}.$$

Set $x^0(\mathrm{J}^{(i)})_j = x^{0(i)}_j$, $x^1(\mathrm{J}^{(i)})_j = x^{1(i)}_j$. Define

$$x'^{(i)}_j = x^{0(i)}_j, \quad |j| \leq \mathrm{N},$$
$$\quad = x^{1(i)}_j, \quad |j| > \mathrm{N}.$$

We have

$$x^0_j = \lim_{i\to\infty} x^{0(i)}_j,$$
$$x^1_j = \lim_{i\to\infty} x^{1(i)}_j,$$

although convergence is not uniform in j. Also $h_i \to h$ uniformly as $i \to \infty$. We have

$$G_i(x'^{(i)}) - G_i(x^{1(i)}) = \sum_{j=-N-1}^{N} (h_i(x_j'^{(i)}, x_{j+1}'^{(i)}) - h_i(x_j^{1(i)}, x_{j+1}^{1(i)})),$$

and

$$G(x') - G(x^1) = \sum_{j=-N-1}^{N} (h(x_j', x_{j+1}') - h(x_j^1, x_{j+1}^1)),$$

so it follows that

$$\lim_{i \to \infty} G_i(x'^{(i)}) - G_i(x^{1(i)}) = G(x') - G(x^1) > 0.$$

Moreover, $G_i(x^{1(i)}) = 0$, for large enough i, by the defining property of $M^{(i)}$. Hence, $G_i(x'^{(i)}) > 0$, for large enough i. But, this contradicts Lemma (6.2). □

The following is a slight variation on the notation introduced in § 19.

Notation. — If $x \in \pi_1 M$, we define $\hat{x} = (\ldots, \hat{x}_j, \ldots)$ by choosing y so that $(x, y) \in M$, and setting $\hat{x}_j = \pi_1 f^j(x, y)$.

If $x \in \pi_1 M$, then either $x \in \pi_1 M_{f,\omega}$ or $x \in J$, for some complementary interval J of $\pi_1 M_{f,\omega}$. In the latter case, $J = [\varphi(t-), \varphi(t+)]$, for some point t of discontinuity of $\varphi = \varphi_{f,\omega}$. Since $M_{f,\omega} \subset M$ and M is f-monotone, it follows that $\hat{x} \in \mathscr{X}_{\omega,t}$, when $x \in J = [\varphi(t-), \varphi(t+)]$.

Proof that $M \subset M'_{f,\omega}$. Consider $x \in J \cap \pi_1 M$, where $J = [\varphi(t-), \varphi(t+)]$ and t is a point of discontinuity of $\varphi = \varphi_{f,\omega}$. It is enough to show that $G(\hat{x}) = 0$.

For arbitrary $x \in J$, we define $\hat{x} = (\ldots, \hat{x}_j, \ldots)$, as follows. If $x \in J \cap \pi_1 M$, we have already defined x. Otherwise, x is contained in a complementary interval $[x^0, x^1]$ of $\pi_1 M$. We have $x^0, x^1 \in J \cap \pi_1 M$, and we define

$$\hat{x}_j = (1 - \lambda)\, \hat{x}_j^0 + \lambda \hat{x}_j^1, \quad \text{where } x = (1 - \lambda)\, x^0 + \lambda x^1.$$

It follows easily from Lemma (23.1) that $\hat{x}_j$ is a Lipschitz function of x. Moreover,

$$\sum_{j=-\infty}^{\infty} \int_{\varphi(t-)}^{\varphi(t+)} \left| \frac{d\hat{x}_j}{dx} \right| dx = \sum_{j=-\infty}^{\infty} (\varphi(t + \omega j +) - \varphi(t + \omega j -)) \leq 1.$$

Now we may reason exactly as in the proof of Lemma (20.1) and conclude that $G(\hat{x})$ is an absolutely continuous function of $x \in J$, and

$$\frac{dG(\hat{x})}{dx} = \sum_{i=-\infty}^{\infty} (h_2(\hat{x}_{i-1}, \hat{x}_i) + h_1(\hat{x}_i, \hat{x}_{i+1}))\, \frac{d\hat{x}_i}{dx},$$

on $J \setminus Z$, where Z is a set of measure 0 in J.

For $x \in \pi_1 M \cap J$, we have that $\hat{x}$ is an equilibrium sequence, by the f-invariance of M. Hence

$$\frac{dG(\hat{x})}{dx} = 0, \quad \text{on } (J \cap \pi_1 M) \setminus Z.$$

By Lemma (23.2), the value of the function $x \mapsto G(\hat{x})$ is the same on both endpoints of a complementary interval of $\pi_1 M$ in J. These facts together with the absolute continuity of $x \mapsto G(\hat{x})$ imply that $x \mapsto G(\hat{x})$ is constant on $\pi_1 M \cap J$. Since it is zero on the endpoints of J, we get that it is zero everywhere on $\pi_1 M \cap J$.

This finishes the proof of Proposition (9.2). □

Another proof of Proposition (9.2) follows from Aubry's work. For, by Aubry's work, there is an identification of $\pi_1 M'_{f,\omega}$ with the set of m.e. states. This is a deep result, and Proposition (9.2) is an immediate consequence of it.

24. Proof of Proposition (9.4)

Note that Propositions (5.1) and (6.4) follow immediately from Proposition (9.4).

Let $(f, \omega) \in \mathscr{D}$ and let ω be irrational. Since $\mathscr{D}$ has a countable basis for its topology, it is easily seen that in order to prove that ΔW is continuous at (f, ω), it is sufficient to prove that a sequence $(f_1, \omega_1), (f_2, \omega_2), \ldots$ which tends to (f, ω) has a subsequence $(f_{i(1)}, \omega_{i(1)}), (f_{i(2)}, \omega_{i(2)}), \ldots$ such that

$$\Delta W_{f,\omega} = \lim_{j\to\infty} \Delta W_{f(i(j)),\,\omega(i(j))}.$$

It follows from Lemma (22.1) that any sequence $(f_1, \omega_1), (f_2, \omega_2), \ldots$ which tends to (f, ω) has a subsequence $(f_{i(1)}, \omega_{i(1)}), \ldots$ such that $\lim_{j\to\infty} M'_{f(i(j)),\,\omega(i(j))}$ exists. Therefore, it will be enough to prove:

Lemma **(24.1)**. — *Let $(f_1, \omega_1), (f_2, \omega_2), \ldots$ be a sequence in $\mathscr{D}$ which converges to $(f, \omega) \in \mathscr{D}$, where ω is irrational, and suppose*

(24.1) $$M = \lim_{i\to\infty} M'_{f(i),\,\omega(i)}$$

exists. Then

$$\Delta W_{f,\omega} = \lim_{i\to\infty} \Delta W_{f(i),\,\omega(i)}.$$

In the rest of this section, we will use the following abbreviations, which we introduced in § 23: $M^{(i)} = M'_{f(i),\,\omega(i)}$, $G = G_{f,\omega}$, $G_i = G_{f(i),\,\omega(i)}$, $h = h_f$, $h_i = h_{f(i)}$, and the further abbreviations: $\Delta W = \Delta W_{f,\omega}$, $\Delta W_i = \Delta W_{f(i),\,\omega(i)}$.

Lemma **(24.2)**. — *Under the hypotheses of Lemma* (24.1), *we have*

$$\Delta W \leq \liminf_{i\to\infty} \Delta W_i.$$

Proof. — By Proposition (19.1) and the definition of ΔW, we may choose a complementary interval J of $\pi_1 M$ so that

$$\Delta W = \Delta W_J.$$

From (24.1), one easily deduces

(24.2) $$\pi_1 M = \lim_{i\to\infty} \pi_1 M^{(i)}.$$

Since $\pi_1 M$ and $\pi_1 M^{(i)}$ are invariant under the translation $x \mapsto x + 1$, it follows from Lemma (22.2) that (24.2) may be interpreted as convergence in the Hausdorff metric. Consequently, we may choose complementary intervals $J^{(i)}$ to $\pi_1 M^{(i)}$ such that

$$J^{(i)} \to J, \quad \text{as } i \to \infty.$$

Let $x^0 = x^0(J)$, $x^1 = x^1(J)$, $x^{0(i)} = x^0(J^{(i)})$, $x^{1(i)} = x^1(J^{(i)})$. Obviously,

(24.3) $$x_j^{0(i)} \to x_j^0, \quad x_j^{1(i)} \to x_j^1, \quad \text{as } i \to \infty,$$

although convergence is not uniform in j.

Let $\delta > 0$. For each i, we may choose a continuous path $x^{t,(i)}$ in $\mathscr{X}_{J^{(i)}}$ connecting $x^{0(i)}$ and $x^{1(i)}$ such that

(24.4) $$G_i(x^{t,(i)}) \geq -\Delta W_{J^{(i)}} - \delta/10,$$

for all $t \in [0, 1]$. This is a consequence of the definition of minimax value (§ 3) and the definition of ΔW_J (§ 18).

Choose N so large that

(24.5) $$\sum_{|j|>N} (x_j^1 - x_j^0) < \delta/20C,$$

where C is given by (13.3). It is possible to find such an N by (13.1).

Since $\omega_i \to \omega$ and ω is irrational, there exists i_0 such that if $i \geq i_0$ and ω_i is rational, then

(24.6) $$q_i \geq 2(N + 100),$$

where $\omega_i = p_i/q_i$ in lowest terms.

Since $f_i \to f$ in the C^1 topology and f_i commutes with T, we obtain that the first and second partial derivatives of h_i converge uniformly to the corresponding partial derivatives of h. From (13.3), it follows that there exists $i_1 \geq i_0$ such that

(24.7) $$\sup_{B_i} \frac{\partial h_i(x, x')}{\partial x} < 2C, \quad \sup_{B_i} \frac{\partial h_i(x, x')}{\partial x'} < 2C,$$

for $i \geq i_1$, where B_i is the domain of h_i, i.e. $B_i = \{(x, x') \in \mathbf{R}^2 : f_{i0}(x) \leq x' \leq f_{i1}(x)\}$.

From now on, we fix $i \geq i_1$. In the case ω_i is irrational, we set

$$\begin{aligned} x^{t,(i),*} &= x_j^{t,(i)}, & |j| &\leq N + 2, \\ &= x_j^{0,(i)}, & |j| &> N + 2, \\ x_j^{t,(i),**} &= x_j^{0,(i)}, & |j| &\leq N + 2, \\ &= x_j^{t,(i)}, & |j| &> N + 2. \end{aligned}$$

We have $x^{t,(i),*}, x^{t,(i),**} \in \mathscr{X}_{J^{(i)}}$, provided δ is small enough, by (24.5) and the fact that $M^{(i)}/T$ is uniformly bounded away from the boundary of the annulus. In the case ω_i is rational, we write $\omega_i = p_i/q_i$ in lowest terms, and define $x_j^{t,(i),*}$ and $x_j^{t,(i),**}$ by the same formulas as above when

$$-\frac{q_i}{2} < j \leq \frac{q_i}{2},$$

and for other j define $x_j^{t,(i),*}$ and $x_j^{t,(i),**}$ so that $x_{j+q(i)}^{t,(i),*} = x_j^{t,(i),*} + p_i$ and $x_{j+q(i)}^{t,(i),**} = x_j^{t,(i),**} + p_i$. With this definition, we again have $x^{t,(i),*}, x^{t,(i),**} \in \mathscr{X}_{J(i)}$. Whether ω_i is rational or irrational, we have

(24.8)
$$\begin{aligned} &| G_i(x^{t,(i),*}) + G_i(x^{t,(i),**}) - G_i(x^{t,(i)}) - G_i(x^{0,(i)}) | \\ &= | \sum_{j=-N-3,\, N+2} \{ h_i(x_j^{t,(i)}, x_{j+1}^{0,(i)}) + h_i(x_j^{0,(i)}, x^{t,(i)}) \\ &\quad - h_i(x_j^{t,(i)}, x_{j+1}^{t,(i)}) - h_i(x_j^{0,(i)}, x_{j+1}^{0,(i)})\} | \leq 2C \sum_{j=-N-2,\, N+2} (x_j^{1,(i)} - x_j^{0,(i)}), \end{aligned}$$

by (24.7) and the Mean Value Theorem.

From (24.3) and (24.5), we get that there exists $i_2 \geq i_1$ such that

(24.9) $$x_j^{1,(i)} - x_j^{0,(i)} < \delta/20C,$$

when $|j| = N + 2$ and $i \geq i_2$. From now on, we suppose $i \geq i_2$.

Since $G_i(x^{0,(i)}) = 0$ and $G_i(x^{t,(i),**}) \leq 0$, inequalities (24.8) and (24.9) imply

$$G_i(x^{t,(i),*}) \geq G_i(x^{t,(i)}) - \delta/5.$$

Using (24.4), we get

(24.10) $$G_i(x^{t,(i),*}) \geq -\Delta W_{J(i)} - 3\,\delta/10.$$

Let $\delta_1 = \delta/80C(N + 3)$. By (24.3), we may choose $i_3 \geq i_2$ such that

$$|x_j^{\varepsilon,(i)} - x_j^{\varepsilon}| < \delta_1, \quad \varepsilon = 0, 1, \; |j| \leq N + 3,$$

for $i \geq i_3$. Furthermore, we may choose $i_4 \geq i_3$ such that

(24.11) $$\sup |h_i - h| < \delta/40(N + 3),$$

for $i \geq i_4$. From now on, we suppose $i \geq i_4$.

We may choose a continuous curve $t \mapsto x^{t,*}$ in $\mathscr{X}_J$ such that

$$\begin{aligned} |x_j^{t,(i),*} - x_j^{t,*}| &< \delta_1, \quad |j| \leq N + 3, \; t \in [0, 1] \\ x_j^{t,*} &= x_j^0, \quad |j| > N + 2, \\ x_j^{0,*} = x_j^0, \quad x_j^{1,*} &= x_j^1, \quad |j| \leq N + 2. \end{aligned}$$

Note that the curve $t \mapsto x^{t,*}$ depends on i. But, we may suppress i from the notation, because i is fixed throughout the discussion.

We have

$$\begin{aligned} G(x^{t,*}) &= \sum_{j=-\infty}^{\infty} (h(x_j^{t,*}, x_{j+1}^{t,*}) - h(x_j^0, x_{j+1}^0)) \\ &= \sum_{j=-N-3}^{N+2} (h(x_j^{t,*}, x_{j+1}^{t,*}) - h(x_j^0, x_{j+1}^0)), \end{aligned}$$

since $x_j^{t,*} = x_j^0$, for $|j| > N + 2$. Likewise,

$$G_i(x^{t,(i),*}) = \sum_{j=-N-3}^{N+2} (h_i(x_j^{t,(i),*}, x_{j+1}^{t,(i),*}) - h_i(x_j^{0,(i)}, x_{j+1}^{0,(i)})).$$

Hence,

$$\begin{aligned}|G(x^{t,*}) - G_i(x^{t,(i),*})| &\leq 4(N+3) \sup |h_i - h| \\ &+ \sum_{j=-N-3}^{N+2} |h(x_j^{t,(i),*}, x_{j+1}^{t,(i),*}) - h(x_j^{t,*}, x_{j+1}^{t,*})| \\ &+ \sum_{j=-N-3}^{N+2} |h(x_j^{0,(i)}, x_{j+1}^{0,(i)}) - h(x_j^0, x_{j+1}^0)| \\ &\leq 4(N+3) \sup |h_i - h| + 8C(N+3)\, \delta_1 \leq \delta/5.\end{aligned}$$

Combining this with (24.10), we get

(24.12) $$G(x^{t,*}) \geq -\Delta W_{J(i)} - \delta/2,$$

for all $t \in [0, 1]$.

We define a continuous curve $t \mapsto x^t$ in $\mathscr{X}_J$ by

$$\begin{aligned} x_j^t &= x_j^{t,*} & |j| \leq N+2, \\ &= (1-t)\, x_j^0 + t x_j^1 & |j| > N+2. \end{aligned}$$

By (24.5), (13.3), and the Mean Value Theorem, we have

$$|G(x^t) - G(x^{t,*})| < \delta/10.$$

From this and (24.12), we get

$$G(x^t) \geq -\Delta W_{J(i)} - \delta,$$

for all $t \in [0, 1]$. Obviously, $\Delta W_{J(i)} \leq \Delta W_i$.

We have, in this way, found a continuous curve x^t in $\mathscr{X}_J$, connecting x^0 and x^1, such that

$$G(x^t) \geq -\Delta W_i - \delta,$$

so $\Delta W \leq \Delta W_i + \delta$. Since $\delta > 0$ was arbitrary and $i \geq i_4$ was arbitrary, this proves Lemma (24.2). □

Lemma **(24.3)**. — *Under the hypotheses of Lemma* (24.1), *we have*

$$\Delta W \geq \limsup_{i \to \infty} \Delta W_i.$$

Proof. — By the hypotheses we assumed on f, we have that h is twice continuously differentiable on the interior of B, and its second partial derivatives extend continuously to the boundary of B. Let h_{11}, h_{12}, h_{22} denote the appropriate second partial derivatives of h. Let

$$C_1 = \max(|h_{11}| + 2\,|h_{12}| + |h_{22}|).$$

Note that the maximum of the function in the parenthesis actually exists, because $h(x, x') = h(x+1, x'+1)$ and the quotient of B by the equivalence relation generated by $(x, x') \sim (x+1, x'+1)$ is compact. Let $\delta > 0$ and let

$$\delta_1 = \delta/20 \max\{C, C_1\},$$

where C is given by (13.3).

Let $J_1, J_2, \ldots$ be a maximal ω-independent set of complementary intervals of $\pi_1 M$. (We defined the notion of ω-independent intervals in § 19.) Such a maximal ω-independent set may be finite (and possibly empty) or countably infinite. For each J_i in this set, and each integer j, we let $J_{ij} = [x^0(J_i)_j, x^1(J_i)_j]$. Obviously, the collection of all $\rho(J_{ij})$ is the set of all complementary intervals of $\rho\pi_1 M$ in $\mathbf{R}/\mathbf{Z}$. Since $\rho(J_{ij}) \cap \rho(J_{i'j'}) = \emptyset$, when $(i, j) \neq (i', j')$, we have

$$\Sigma |J_{ij}| \leq 1,$$

where $|J|$ denotes the length of J. Hence we may find a non-negative integer n and positive integer N such that

$$\textbf{(24.13)} \qquad \Sigma\{|J_{ij}| : i > n \text{ or } |j| > N\} < \delta_1.$$

Of course, n may be taken no larger than the number of complementary intervals in a maximal ω-independent set. For example, if there are no such complementary intervals (i.e. M is a curve), then we may, and will, take $n = 0$.

Let $\delta_2 = \delta/40(2N+2)\,C$.

Consider a positive integer i, let J' be a complementary interval of $\pi_1 M^{(i)}$, and let $J'_j = [x^0(J')_j, x^1(J')_j]$. Since (24.1) and (24.2) are valid for the Hausdorff metric, we have that for i_0 large enough, if $i \geq i_0$ then there exist integers $a_1, \ldots, a_k$ with $a_{\alpha+1} > a_\alpha + 2N + 100$, such that:

1) For $\alpha = 1, \ldots, k$, there exist $b_\alpha, c_\alpha \in \mathbf{Z}$, $1 \leq b_\alpha \leq n$, such that

$$|x^\varepsilon(J'_j) - x^\varepsilon(J_j^{(\alpha)})| < \delta_2,$$

when $a_\alpha - 1 \leq j \leq a_\alpha + 2N + 1$, $\varepsilon = 0$ or 1, where we set

$$J_j^{(\alpha)} = (J_{b(\alpha)})_{j - a(\alpha) - N} + c_\alpha.$$

2) $b_\alpha \neq b_{\alpha'}$, if $\alpha \neq \alpha'$.

3) Suppose either: *a)* ω_i is irrational and j does not satisfy $a_\alpha \leq j \leq a_\alpha + 2N$, for any $\alpha = 1, \ldots, k$, or *b)* ω_i is rational, $\omega_i = p_i/q_i$ in lowest terms and j does not satisfy $a_\alpha \leq j + q_i \ell \leq a_\alpha + 2N$, for any $\ell \in \mathbf{Z}$ and any $\alpha = 1, \ldots, k$. In either case,

$$|J'_j| < \delta_1.$$

In other words, since $\pi_1 M^{(i)}$ differs by very little from $\pi_1 M$ in the Hausdorff metric, we have that each J'_j either has length $< \delta_1$ or $\rho(J'_j)$ differs by very lit'le from one of the $\rho(J_{i\ell})$ for $1 \leq i \leq n$, and $|\ell| \leq N$. This is because there exists $\eta > 0$ such that $\rho(J_{i\ell})$ has length $< \delta_1 - \eta$ when $i > n$ or $|\ell| < N$, by (24.13). Moreover, when $\rho(J'_j)$ differs by very little from one of the $\rho(J_{i\ell})$ then $\rho(J'_{j+j'})$ differs by very little from $\rho(J'_{i,\ell+j'})$ when $|\ell + j'| \leq N + 1$, since $M^{(i)}$ differs by very little from M in the Hausdorff metric, and f_i is uniformly close to f.

From now on, we suppose $i \geq i_0$.

Notation. — We set $J^{(\alpha)} = J_0^{(\alpha)}$,

$$x_j^{\varepsilon(\alpha)} = x^\varepsilon(J_j^{(\alpha)}) = x^\varepsilon(J^{(\alpha)})_j, \quad \varepsilon = 0, 1,$$
$$x^{\varepsilon(\alpha)} = x^\varepsilon(J^{(\alpha)}),$$
$$x_j^{\varepsilon'} = x^\varepsilon(J'_j) = x^\varepsilon(J')_j, \quad \varepsilon = 0, 1,$$
$$x^{\varepsilon'} = x^\varepsilon(J').$$

For each $\alpha = 1, \ldots, k$, we choose a continuous curve γ_α in $\mathscr{X}_{J^{(\alpha)}}$ connecting $x^{0(\alpha)}$ and $x^{1(\alpha)}$ such that

$$\textbf{(24.14)} \qquad G\gamma_\alpha(t) \geq -\Delta W_{J^{(\alpha)}} - \delta/10$$
$$\geq -\Delta W - \delta/10,$$

for all $t \in [0, 1]$. We may do this by the definition of ΔW_J (§ 18) and the definition of minimax value (§ 3). We define $\gamma'_\alpha : [0, 1] \to \mathscr{X}_{J^{(\alpha)}}$ by

$$\gamma'_\alpha(t)_j = \gamma_\alpha(t)_j, \quad \text{if } |j - a_\alpha - N| \leq N,$$
$$= x_j^{0(\alpha)}, \quad \text{if } j - a_\alpha - N > N,$$
$$= x_j^{1(\alpha)}, \quad \text{if } j - a_\alpha - N < -N.$$

By definition of G,

$$G\gamma'_\alpha(t) - G\gamma_\alpha(t) = \sum_{j=-\infty}^{\infty} h(\gamma'_\alpha(t)_j, \gamma'_\alpha(t)_{j+1}) - h(\gamma_\alpha(t)_j, \gamma_\alpha(t)_{j+1}).$$

From (24.13) and the definition of γ'_σ, we obtain

$$\sum_{j=-\infty}^{\infty} |\gamma'_\alpha(t)_j - \gamma_\alpha(t)_j| = \sum_{|j-a(\alpha)-N|>N} |\gamma'_\alpha(t) - \gamma_\alpha(t)| < \delta_1.$$

From (13.3), the Mean Value Theorem, and the definition of δ_1, we then obtain

$$|G\gamma'_\alpha(t) - G\gamma_\alpha(t)| < 2C\,\delta_1 < \delta/10,$$

for all $t \in [0, 1]$. From this and (24.14), we obtain

$$\textbf{(24.15)} \qquad G\gamma'_\alpha(t) \geq -\Delta W - \delta/5,$$

for all $t \in [0, 1]$.

If ω_i is rational, we set $\omega_i = p_i/q_i$, in lowest terms, $q_i > 0$. If ω_i is irrational, we set $q_i = +\infty$. We choose $i_1 \geq i_0$ such that if $i \geq i_1$, then

$$\textbf{(24.16)} \qquad -q_i/2 < a_1 - 100 < a_k + 2N + 100 < q_i/2.$$

It is possible to choose such an i_1 since ω is irrational and $\omega_i \to \omega$ as $i \to \infty$. From now on, we will suppose that $i \geq i_1$.

There exists a continuous curve $\gamma''_\alpha : [0, 1] \to \mathscr{X}_{J'}$, such that

$$(24.17) \qquad |\gamma''_\alpha(t)_j - \gamma'_\alpha(t)_j| < \delta_2,$$

for $a_\alpha - 1 \leq j \leq a_\alpha + 2N + 1$,

$$\gamma''_\alpha(\varepsilon)_j = x^{\varepsilon'}_j, \quad \varepsilon = 0, 1,$$

for $a_\alpha \leq j \leq a_\alpha + 2N$,

$$\gamma''_\alpha(t)_j = x^{0'}_j, \quad \text{if } q_i/2 \geq j > a_\alpha + 2N,$$

and $$\gamma''_\alpha(t)_j = x^{1'}_j, \quad \text{if } -q_i/2 < j < a_\alpha.$$

This is because J'_j is very close to $J^{(\alpha)}_j$ in the appropriate range (condition 1 in the definition of $a_1, \ldots, a_k$).

Note that J', $J^{(\alpha)}$, γ'_α, γ''_α, etc., depend on i. However, we may suppress i from the notation, because we keep it fixed throughout the discussion.

Choose $i_2 \geq i_1$, so that if $i \geq i_2$, then (24.7) holds. From now on, we suppose $i \geq i_2$. To continue the proof of Lemma (24.3), we will need the following result:

Sublemma 1. — *Let* $A \in \{-\infty\} \cup \mathbf{Z}$ *and* $B \in \mathbf{Z} \cup \{+\infty\}$, *and suppose* $A < B$. *Suppose* $\max\{|J'_A|, |J'_B|\} < \eta$. *Then*

$$\Big|\sum_{j=A}^{B-1} h_i(x^{1'}_j, x^{1'}_{j+1}) - h_i(x^{0'}_j, x^{0'}_{j+1})\Big| < 4C\eta,$$

provided η *is small enough.*

Proof. — In the case ω_i is rational, we have that the above sum vanishes when $B - A$ is an integral multiple of q_i. It follows that we may subtract $\sum_{A(0)}^{A(0)+q(i)\ell}$ from $\sum_{A}^{B-1}$, for any integer A_0 and any positive integer ℓ, without changing the value of the sum. Consequently, we may assume, without loss of generality, that $|B - A| \leq q_i/2$. We will assume this in the following argument.

Define $x^!$ by

$$x^!_j = x^{1'}_j \quad \text{if } A \leq j \leq B,$$
$$= x^{0'}_j, \quad \text{otherwise},$$

in the case ω_i is irrational, and

$$x^!_j = x^{1'}_j, \quad \text{if } \ell \in \mathbf{Z} \quad \text{such that } A \leq j + \ell q_i \leq B,$$
$$= x^{0'}_j, \quad \text{otherwise},$$

in the case $\omega_i = p_i/q_i$. Since M_i/T is uniformly bounded away from the boundary of the annulus, we have $x^! \in X_{J'}$, provided that η is small enough.

By Lemma (6.2), we have $G_i(x^!) \leq 0$. By the definition of $x^!$, this is the same as

$$\sum_{j=A}^{B-1} (h_i(x_j^{1'}, x_{j+1}^{1'}) - h_i(x_j^{0'}, x_{j+1}^{0'})) \leq h_i(x_{A-1}^{0'}, x_A^{0'}) - h_i(x_{A-1}^{0'}, x_A^{1'}) + h_i(x_B^{0'}, x_{B+1}^{0'}) - h_i(x_B^{1'}, x_{B+1}^{0'}).$$

By (24.7), the Mean Value Theorem, and the hypothesis of the sublemma, the right side of this inequality is $< 4C\eta$. Hence,

$$\sum_{j=A}^{B-1} h_i(x_j^{1'}, x_{j+1}^{1'}) - h_i(x_j^{0'}, x_{j+1}^{0'}) < 4C\eta.$$

Interchanging $x^{0'}$ and $x^{1'}$ in this argument (including the definition of $x^!$) and using the fact that $G_i(x^{1'}) = 0$ (by Lemma (6.2)), we obtain

$$\sum_{j=A}^{B-1} h_i(x_j^{0'}, x_{j+1}^{0'}) - h_i(x_j^{1'}, x_{j+1}^{1'}) < 4C\eta.$$

Combining these two inequalities gives the required result. □

Remarks. — In the case ω_i is irrational, if $A = -\infty$ or $B = +\infty$, the right side of the estimate may be improved to $2C\eta$. If J' is replaced by $J^{(\alpha)}$ and h_i by h, the estimate of Sublemma 1 still holds, with the right side half as large, because we may use (13.3) in place of (24.7).

Choose $i_3 \geq i_2$ so that if $i \geq i_3$, then (24.11) holds. From now on, we suppose $i \geq i_3$.

By the definition of G and γ'_α, we have

$$G\gamma'_\alpha(t) = \sum_{j=-\infty}^{a(\alpha)-2} [h(x_j^{1(\alpha)}, x_{j+1}^{1(\alpha)}) - h(x_j^{0(\alpha)}, x_{j+1}^{0(\alpha)})] + \sum_{j=a(\alpha)-1}^{a(\alpha)+2N} [h(\gamma'_\alpha(t)_j, \gamma'_\alpha(t)_{j+1}) - h(x_j^{0(\alpha)}, x_{j+1}^{0(\alpha)})].$$

By Sublemma 1 and the remarks following it,

$$\left|\sum_{j=-\infty}^{a(\alpha)-2} [h(x_j^{1(\alpha)}, x_{j+1}^{1(\alpha)}) - h(x_j^{0(\alpha)}, x_{j+1}^{0(\alpha)})]\right| \leq C\,|J_{a(\alpha)-1}^{(\alpha)}| < C\,\delta_1,$$

where the last inequality is a consequence of $J_{a(\alpha)-1}^{(\alpha)} = J_{b(\alpha),-N-1}$ and (24.13). By the definition of G_i and γ''_α, we have

$$G_i\,\gamma''_\alpha(t) = A + A' + A'',$$

where, setting $L = -\left[\dfrac{q_i+1}{2}\right]$, we define

$$A = h_i(x_L^{0'}, x_{L+1}^{1'}) - h_i(x_L^{0'}, x_{L+1}^{0'}), \quad \text{if } q_i < \infty,$$
$$= 0, \quad \text{if } q_i = \infty,$$
$$A' = \sum_{j=L+1}^{a(\alpha)-2} (h_i(x_j^{1'}, x_{j+1}^{1'}) - h_i(x_j^{0'}, x_{j+1}^{0'})),$$
$$A'' = \sum_{j=a(\alpha)-1}^{a(\alpha)+2N} (h_i(\gamma''_\alpha(t)_j, \gamma''_\alpha(t)_{j+1}) - h_i(x_j^{0'}, x_{j+1}^{0'})).$$

Then $|A| \leq 2C|J'_{L+1}|$, by (24.7) and the Mean Value Theorem, and $|A'| \leq 4C\eta$, where $\eta = \max\{|J'_{L+1}|, |J'_{a(\alpha)-1}|\}$, by Sublemma 1. By condition 3 in the definition of $a_1, \ldots, a_k$ and (24.16), $\eta \leq \delta_1$. Hence

$$|A| + |A'| \leq 6C\,\delta_1.$$

Consequently,

$$\begin{aligned}|G_i\gamma''_\alpha(t) - G\gamma'_\alpha(t)| &\leq 8C\,\delta_1 + 2(2N+2)\sup|h - h_i| \\ &+ \sum_{j=a(\alpha)-1}^{a(\alpha)+2N} |h(\gamma''_\alpha(t)_j, \gamma''_\alpha(t)_{j+1}) - h(\gamma'_\alpha(t)_j, \gamma'_\alpha(t)_{j+1})| \\ &+ \sum_{j=a(\alpha)-1}^{a(\alpha)+2N} |h(x_j^{0'}, x_{j+1}^{0'}) - h(x_j^{0(\alpha)}, x_{j+1}^{0(\alpha)})| \\ &\leq 8C\,\delta_1 + 2(2N+2)\sup|h - h_i| + 4(2N+2)\,C\,\delta_2 \leq 6\,\delta/10,\end{aligned}$$

where the second inequality is a consequence of (24.17), condition 1 in the definition of $a_1, \ldots, a_k$, (13.3), and the Mean Value Theorem, and the last inequality follows from (24.11) and the definitions of δ_1 and δ_2. Combining the inequality which we have just derived with (24.15), we obtain

(24.18) $$G_i\gamma''_\alpha(t) \geq -\Delta W - \delta,$$

for all $t \in [0, 1]$.

For $1 \leq \alpha \leq k+1$, we define $\gamma'''_\alpha : [0, 1] \to \mathscr{X}_{J'}$ by setting

$$\gamma'''_\alpha(t)_j = (1-t)\,x_j^{0'} + t x_j^{1'},$$

when any one of the following conditions holds:

1) $\alpha = 1$ and $-q_i/2 < j < a_1$, or
2) $1 < \alpha < k+1$ and $a_{\alpha-1} + 2N < j < a_\alpha$, or
3) $\alpha = k+1$ and $a_k + 2N < j \leq q_i/2$,

and by setting

$$\gamma'''_\alpha(t)_j = x_j^{1'}, \quad \text{if } \alpha > 1 \text{ and } -q_i/2 < j \leq a_{\alpha-1} + 2N,$$

$$\gamma'''_\alpha(t)_j = x_j^{0'}, \quad \text{if } \alpha < k+1 \text{ and } a_\alpha \leq j \leq q_i/2.$$

In order to complete the proof of Lemma (24.3), we need the following elementary result from calculus.

Sublemma 2. — Let u be a C^2 real valued function on an interval $[a, b]$. Then

$$|u((1-t)\,a + tb) - (1-t)\,u(a) - tu(b)| \leq \sup_{[a,b]} |u''|\,(b-a)^2/8,$$

everywhere on $[a, b]$.

Proof. — Let

$$v(t) = u((1-t)\,a + tb) - (1-t)\,u(a) - tu(b).$$

Obviously, $v(a) = v(b) = 0$. Let t_0 be a point in $[0, 1]$ where $|v(t)|$ takes its maximum. Then $v'(t_0) = 0$ and

$$|v(t_0)| = \left| \int_0^{t(0)} ds \int_s^{t(0)} dt v''(t) \right| \leq \sup |v''| \, t_0^2/2.$$

Similarly,

$$|v(t_0)| \leq \sup |v''| \, (1 - t_0)^2/2.$$

Since $v''(t) = (b-a)^2 u''((1-t)a + tb)$, and $t_0 \leq 1/2$ or $1 - t_0 \leq 1/2$, we get

$$|v(t_0)| \leq \sup |u''| \, (b-a)^2/8. \quad \square$$

Since $f_i \to f$ in the C^1 topology, it follows that $h_i \to h$ in the C^2 topology. From this and the definition of C_1, it follows that we may choose $i_4 \geq i_3$ such that if $i \geq i_4$ then

$$\textbf{(24.19)} \qquad |h_{i,11}| + 2\,|h_{i,12}| + |h_{i,22}| < 2C_1,$$

everywhere on B_i, where $h_{i,11}$, etc., denote the appropriate second partial derivatives of h_i. From now on, we suppose $i \geq i_4$. Applying Sublemma 2 to

$$G_i \gamma_\alpha'''(t) = \sum_j \left(h_i(\gamma_\alpha'''(t)_j, \gamma_\alpha'''(t)_{j+1}) - h_i(x_j^{1'}, x_{j+1}^{1'})\right),$$

we obtain

$$|G_i \gamma_\alpha'''(t) - (1-t)\, G_i \gamma_\alpha'''(0) - t G_i \gamma_\alpha'''(1)|$$
$$\leq \frac{1}{8} \sum_j \sup_{t \in [0,1]} \left| \frac{d^2}{dt^2} h_i(\gamma_\alpha'''(t)_j, \gamma_\alpha'''(t)_{j+1}) \right| \leq \frac{1}{4} \sum_j{}' C_1 \, |J_j'|^2,$$

where $\sum_j'$ denotes summation over the following set of j's:

$$\begin{aligned} &\{-q_i/2 < j < a_1\}, && \text{if } \alpha = 1, \\ &\{a_{\alpha-1} + 2N < j < a_\alpha\}, && \text{if } 1 < \alpha < k+1 \\ &\{a_\alpha < j \leq q_i/2\}, && \text{if } \alpha = k+1. \end{aligned}$$

The last inequality above is a consequence of the fact that the second derivative of $\gamma_\alpha'''(t)_j$ vanishes, by the definition of $\gamma_\alpha'''(t)_j$, of the fact that the first derivative of $\gamma_\alpha'''(t)_j$ equals $x_j^{1'} - x_j^{0'} = |J_j'|$, for j appearing in the summation, and vanishes for other j, and of (24.19). By condition 3 in the definition of $a_1, \ldots, a_k$, we have $|J_j'| < \delta_1$ for those j which appear in the sum $\sum_j'$. By (13.1) and its obvious analogue for rational ω_i, we have

$$\Sigma\{|J_j'| : -q_i/2 < j \leq q_i/2\} \leq 1.$$

Hence,

$$\sum_j{}' |J_j'|^2 \leq \delta_1 \sum_j{}' |J_j'| \leq \delta_1.$$

Together with our previous inequality and the definition of δ_1, this implies

(24.20) $$|G_i\gamma_\alpha'''(t) - (1-t)\,G_i\gamma_\alpha'''(0) - tG_i\gamma_\alpha'''(1)| \leq C_1\,\delta_1/4 < \delta/10.$$

Setting $L = -[(q_i+1)/2]$, we have $G_i\gamma_\alpha''(0) = 0$, and

$$G_i\gamma_\alpha'''(0) = A + A' + A'',$$

if $\alpha > 1$, where

$$\begin{aligned} A &= h_i(x_L^{0'}, x_{L+1}^{1'}) - h_i(x_L^{0'}, x_{L+1}^{0'}) \quad \text{if } q_i < \infty,\\ &= 0 \quad \text{if } q_i = \infty,\\ A' &= \sum_{j=L+1}^{a(\alpha-1)+2N} (h_i(x_j^{1'}, x_{j+1}^{1'}) - h_i(x_j^{0'}, x_{j+1}^{0'})),\\ A'' &= h_i(x_{a(\alpha-1)+2N}^{1'}, x_{a(\alpha-1)+2N+1}^{0'}) - h_i(x_{a(\alpha-1)+2N}^{1'}, x_{a(\alpha-1)+2N+1}^{1'}). \end{aligned}$$

Furthermore,

$$|A|, |A''| \leq 2C\,\delta_1,$$

by (24.7), the Mean Value Theorem, and condition 3 in the definition of $a_1, \ldots, a_k$, and

$$|A'| \leq 4C\,\delta_1,$$

by Sublemma 1 and condition 3 in the definition of $a_1, \ldots, a_k$. Hence

(24.21) $$|G_i\gamma_\alpha'''(0)| \leq 8C\,\delta_1 \leq 2\,\delta/5,$$

by the definition of δ_1. A similar argument shows that

(24.22) $$|G_i\gamma_\alpha'''(1)| \leq 2\,\delta/5.$$

From (24.20-22), we get

(24.23) $$|G_i\gamma_\alpha'''(t)| < \delta,$$

for all $t \in [0, 1]$.

We define $\gamma : [0, 1] \to \mathscr{X}_{J'}$ as the cocatenation of $\gamma_1''', \gamma_1'', \gamma_2''', \gamma_2'', \ldots, \gamma_k'', \gamma_{k+1}'''$, in that order. In other words,

$$\begin{aligned} \gamma(t) &= \gamma_\alpha'''((2k+1)\,t - 2\alpha + 2), \quad \frac{2\alpha-2}{2k+1} \leq t \leq \frac{2\alpha-1}{2k+1}\\ &= \gamma_\alpha''((2k+1)\,t - 2\alpha + 1), \quad \frac{2\alpha-1}{2k+1} \leq t \leq \frac{2\alpha}{2k+1}. \end{aligned}$$

It is obvious from the definitions that γ is well-defined and continuous, i.e. the two definitions of it given at the common endpoints of adjacent intervals are equal. Thus, γ is a continuous curve in $\mathscr{X}_{J'}$, connecting $x^0(J')$ and $x^1(J')$. By (24.18) and (24.23),

$$G_i\gamma(t) \geq -\Delta W - \delta,$$

for all $t \in [0, 1]$. Hence,

$$\Delta W_i \leq \Delta W + \delta.$$

Since δ is an arbitrary positive number, we thereby obtain the conclusion of Lemma (24.3). □

Proof of Lemma (24.1). — Immediate from Lemmas (24.2) and (24.3). □

Proof of Proposition (9.4). — From Lemma (24.1), by the discussion at the beginning of this section. □

25. The Peierls Energy Barrier

Our purpose in this section is to relate what we have done to something Aubry, Le Daeron, and André did in [7]. We will explain a version of a result of these authors, which makes sense in our context.

We suppose ω is an irrational number such that $\rho(f_0) < \omega < \rho(f_1)$. We will suppose that M_ω/T is a Cantor set, not a circle. Equivalently, $\pi_1 M_\omega \neq \mathbf{R}$. Let J be a complementary interval of $\pi_1 M_\omega$. For $x \in J$, we define

$$V(x) = \max\{G_\omega(\hat{x}) : \hat{x} \in X_J, \hat{x}_0 = x\}.$$

Here, X_J and G_ω are as defined in § 18. By Lemma (6.2), $V(x) \leq 0$.

Definition. — The Peierls energy barrier associated to J *is* $\max\{-V(x) : x \in J\}$.

Note that $V(x)$ is a continuous function of x, so $-V(x)$ actually takes a maximum in J. Obviously, the Peierls energy barrier associated to J is a lower bound for ΔW_J.

However, the Peierls energy barrier associated to $J_i = [x^0(J)_i, x^1(J_i)]$ depends on i, in contrast to $\Delta W_{J(i)}$, which is independent of i. In fact, we have $|J_i| \to 0$, as $i \to \pm\infty$. From this, it follows easily that the Peierls energy barrier associated to J_i tends to 0 as $i \to \pm\infty$. Moreover, one can produce examples where ΔW_J is greater than the maximum of the Peierls energy barrier associated to the J_i.

The definition of the Peierls energy barrier which we have given above is not the same as that which Aubry, Le Daeron, and André gave in [7]. However, it is equivalent to their definition, as may easily be shown by means of a suitable variant of the Fundamental Lemma of [7]. The difference between the two definitions is that in defining $V(x)$, the authors of [7] allow $\hat{x}$ to vary over a larger space than our X_J. However, their methods show that the maximum over this larger space is still achieved for $\hat{x} \in X_J$, so the different definitions arrive at the same result.

The analogue of our Proposition (5.2) which is proved in [7] may be stated in our terminology, as follows: $J \subset \pi_1(M'_\omega)$ if and only if the Peierls energy barrier associated to J vanishes. Aubry, Le Daeron, and André also have a result for rational ω (which we do not have).

BIBLIOGRAPHY

[1] AUBRY, S., The devil's staircase transformation in incommensurate lattices, in *The Riemann problem, complete integrability and arithmetic applications*, *Lecture Notes in Math.*, **925**, Berlin-New York, Springer, 1982, 221-245.

[2] AUBRY, S., *On the Dynamics of Structured Phase Transition, Lattice Locking and Ergodic Theory*, Preprint.

[3] AUBRY, S., *On Modulated Crystallographic Structures*, II : *Exact Results on the Classical Ground States of a One-dimensional Model*, Preprint, 1978.

[4] AUBRY, S., Many Defect Structures, Stochasticity and Incommensurability, in *Les Houches Session XXXV*, 1980, *Physics of Defects*, R. Baltan et al. eds., North-Holland Publ. Co., 1981.

[5] AUBRY, S., The New Concept of Transition of Breaking of Analyticity in a Crystallographic Model, in *Solitons in Condensed Matter*, A. Bishop and T. Schneider eds., Springer Series in Solid State Sci. 8, Berlin, Springer, 1978, 264-277.

[6] AUBRY, S. and ANDRÉ, G., Analyticity Breaking and Anderson Localization in Incommensurate Lattices, *Annals of Israel Physical Society*, Vol. 3, Group Theoretical Methods in Physics, L. P. Horowitz and Y. Ne'eman eds., 1980, 133-164.

[7] AUBRY, S., LE DAERON, P. Y. and ANDRÉ, G., *Classical Ground-States of a One-dimensional Model for Incommensurate Structures*, Preprint. (A revised version of the first part of this preprint was published as Aubry, S. and Le Daeron, P. Y., The Discrete Frenkel-Kontorova Model and its Extensions, *Physica*, 8D (1983), 381-422. However this does not include the material referred to in §25.)

[8] AXEL, F. and AUBRY, S., Spatially modulated phases of a one-dimensional lattice model with competing interactions, *J. Phys. C. : Solid State Physics*, **14** (1981), 5433-5451.

[9] BIRKHOFF, G. D., On the Periodic Notions of Dynamical Systems, *Acta Math.*, **50** (1928), 359-379. Reprinted in *Collected Mathematical Papers*, Vol. II, *Amer. Math. Soc.*, New York, 1950, 333-353.

[10] BIRKHOFF, G. D., Surface Transformations and their Dynamical Applications, *Acta Math.*, **43** (1922), 1-119. Reprinted in *Collected Mathematical Papers*, Vol. II, *Amer. Math. Soc.*, New York, 1950, 111-229.

[11] BIRKHOFF, G. D., Sur quelques courbes fermées remarquables, *Bull. Soc. Math. de France*, **60** (1932), 1-26. Reprinted in *Collected Mathematical Papers*, Vol. II, *Amer. Math. Soc.*, New York, 1950, 418-443.

[12] BOURBAKI, N., *Topologie générale*, Chapitres 1 à 4, Paris, Hermann, 1971.

[13] BRISAC, R., Sur les fonctions multiformes, *C. R. Acad. Sci.*, Paris, **244** (1947), 92-94.

[14] BRUHAT, F. et CARTAN, H., Sur la structure des sous-ensembles analytiques réels, *C. R. Acad. Sci.*, Paris, **244** (1957), 988-990.

[15] CARTAN, H., Théorie des filtres; Filtres et ultrafiltres, *C. R. Acad. Sci.*, Paris, **205** (1937), 595-598 et 777-779.

[16] HERMAN, M. R., Contre-exemples de classe $C^{3-\varepsilon}$ et à nombre de rotation fixé au théorème des courbes invariantes, handwritten notes from lectures given in Brazil (1979), published in *Astérisque*, **103-104** (1983).

[17] KATOK, A., Some remarks on Birkhoff and Mather twist map theorems, to appear in *Ergodic theory and dynamical systems*.

[18] KELLEY, J. L., *General Topology*, New York, Van Nostrand, 1955.

[19] LOJASIEWICZ, S., Sur les ensembles semi-analytiques, *International Congress of Mathematicians*, Vol. II, Nice (1970), Paris, Gauthier-Villars, 1971, 237-241.

[20] LOJASIEWICZ, S., *Ensembles semi-analytiques*, Preprint, Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France, 1964.

[21] MATHER, J. N., Existence of Quasi-Periodic Orbits for Twist Homeomorphisms of the Annulus, *Topology*, **21** (1982), 457-467.

[22] MATHER, J. N., Concavity of the Lagrangian for Quasi-Periodic Orbits, *Comment. Math. Helv.*, **57** (1982), 356-376.

[23] MATHER, J. N., Glancing Billiards, *Ergodic theory and dynamical systems*, **2** (1982), 397-403.

[24] MATHER, J. N., Non-existence of invariant circles, *Ergodic theory and dynamical systems*, **4** (1984), 301-309.

[25] MICHAEL, E., Topologies on spaces of subsets, *Trans. Amer. Math. Soc.*, **71** (1951), 151-182.

[26] NEWMAN, R. P. A. C. and Percival, I. C., Definite Paths and Upper Bounds on Regular Regions of Velocity Phase Space, *Physica*, D6 (1982-1983), 249-259.

[27] Percival, I. C., Variational Principles for Invariant Tori and Cantori, in *Symposium on Nonlinear Dynamics and Beam-Beam Interactions*, *Amer. Inst. of Physics*, *Conf. Proc.*, No. 57, M. Month and J. C. Herrara eds., 1980, 302-310.

[28] Percival, I. C., A rational principal for invariant tori of fixed frequency, *J. Phys. A.*, **12** (1979), 157-160.

[29] Sinai, Ya. G., *Introduction to Ergodic Theory*, *Math. Notes*, Princeton Univ. Press, 1977 (translated from Russian by V. Scheffer).

[30] Titchmarsh, E. C., *The Theory of Functions*, Clarendon Press, Oxford, 1932.

Department of Mathematics
Princeton University
Princeton, New Jersey 08544

Manuscrit reçu le 25 août 1985.