SIAM REVIEW
Vol. 28, No. 4, December 1986

RECENT DEVELOPMENTS IN THE THEORY OF HAMILTONIAN SYSTEMS*

JÜRGEN MOSER†

Abstract. Area-preserving mappings of an annulus occur as Poincaré mappings of Hamiltonian systems; they were studied extensively by G. D. Birkhoff. Recently Aubry and Mather investigated the subclass of so-called monotone twist mappings for which they constructed independently closed invariant Cantor sets. Their work led to important new results. Their theory is discussed and related to Hamiltonian systems satisfying a Legendre condition. The connection of this theory with the stability problem, KAM theory, and in particular, the disintegration of invariant tori is discussed. Various constructions of Mather sets are explained, in which minimal solutions of variational problems play a central role. This theory has a close relation to the differential geometric investigations by Morse and Hedlund on geodesics on two-dimensional surfaces.

Key words. stability theory, theory of Aubry and Mather, area-preserving twist mapping, breakdown of stability

1. Introduction. In the last twenty years one of the most fascinating developments in the theory of Hamiltonian systems is connected with the discovery of the new integrable systems, like the Toda lattice, the Korteweg–de Vries equations and various other systems of infinite degrees of freedom. This subject grew very rapidly. Although it originated from applied problems, it has in the meantime spread to a variety of other more abstract fields such as Lie algebras and algebraic geometry. The topic of integrable systems has already been described by many authors and will not be discussed here. Moreover, integrable sytems are highly exceptional among all Hamiltonian systems and therefore are less relevant for applications, where perturbations of integrable systems are more frequently encountered.

In this lecture, however, I should like to draw attention to a *different* interesting new development in this field which is connected with the phenomenon of "breakdown of stability." I am referring to the works by S. Aubry and J. Mather which were begun independently and with different motivations but led to similar results by different methods. While Aubry investigated certain models of solid state physics related to dislocations in one-dimensional crystals Mather studied area-preserving annulus mappings, as they occur as section mappings for Hamiltonian systems of two degrees of freedom. Such annulus mappings play a role in the stability theory. They occurred already in Poincaré's work on the restricted three body problem.

Before delving into the subject let me try to sketch the larger picture of the stability problem for Hamiltonian systems. Speaking quite intuitively, we are familiar with systems with very unstable, even ergodic, behavior, as it is demanded by statistical mechanics. Orbits may come arbitrarily close to any state on a given energy surface for these unstable systems (e.g. Anosov systems). In fact, it has been the common belief that these unstable systems constitute the "general case." On the other hand, there are other systems exhibiting clearly stable behavior, as, for example, the planetary motion of the solar system. The problem is to decide which systems have stable and which unstable behavior.

At this stage of the knowledge one can establish stability of a Hamiltonian system only if a system is sufficiently close to so-called *integrable systems*—these are systems

*Received by the editors September 30, 1985; accepted for publication February 10, 1986. This paper is an expanded version of the 1984 John von Neumann Lecture presented at the SIAM 1984 National Meeting in Seattle, Washington, July 16–20, 1984.

† Forschungsinstitut für Mathematik, ETH Zentrum, Zürich, Switzerland.

for which so many integrals are known that stability is evident. Moreover small perturbations of such systems do not affect the stability behavior too much. This is the meaning of the KAM theory, which by now is over 20 years old. But as the system recedes further from the exceptional integrable system stability will deteriorate and ultimately get lost. This is the expected phenomenon which one would like to understand. For example, the planetary system has the special feature that the masses of the planets are small compared to that of the sun, so that the forces between the planets are much smaller than those between the planets to the sun. If one neglects the former, one has the Kepler approximation, in which all planets move on ellipses. This is the integrable approximation which is evidently stable, and this system will retain its stability for small masses of the planets. If we enlarge the masses of the planets sufficiently, we would expect the system to become unstable.

More generally one studies small perturbations $H = H_0 + \varepsilon P$ of an integrable Hamiltonian H_0. For such an integrable system of n degrees of freedom one has n independent integrals in involution whose level sets are n-dimensional tori. In this case stability is obvious. After small perturbation of such a system, i.e., for small ε, the system corresponding to the Hamiltonian $H = H_0 + \varepsilon P$ still possesses a large set of invariant tori. This is the content of the so-called KAM theory. In other words, for most initial points in phase space the orbits lie on such tori and exhibit stable behavior. For $n \geqq 3$ the exceptional orbits not lying on such tori may leak out and escape slowly. For $n = 2$ such slow escape is not possible since the exceptional orbits on a three-dimensional energy surface are trapped between the two-dimensional tori. Nevertheless, the set of these tori generally forms a rather complicated Cantor set, even for simple and smooth Hamiltonians.

We will restrict attention to the case of two degrees of freedom since the theory to be discussed is applicable only if $n = 2$. Although for *small* ε we have stable behavior one knows from numerical experiments and from experience that the system will, in general, become unstable as the perturbation gets *large*. This phenomenon leads to many difficult and important questions. How does this breakdown occur? How do the invariant curves disintegrate? Can one give realistic quantitative estimates for the size of the perturbation still giving rise to stability?

To understand this phenomenon of "breakdown of stability," one turns to *simpler* models. In fact, the simplest systems are area-preserving mappings which have been studied thoroughly over the years. After Poincaré it was mainly G. D. Birkhoff who, around 1920 [3], studied periodic orbits of these mappings and also so-called discontinuous invariant sets. In recent years there has been a renewed interest in iteration of mappings, partly because these problems lend themselves to *numerical experiments*. Many studies deal with iteration of *conformal mappings*, such as iteration of rational functions on the complex sphere. In this theory one has many tools available from complex analysis lacking in the theory of area-preserving mappings. But the latter have more importance for a variety of applications, and they will be the subject of this paper.

We begin with a simple nonlinear area-preserving map φ_0 of the u-v-plane taking (u, v) into (u_1, v_1) where

$$u_1 = u\cos\psi - v\sin\psi, \qquad v_1 = u\sin\psi + v\cos\psi$$

where $\psi = \psi(u^2 + v^2)$, e.g., $\psi = \alpha + \beta(u^2 + v^2)$. This mapping with the fixed point $(u, v) = (0, 0)$ preserves concentric circles and rotates them by an angle which depends on the radius. This "twist mapping" corresponds to the integrable case. It occurs in the

stability theory of elliptic periodic orbits of Hamiltonian systems. More precisely, one is led to consider area-preserving mappings φ_ε close to φ_0 and one asks for the stability of its fixed point, which can be taken at the origin again. It turns out that under the twist condition $\psi'(0)\neq 0$ or $\beta\neq 0$ one can establish the existence of a sequence of invariant curves surrounding the fixed point and converging to it, which implies its stability [34].

In this situation our questions amount to finding an estimate for the region of stability and to a description of how invariant curves can disintegrate when the parameter ε increases beyond a critical value.

In the following I shall try to give a rough description of the results of Aubry and Mather which throw some light on the above questions, even though they do not solve them. It is not the purpose of this lecture to give a complete exposition or even proofs; such expository accounts are going to appear soon (see [2], [20]). Instead we will try to show the results in their different guises, their connection with mechanics, calculus of variations and differential geometry. In particular, we wish to emphasize some of the underlying principles of this theory which allow to single out a subset of solutions by a variational principle (and not by initial conditions). The set of these solutions constitutes a closed invariant set, including the set of invariant curves as well as the so-called "Mather sets." In §§2 and 3 we will introduce the monotone twist mappings and describe the invariant Mather sets and in §4 interpret them as "generalized" invariant curves. In §§5–6 we show their relations to variational problems on a two-dimensional torus and to mechanics. It turns out that the monotone twist property is related to the Legendre condition of calculus of variations, and the invariant curves to extremal fields on the torus. In §7 we show how such extremal fields, or Mather sets can be obtained from minimal solutions, following and generalizing the ideas of Aubry and Le Daeron [1]. Actually the proof of the results of this section have not yet appeared in the form given, but they are contained in [29] where a much more general problem concerning elliptic partial differential equation is treated. Finally, in §8 we recall some earlier results of G. Hedlund [16] who studied geodesics of class A on a two-dimensional torus. His work was based on a paper by M. Morse [27] on such geodesics on two-dimensional manifolds of genus $g\geqq 2$.

2. Monotone twist mappings and stability.

2.1. Area-preserving annulus mapping. We begin with an area-preserving mapping φ of an annulus

$$A' : a' \leqq u^2+v^2 \leqq a''$$

onto itself. Analytically it is given by the formulae $\varphi:(u,v)\to(u_1,v_1)=(F(u,v),\, G(u,v))$ with Jacobian equal to 1:

$$\frac{\partial(F,G)}{\partial(u,v)}=1.$$

We will always assume that F, G are smooth functions although this is not necessary for most of the following. Moreover, φ should be invertible and preserve the boundaries. We will allow the case $a'=0$ as well as $a''=\infty$.

We mentioned already that such mappings occur in celestial mechanics as Poincaré maps (Fig. 1). However, they also are familiar to physicists studying orbits in accelerator theory. They are also basics for Aubry's work on simplified models describing one-dimensional crystals. In fact, it is amazing how frequently such mappings occur, and that very simple mappings can give rise to complicated behavior of points under iteration of φ.

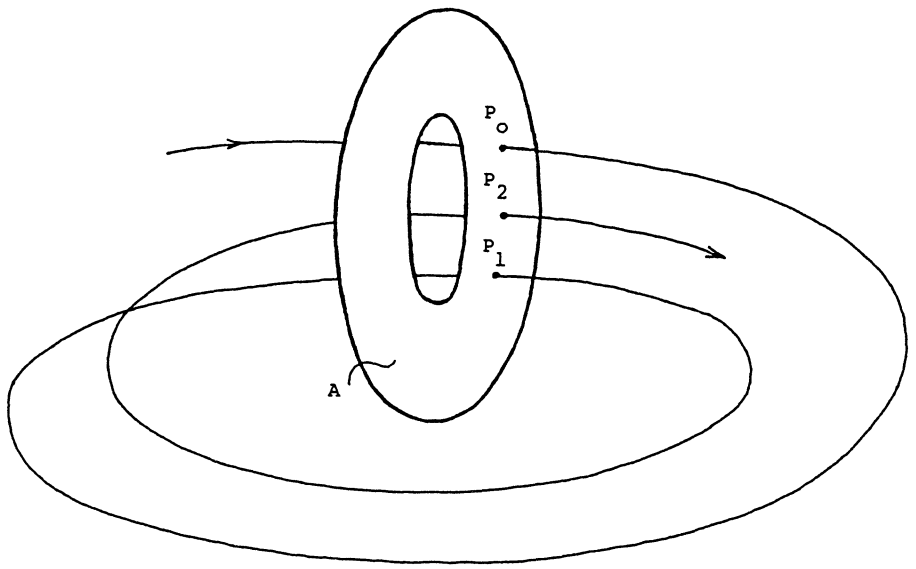

FIG. 1. *Poincaré map.*

Using "polar coordinates" x, y, defined by

$$u=\sqrt{y}\cos 2\pi x, \qquad v=\sqrt{y}\sin 2\pi x,$$

the annulus A' is taken into the strip

$$A: a' \leqq y \leqq a'' \tag{2.1}$$

where x is an angular variable mod 1. Since

$$du \wedge dv = -\pi\, dx \wedge dy$$

the transformed mapping, denoted by φ again, is again area-preserving. We write it in the form

$$x_1 = f(x,y), \qquad y_1 = g(x,y) \tag{2.2}$$

where

$$\begin{aligned} &f(x+1,y)=f(x,y)+1, \qquad g(x+1,y)=g(x,y), \\ &g(x,y)-y=0 \quad \text{for } y=a',a''. \end{aligned} \tag{2.3}$$

In this formulation we will also allow $a'=-\infty$ or $a''=+\infty$ or both.

For such a mapping φ we can define the iterates φ^j for $j=1,2,\cdots$, and since φ is invertible with inverse φ^{-1}, also for $j=-1,-2,\cdots$, φ^0 is defined as the identity. A sequence of points $P_j=\varphi^j(P)$ for $P\in A$ is called an orbit, where j plays the role of the discrete time.

2.2. Stability. We will call φ stable in A if the two boundaries of A have neighborhoods such that no orbit starting in one neighborhood reaches the other (Fig. 2). In the other case, when φ is not stable in A, one sometimes calls the annulus A, following G. D. Birkhoff, a "ring of instability." This concept is inspired by the stability of a fixed point, which would constitute one boundary of the annulus $0=a'\leqq y \leqq a''$. Of course, if there exists an invariant closed curve

$$\Gamma: y=w(x)=w(x+1) \tag{2.4}$$

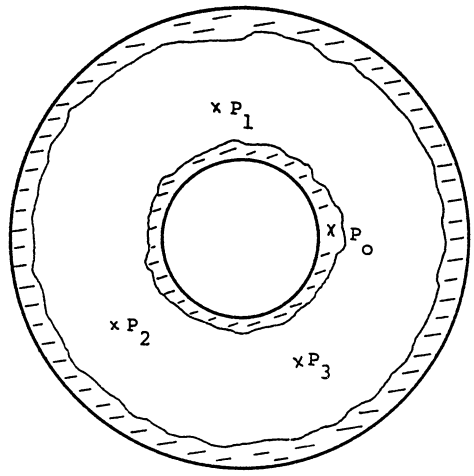

FIG. 2. *Stability of an annulus map.*

encircling the inner boundary such that φ maps Γ into itself, then φ is obviously stable. The converse of this statement holds for a special class of mappings which we discuss now.

2.3. Monotone twist mappings. An area preserving mapping (2.1) is called a monotone twist mapping if

$$\frac{\partial f}{\partial y} \neq 0$$

in A. We may assume that

$$\frac{\partial f}{\partial y} > 0 \tag{2.5}$$

since in the other case we consider φ^{-1}. Geometrically this means that a radial vector is turned to the right under the mapping (Fig. 3).

It is this class of mappings to which the theory to be discussed applies. For example the twist mapping of the introduction takes the form

$$(x,y) \to \left(x + \frac{1}{2\pi}\psi(y), y\right)$$

and the condition (2.5) amounts to $\psi'(y) > 0$. The mapping twists the annulus the more to the right the bigger the radius.

For monotone twist mappings one has the following interesting theorem.

THEOREM (G. D. Birkhoff [3]).[1] *If φ is a monotone twist mapping which is stable in A then there exists a closed invariant curve of the form* (2.4) *separating the boundary components. Moreover, the function $w(x)$ is Lipschitz continuous.*

This shows the close relation between stability and invariant curves. We mention a recent sharper result by J. Mather [20]: He showed that in a ring of instability A there exists an orbit $\varphi^j(P)$ which for $j \to +\infty$ approaches one boundary component and for $j \to -\infty$ the other.

[1]See also reference [35].

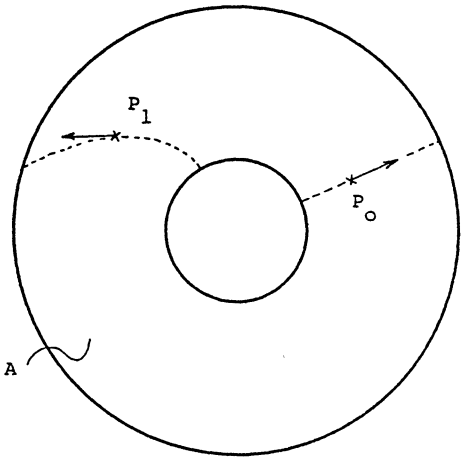

FIG. 3. *Monotone twist map.*

2.4. Rotation number. At this point we recall some well-known results by Denjoy [8] about homeomorphisms of a circle into itself. If Γ is a closed curve invariant under φ, then the restricted map $\varphi|_\Gamma$ is an invertible circle mapping. With such a mapping one can associate a rotation number

$$\lim_{j\to\infty} \frac{x_j}{j} = \alpha = \alpha(\Gamma)$$

where $\varphi^j(x,y) = (x_j, y_j)$. This limit is independent of the initial point (x,y) on Γ and represents an average rotation angle. In particular, the boundary circles $y = a', a''$ are such invariant curves and we will denote their rotation number by α', α'', respectively. For any invariant curve Γ in A one has

$$\alpha' \leqq \alpha(\Gamma) \leqq \alpha''.$$

In the next section we will consider the question whether for any $\alpha \in (\alpha', \alpha'')$ one can find an invariant curve Γ for which $\alpha(\Gamma) = \alpha$.

Denjoy proved the remarkable fact that if $\varphi|_\Gamma$ is twice continuously differentiable and α irrational then every orbit is dense on Γ. On the other hand he constructed once continuously differentiable circle maps with irrational α for which none of its orbits are dense on Γ but whose limit set L is a Cantor set which is independent of the choice of the orbit. One sometimes speaks of a "Denjoy minimal set" in this case. This fact will be of great importance for the following; it may be surprising that such subtleties are significant. Since we assume φ to be smooth, we will have dense orbits in Γ if Γ is in C^2 and the limit set L can be different from Γ only if Γ loses smoothness! This fact is important for understanding how an invariant curve can disintegrate.

2.5. Generating functions. A monotone twist mapping can be represented in terms of a Legendre generating function. Since $f_y > 0$ one can in (2.2) express y and y_1 as functions of x and x_1. On account of the area-preserving character there exists a function $h = h(x_0, x_1)$ such that

$$y_0 = -\frac{\partial h}{\partial x_0}, \qquad y_1 = \frac{\partial h}{\partial x_1} \tag{2.6}$$

where we replaced x, y by x_0, y_0. Here $h(x_0+s, x_1+s)$ has period 1 in s and the twist condition (2.5) translates into

$$\frac{\partial^2 h}{\partial x_0 \partial x_1} < 0.$$

The domain of definition of h has the more complicated form

$$f(x_0, a') \leqq x_1 \leqq f(x_0, a''),$$

which by the periodicity of $f(x,y)-x$ corresponds to an annulus again.

2.6. Examples. We look at some examples which have been studied numerically. For any two functions f, g where $f(y)-y$, $g(x)$ have period 1, the mapping

$$x_1 = x + f(y_1), \qquad y_1 = y + g(x), \tag{2.7}$$

is area-perserving. These mappings can be viewed on the torus since, also, y can be identified mod 1. They are monotone twist mappings if $f'(y) > 0$. The so-called standard mapping (Chirikov, Greene) is given by

$$f(y) = y, \qquad g(x) = \frac{\lambda}{2\pi} \sin 2\pi x. \tag{2.8}$$

An example studied by F. Rannou [33] numerically is given by

$$\begin{aligned} f(y) &= y + \frac{1}{2\pi}(1 - \cos 2\pi y), \\ g(x) &= \frac{\lambda}{2\pi}(1 + \sin 2\pi x - \cos 2\pi x). \end{aligned} \tag{2.9}$$

The twist condition holds only in the weak sense $f' \geqq 0$ here.

These mappings are obtained by composing the "shear" mappings

$$\psi_1 : (x,y) \to (x + f(y), y), \qquad \psi_2 : (x,y) \to (x, y + g(x)).$$

Actually Rannou's example is given by $\psi_2 \circ \psi_1$ while (2.7) is of the form $\psi_1 \circ \psi_2$, but both mappings are equivalent.

These mappings (2.8), (2.9) have for $\lambda = 0$ all circles $y = \text{const}$ invariant. Also for small values of $|\lambda|$ these mappings possess many invariant curves (2.4). However, if λ is sufficiently large, no such curve exists. For example (2.8) Mather showed that this certainly holds for $\lambda > \frac{4}{3}$; hence for $\lambda > \frac{4}{3}$ there are orbits for which the y_j are unbounded. These estimates were improved by Percival and MacKay. Thus for small $\lambda > 0$ one has boundedness of the y_j for all orbits and for $\lambda > \lambda^*$ the y_j are unbounded for some orbits. This illustrates the breakdown of stability. Of course, the precise ranges of the λ in the two cases are not known.

Greene investigated invariant curves for a rotation number $\alpha = \frac{1}{2}(\sqrt{5}+1)$, which is very badly approximable by rationals. Therefore the corresponding curves are expected to survive for larger values of λ than those with other rotation numbers. He gave numerical evidence for such curves to exist for

$$0 \leqq \lambda < \lambda^* = 0.971635 \cdots$$

but finds breakup for $\lambda > \lambda^*$ [13] (Fig. 4).

The pictures of Rannou's calculation for (2.9) are particularly striking. They illustrate the breakdown of stability in a most dramatic way (Figs. 5–7).

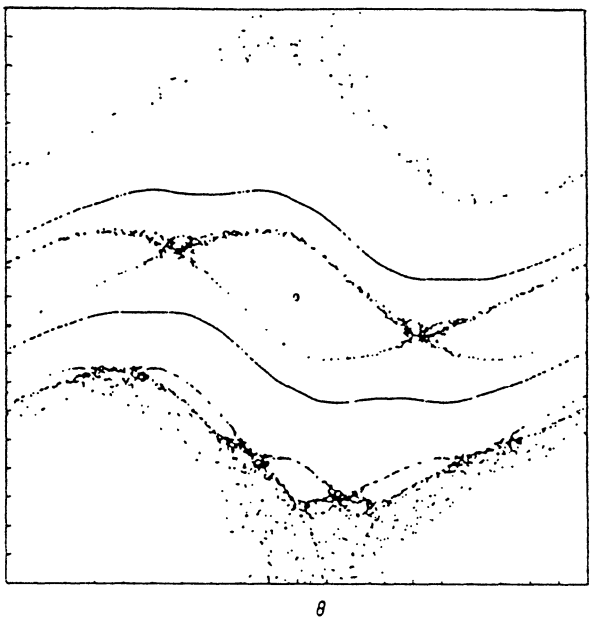

FIG. 4 [13]. *Five orbits for* (2.8) *with* $\lambda = 0.97$.

Finally we mention that the billiard problem gives rise to a monotone twist map (Fig. 8). One considers the motion of a masspoint in a plane region bounded by a smooth closed curve of positive curvature. The point moves at unit velocity along a straight line and gets reflected at the boundary so that the angle of incidence equals the angle of reflection. This problem gives rise to a monotone twist mapping by going from one collision with the boundary to the next. We normalize the length of the boundary curve to 1. To describe a point P on the boundary we measure the arc length from a fixed point P_0 on the boundary to P and call it x. If we denote the angle of incidence of a shot through P with the oriented boundary tangent by θ, $0 \leqq \theta \leqq \pi$ and set $y = -\cos\theta$, then the mapping $(x,y) \to (x_1,y_1)$ is a monotone twist mapping of the annulus $-1 \leqq y \leqq +1$ into itself. The generating function h is the negative of the Euclidean distance between P_0 and P_1.

Also for this example the question of stability can be asked. If the boundary is close to a circle or an ellipse one has stability, but J. Mather [22] showed that for a boundary curve with one flat point, where the curvature vanishes, one has instability. That means, in this case there exists a billiard shot which for $t \to +\infty$ skims along the boundary in one direction and for $t \to -\infty$ in the opposite direction. This corresponds to an orbit for which $y_j \to +1$ for $j \to +\infty$ and $y_j \to -1$ for $j \to -\infty$.

3. Mather sets.

3.1. Parametrized invariant curves. We turn to the question raised in the last section: If α is a given number in $\alpha' < \alpha < \alpha''$ does there exist an invariant curve of rotation number α? The answer to this question is negative, in general. For rational values of α one finds, by a fixed point theorem of Birkhoff, periodic orbits of this rotation number, but generally they are isolated and a curve with this rotation number

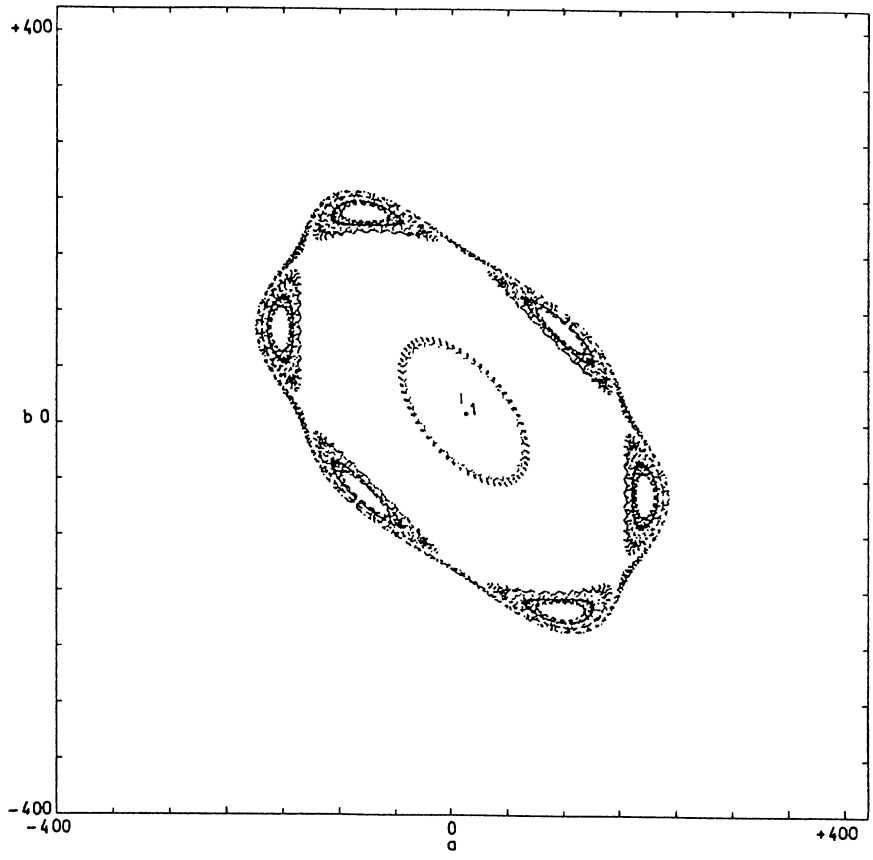

FIG. 5 [33]. *Two orbits for* (2.9) *with* $\lambda = 1.3$.

does not exist. However, if the concept of the invariant curve is appropriately generalized the answer to our question is positive. This leads to the concept of Mather sets which are closed invariant sets whose orbits have a rotation number. The precise definition will be given below.

We begin with a different description of an invariant curve Γ. Instead of writing Γ in the form $y = w(x)$ we ask for a parametric representation

$$x = u(\theta), \qquad y = v(\theta) \tag{3.1}$$

where $u(\theta)$ is monotone increasing and $u(\theta) - \theta$, $v(\theta)$ are continuous functions of period 1. We require that the parameter is introduced in such a way that the mapping $\varphi|_\Gamma$ corresponds to a rotation $\theta \to \theta_1 = \theta + \alpha$, i.e.

$$x_1 = u(\theta + \alpha), \qquad y_1 = v(\theta + \alpha). \tag{3.2}$$

In order to determine such u, v we have to solve the difference equation

$$u(\theta + \alpha) = f(u, v), \qquad v(\theta + \alpha) = g(u, v), \tag{3.3}$$

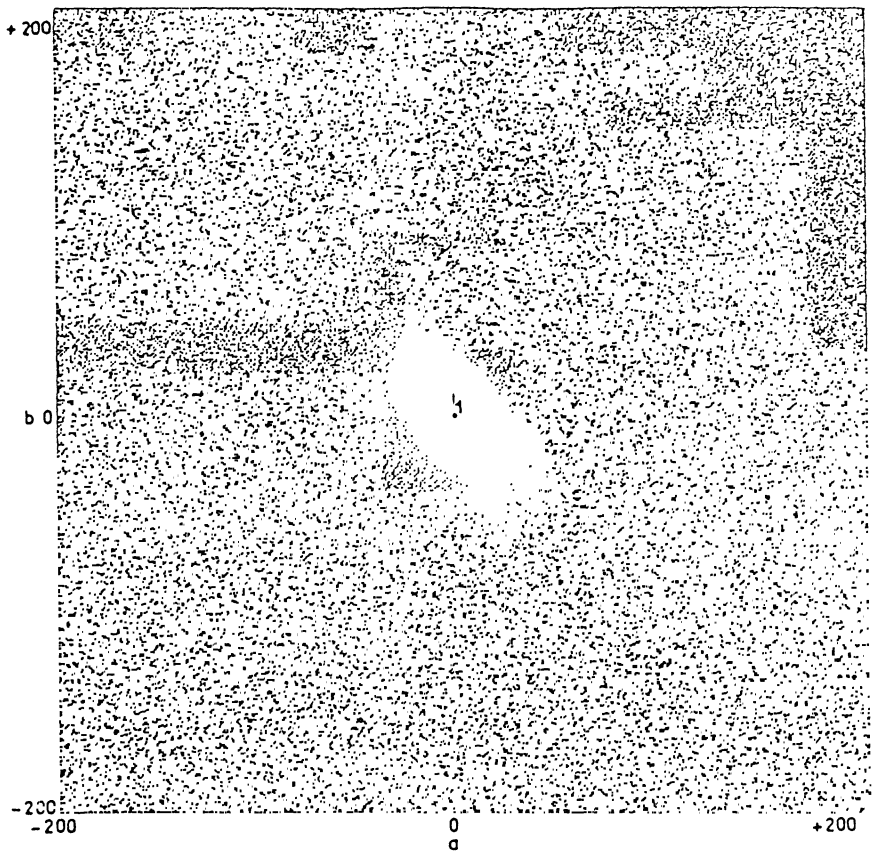

FIG. 6 [33]. *Orbit for* (2.9) *with* $\lambda = 1.3$.

with the above qualifications for u, v. Observe that it suffices to determine the function u; indeed if u is known we can determine v from the first equation of (3.3), since $f_y > 0$. Therefore it is useful to eliminate v from the equations (3.3). This can be done most effectively by using equations (2.6) with which (3.3) takes the form

$$\begin{aligned} v(\theta) &= -\frac{\partial h}{\partial x_0}(u(\theta), u(\theta+\alpha)), \\ v(\theta+\alpha) &= +\frac{\partial h}{\partial x_1}(u(\theta), u(\theta+\alpha)); \end{aligned} \tag{3.4}$$

hence

$$\frac{\partial h}{\partial x_0}(u(\theta),\ u(\theta+\alpha)) + \frac{\partial h}{\partial x_1}(u(\theta-\alpha),\ u(\theta)) = 0 \tag{3.5}$$

is the desired nonlinear difference equation for u. We require again that $u(\theta)-\theta$ has period 1 and is monotone increasing but will now drop the requirement for u to be continuous. If u is a continuous solution, then v will, by (3.4), also be continuous and

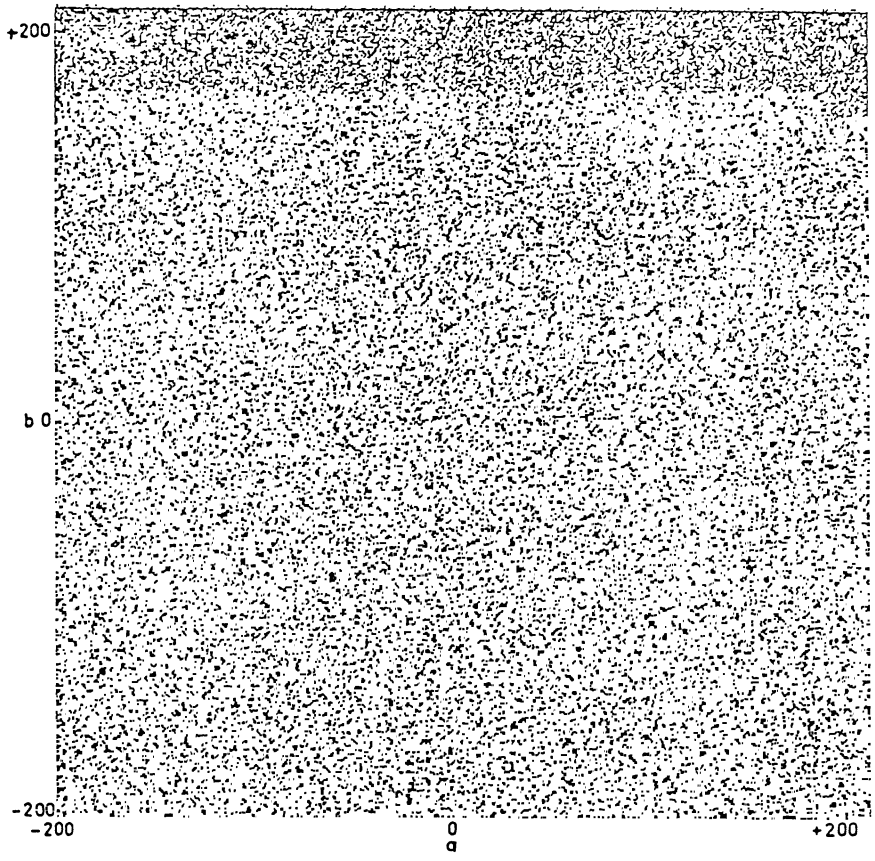

FIG. 7 [33]. *Orbit for* (2.9) *with* $\lambda = 10$.

(3.1) represents an invariant curve. If, however, u is merely monotone, hence has at most countably many discontinuities, then also v will have discontinuities and (3.1) does not represent a curve anymore.

3.2. Mather sets. Now we define Mather sets of rotation number α as the closed invariant sets given by (3.1) where u is a monotone function of θ, $u-\theta$, v have period 1 and equations (3.3) hold (but u, v need not be continuous).

THEOREM 3.1 (J. Mather [21], Aubry and Le Daeron [1]). *If* φ *is a monotone twist mapping and* α *a given number in* (α', α'') *then there exists an invariant Mather set* Γ_α *with rotation number* α. *Moreover,* Γ_α *is a subset of a closed curve*

$$y = w(x) \tag{3.6}$$

where $w(x)$ *has period* 1 *and is Lipschitz continuous, i.e.* $v(\theta) = w(u(\theta))$.

For rational $\alpha = p/q$ this theorem provides periodic orbits x_j, y_j satisfying

$$x_{j+q} = x_j + p, \qquad y_{j+q} = y_1 \quad \text{for } j \in \mathbb{Z}.$$

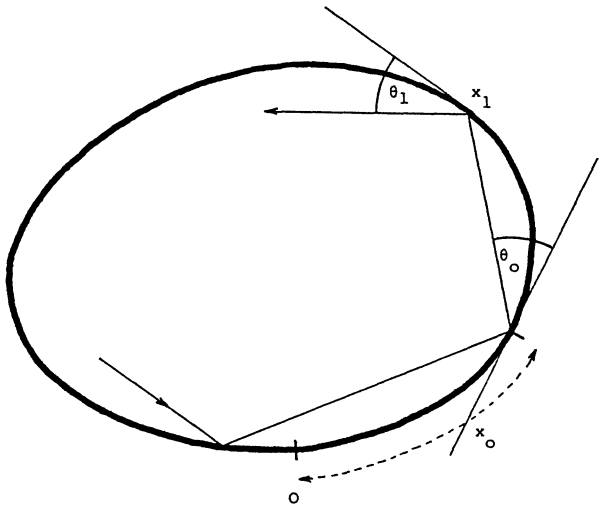

FIG. 8. *Billiard problem.*

If these are isolated, the functions u, v are piecewise constant taking on just the discrete values $(u,v)=(x_j,y_j)$ of the periodic orbits. But it is worth mentioning that the periodic orbits on Γ_α have a special property: Since u is monotone increasing the mapping

$$j\alpha-k\to u(j\alpha-k)=x_j-k \tag{3.7}$$

is monotone for all integers j, k. It is easy to see that in general there may be, for example, elliptic periodic orbits lacking this monotonicity property.

More interesting is the case of irrational α. Also in this case one has the monotonicity property (3.7) and since the set $\{j\alpha-k;\ j,k\in\mathbb{Z}\}$ is dense one can define $u=u(\theta)$ from the knowledge of one orbit $\{x_j\}$. The set Γ_α is the limit set of any orbit on it.

We have to distinguish two cases: A) u, v are continuous functions, i.e., Γ_α defines an invariant curve which, in fact, is Lipschitz continuous. Moreover every orbit on Γ_α is dense on Γ_α. B) u, v have countably many discontinuities. In this case, Γ_α can be viewed as a Cantor set on the curve (3.6) and the gaps of this Cantor set correspond to the discontinuities of u, v. Of course, the curve (3.6) has no invariant meaning but it can be obtained from Γ_α by first defining $w(x)$ on the projection of Γ_α on the x-axis and extrapolating it linearly in the gaps. By the theorem this gives a Lipschitz continuous function.

This case distinction corresponds to that in Denjoy's theory: In case A) every orbit on the invariant curve is dense while case B) corresponds to a Denjoy minimal set. Of course, one has to keep in mind that φ does not leave the curve (3.6) invariant, but by projecting the mapping φ on the x-axis such a homeomorphism of the circle can be obtained and the above comparison becomes relevant.

Of course, Cantor sets are complicated objects which frequently come up as limit sets of orbits. But one has to keep in mind that the limit sets Γ_α have several simple properties, special to Mather sets: They are subsets of one-dimensional closed Lipschitz continuous curves (3.6), hence their orbits cannot wander all over. Secondly, the "ordering" of an orbit $\{x_j\}$ on Γ_α is by (3.7) the same as for the rotation of a circle by the angle α. These are the significant properties of Mather sets.

We have to add that Mather proved his theorem in greater generality than we indicated. He does not require any smoothness of φ but assumes that φ is an area-preserving homeomorphism for which $f(x,y)$ is strictly monotone in y.

3.3. Variational problem. Mather based his proof on the variational problem

$$\int_0^1 h(u(\theta),\, u(\theta+\alpha))\, d\theta, \tag{3.8}$$

which had been used before by Percival [31], [32] for numerical purposes. Similar variational problems for invariant tori had been mentioned even earlier (see [19]) as I learned from Percival. But in these references the variational problems were used formally while Mather based his existence proof on (3.8).

Mather minimizes the above functional in the class of weakly monotone functions $u=u(\theta)$ for which $u(\theta)-\theta$ had period 1 and for which $f(u(\theta),a')\leqq u(\theta+\alpha)\leqq f(u(\theta),a'')$ holds. Moreover, he imposes the normalization that $u(\theta)$ is continuous from the left and $u(\theta)\leqq 0$ for $\theta\leqq 0$, $u(\theta)\geqq 0$ for $\theta\geqq 0$. The latter normalization is compatible with the problem since with $u(\theta)$ also $u(\theta+c)$ is a solution. Incidentally, in Mather's paper the functional is maximized, which is due to the fact that we replaced h by $-h$.

It is quite easy to show that the function (3.8) has a minimum in this class. But the difficulty is to verify that such a minimal satisfies the corresponding Euler equation which corresponds precisely to the difference equation (3.5) to be solved. The problem is that the set of admissible functions has boundary points, e.g. functions which are constant on an interval, or have discontinuities. Therefore one would expect just an inequality instead of the equation (3.5), but Mather shows by use of appropriate variations that (3.5) is actually satisfied for a minimal of (3.8).

We suggest an alternate approach which avoids the above problems and uses the familiar technique of regularization. We replace the functional (3.8) by the regularized functional

$$\int_0^1 \varepsilon\left(\frac{\partial u}{\partial\theta}\right)^2 + h(u(\theta),u(\theta+\alpha))\, d\theta$$

for $\varepsilon>0$. We minimize this functional over all functions $u(\theta)=\theta+\hat{u}(\theta)$ where $\hat{u}\in H^1_{\text{per}}[0,1]$ belongs to the Sobolev space of periodic functions for which $d\hat{u}/d\theta\in L^2[0,1]$. It is standard to show that for $\varepsilon>0$ a minimal $u=u_\varepsilon(\theta)$ exists, is smooth and satisfies the Euler equation

$$-\varepsilon\left(\frac{d}{d\theta}\right)^2 u+\frac{\partial h}{\partial x_0}(u,u^+)+\frac{\partial h}{\partial x_1}(u^-,u)=0, \qquad u^\pm=u(\theta\pm\alpha) \tag{3.9}$$

which is a difference-differential equation. The crucial point is that on account of the maximum principle for the linearized Euler equation one proves

$$\frac{du_\varepsilon}{d\theta}>0 \quad \text{for all } \theta \quad \text{if } \varepsilon>0;$$

that is, the monotonicity is forced by our regularization.

Now using Helly's theorem it is easy to find a sequence $\varepsilon=\varepsilon_\nu\to 0$ for which u_ε converges almost everywhere to the desired solution of (3.5).

By this procedure one obtains the discontinuous solution u as the limit of the smooth solutions u_ε of (3.9). This is reminiscent of the smoothing of shock-solutions in fluid mechanics by adding dissipation. The variational problem (3.8) thus appears as a degenerate problem. The complicated feature of this functional can be gleaned from the fact that near such a minimal representing a Cantor set Γ_α there exist uncountably many local minima of (3.8). This was shown recently by J. Mather [23]. In the proof of this result the so-called "Peierls energy barrier," inspired by the physical model, plays a central role.

4. Relation to KAM theory. These results are, of course, closely related to the existence theory of invariant curves obtained via KAM theory. In that theory one considers a family of area-preserving maps φ_ε close to a simple mapping

$$\varphi_0 : (x,y) \to (x+\psi(y),\ y), \qquad \psi'(y) \neq 0,$$

preserving the circles $y = \text{const}$. This is the integrable mapping discussed in the introduction. The theory guarantees the existence of smooth closed curves invariant under φ_ε with rotation number α provided

$$|q\alpha - p| \geqq \gamma |q|^{-\nu} \quad \text{for all integer pairs } (q,p) \neq (0,0), \tag{4.1}$$

$$|\varepsilon| \leqq \varepsilon_0(\alpha). \tag{4.2}$$

In contrast to Theorem 3.1 this theory assures us that the curves are smooth and certainly not Cantor sets. But the prices are: very restrictive number theoretical conditions and a severe smallness condition. One may say that Mather's theory provides, without such restrictions, a generalized or weak solution of the difference equation while the KAM theory yields classical solutions under the above restrictions. Thus it appears that KAM theory provides the regularity of the weak solutions whose existence in ensured quite generally.

On the other hand one has to observe that KAM theory is applicable for systems of more than two degrees of freedom for which the theory of Aubry and Mather breaks down.

For two-dimensional mappings the invariant curves enclose invariant sets and therefore are useful for stability theory—which was the original motivation for their discovery. For the Mather sets this cannot be said, and orbits can leak out through the gaps of the Cantor sets. But since we are interested in the breakdown of stability and the disintegration of invariant curves this theory provides the appropriate framework.

We describe the picture provided by this theory. Fixing a rotation number α satisfying (4.1) we ask for the corresponding invariant sets $\Gamma_\alpha(\varepsilon)$. For sufficiently small ε these are invariant curves provided by the KAM theory, but for larger values of ε they may become Cantor sets. Let ε^* be the largest value for which $\Gamma_\alpha(\varepsilon)$ are curves for all ε in $0 \leqq \varepsilon \leqq \varepsilon^*$. Then we have in $\Gamma_\alpha(\varepsilon^*)$ an invariant curve which under arbitrarily small increase of ε disintegrates to a Cantor set. From this we conclude that $\Gamma_\alpha(\varepsilon^*)$ cannot be very smooth, since otherwise one could prove with the help of KAM theory that $\Gamma_\alpha(\varepsilon)$ would remain a curve for sufficiently small $|\varepsilon - \varepsilon^*|$. Thus $\Gamma_\alpha(\varepsilon)$ breaks up through loss of smoothness.

One has to keep in mind that the smallness condition (4.2) depends on the choice of α satisfying (4.1) and ε_0 will become zero for several α violating (4.1). These α form an exceptional set of small measure if γ is small and $\nu > 2$. This means that for such exceptional values of α the invariant circle may disintegrate for any $\varepsilon > 0$.

We summarize: For a family of monotone twist maps φ_ε as above one obtains for sufficiently small ε, say $0 \leqq \varepsilon \leqq \varepsilon_1$ a set of invariant curves forming, in general, a Cantor set, i.e., if we intersect this set of invariant curves with the y-axis we obtain a Cantor set. We refer to this as a Cantor set "in the radial direction," since y corresponds to the radial variable. Following such an invariant curve $\Gamma_\alpha(\varepsilon)$ to a larger value of ε it will, in general, disintegrate to a Cantor set on a Lipschitz curve, i.e., loosely speaking it forms a Cantor set in the horizontal direction (Fig. 9). This can go so far that ultimately no invariant curve of the form (3.6) is left and all Mather sets are Cantor sets. (Of course, other invariant curves surrounding elliptic fixed points may still exist.) In this case we have lost stability and the annulus represents a ring of instability.

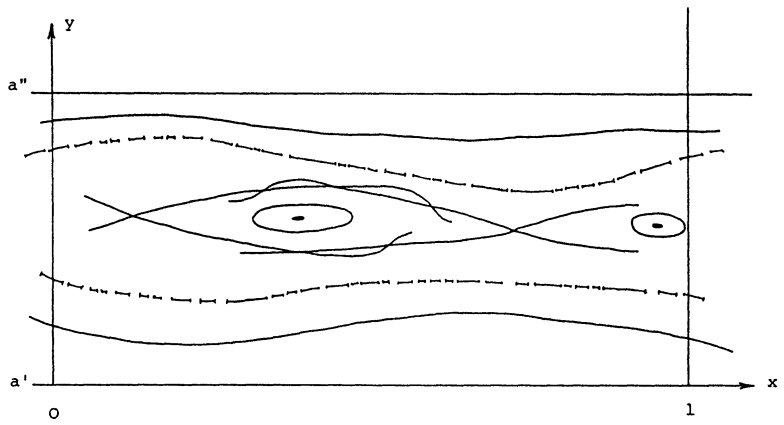

FIG. 9. *Invariant curves and Mather sets.*

This latter possibility can certainly occur as we have discussed at the end of §2. Other examples can be found in [30] and another will be discussed in §8. The underlying principle, which allows one to prove the absence of invariant curves, is related to Aubry's construction. He characterizes the orbits $\{x_j, y_j\}$ lying on a Mather set by another variational principle. He calls an orbit $\{x_j, y_j\}$ a "minimal energy orbit" if

$$\sum_{j=-\infty}^{+\infty} \left(h(x_j+\xi_j, x_{j+1}+\xi_{j+1}) - h(x_j, x_{j+1})\right) \geqq 0$$

holds for all sequences ξ_j vanishing for sufficiently large $|j|$. The corresponding y_j are determined by the relations (2.6).

Aubry and Le Daeron [1] construct the Mather sets out of such minimal orbits. We will describe this procedure in §7 in an analogous problem. Since the orbits on an invariant curve are minimal energy orbits we can conclude: If there exist points through which passes no minimal energy orbit then there exists no invariant curve. It is often easy to verify that minimal energy orbits have to avoid certain regions where h is too large. On this basis mappings without invariant curves can be found. The same argument is used in [22].

5. Connection with mechanics. The above mentioned theory is closely related to variational problems of the form

$$\int_{t_1}^{t_2} F(t, x, \dot{x})\, dt \tag{5.1}$$

where $F=F(t,x,\dot{x})$ satisfies the Legendre condition

$$F_{\dot{x}\dot{x}}>0 \tag{5.2}$$

and has period 1 in t and x. One can view (5.1) as a variational problem on the two-dimensional torus $T^2=\mathbb{R}^2/\mathbb{Z}^2$. For example, if $V=V(t,x)$ is a potential having period 1 in t, x then

$$F=\tfrac{1}{2}\dot{x}^2-V(t,x) \tag{5.3}$$

is an admissible integrand for (5.1) and the corresponding Euler equations are

$$\ddot{x}+V_x(t,x)=0. \tag{5.4}$$

This can be considered as the equation of motion of a mechanical problem, and (5.1) the corresponding Hamilton principle of mechanics.

In the special case where $V=V(x)=V(x+1)$ is independent of t the system (5.4) is integrable and most solutions are periodic. We may assume that

$$\max_x V(x)=0.$$

The solution of (5.4) for the time independent case is obtained from the energy relation

$$\tfrac{1}{2}\dot{x}^2+V(x)=E, \qquad E\geqq\min_x V(x).$$

For $E>0$ we obtain

$$\int_0^x \frac{dy}{\sqrt{2(E-V(y))}}=t-t_0$$

and the left-hand side is a monotone function of the form

$$\gamma x+q(x), \qquad q(x+1)=q(x)$$

where $\gamma>0$ is the above integral taken over $0\leqq x\leqq 1$. Inverting this relation we obtain

$$x=\theta+p(\theta), \qquad \theta=\alpha t+\beta$$

where p has period 1 and

$$\alpha=\gamma^{-1}, \qquad \beta=-\alpha t_0.$$

This trivial observation is to be extended to the nonautonomous case (5.4). As a consequence of the above theory by Aubry and Mather we can assert that for every real number α there exists a function $p(t,\theta)$ of period 1 in t and θ such that

$$U(t,\theta)=\theta+p(t,\theta)$$

is strictly monotone in θ and such that

$$x(t)=U(t,\alpha t+\beta) \tag{5.5}$$

is a solution of (5.4). Of course, the functions p and U depend also on α.

If $p(t,\theta)$, hence also U, is a *continuous* function one calls the corresponding solutions *quasi-periodic* of frequencies 1 and α. However, in general, U will have discontinuities; since U is monotone in θ one has for every fixed x only denumerably many such discontinuities. On the other hand, since the solutions $x(t)$ are smooth, we see from (5.5) that $U(t,\alpha t+\beta)$ is a smooth function of t, if β is fixed.

Thus the solutions (5.5) can be viewed as generalized quasi-periodic functions.

Representing p as a Fourier series

$$p(t,\theta)=\sum_{k,l} c_{kl} e^{i(kt+l\theta)}$$

the formula (5.5) takes the form

$$x(t)=\alpha t+\beta+\sum_{k,l} c_{kl} e^{i(k+l\alpha)t}.$$

For rational α these represent periodic solutions, but for irrational α we shall call them generalized quasi-periodic solutions.

If we view (5.4) as a system with periodic forcing of frequency 1 then our statement asserts the existence of generalized quasi-periodic responses of frequencies 1, α, where α is any prescribed number!

A similar result holds for the Euler equation of a more general variational problem (5.1). We will assume that the integrand $F\in C^2$ satisfies

$$\text{(5.6)}\qquad \begin{aligned} &\text{i)} && F(t,x,\dot{x}) \text{ has period 1 in } t,\ x,\\ &\text{ii)} && \delta^{-1}\geqq F_{\dot{x}\dot{x}}(t,x,\dot{x})\geqq\delta>0,\\ &\text{iii)} && |F_{x\dot{x}}|+|F_{t\dot{x}}|=\mathcal{O}(|\dot{x}|), \qquad F_x=\mathcal{O}(\dot{x}^2). \end{aligned}$$

THEOREM 5.1. *If F satisfies the above conditions* (5.6) *and α is a given real number, then there exists a function $U(t,\theta)$ such that $U-\theta$ has period* 1 *in t, θ, U is strictly monotone increasing in θ, and such that for any β*

$$\text{(5.7)}\qquad x(t)=U(t,\alpha t+\beta)$$

is a solution of the Euler equation

$$\frac{\partial}{\partial t}F_{\dot{x}}=F_x.$$

The family of solutions (5.7) can be described more geometrically, at least if the function $U=U(t,\theta)$ is continuous. Indeed, in this case these solutions cover the plane simply, i.e., through every point $(t,x)\in\mathbb{R}^2$ goes exactly one curve. They form the solutions of a vector field

$$\text{(5.8)}\qquad \dot{x}=\psi(t,x)$$

where ψ has period 1 in t and x. This function is determined by the equation

$$\psi(t,x)=\left(\frac{\partial}{\partial t}+\alpha\frac{\partial}{\partial\theta}\right)U(t,\theta), \qquad x=U(t,\theta);$$

since U is strictly monotone in θ the variable θ is uniquely determined by the last equation.

Thus in case U is a continuous function, the family (5.7) can be viewed as a set of solutions of a vector field (5.8) which, of course, still depends on α, and whose solutions are also solutions of the Euler equation. One observes, by the way, that for irrational α the function U is continuous in t, θ if and only if each orbit (5.7) is dense on the torus.

In case $U(t,\theta)$ is not continuous, it has for each fixed t a denumerable number of discontinuities. If α is irrational the range of $U(t,\theta)$ for fixed t and variable θ forms a Cantor set and the family of solutions (5.7) constitutes a Cantor set $\mathscr{L}$ on which one

can define again a vector field (5.8). This Cantor set is invariant under the translations $(t,x)\to(t+j,x+k)$, for integer j, k, and therefore can be viewed as a Cantor set on the torus T^2.

This Cantor set with its vector field corresponds to the Mather set of §3 while the vector field (5.8) in the case of a continuous U corresponds to an invariant curve. Indeed, for $t=0$ one has $\dot{x}=\psi(0,x)$ or

$$y=F_{\dot{x}}(0,x,\psi(0,x))=w(x),$$

which defines a continuous curve (of period 1) invariant under the map, since for $t=1$, $x_1=x(1)$

$$y_1=F_{\dot{x}}(1,x_1,\psi(1,x_1))=w(x_1).$$

We summarize: For the variational problem (5.1) an invariant curve of rotation number α corresponds to a vector field (5.8) on the torus whose solutions are extremals $x(t)$ satisfying $t^{-1}x(t)\to\alpha$ for $t\to\pm\infty$. For α irrational these solutions are dense on the torus if the curve is smooth. The Mather sets of rotation number α correspond to such vector fields on a Cantor set and the corresponding orbits are *not dense* on the torus.

For a variational problem with $F=F^0(\dot{x})$ independent of t and x the extremals are linear functions of t and therefore the solutions (5.7) are of the form

$$x=\alpha t+\beta,$$

i.e., $U(\theta)=\theta$. The corresponding vector fields (5.8) are given by

$$\dot{x}=\alpha. \tag{5.9}$$

This represents the integrable case, with the integral $\dot{x}$, and the vector fields correspond to the invariant curves. As a consequence of KAM theory, for any α that satisfies

$$\left|\alpha-\frac{p}{q}\right|\geqq cq^{-\tau} \quad \text{for all integer } p,q; \quad q>0 \tag{5.10}$$

and any $F(t,x,\dot{x})$ sufficiently those to F^0, one can find again a vector field on the torus with solutions (5.7) represented by a continuous function U. If $F-F^0$ gets larger the torus will develop gaps and the vector field is defined only on a Cantor set on T^2. In particular, the equation

$$\ddot{x}+\lambda V_x(t,x)=0$$

will possess a vector field corresponding to α satisfying (5.10) if $\lambda>0$ is sufficiently small, but for larger λ the torus generally breaks up into a Cantor set. If V is real analytic also the function $U=U(t,\theta;\lambda,\alpha)$ will be analytic for small values of λ but become discontinuous in θ, in general, if λ gets larger. The phenomenon of loss of analyticity, or "phase transition," will be illustrated in a geometrical example in §8.

6. The relation between discrete and continuous systems. There is a close relation between monotone twist maps on a cylinder and the variational problems (5.1). For example, the twist map given by the generating function

$$h(x_0,x_1)=\tfrac{1}{2}(x_1-x_0)^2-g(x_0) \tag{6.1}$$

corresponds to a variational problem (5.1) with

$$F(t,x,p)=\tfrac{1}{2}p^2-\delta_{\text{per}}(t)g(x) \tag{6.2}$$

where

$$\delta_{\text{per}}(t)=\sum_{j=-\infty}^{+\infty}\delta(t-j)$$

is the periodic Dirac-δ-function. The Poincaré mapping taking the initial values $x(0)=x_0$, $\dot{x}(0)=y_0$ into $x(1)=x_1$, $\dot{x}(1)=y_1$ clearly agrees with the given monotone twist map.

In this example the integrand (6.2) is singular, but this drawback can be avoided. As a matter of fact *any* monotone twist mapping can be obtained as a Poincaré mapping for the Euler equation of such a variational problem on a torus.

We want to formulate this statement more precisely, since it reveals the important fact that the *monotone character* of a twist map corresponds to the *Legendre condition* of the variational problem.

We consider a monotone twist mapping φ of an annulus $A=\{x,y\,|\,a'\leqq y\leqq a'',\ x \bmod 1\}$ as described in §2 and ask whether we can find a family of area-preserving mappings φ^t, $0\leqq t\leqq 1$, of A into itself such that

$$\varphi^0=\text{id}, \qquad \varphi^1=\varphi$$

and, moreover, such that also

$$\varphi^{t+\varepsilon_0}(\varphi^t)^{-1} \quad \text{for all } t\in[0,1] \tag{6.3}$$

is a monotone twist mapping if $\varepsilon>0$ is small enough. Clearly any such family of area-preserving mappings is generated by a nonautonomous Hamiltonian system

$$\dot{x}=H_y(t,x,y), \qquad \dot{y}=-H_x(t,x,y)$$

by taking the initial values $(x(0),y(0))$ into $(x(t),y(t))=\varphi^t(x(0),y(0))$. The condition (6.3) amounts to

$$H_{yy}>0$$

since the "twist" can be calculated from

$$x(t+\varepsilon)=x(t)+\varepsilon H_y+O(\varepsilon^2)$$

to be εH_y. Therefore we will look for a smooth Hamiltonian $H=H(t,x,y)$ defined for $a'\leqq y\leqq a''$ and all t, x with the properties

$$\begin{aligned} &\text{i)} \quad H(t+1,x,y)=H(t,x,y)=H(t,x+1,y),\\ &\text{ii)} \quad H_x(t,x,y)=0 \quad \text{for } y=a',a'',\\ &\text{iii)} \quad H_{yy}>0. \end{aligned} \tag{6.4}$$

THEOREM 6.1 [28]. *Given a smooth monotone twist mapping φ of the annulus A into itself then there exists a smooth Hamiltonian $H=H(t,x,y)$ defined for $a'\leqq y\leqq a''$ and all real t, x satisfying* (6.4) *such that the Poincaré mapping $(x(0),y(0))\to(x(1),y(1))$ agrees with the given mapping φ.*

The mapping $\varphi^t:(x(0),y(0))\to(x(t),y(t))$ gives rise to the desired interpolation of $\varphi=\varphi^1$ and the identity $\text{id}=\varphi^0$, which satisfies the condition (6.3).

On the other hand, it is standard that such a Hamiltonian system corresponds to the Euler equation of a variational problem (5.1). Here $F=F(t,x,p)$ is defined via the invertible Legendre transformation

$$p=H_y(t,x,y)$$

by the formula

$$F(t,x,p)=yp-H(t,x,y)$$

and the condition $H_{yy}>0$ translates into the Legendre condition

$$F_{pp}(t,x,p)>0. \tag{6.5}$$

In other words, any monotone twist mapping is the Poincaré mapping corresponding to a variational problem (5.1) satisfying the Legendre condition (6.5). Moreover, $F(t,x,p)$ has period 1 in t, x. However, the domain of definition becomes more complicated in the t, x, p variables, and we avoid this difficulty by assuming $a'=-\infty$, $a''=+\infty$, i.e. that A is a cylinder. Then $H(t,x,y)$, $F(t,x,p)$ are defined for *all* real arguments.

We ask whether conversely every mapping φ generated in this way by a variational problem is a monotone twist mapping. This is definitely not so; one has to require, in addition, that the solutions $x=x(t)$ of the Euler equation for (5.1) have no conjugate points for $0\leqq t\leqq 1$. This means, if $x=x(t,\lambda)$ is a family of solutions of the Euler equation satisfying

$$\begin{aligned} &x(0,\lambda)=x_0 \quad \text{independent of } \lambda,\\ &\dot{x}(0,\lambda)=\lambda \end{aligned}$$

then

$$\frac{\partial x}{\partial \lambda}>0 \quad \text{for } 0<t\leqq 1$$

for any choice of x_0,λ.

Thus this family $x(t,\lambda)$ depends monotonically on λ if $t\in(0,1]$ and does not focus on any point. In this case the generating function $h(x_0,x_1)$ of the mapping φ^1 is obtained by

$$h(x_0,x_1)=\int_0^1 F(t,x(t),\dot{x}(t))\,dt$$

where $x=x(t)$ is the solution of the Euler equation satisfying

$$x(0)=x_0, \qquad x(1)=x_1.$$

This formula is well known from calculus of variations, or, say, in geometrical optics where h plays the role of the characteristic function of Hamilton.

7. Minimal solutions. We indicate the ideas on which the proof of the results mentioned in §5 is founded. For this we need

DEFINITION 7.1. A function $x\in C^1(\mathbb{R})$ is called a *minimal solution* of the variational problem (5.1) if

$$\int_{-\infty}^{+\infty}(F(t,x+\xi,\dot{x}+\dot{\xi})-F(t,x,\dot{x}))\,dt\geqq 0 \tag{7.1}$$

for all functions $\xi\in C^1_{\text{comp}}(\mathbb{R})$ vanishing for large values of $|t|$.

It is clear that any minimal solution is a solution of the Euler equation, but not every solution of the Euler equation is a minimal solution. For example, a periodic solution of the Euler equation, satisfying

$$x(t+q)=x(t)+p \quad \text{for all } t$$

and some integers p, q, is minimal only if its Floquet multiplier λ is real. In other words a stable periodic solution ($|\lambda|=1$, $\lambda \neq \pm 1$) is never a minimal solution [7]. Moreover, minimal solutions do not possess conjugate points by a well-known criterion due to Jacobi. Thus by the quest for minimal solutions we select a special class of solutions. It is important to observe that solutions (5.7) corresponding to a vector field (5.8), in the case that U is continuous, are always minimal. This is a standard fact of calculus of variations where one shows that an extremal that can be embedded in an extremal field is minimal (see e.g. [9]). In our case the solutions (5.7) form, for fixed α and variable β, such an extremal field covering $\mathbb{R}^2$ simply. One can express this fact also as follows: A field of extremals, covering $\mathbb{R}^2$ simply, is in fact a field of minimals in the sense of Definition 7.1.

Therefore in the search for invariant curves of twist mappings or for the corresponding extremal fields (5.7) we can restrict ourselves to the class of minimal solutions. Of course, at first it is not at all clear whether such minimal solutions exist. We list some properties of these minimals.

THEOREM 7.2. *If F satisfies the conditions* (5.6) *and x is a minimal solution for* (5.1) *then there exists a real number α such that*

$$\sup_{-\infty<t<+\infty} |x(t)-\alpha t|<\infty.$$

Moreover, there exists a constant c depending on F only and not the specific minimal $x(t)$ such that for all real s, t

$$|x(t+s)-x(t)-\alpha\cdot s|\leqq c(1+|\alpha|)$$

holds. Moreover there exist constants γ_1, γ_2 depending on F and monotonically increasing on $|\alpha|$ such that

$$|\dot{x}(t)|\leqq\gamma_1, \quad |\ddot{x}(t)|\leqq\gamma_2 \quad \textit{for all real } t. \tag{7.2}$$

The number α is uniquely determined by the minimal $x=x(t)$. It represents an average slope of $x=x(t)$ and clearly corresponds to the rotation number of §2. We will denote the set of minimal solutions by $\mathcal{M}$, and those with rotation number α by $\mathcal{M}(\alpha)$.

THEOREM 7.3. *Given any real α there exists a minimal solution with rotation number α, i.e. $\mathcal{M}(\alpha)\neq\varnothing$.*

In other words, for any straight line $x=\alpha t+\beta$ there is a minimal solution $x=x(t)$ that has a distance

$$1+\frac{1+|\alpha|}{\sqrt{1+\alpha^2}}c\leqq 1+\sqrt{2}\,c=D$$

from the straight line; here D is independent of the particular solution, and just depends on the integrand F.

The inequalities of Theorem 7.2 provide for a compactness theorem on which the proof of the existence Theorem 7.3 is based. We consider the set of minimal solutions

$$\mathcal{M}_A=\bigcup_{|\alpha|\leqq A}\mathcal{M}(\alpha)$$

and notice that by (7.2) $|\dot{x}|+|\ddot{x}|$ is uniformly bounded on $\mathscr{M}_A$. Moreover, we can replace $x(t)$ by $x(t)+$integer to achieve

$$0 \leqq x(0) < 1$$

so that $|x(t)|$ is uniformly bounded in an interval $|t| \leqq T$ for all such $x \in \mathscr{M}_A$. By the theorem of Ascoli–Arzela one sees that any sequence of this type has a subsequence converging in $C^1[-T, T]$. From this one can extract a subsequence converging in $C^1[-T, +T]$ for every interval $|t| \leqq T$. It turns out, and this is easy to prove, that the limit is in $\mathscr{M}_A$ again. Thus we have

THEOREM 7.4. *The set $\mathscr{M}_A/\mathbb{Z}$ is compact with respect to the topology given by C^1-convergence on every compact interval.*

For rational α one readily establishes the existence of periodic minimal solutions in $\mathscr{M}(\alpha)$ by direct methods of calculus of variations. For an irrational α one constructs $x \in \mathscr{M}(\alpha)$ by approximating α by rationals α_ν and x as limit of a subsequence of $x_\nu \in \mathscr{M}(\alpha_\nu)$, using Theorem 7.4. One can view Theorems 7.2 and 7.3 as a comparison statement between the minimal solutions of (5.1) and the trivial variational problem

$$\int \dot{x}^2 dt, \qquad F^0 = \dot{x}^2,$$

whose minimal solutions are straight lines $x = \alpha t + \beta$. According to Theorem 7.2 every minimal lies in a strip of width 2D from such a line and conversely every such strip contains such a minimal.

There is another close relation between the variational problems with the integrands F and F^0. For this purpose we note that the translations

$$(t, x) \to (t+j, x+k)$$

with integers j, k leave the variational problem invariant. Therefore, if $x(t)$ is a minimal solution so is

$$x(t+j)-k.$$

On the torus this represents, of course, the same curve. We will say that x has a selfintersection on the torus if

$$x(t+j)-k-x(t)$$

changes sign for some choice of integers j, k.

THEOREM 7.5. *The minimal solutions of* (5.1) *have no selfintersections on the torus, i.e. for every pair of integers j, k the function*

$$x(t+j)-k-x(t) \tag{7.3}$$

is for all real t either positive or negative or identically zero.

The last case corresponds to periodic minimal solutions with $\alpha = k/j$ being rational. If α is irrational the function (7.3) is never zero if $(j, k) \neq (0, 0)$ and therefore for irrational α Theorem 7.3 provides an ordering of the lattice $\mathbb{Z}^2$: We set $(j, k) < (j', k')$ if

$$x(t+j)-k < x(t+j')-k'.$$

The remarkable fact is that this ordering is "universal," i.e., independent of the choice of the integrand F. In particular, this ordering agrees with that for the parallel straight lines $x = \alpha t + \beta$ of fixed slope α.

THEOREM 7.6. *If α is irrational and $x \in \mathscr{M}(\alpha)$ then*

$$x(t+j)-k-x(t)>0 \quad \textit{iff } \alpha j-k>0.$$

This fact can be expressed in a different way: If we define a function $U(t,\theta)$ for $\theta=\alpha t+\alpha j-k$ by

$$U(t,\alpha t+\alpha j-k)=x(t+j)-k$$

then the function $U(t,\theta)$ is strictly monotone in θ, at least on the dense set on which it is defined. One can extend U to functions U^+, U^- by

$$U^+(x,\theta)=\lim_{\theta' \searrow \theta} U(x,\theta'), \qquad U^-(x,\theta)=\lim_{\theta'' \nearrow \theta} U(x,\theta'')$$

where θ', θ'' are decreasing, resp. increasing, sequences taken from the dense set on which U is defined. Thus $U^+=U^-$ except on a denumerable set of lines $x=\alpha t+\beta$. These functions $U^{\pm}(t,\theta)-\theta$ have period 1 in t, θ, as is clear from the definition. Moreover, using the compactness theorem, one obtains that

$$U^{\pm}(t,\alpha t+\beta) \in \mathscr{M}(\alpha)$$

for every choice of β. This gives an outline of the ideas entering the proof of Theorem 5.1, at least for irrational α. This general approach is due to S. Aubry and Le Daeron [1] who developed it for the discrete systems, namely the monotone twist maps. For the continuous case one has to supply some additional estimates which have been carried out in greater generality for multiple integrals [29].

There are many important consequences which we cannot describe here. We mention that U^+, U^- are actually independent of the choice of $x \in \mathscr{M}(\alpha)$, if α is irrational. Moreover, again for irrational α, the image of the mapping

$$(t,\theta) \to (t, U^{\pm}(t,\theta))$$

is $\mathbb{R}^2$ if and only if $U^+=U^-$ is continuous; otherwise it defines the Cantor set $\mathscr{L}(\alpha)$ mentioned in §5, and the gaps of the Cantor set correspond to the sets

$$U^-(t,\theta)<x<U^+(t,\theta)$$

at the discontinuities. In this case the orbits in $\mathscr{M}(\alpha)$ are clearly not dense on the torus, while in the first case, when $U^+=U^-$ is continuous, every orbit in $\mathscr{M}(\alpha)$ is dense on the torus.

We mention that for rational $\alpha=p/q$, $\mathscr{M}(\alpha)$, in general, does not only contain periodic minimals satisfying $x(t+q)=x(t)+p$, but also homoclinic orbits connecting neighboring periodic minimal solutions. This remark can be used for the construction of homoclinic orbits [18], [1].

Finally we mention that the function $U=U(t,\theta)$ can also be obtained from a regularized variational principle

$$\int_0^1 \int_0^1 \left(\varepsilon \left(\frac{\partial U}{\partial \theta} \right)^2 + F(t,U,DU) \right) dx\, d\theta$$

where

$$D=\frac{\partial}{\partial t}+\alpha\frac{\partial}{\partial\theta},$$

and where the admissible functions U are such that

$$U-\theta\in H^1(\mathbb{R}^2/\mathbb{Z}^2).$$

One verifies again that for $\varepsilon>0$ smooth minimals U^ε of this functional exist and satisfy the Euler equation

$$2\varepsilon\left(\frac{\partial}{\partial\theta}\right)^2 U+DF_p(t,U,DU)=F_x(t,U,DU),$$

which is now an elliptic partial differential equation. Moreover, one concludes from the maximum principle that

$$\frac{\partial U^\varepsilon}{\partial\theta}>0$$

and obtains the desired solution as limit of U^ε as $\varepsilon\to 0$ through some sequence.

8. Connection with differential geometry. The same ideas underlying the minimal solutions of the previous section can be found in a differential geometric context in the paper of G. Hedlund [16] which is based on Morse's earlier work [27]. The complex structure of geodesics on a two-dimensional manifold of negative curvature had been known to J. Hadamard [14] and he already encountered Cantor sets. This work was continued by M. Morse, who inquired what happens if the manifold has curvature of variable sign. This led him to the study of geodesics of class A. For our purposes it suffices to define this concept for the torus, although it is meaningful for other manifolds of higher genus.

We represent the torus T^2 as usual by identifying points in the u-v-plane mod 1 and introduce a metric

$$ds^2=E(du)^2+2F\,du\,dv+G\,dv^2 \tag{8.1}$$

where E, F, G are smooth functions of period 1 in u, v, and $EG-F^2>0$. It is well known that geodesic arcs minimize the arc length of all curves with the same endpoints, provided the endpoints are sufficiently close. If the endpoints are far apart, this need not be true anymore; for example, if the arc contains conjugate points. Now Morse and Hedlund considered unending geodesics in the plane which minimize the length for *any two points* on them. They called such geodesics of "class A." In particular, they do not have conjugate points.

This concept of geodesics of class A is completely analogous to Aubry's minimal energy orbits (end of §4) or the minimal solutions of §7, and they also have similar properties. However, these geodesics of class A are given as parametrized curves $u=u(s)$, $v=v(s)$ and are, in general, not representable as the graph $v=f(u)$ of a function, as was required throughout §7. In a geometrical problem such a restriction would be artificial and meaningless. In this respect the results of Hedlund are more general than those of §7, although the functional given by the metric (8.1) is more special. We will point out some of the analogous results which one can find in Hedlund's paper; for more information I refer to the expository paper by Bangert [2] which contains a full discussion of this problem.

Hedlund derives for geodesics of class A a number of properties that show their similarities with straight lines, the class A geodesics of the flat metric, $E=G=1$, $F=0$.

i) There exists a constant $D>0$ depending just on the metric (8.1) such that any geodesic of class A stays within the distance D of a straight line.

ii) Conversely, given any straight line

$$Au+Bv+C=0, \qquad A^2+B^2=1 \tag{8.2}$$

there exists a geodesic of class A $u=u(s)$, $v=v(s)$ within the strip $|Au+Bv+C|<D$.

iii) Any two different geodesics of class A intersect at most once in the plane.

iv) Any geodesic of class A has no selfintersections on the torus.

These properties may suffice to show the similarity. The first two properties show that we can associate with any such geodesic of class A the direction of the line (8.2). If $B\neq 0$ we could define the rotation number

$$\alpha=-\frac{A}{B}=\lim_{s\to\infty}\frac{v(s)}{u(s)}$$

and if $B=0$ we set $\alpha=\infty$.

v) Two geodesics of class A with the same irrational α do not intersect in the plane; hence they form an ordered set.

If α is rational or ∞ then one has geodesics of class A which are periodic and belong to the same free homotopy class as the corresponding straight line. However, one has also homoclinic orbits conecting neighboring periodic geodesics of class A and of the same α (see [27]).

What is the analogue of a Mather set or an invariant curve? We begin with the latter and assume that we have a class A geodesic $(u(s),v(s))$ with irrational α which is dense on the torus, or equivalently the set of translated curves

$$(u(s)+j,\ v(s)+k) \quad \text{for } j,k\in\mathbb{Z}$$

is dense in the plane. Then their closure defines a field of geodesics covering the torus and hence a vectorfield on the torus whose solutions are these geodesics.

This is much the same as the picture in §7.

If, however, a class A geodesic $(u(s),v(s))$ with irrational α is not dense on T^2 then the limit set, i.e. the set of cluster points, of its translates define a Cantor set on the torus which corresponds to a Mather set.

Clearly for the flat metric all geodesics are of class A and those of rotation number α belong to a vectorfield, better a "line field"

$$A\frac{du}{ds}+B\frac{dv}{ds}=0.$$

This corresponds to the integrable case in which all Mather sets are curves.

Can one also have a metric in which one has no invariant curves at all? This corresponds to a situation in which no class A geodesic is dense on the torus. Such a case can be illustrated with a simple example which I owe to Bangert (see [2]). We consider a torus embedded in $\mathbb{R}^3$ that looks like a "tire with a bubble" (Fig. 10). It is geometrically evident that a class A geodesic must avoid running in the bubble, because, if it did, it could be shortened.

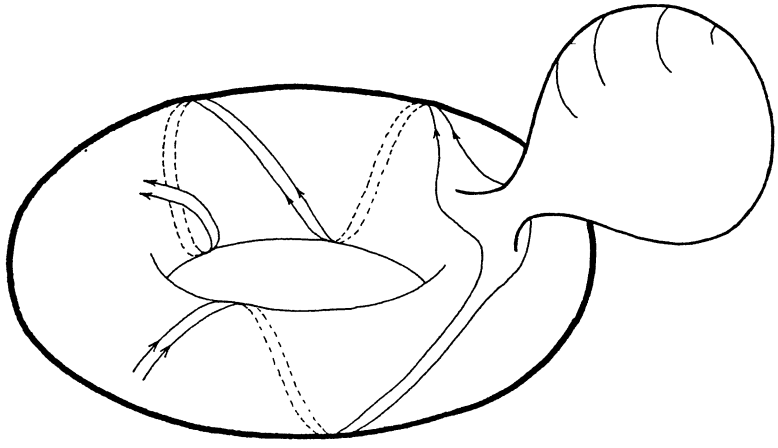

FIG. 10. *Unstable geodesic flow on a torus.*

One can ask whether there is a metric other than the flat one for which all geodesics are of class A, and therefore free from conjugate points. The answer is negative. It was given by E. Hopf [17], who showed: If all geodesics have no conjugate points then the curvature is identically zero, hence the metric is flat.

With this brief excursion into differential geometry we wanted to illustrate the interaction of mathematical ideas in various specialities and the fruitfulness of looking at a problem from different angles. In conclusion we mention that in our discussion many points had to be left out and several problems are left open. We pointed out already that the results are restricted to Hamiltonian systems of two degrees of freedom. No generalization to more degrees of freedom have been found so far, except for the perturbation theory. On the other hand, there is another generalization of this theory to higher dimensions, where the one dimensional orbits are replaced by hypersurfaces of codimension 1 that satisfy nonlinear elliptic partial differential equations (see [29]).

As far as I am aware one has not been able to describe the process of disintegration of an invariant curve into a Cantor set in detail. Efforts in this direction are made with the attempt to use renormalization methods [12]. Also the recent results of Boyland and Hall [5], [15] were not discussed. They established some of the results about periodic orbits with monotonicity property of two-dimensional mappings without use of the area-preserving property, replacing it by topological properties. But for our exposition the variational problem (which leads to the area-preserving property) is central and the concept of the minimal solution is the decisive one with which a subset of orbits can be singled out.

REFERENCES

[1] S. AUBRY AND P. Y. LE DAERON, *The discrete Frenkel–Kontorova model and its extensions* I, Phys. D, 8 (1983), pp. 381–422.

[2] V. BANGERT, *Mather sets for twist maps and geodesics on tori*, Dynamics Reported, 1 (1987), to appear.

[3] G. D. BIRKHOFF, *Surface transformations and their dynamical applications*, Acta Math., 43 (1920), pp. 1–119 resp. §44.

[4] _____, *Sur quelques courbes fermées remarquables*, Bull. Soc. Math. France, 60 (1932), pp. 1–26.

[5] P. L. BOYLAND AND G. R. HALL, *Invariant circles and the order structure of periodic orbits in monotone twist maps*, Preprint of the Boston University, 1985.

[6] S. E. BURKOV AND YA. G. SINAI, *Non-KAM (Kolmogorov–Arnold–Moser) incommensurate states of the discrete Frenkel–Kontorova chain*, J. Physique Lett., 45 (1984), pp. L-409-L-412.

[7] C. CARATHEODORY, *Variationsrechnung*, Teubner, 1935; in particular §443.

[8] A. DENJOY, *Sur les courbes définés par les équations différentielles à la surface du tore*, J. Math. Pures Appl. (9), 11 (1932), pp. 333–375.

[9] I. M. GELFAND AND S. V. FOMIN, *Calculus of Variations*, Prentice–Hall, Englewood Cliffs, NJ, 1963, (esp. Chap. 6, §34).

[10] J. M. GREENE, *Two-dimensional measure-preserving mappings*, J. Math. Phys., 9 (1968), pp. 760–768.

[11] _____, *The Calculation of* KAM-*surfaces*, Ann. New York Acad. Sci., 357 (1980), pp. 80–89.

[12] J. M. GREENE, R. S. MACKAY, F. VIVALDI AND M. J. FEIGENBAUM, *Universal behavior in families of area-preserving maps*, Phys. D, 3 (1981), pp. 468–486.

[13] J. M. GREENE, *A method for determining a stochastic transition*, J. Math. Phys., 20 (1979), pp. 1183–1201.

[14] J. HADAMARD, *Sur les lignes geodesiques*, Proc. verbaux des séances de la Soc. sciences phys. et naturelles, Bordeaux, Paris 8°, 1896/97, pp. 115–117.

[15] G. R. HALL, *A topological version of a theorem of Mather on twist maps*, Ergodic Theory Dynamical Systems, 4 (1982), pp. 583–603.

[16] G. A. HEDLUND, *Geodesics on a two-dimensional Riemannian manifold with periodic coefficients*, Ann. of Math., 33 (1984), pp. 719–739.

[17] E. HOPF, *Closed surfaces without conjugate points*, Proc. Nat. Acad. Sci. U.S.A., 34 (1948), pp. 47–51.

[18] A. KATOK, *Some remarks on Birkhoff and Mather twist map theorems*, Ergodic Theory Dynamical Systems, 2 (1982), pp. 185–194.

[19] A. KLEIN AND LI CHING-TEH, *Semiclassical quantization of nonseparable systems*, J. Math. Phys., 20 (1979), pp. 572–578.

[20] J. N. MATHER, *Existence of asymptotic orbits for area-preserving monotone twist diffeomorphisms*, preliminary manuscript.

[21] _____, *Existence of quasi-periodic orbits for twist homeomorphisms of the annulus*, Topology, 21 (1982), pp. 457–467.

[22] _____, *Nonexistence of invariant circles*, Ergodic Theory Dynamical Systems, 4 (1984), pp. 301–309.

[23] _____, *More Denjoy minimal sets for area-preserving diffeomorphisms*, Comment. Math. Helv., 60 (1985), pp. 508–557.

[24] _____, *Area-preserving twist homeomorphisms of the annulus*, Comment. Math. Helv., 54 (1979), pp. 397–404.

[25] _____, *Concavity of the Lagrangian for Quasi-Periodic Orbits*, Comment. Math. Helv., 57 (1983), pp. 356–376.

[26] _____, *Non-uniqueness of solutions of Percival's Euler–Lagrange equations*, Comm. Math. Phys., 86 (1983), pp. 465–473.

[27] M. MORSE, *A fundamental class of geodesics on any closed surface of genus greater than one*, Trans. Amer. Math. Soc., 26 (1924), pp. 25–60.

[28] J. MOSER, *Monotone twist mappings and the calculus of variation*, Ergodic Theory Dynamical Systems (1986), to appear.

[29] _____, *Minimal solutions of variational problems on a torus*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 3 (1986), pp. 229–272.

[30] _____, *Break-down of stability*, in Lecture Notes in Phys., 247, J. M. Jowett, M. Month and S. Turner, eds., Springer, Berlin, New York, 1986, pp. 492–518.

[31] I. C. PERCIVAL, *A variational principle for invariant tori of fixed frequency*, J. Phys. A, 12 (1979), p. L57.

[32] _____, *Variational principles for invariant tori and Cantori*, in AIP Conf. Proc. No. 57, M. Month and J. C. Herrera, eds., 1980, pp. 302–310.

[33] F. RANNOU, *Numerical study of discrete plane area-preserving mappings*, Astronom. and Astrophys., 31 (1974), pp. 289–301.

[34] C. L. SIEGEL AND J. MOSER, *Lectures on Celestial Mechanics*, Springer–Verlag, New York, Berlin, 1971.

[35] M. R. HERMAN, *Sur les courbes invariantes par les difféomorphismes de l'anneau*, Astérisque, vol. 103–104, 1983, pp. 1–221.