Math. Z. 207, 169–207 (1991)

Mathematische Zeitschrift

Action minimizing invariant measures for positive definite Lagrangian systems

John N. Mather*

Mathematik, ETH-Zentrum, CH-8092, Zürich, Switzerland

Received November 6, 1989; in final form March 12, 1990

In recent years, several authors have studied "minimal" orbits of Hamiltonian systems in two degrees of freedom and of area preserving monotone twist diffeomorphisms. Here, "minimal" means action minimizing. This class of orbits has many interesting properties, as may be seen in the survey article of Bangert [4]. It is natural to ask if there is any generalization of this class of orbits to Hamiltonian systems in more degrees of freedom.

In this article, we propose a generalization to periodic Hamiltonian systems in more degrees of freedom. However, we generalize not the notion of minimal orbit, but the closely related notion of minimal measure, which we introduced in [18].

We obtain two basic results here: an existence theorem for minimal measures, and a regularity theorem which asserts that the minimal measures can be expressed as (partially defined) Lipschitz sections of the tangent bundle.

In the sort of generalization that we do here, a major difficulty is finding the right setting. The setting which we propose here has two important features: the results are valid for periodic positive definite Lagrangian systems, and the results are formulated in terms of invariant measures.

I am indebted to J. Moser for pointing out to me several years ago that periodic positive definite Lagrangian systems in one degree of freedom provide a setting in which it is possible to formulate results which generalize both the author's results [17] (and the closely related results of Aubry and Le Daeron [1]) and the results of Hedlund [12] concerning "class A" geodesics on a Riemannian manifold diffeomorphic to the 2-torus. Indeed, Moser has proved [20] that every twist diffeomorphism is the time one map associated to a suitable periodic positive definite Lagrangian system. Denzler [10] has carried out Moser's program in one degree of freedom. This remark of Moser suggested to me that periodic positive definite Lagrangian systems should provide the right setting in more degrees of freedom.

There is some earlier work in the direction of this paper. Bernstein and Katok [6] obtained results concerning periodic orbits near invariant tori, using a variational method related to the variational method of this paper.

* Supported by NSF grant DMS-8806067.01 and a Guggenheim Fellowship

Also, the recent article of Katok [14] contains results about minimal orbits in more degrees of freedom. Bangert [5] studies minimal (or "class A") geodesics on higher dimensional manifolds. We should also mention the important recent results of M. Herman [13]. Although his methods are not variational, he gives examples showing that the Lipschitz graph property of invariant tori holds only for positive (or negative) definite invariant tori, thus showing that the positive definiteness condition is not just a convenience for the proof, but actually makes a difference in the dynamics.

1 Periodic positive definite Lagrangian systems

Throughout this paper, we let M denote a compact, connected C^∞ manifold, TM its tangent bundle, and $L: TM \times \mathbb{R} \to \mathbb{R}$ a C^2 function, called the "Lagrangian". In all the examples which will be of interest to us, M is a torus, but everything works for arbitrary compact, smooth manifolds. We impose various conditions on L, once and for all.

Periodicity. We suppose that L is periodic in the $\mathbb{R}$ factor. For simplicity, we will suppose that the period is one:

$$L(\xi, t+1) = L(\xi, t), \qquad \xi \in TM, \qquad t \in \mathbb{R}.$$

Positive definiteness. We suppose that L has positive definite fiberwise Hessian second derivative, everywhere. This condition may be expressed in two ways. Here is the simpler: For $m \in M$, let TM_m denote the tangent space to M at m. Our condition is simply that for each $m \in M$ and $t \in \mathbb{R}$, the restriction $L | TM_m \times t$ has everywhere positive definite Hessian second derivative. Here, the Hessian second derivative is taken with respect to any *linear* system of coordinates on T_M. Since TM_m is a vector space, it is meaningful to speak of linear coordinates on TM_m. It is an elementary exercise to show that if the Hessian second derivative is positive definite with respect to one such system, it is positive definite with respect to every such system.

The more classical way to express this condition is to introduce a C^∞ system of local coordinates $x_1, \ldots, x_n$ for an open set U in M. One has local coordinates $x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n$ in $\pi^{-1} U$ canonically associated to these coordinates, where $\pi: TM \to M$ denotes the projection. The more classical form of the positive definiteness condition is to require that the matrix $L_{\dot{x}_i \dot{x}_j}$ of second partial derivatives should be positive definite everywhere that it is defined, and this should hold true for every C^∞ local coordinate system $x_1, \ldots, x_n$.

Superlinear growth. We suppose that L has fiberwise superlinear growth:

$$L(\xi, t)/\|\xi\| \to +\infty, \quad \text{as} \quad \|\xi\| \to +\infty, \quad \text{for } \xi \in TM, t \in \mathbb{R}.$$

Here, $\| \ \|$ denotes the norm associated to a Riemannian metric on M. Since M is compact, this condition is independent of which Riemannian metric is chosen.

The fourth condition is the completeness of the Euler-Lagrange flow associated to L. To explain this condition we must first explain the Euler-Lagrange vector field.

We seek C^1 curves $\gamma: [t_0, t_1] \to M$ which satisfy the variational condition

$$\delta \int L(d\gamma(t), t)\, dt = 0,$$

for the fixed endpoint problem. Here, $d\gamma: [t_0, t_1] \to TM$ denotes the differential of γ. Let us be explicit about what this variational condition means. Consider a C^1 mapping

$$\Gamma: [-\varepsilon, \varepsilon] \times [t_0, t_1] \to M$$

such that $\Gamma(0, t) = \gamma(t)$, for all $t \in [t_0, t_1]$ and $\Gamma(s, t_0) = \gamma(t_0)$ and $\Gamma(s, t_1) = \gamma(t_1)$ for all $s \in [-\varepsilon, \varepsilon]$. The variational condition means that

$$\frac{d}{ds} \int L\left(\frac{\partial \Gamma}{\partial t}(s, t), t\right) dt \bigg|_{s=0} = 0,$$

and this holds for every such Γ. The best known result in the calculus of variations is that such a C^1 curve γ satisfies the variational condition if and only if it is C^2 and satisfies a certain second order differential equation, called the Euler-Lagrange equation. If $x = (x_1, \ldots, x_n)$ is a C^∞ system of local coordinates in an open set U, then the Euler-Lagrange equation has the well known form

$$\frac{d}{dt} L_{\dot{x}} = L_x,$$

where we use L_x as a shorthand expression for the n-tuple $(L_{x_1}, \ldots, L_{x_n})$. It follows from the positive definiteness condition that this equation defines a (time dependent) vector field $E = E_L$ on TM with the property that a C^1 curve $\gamma: [t_0, t_1] \to M$ satisfies the variational condition if and only if $d\gamma: [t_0, t_1] \to TM$ is an integral curve of E. We will call E the *Euler-Lagrange vector field on TM associated to L.*

By our assumption that L is C^2 and has superlinear growth, the Legendre transformation associated to L is a C^1 diffeomorphism of TM onto T^*M. Moreover, the Euler-Lagrange vector field corresponds, under the Legendre transformation, to a vector field on T^*M given by Hamilton's equation. It is easily seen that this vector field is C^1 (see [7], p. 207) even though the Euler-Lagrange vector field may be only C^0. Consequently, the fundamental existence and uniqueness theorem of ordinary differential equations applies, i.e., for each initial condition $(\xi_0, t_0) \in TM \times \mathbb{R}$, there is an integral curve γ of E satisfying the initial condition $\gamma(t_0) = \xi_0$. Moreover, there is a maximal such integral curve $\gamma: (a, b) \to TM$, in the sense that if $\mu: (a', b') \to TM$ is any other integral curve of E satisfying the initial condition $\mu(t_0) = \xi_0$, then $t_0 \in (a', b') \subset (a, b)$, and μ is the restriction of γ to (a', b').

Now we state the fourth condition.

Completeness. Every maximal integral curve of E_L has all of $\mathbb{R}$ as its domain of definition.

For the study of dynamics, it is convenient to introduce a time independent vector field $\tilde{E}_L$ on $P = TM \times (\mathbb{R}/\mathbb{Z})$. This is definied by

$$\tilde{E}_L(\xi, \theta) = (E_L(\xi, t), \partial/\partial\theta),$$

if $\theta \equiv t \pmod 1$. This is well defined in view of the periodicity of L (and therefore of E_L).

Completeness means that the flow $\Phi = \Phi_E = \Phi_L$ associated to $\tilde{E}_L$ is defined on all of $P \times \mathbb{R}$; it is the C^1 mapping $\Phi: P \times \mathbb{R} \to P$ uniquely defined by the conditions $\Phi | P \times 0 = \text{identity}$ and

$$d\Phi(p, t)/dt = \tilde{E}_L(\Phi(p, t), t),$$

for $p \in P$ and $t \in \mathbb{R}$. We call Φ_L the *Euler-Lagrange flow* associated to L.

In this paper, we will obtain certain properties of the dynamics of Φ_L, assuming that L satisfies the conditions listed above. These generalize results previously obtained by the author for twist maps. Our basic results are an existence theorem and a regularity theorem for minimal measures. We state and prove the existence theorem in § 2 and the regularity theorem in § 4. In § 5, we give an application to the dynamics of a perturbation of a system which has an invariant torus. In § 6, we discuss how certain results concerning twist diffeomorphisms follow from the results obtained in §§ 2, 3, 4.

2 Minimal measures

Let $P^* = P \cup \infty$ denote the one point compactification of $P = TM \times (\mathbb{R}/\mathbb{Z})$. The Euler-Lagrange flow Φ_L extends to a flow on P^* which fixes ∞. We continue to denote this extended flow by Φ_L. We let $\mathfrak{M}_L$ denote the set of Φ_L-invariant probability measures on P^*. In this section, we will prove the existence of elements of $\mathfrak{M}_L$ which minimize various functions on $\mathfrak{M}_L$.

A basic result in functional analysis (the Riesz representation theorem) states that the set of Borel probability measures on a compact metric space X is a subset of the dual $C(X)^*$ of the Banach space $C(X)$ of continuous functions on X. (See Lanford [16] for a nice exposition of this and related results from functional analysis which we will be using.) It is obviously a convex set and it is well known to be metrizable and compact with respect to the weak topology on $C(X)^*$ defined by $C(X)$, also called the weak-* topology. The restriction of this topology to the set of Borel measures is frequently called the *vague* topology on measures.

Since P^* is metrizable, as well as compact, it follows that the set of Borel probability measures on P^* is a metrizable, compact, convex subset of the dual of the Banach space of continuous functions on P^*. The set $\mathfrak{M}_L$ is obviously a compact, convex subset of this set.

A result of Kryloff and Bogoliuboff [15] states that any flow Ψ on a compact metric space X has an invariant measure. (See also [22, Chapt. VI, § 9].) For the case we are considering, i.e. the Euler-Lagrange flow Φ_L on P^*, this result tells us nothing, since ∞ is a fixed point, so the atomic measure supported on ∞ is invariant.

Nonetheless, its proof will be useful to us, so we repeat it here. Let γ_n be a trajectory of the flow Ψ defined on a time interval of length n and let μ_n be the probability measure evenly distributed along γ_n. Clearly,

$$\| \Psi_{t*} \mu_n - \mu_n \| \leqq 2t/n.$$

Let μ be a point of accumulation of μ_n, as $n \to \infty$, with respect to the vague topology. For any continuous function u on X, any $t \in \mathbb{R}$, and any $n_0, \varepsilon > 0$, there exists $n \geqq n_0$ such that

$$|\int u \circ \Psi_s \, d\mu - \int u \circ \Psi_s \, d\mu_n| < \varepsilon, \qquad \text{for } s = 0, t.$$

It follows that

$$\begin{aligned} |\int u \circ \Psi_t \, d\mu - \int u \, d\mu| &\leqq 2\varepsilon + |\int u \circ \Psi_t \, d\mu_n - \int u \, d\mu_n| \\ &\leqq 2\varepsilon + \|u\| \, \| \Psi_{t*} \mu_n - \mu_n \| \leqq 2\varepsilon + 2t \|u\|/n. \end{aligned}$$

Since n_0 may be taken arbitrarily large and ε arbitrarily small, it follows that

$$|\int u \circ \Psi_t \, d\mu - \int u \, d\mu| = 0,$$

i.e., μ is Ψ-invariant.

This argument shows that any point of accumulation of the μ_n, as $n \to \infty$, is a Ψ-invariant measure.

We may apply this argument to the Euler-Lagrange flow Φ_L to obtain invariant measures other than the atomic measure supported at ∞. Specifically, for $\mu \in \mathfrak{M}_L$, we define the *average action* of μ as

$$A(\mu) = \int L \, d\mu,$$

where we set $L(\infty) = \infty$. Since L is bounded below, this integral exists, although it may be $+\infty$. We will prove the existence of $\mu \in \mathfrak{M}_L$ for which $A(\mu) < \infty$.

In fact, the existence of such a μ follows almost immediately from the above argument and a theorem of Tonelli which guarantees the existence of curves $\gamma: [a, b] \to M$ which minimize $\int_a^b L(d\gamma(t), t) \, dt$ subject to a fixed boundary condition. For our purposes it is useful to have a form of Tonelli's theorem concerning curves on a covering space of M.

Let $\tilde{M}$ be a covering space of M. If $[a, b]$ is a finite interval and $\gamma: [a, b] \to \tilde{M}$ is an absolutely continuous curve, we define its *action* as

$$A(\gamma) = \int_a^b L(d\pi\gamma(t), t) \, dt,$$

where $\pi: \tilde{M} \to M$ denotes the projection. (In what follows, we will omit π.) Since γ is absolutely continuous, $d\gamma(t)$ exists for almost all t, and is a measurable function of t. Since L is bounded below, the above integral exists, although it may be $+\infty$.

Tonelli's theorem. *Let $a < b \in \mathbb{R}$ and let $x_a, x_b \in \tilde{M}$. The action takes a finite minimum value over the set of absolutely continuous curves $\gamma: [a, b] \to \tilde{M}$ such that $\gamma(a) = x_a$, $\gamma(b) = x_b$.*

For Tonelli's theorem, it is enough to assume the hypotheses of positive definiteness and superlinear growth. However, Ball and Mizel [3] have shown that if only these hypotheses are assumed, a (Tonelli) minimizer need not be C^1. On the other hand, the hypothesis of completeness guarantees that the minimizers are C^1 and therefore satisfy the Euler-Langrange equation. We will explain why this is so shortly.

The proof of Tonelli's theorem is based on a lower semi-continuity property of the action. We will also state below an addendum to the lower semi-continuity property which will be useful later.

For this we need to introduce a couple of metrics on the space of absolutely continuous curves in $\tilde{M}$. We choose, once and for all, a C^∞ Riemannian metric on M. This gives rise, in a canonical way, to a Riemannian metric on TM: if $\xi \in TM$, then the Riemannian connection on M gives rise to a direct sum splitting $T_\xi(TM) = T_x M \oplus T_x M$, where x is the projection of ξ on M. Here, the first summand is the tangent space at ξ to the fiber over x of the projection $TM \to M$ and the second summand is the image of the linear mapping $T_x M \to T_\xi(TM)$ given by the Riemannian connection. We provide TM with the unique Riemannian metric for which the summands are orthogonal and the restriction of the metric to each summand is the inner product given by the Riemannian metric on M. Given $\gamma_0, \gamma_1 : [a, b] \to M$, we set

$$d_0(\gamma_0, \gamma_1) = \sup\{\operatorname{dist}(\gamma_0(t), \gamma_1(t)) \colon t \in [a, b]\}$$

$$d_{ac}(\gamma_0, \gamma_1) = \int_a^b \operatorname{dist}(d\gamma_0(t), d\gamma_1(t))\, dt.$$

In the first formula "dist" means the distance function defined by the Riemannian metric on M; in the second formula, it means the distance function defined by the Riemannian metric on TM.

Clearly, d_0 is a metric on the space $C^0([a, b], M)$ of continuous curves $[a, b] \to M$; its underlying topology is what is variously called the uniform topology, the compact-open topology, or the C^0-topology. Likewise, d_{ac} is a metric on the space space $C^{ac}([a, b], M)$ of absolutely continuous curves $[a, b] \to M$; we will call its underlying topology the C^{ac}-topology.

Note that changing the Riemannian metric on M changes d_0 and d_{ac} only within their equivalence classes, i.e. there exists a constant C such that

$$C^{-1} d_0 \leqq d_0' \leqq C d_0$$
$$C^{-1} d_{ac} \leqq d_{ac}' \leqq C d_{ac},$$

where d_0' and d_{ac}' are the new metrics.

Tonelli's theorem follows immediately from:

Lemma. *Let $K \in \mathbb{R}$. The set $\{A \leqq K\}$, consisting of all $\gamma \in C^{ac}([a, b], M)$ for which $A(\gamma) \leqq K$, is compact in the C^0-topology.*

To obtain Tonelli's theorem from this lemma, we remark that it is obvious that the set S_K of absolutely continuous curves $\gamma \colon [a, b] \to \tilde{M}$ such that $A(\gamma) \leqq K$, $\gamma(a) = x_a$, and $\gamma(b) = x_b$ is non-empty for large enough K. As K decreases, S_K decreases. It follows easily from the lemma that these sets are compact; consequently, there is a smallest one. Any member of the smallest one is a (Tonelli) minimizer.

The lemma is a semi-continuity result: it implies that if $\gamma_1, \gamma_2, \ldots$ is a sequence in $C^{ac}([a, b], M)$ which converges C^0 to γ, then $\gamma \in C^{ac}([a, b], M)$ and $A(\gamma) \leqq \liminf A(\gamma_i)$. The following addendum to this lemma will be useful later.

Addendum. *If $\gamma_1, \gamma_2, \ldots$ is a sequence in $C^{ac}([a,b], M)$ which converges C^0 to γ and $A(\gamma_i)$ converges to $A(\gamma)$, then $\gamma_1, \gamma_2, \ldots$ converges in the C^{ac} – topology to γ.*

We will prove the lemma and its addendum in Appendix 1.

In addition to Tonelli's theorem, we need the more ancient result due to Weierstrass that sufficiently short solutions of the Euler-Lagrange equation are *strict* minimizers, i.e. any sufficiently short solution has the property that it not only minimizes the action subject to the boundary conditions, but it is the unique curve to do so.

Theorem. (Weierstrass). *For any $K>0$, there exist ε, $C_0, C_1>0$, such that if $a<b\leqq a+\varepsilon$, and $\gamma: [a,b] \to \tilde{M}$ is a solution of the Euler-Lagrange equation satisfying $\|d\gamma(t)\| \leqq K$, for all $t \in [a,b]$, then*

$$A(\gamma_1) \geqq A(\gamma) + F(d_{ac}(\gamma, \gamma_1))$$

for any absolutely continuous curve $\gamma_1: [a,b] \to \tilde{M}$ such that $\gamma_1(a)=\gamma(a)$ and $\gamma_1(b)=\gamma(b)$. Here,

$$F(t)=\min(C_0\, t^2, C_1\, t).$$

Moreover, still assuming $b-a\leqq\varepsilon$, we have that for any $x_a, x_b \in \tilde{M}$ such that $\operatorname{dist}(x_a, x_b) \leqq K(b-a)/2$, there exists a solution γ of the Euler-Lagrange equation satisfying $\gamma(a)=x_a$, $\gamma(b)=x_b$, and $\|d\gamma(t)\| \leqq K$, for all $t \in [a,b]$.

We sketch a proof in Appendix 2.

Now let $\gamma: [a,b] \to \tilde{M}$ be a minimizer. Let $t \in [a,b]$. We have one of the following two alternatives:

1) $\operatorname{dist}(\gamma(s), \gamma(t))/|t-s| \to \infty$ as $s \to t$, and $\|d\gamma(s)\| \to \infty$ as $s \to t$ over the set of points where $d\gamma(s)$ exists, or

2) γ is C^1 and satisfies the Euler-Lagrange equation in a neighborhood of t.

For, if $\operatorname{dist}(\gamma(s), \gamma(t))/|t-s|$ does not tend to ∞ as $s \to t$, then there exists $K>0$, and $a<t<b$ with $b-a$ arbitrarily small, such that $\operatorname{dist}(\gamma(a), \gamma(b)) \leqq K(b-a)/2$. By Weierstrass's theorem, there exists a solution γ_1 of the Euler-Lagrange equation with $\gamma_1(a)=\gamma(a)$, $\gamma_1(b)=\gamma(b)$, and γ_1 is a strict minimizer. Since γ is a minimizer, it follows that $\gamma_1(t)=\gamma(t)$ for $a \leqq t \leqq b$, and we have the second alternative.

Also, if $\|d\gamma(s)\|$ has a finite point of accumulation as $s \to t$, then $d\gamma(s)$ has a point $v \in T_{\gamma(t)} M$ of accumulation as $s \to t$. Let $s_i \to t$ and $d\gamma(s_i) \to v$. By the existence theorem for ordinary differential equations, there exists $\delta>0$ and for eauh i a solution $\gamma_i: [s_i-\delta, s_i+\delta] \to \tilde{M}$ of the Euler-Lagrange equation such that $d\gamma_i(s_i)=d\gamma(s_i)$. By choosing i large enough, we may suppose that $t \in [s_i - \delta, s_i+\delta]$ There is a maximal interval containing s_i on which γ_i and γ coincide. By what we showed in the previous paragraph, this interval is open in $[s_i-\delta, s_i+\delta]$. By continuity, it is closed in $[s_i-\delta, s_i+\delta]$. Therefore, $\gamma=\gamma_i$ on $[s_i-\delta, s_i+\delta]$, so we again have the second alternative.

For all this, we need only suppose positive definiteness and superlinear growth of L. Now, if we add the hypothesis of completeness of the Euler-Lagrange flow, we see that we cannot have the first alternative; specifically, we cannot have that $\|d\gamma(s)\| \to \infty$ as $s \to t$ over the set of points where $d\gamma(s)$ exists. Consequently, any minimizer is a solution of the Euler-Lagrange equation.

Now we prove a preliminary result:

Proposition. *There exists $\mu \in \mathfrak{M}_L$ such that $A(\mu) < \infty$.*

Proof. Let α_n be an absolute mimimizer (i.e. with free boundaries) defined on a time interval of length n. Let $\gamma_n(t) = (d\alpha_n(t), t)$. By what we have just shown, γ_n is a trajectory of the Euler-Lagrange flow. Let μ_n be the probability measure evenly distributed along γ_n. Let μ be a (vague) point of accumulation of μ_n as $n \to \infty$. Our previous argument shows that μ is Φ_L-invariant.

Clearly, there exists $C > 0$ and for each n a C^1 curve β_n defined on a time interval of length n such that $A(\beta_n) \leqq Cn$. Hence,

$$A(\mu_n) = n^{-1} A(\alpha_n) \leqq n^{-1} A(\beta_n) \leqq C.$$

Therefore, it will be enough to show that $A(\mu) \leqq C$. But this is an immediate consequence of the following:

Lemma. *$A(\mu) = \int L\, d\mu$ is a lower semi-continuous function on the set of probability measures on P^*, provided with the vague topology.*

Proof. Let $A_K(\mu) = \int \min(L, K)\, d\mu$, for $K \in \mathbb{R}$. Then A_K is continuous, and $A_K \uparrow A$ as $K \uparrow \infty$. □

The lemma has the following immediate consequence: there exists $\mu \in \mathfrak{M}_L$ which minimizes A over $\mathfrak{M}_L$.

Next, we define the rotation vector $\rho(\mu)$ of a Φ_L-invariant probability measure μ: Let λ be a C^∞ 1-form on M. We may regard λ as a mapping $TM \to \mathbb{R}$ which is linear on the fibers. We will continue to denote the composition of this mapping with the projection $P = TM \times (\mathbb{R}/\mathbb{Z}) \to TM$ by the same symbol. If μ is a Borel probability measure on P such that $\int L\, d\mu < \infty$, then $\lambda \in L^1(\mu)$, since L has fiberwise superlinear growth.

Lemma. *If λ is exact and μ is Φ_L-invariant, then $\int \lambda\, d\mu = 0$.*

Proof. Let $\lambda = du$, where u is a C^∞ function on M. Let $v \in TM$, $\theta \in \mathbb{R}/\mathbb{Z}$, $s \in \mathbb{R}$ and let v_s denote the TM component of $\Phi_s(v, \theta) \in TM \times (\mathbb{R}/\mathbb{Z})$, where Φ (as usual) denotes the Euler-Lagrange flow. Then $\lambda(v_s) = v_s \cdot u$ (i.e. the directional derivative of u in the direction $v_s) = du(\pi v_s)/ds$, where $\pi\colon TM \to M$ denotes the projection. This last equation follows from $d(\pi v_s)/ds = v_s$, which is a consequence of the definition of the Euler-Lagrange flow. Hence

$$\begin{aligned}\int \lambda\, d\mu &= T^{-1} \int_0^T ds \int \lambda \Phi_s^*\, d\mu = T^{-1} \int_0^T ds \int (\lambda \circ \Phi_s)\, d\mu \\ &= T^{-1} \int_0^T ds \int \lambda(v_s)\, d\mu(v) = T^{-1} \int_0^T ds \int (du(\pi v_s)/ds)\, d\mu(v) \\ &= T^{-1} \int [u(\pi v_T) - u(\pi v)]\, d\mu(v) \to 0, \quad \text{as} \quad T \to \infty.\end{aligned}$$

Since $\int \lambda\, d\mu$ is independent of T, we obtain $\int \lambda\, d\mu = 0$. □

Corollary. *If μ is Φ_L-invariant, there exists $\rho(\mu) \in H_1(M, \mathbb{R})$ such that*

$$\langle [\lambda], \rho(\mu) \rangle = \int \lambda \, d\mu,$$

for every closed 1*-form λ on M, where $[\lambda]$ denotes the de Rham cohomology class of λ, and $\langle , \rangle$ denotes the canonical pairing between cohomology and homology.* □

Thus, to every $\mu \in \mathfrak{M}_L$ such that $A(\mu) < \infty$, we have associated $\rho(\mu) \in H_1(M, \mathbb{R})$. This is called the *rotation vector* of μ. It is similar to the rotation vector defined by Schwartzman [26].

If $c \in H^1(M, \mathbb{R})$, we set

$$A_c(\mu) = A(\mu) - \langle c, \rho(\mu) \rangle = \int (L - \lambda) \, d\mu,$$

where λ is a closed 1-form on M such that $[\lambda] = c$. This is defined for any $\mu \in \mathfrak{M}_L$ such that $A(\mu) < \infty$. We extend it to the case $A(\mu) = \infty$, by setting $A_c(\mu) = \infty$, in this case.

Note that $L - \lambda$ satisfies the conditions we have imposed in § 1 on L and that the Euler-Lagrange flow of $L - \lambda$ is the same as that of L. The last point may be seen by observing that the variational equations $\delta \int L \, dt = 0$ and $\delta \int (L - \lambda) \, dt = 0$ for the fixed endpoint problem clearly have the same solutions, since λ is closed.

Consequently, A_c is lower semi-continuous (for the same reason A is). Therefore A_c takes a minimum value, which we denote by $-\alpha(c)$.

It is easily verified that $\alpha(c)$ is a convex function on $H^1(M, \mathbb{R})$, in the sense that its *epigraph* $\{(c, z) : z \geqq \alpha(c)\}$ is a convex subset of $H^1(M, \mathbb{R})$.

Let $\alpha^* : H_1(M, \mathbb{R}) \to \mathbb{R}$ denote the *conjugate* function of α in the sense of convex analysis, i.e.

$$-\alpha^*(h) = \min \{\alpha(c) - \langle c, h \rangle\},$$

where c ranges over $H^1(M, \mathbb{R})$. Clearly, α^* takes values in $\mathbb{R} \cup \{+\infty\}$; we will prove in a moment that it takes its values in $\mathbb{R}$.

It follows from the definitions that if $\mu \in \mathfrak{M}_L$ and $A(\mu) < \infty$, then $\alpha^*(\rho(\mu)) \leqq A(\mu)$, i.e. $\alpha^*(h)$ is a lower bound for invariant probability measures of rotation vector h. Thus, to prove that α^* takes its values in $\mathbb{R}$, it is enough to prove that for every $h \in H_1(M, \mathbb{R})$ there is an invariant probability measure μ with $A(\mu) < \infty$ of rotation vector h. First, we prove a technical result:

Lemma. *If $C \in \mathbb{R}$ and λ is a* 1*-form on M, then the mapping $\mu \to \int \lambda \, d\mu$ is a continuous function on the set of probability measures μ on P^* such that $\int L \, d\mu \leqq C$.*

Proof. Let $\varepsilon > 0$. Since L has superlinear growth, there is a continuous function λ_ε on P^* such that $|\lambda - \lambda_\varepsilon| \leqq \varepsilon (C + B)^{-1} (L + B)$ everywhere on P^*, where $-B$ is a lower bound for L. Then

$$\int |\lambda - \lambda_\varepsilon| \, d\mu \leqq \varepsilon (C + B)^{-1} \int (L + B) \, d\mu \leqq \varepsilon,$$

for every probability measure μ such that $\int L \, d\mu \leqq C$. Since λ_ε is a continuous function on P^*, $\mu \to \int \lambda_\varepsilon \, d\mu$ is continuous (with respect to the vague topology).

We have shown that $\mu \to \int \lambda d\mu$ may be uniformly approximated by continuous functions on the set of probability measures μ for which $\int L d\mu \leqq C$. Hence, it is continuous on that set. □

Proposition. *Let $h \in H_1(M, \mathbb{R})$. There exists $\mu \in \mathfrak{M}_L$ with $A(\mu) < \infty$ and $\rho(\mu) = h$.*

Proof. This time, we apply Tonelli's theorem, not on M, but on the covering space $\tilde{M}$ of M defined by $\pi_1(\tilde{M}) = \ker(\mathfrak{H}: \pi_1(M) \to H_1(M, \mathbb{R}))$. Here $\mathfrak{H}$ denotes the Hurewicz homomorphism. The group of Deck transformations of this covering space is

$$H = im(\mathfrak{H}: \pi_1(M) \to H_1(M, \mathbb{R})).$$

Let $T_1, \ldots, T_n$ be a sequence of Deck transformations such that

$$n^{-1} T_n \to h \in H_1(M, \mathbb{R}), \quad \text{as} \quad n \to +\infty.$$

Let $\tilde{x}_0 \in \tilde{M}$. Let $\tilde{x}_n = T_n \tilde{x}_0$. Let $\tilde{\alpha}_n: [0, n] \to \tilde{M}$ minimize $\int_0^n L(d\alpha_n(t), t)\, dt$ subject to the boundary conditions $\tilde{\alpha}_n(0) = \tilde{x}_0$ and $\tilde{\alpha}_n(n) = \tilde{x}_n$, where α_n is the projection of $\tilde{\alpha}_n$ on M. As before, let $\gamma_n(t) = (d\alpha_n(t), t \bmod 1)$. Note that $\tilde{\alpha}_n$ exists by Tonelli's theorem and is C^1 by the completeness hypothesis.

Now we proceed just as before: we let μ_n denote the probability measure evenly distributed along γ_n and we let μ be a point of accumulation of μ_n as $n \to \infty$. Just as before, there exists C such that $\int L d\mu_n \leqq C$ for all n, and hence $A(\mu) = \int L d\mu \leqq C$, by the lower semi-continuity of A. From the specification of the endpoints of $\tilde{\alpha}_n$, it follows that $\int \lambda d\mu_n = \langle [\lambda], n^{-1} T_n \rangle$. By the lemma, we may pass to the limit: $\int \lambda d\mu = \langle [\lambda], h \rangle$, and hence $\rho(\mu) = h$. □

Corollary. *α^* is finite everywhere on $H_1(M, \mathbb{R})$.* □

Now we recall the basic results of convex analysis [23]: A function f on a finite dimensional vector space with values in $\mathbb{R} \cup \{\infty\}$ is said to be *convex* if its epigraph is convex. We may define its conjugate as before. Such a function is said to have *superlinear growth* if $f(x)/\|x\| \to +\infty$ as $\|x\| \to \infty$. It is easy to see that f is everywhere finite if and only if f^* has superlinear growth and f^* is everywhere finite if and only if f has superlinear growth. The epigraph of f^{**} is the closure of the epigraph of f.

In our case α and α^* are everywhere finite, so both have superlinear growth and $\alpha^{**} = \alpha$.

Let $E \subset H_1(M, \mathbb{R}) \times \mathbb{R}$ denote the set of all pairs $(\rho(\mu), z)$ such that $\mu \in \mathfrak{M}_L$ and $A(\mu) \leqq z$. Since L has a lower bound, the projection of E on $\mathbb{R}$ has the same lower bound. Obviously, E is convex.

Note that $\mu \to \rho(\mu)$ is continuous on $\{\mu \in \mathfrak{M}_L : A(\mu) \leqq C\}$, by the previous lemma. Since $\mathfrak{M}_L$ is compact, and A is lower semi-continuous, it follows easily that E is closed. Since E is closed, convex, and bounded below, it is the epigraph of a convex function $\beta: H_1(M, \mathbb{R}) \to \mathbb{R}$. It follows directly from the definitions that $\alpha = \beta^*$. By duality $\beta = \alpha^*$.

From the above discussion, we obtain:

Theorem 1. *The functions $\alpha: H^1(M, \mathbb{R}) \to \mathbb{R}$ and $\beta: H_1(M, \mathbb{R}) \to \mathbb{R}$ are conjugate convex functions and have superlinear growth. For $h \in H_1(M, \mathbb{R})$, we have*

$$\beta(h) = \min\{A(\mu): \mu \in \mathfrak{M}_L \quad \text{and} \quad \rho(\mu) = h\}.$$

For $c \in H^1(\mathbb{M}, \mathbb{R})$, *we have*

$$-\alpha(c) = \min\{A_c(\mu) : \mu \in \mathfrak{M}_L\}. \quad \square$$

Note that in both formulas the minimum is achieved because E is closed.

Note that if $\mu \in \mathfrak{M}_L$, $A(\mu) = \beta(\rho(\mu))$ if and only if there exists $c \in H^1(M, \mathbb{R})$ such that μ minimizes $A_c(\mu)$. Moreover, c is the subderivative of β at $\rho(\mu)$, i.e. the slope of a supporting hyperplane of the epigraph of β at $\rho(\mu)$. We say that μ is a *minimal measure* if either of these equivalent conditions is satisfied.

We conclude this section with a few words concerning the significance of these results.

The only invariant probability measures which have significance for dynamics are the *ergodic* measures. These are defined by the condition that every invariant Borel set should have measure 0 or 1. It is well known that the extremal points of $\mathfrak{M}_L$ are the ergodic measures for the flow Φ_L on P^*. (More generally, for any flow on a compact metric space, the invariant measures form a compact, convex set with respect to the vague topology and the ergodic measures are the extremal points of this set. See [15], [16], or [22, Chapt. VI, § 9].)

Since β has superlinear growth, its epigraph has infinitely many extremal points. Let $(h, \beta(h))$ denote an extremal point of the epigraph of β. The extremal points of the set of $\mu \in \mathfrak{M}_L$ for which $\rho(\mu) = h$ and $A(\mu) = \beta(h)$ are ergodic measures, since they are extremal points of $\mathfrak{M}_L$. Since this set is compact and convex it has extremal points. In other words, we have shown that if $(h, \beta(h))$ is an extremal point of the epigraph of β, then there exists at least one *ergodic* minimal measure with rotation vector h.

For such an ergodic measure μ, i.e. one with $\rho(\mu) = h$ and $A(\mu) = \beta(h)$, Birkhoff's ergodic theorem implies that μ almost every trajectory γ of Φ_L has rotation vector h, i.e.

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \lambda(\gamma(t))\, dt = \langle [\lambda], \rho(\mu) \rangle.$$

for every closed 1-form λ on M (where, as before, we think of λ as a function from P to $\mathbb{R}$).

3 Minimizers and minimal measures

Let $\tilde{M}$ be the covering space of M defined by $\pi_1(\tilde{M}) = \ker \mathfrak{H}$, where $\mathfrak{H}: \pi_1(M) \to H_1(M, \mathbb{R})$ is the Hurewicz homomorphism. In this section, we consider minimizers on $\tilde{M}$, i.e. curves $\zeta: [a, b] \to \tilde{M}$ which minimize the action $A(\zeta) = \int_a^b L(d\zeta(t), t)\, dt$, over the class of absolutely continuous curves having the same endpoints. Here, and subsequently, we will denote the pull-back to $\tilde{M}$ of a function (such as L) or a form on M by the same symbol. We will describe certain relations between minimizers on $\tilde{M}$ and minimal measures on M.

Let $h_1, \ldots, h_l$ be a free basis of the group $H = \operatorname{im}(\pi_1(M) \to H_1(M, \mathbb{R}))$ of Deck transformations of $\tilde{M}$ over M and let $\lambda_1, \ldots, \lambda_l$ be closed 1-forms on M, whose cohomology classes $[\lambda_1], \ldots, [\lambda_l]$ are the dual basis of $H^1(M, \mathbb{R})$.

If $x, y \in \tilde{M}$ and $\zeta:[a,b] \to \tilde{M}$ is a C^1 curve connecting x to y, we define the *difference vector* $y-x \in H_1(M, \mathbb{R})$ by

$$\langle [\lambda_i], y-x \rangle = \int_a^b \lambda_i(d\zeta(t))\, dt$$

and the *rotation vector*

$$\rho(\zeta) = (y-x)/(b-a).$$

Obviously, the difference vector $y-x$ is independent of the choice of ζ. Of course, these are not intrinsic notions: they depend on the choice of 1-forms $\lambda_1, \ldots, \lambda_l$.

Proposition 1 *Consider a sequence* $\zeta_i:[a_i, b_i] \to \tilde{M}$, $i=1,2,\ldots$ *of minimizers. Suppose that* $\rho(\zeta_i) \to h \in H_1(M, \mathbb{R})$, *and* $b_i - a_i \to \infty$, *as* $i \to \infty$. *Then* $A(\zeta_i)/(b_i - a_i) \to \beta(h)$, *as* $i \to \infty$.

Recall that by Theorem 1, $\beta(h) = A(\mu)$ for any minimal measure μ such that $\rho(\mu) = h$.

Proof. First, suppose that $\liminf A(\zeta_i)/(b_i - a_i) < \beta(h)$. Let $\pi: \tilde{M} \to M$ denote the projection. Let $\gamma_i(t) = (d\pi\, \zeta_i(t), t \bmod 1)$ so that $\gamma_i:[a_i, b_i] \to P$ is a trajectory of the Euler-Lagrange flow. Let μ_i be the probability measure evenly distributed along γ_i. By passing to a subsequence, we may suppose that the sequence $\mu_1, \mu_2, \ldots$ converges vaguely to a probability measure μ and $A(\mu_i) = A(\zeta_i)/(b_i - a_i)$ converges to a number $< \beta(h)$.

By a lemma proved in the last section, ρ is continuous on sets where A is bounded. Consequently, $\rho(\mu) = \lim \rho(\mu_i) = \lim \rho(\gamma_i) = h$. By the semi-continuity of A, also proved in the last section, we have that

$$A(\mu) \leqq \lim A(\mu_i) < \beta(h),$$

which is impossible since $\beta(h)$ is the minimum of $A(\mu)$ for measures for which $\rho(\mu) = h$. This contradiction shows that $\liminf A(\zeta_i)/(b_i - a_i) \geqq \beta(h)$.

Since β has superlinear growth, any point of the epigraph of β may be expressed as a convex combination of extremal points of the epigraph. In particular, there exist $h_1, \ldots, h_k \in H_1(M, \mathbb{R})$ such that $(h_i, \beta(h_i))$ are extremal points of the epigraph of β, for $i=1, \ldots, k$, and $\tau_1, \ldots, \tau_k > 0$, $\Sigma \tau_i = 1$ such that $h = \Sigma \tau_i h_i$, $\beta(h) = \Sigma \tau_i \beta(h_i)$.

Since the $(h_i, \beta(h_i))$ are extremal points of the epigraph of β, there exist ergodic probability measures $\mu_1, \ldots, \mu_k$ such that $\rho(\mu_i) = h_j$ and $A(\mu_j) = \beta(h_j)$. Since $A(\mu_j) < \infty$, we have $L \in L^1(\mu_j)$ and $\lambda_1, \ldots, \lambda_l \in L^1(\mu_j)$. Consequently, we may apply Birkhoff's ergodic theorem to these functions. Let γ_j be a trajectory of the Euler-Lagrange flow and let $\gamma_j^T = \gamma_j | [-T, T]$. According to Birkhoff's ergodic theorem,

$$A(\gamma_j^T)/2T \to A(\mu_j), \rho(\gamma_j^T) \to \rho(\mu_j),$$

as $T \to \infty$, for μ_j-almost every trajectory γ_j. For the subsequent discussion, we choose one such trajectory, for each $j=1, \ldots, k$. Let ξ_j be a lift to $\tilde{M}$ of the projection of γ_j on M.

For each sufficiently large i, we choose

$$a_i < a_{i1} < b_{i1} < a_{i2} < b_{i2} < \ldots < a_{ik} < b_{ik} < b_i.$$

We choose them so that $(b_{ij}+a_{ij})/2$ is an integer. We set $T_{ij}=(b_{ij}-a_{ij})/2$. We let $\xi_{ij}:[a_{ij}, b_{ij}] \to \tilde{M}$ have the form $\xi_{ij}(t)=D_{ij}\,\xi_j(t+(b_{ij}+a_{ij})/2)$, where D_{ij} is a Deck transformation of $\tilde{M}$ over M, chosen to satisfy certain properties which will be stated below. We let $x_{ij}, y_{ij} \in \tilde{M}$ be the endpoints of ξ_{ij}.

We construct $\xi_i^*:[a_i, b_i] \to \tilde{M}$ in the following way. We join $\xi_i(a_i)$ to x_{i1} by a minimizer, x_{i1} to y_{i1} by ξ_{i1}, y_{i1} to x_{i2} by a minimizer, and so on, thus filling in the intervals $[a_{ij}, b_{ij}]$ by the ξ_{ij} and filling in the complementary intervals by minimizers.

We assert that we may make this construction in such a way that $(b_i - a_i)^{-1} A(\xi_i^*) \to \beta(h)$, as $i \to \infty$. Since ζ_i is a minimizer, we have $A(\zeta_i) \leqq A(\xi_i^*)$. Since we have already proved that $\liminf A(\zeta_i)/(b_i - a_i) \geqq \beta(h)$, this will show that $\lim A(\zeta_i)/(b_i - a_i) = \beta(h)$, which is what was to be proved.

To achieve $\lim (b_i - a_i)^{-1} A(\xi_i^*) = \beta(h)$, we make the choices so that $b_{ij} - a_{ij} = \tau_j T_i$, for some number T_i. Since $\Sigma \tau_j = 1$, we have that $T_i < b_i - a_i$. We make the choices so that $T_i/(b_i - a_i) \to 1$, as $i \to \infty$, but so that the rate of convergence is slower than the rate of convergence of $\rho(\gamma_j^{T_{ij}})$ to $\rho(\mu_j)$ and the rate of convergence of $\rho(\zeta_i)$ to h. By making an appropriate choice of the Deck transformations D_{ij} above, and placing the intervals $[a_{ij}, b_{ij}]$ appropriately in $[a, b]$, subject to the above restrictions, we may then arrange that $\|x_{i,j+1} - y_{ij}\|/(a_{i,j+1} - b_{ij})$ is bounded independently of i and j, where we set $b_{i0} = a_i$, $y_{i0} = \xi_i(a_i)$, $a_{i,k+1} = b_i$, $x_{i,k+1} = \xi_i(b_i)$ and $\| \;\|$ denotes a norm on $H_1(M, \mathbb{R})$ which is fixed, once and for all. It is possible to make these choices because $h = \Sigma \tau_j h_j = \Sigma \tau_j \rho(\mu_j)$, and the convergence of $\rho(\gamma_j^{T_{ij}})$ to $\rho(\mu_j)$ and of $\rho(\zeta_i)$ to h is faster than the convergence of $T_i/(b_i - a_i)$ to 1.

If we make the choices in this way, we have $\lim (b_i - a_i)^{-1} A(\zeta_i^*) = \beta(h)$. For, the integral of L over the intervals $[a_{i,j+1}, b_{ij}]$ makes a negligible contribution to $(b_i - a_i)^{-1} A(\zeta_i^*)$, in the limit. Since $A(\gamma_j^T)/2T \to A(\mu_j)$, the limit of the contribution of the integral of L over the other intervals is

$$\Sigma \tau_j A(\mu_j) = \Sigma \tau_j \beta(h_j) = \beta(h). \quad \square$$

Corollary. *For every* K, $\varepsilon > 0$, *there exists* $T > 0$ *such that if* $\zeta:[a, b] \to \tilde{M}$ *is a minimizer, then*

$$|(b-a)^{-1} A(\zeta) - \beta(\rho(\zeta))| < \varepsilon,$$

if $\|\rho(\zeta)\| \leqq K$ *and* $b - a \geqq T$. $\square$

We will say that a curve $\zeta:\mathbb{R} \to \tilde{M}$ is a *minimizer* if its restriction to each finite interval is a minimizer, i.e. it minimizes the action subject to a fixed endpoint condition. For simplicity, we will also say that the associated trajectory $\gamma(t) = (d\zeta(t), t)$ of the Euler-Lagrange flow is a *minimizer*. Finally, a curve on M or in P will be said to be an $\tilde{M}$*-minimizer* if the lift of it to $\tilde{M}$ or $TM \times (\mathbb{R}/\mathbb{Z})$ is a minimizer.

Let $\zeta:\mathbb{R} \to M$ be a C^1 curve and let $\gamma(t) = (d\zeta(t), t \bmod 1)$. Let μ be a (Borel) probability measure on P^*. We will say that μ is a *limit measure* of ζ (or of γ or of a lift $\tilde{\zeta}$ of ζ to the cover $\tilde{M}$) if there is a sequence $[a_i, b_i]$, $i = 1, 2, \ldots$

of closed intervals in $\mathbb{R}$ with $b_i - a_i$ tending to ∞, such that μ_i tends vaguely to μ, where μ_i is the probability measure evenly distributed along $\gamma|[a_i, b_i]$. The set of limit measures of such a curve is obviously compact.

Proposition 2 *Let $\zeta: \mathbb{R} \to \tilde{M}$ be a minimizer and suppose that*

$$\liminf_{b \to +\infty, a \to -\infty} \|\zeta(b) - \zeta(a)\|/(b-a) < \infty.$$

Then there exists $c \in H^1(M, \mathbb{R})$ such that every limit measure of ζ minimizes A_c.

Here, $\zeta(b) - \zeta(a) \in H_1(M, \mathbb{R})$ denotes the difference vector of $\zeta(b)$ and $\zeta(a)$, defined above.

The growth condition on ζ may alternatively be formulated in terms of the Riemannian metric we imposed on M. We may lift this Riemannian metric to $\tilde{M}$ and use the lifted metric to define a distance function on M. Obviously, there exists a constant C such that $\|y-x\| \leqq C\,\mathrm{dist}(x,y)$, for all $x, y \in \tilde{M}$ and $\mathrm{dist}(x,y) \leqq C\|x-y\|$, for all $x, y \in \tilde{M}$ for which $\mathrm{dist}(x,y) \geqq C$. Thus, the growth condition is equivalent to $\liminf \mathrm{dist}(\zeta(a), \zeta(b)/(b-a) < \infty$.

For the proof of Proposition 2, we need:

Lemma. *For every $K>0$, there exists $K'>K$, such that if $\zeta: [a,b] \to \tilde{M}$ is a minimizer and* $\mathrm{dist}(\zeta(a), \zeta(b))/(b-a) \leqq K$, *then for $a \leqq a' < a'+1 \leqq b' \leqq b$, we have* $\mathrm{dist}(\zeta(a'), \zeta(b'))/(b'-a') \leqq K'$.

Proof. For simplicity, we assume that the Lagrangian L is non-negative. There is no loss of generality in assuming this, since we may always add a positive constant to L, without changing its minimizers.

For every $K>0$, there exist non-negative numbers $C_K^{\min}$, $C_K^{\max}$ such that

$$C_K^{\min} K \leqq A(\zeta)/(b-a) \leqq C_K^{\max} K$$

for any minimizer $\zeta: [a,b] \to \tilde{M}$ such that $\mathrm{dist}(\zeta(a), \zeta(b))/(b-a) = K$. Moreover, both $C_K^{\min}$ and $C_K^{\max}$ may be taken to be increasing functions of K which tend to ∞ as K goes to ∞. The fact that $C_K^{\min}$ may be so chosen is a consequence of the superlinear growth of L.

Let K be as given in the statement of the lemma. Choose K'' so that $100\, C_K^{\max} \leqq C_{K''}^{\min}$. Choose K''' so that $100\, C_{100K''}^{\max} < C_{K'''}^{\min}$. Choose $K' \geqq K'''$ so that $100(K'''/K'')\, C_K^{\max} \leqq C_{K'}^{\min}$. We assert that K' satisfies the conclusion of the lemma.

Suppose otherwise. Then $[a,b]$ contains a subinterval $[a',b']$ of length 1 such that $\mathrm{dist}(\zeta(a'), \zeta(b')) > K'$. Note that $b-a \geqq 100(K'''/K'')$, because otherwise the estimate $100(K'''/K'')\, C_K^{\max} \leqq C_{K'}^{\min}$ shows that ζ is not a minimizer. (Here, we use the assumption that $L \geqq 0$ and the fact that $K \leqq K'' \leqq K''' \leqq K'$.) Let b'' be the smallest number $> a'$ such that $\mathrm{dist}(\zeta(a'), \zeta(b'')) = K'''$. Suppose the midpoint of $[a', b'']$ is to the left of the midpoint of $[a,b]$. (The other case may be treated similarly.) Chop up $[b'', b]$ into intervals $[c_i, d_i]$ of length $2K'''/K''$ (with possibly a piece left over). There are at least 24 of them. Since $100\, C_K^{\max} \leqq C_{K''}^{\min}$, we have $\mathrm{dist}(\zeta(c_i), \zeta(d_i))/(d_i - c_i) \leqq K''$ on at least one of these intervals (otherwise ζ would not be a minimizer). Let $[c,d]$ denote this interval.

Let n be the integer most closely approximating (K'''/K''). Let ζ^* be ζ on $[a,a'] \cup [d,b]$. Let $\zeta^*|[b''+n, c+n]$ be the translate (by n in the time coordinate) of $\zeta|[b'', c]$. Let $\zeta^*|[a', b''+n]$ be the minimizer joining $\zeta(a')$ and $\zeta(b'')$. Let $\zeta^*|[c+n, d]$ be the minimizer joining $\zeta(c)$ and $\zeta(d)$. Then

$$A(\zeta)-A(\zeta^*)=A(\zeta|[a', b'']\cup[c, d])-A(\zeta^*|[a', b''+n]\cup[c+n, d])$$
$$\geqq C^{\min}_{K'''} K'''-4C^{\max}_{100K''} K'''>0.$$

This contradicts the assumption that ζ is a minimizer. □

Proof of Proposition 2 Let $\Sigma_\zeta \subset H_1(M, \mathbb{R}) \times \mathbb{R}$ denote the convex hull of the set of pairs $(\rho(\mu), A(\mu))$, where μ is a limit measure of ζ. The existence of c such that every limit measure of ζ mimimizes A_c is easily seen to be equivalent to the statement that $\Sigma_\zeta \subset \operatorname{graph} \beta$.

Now we prove that $\Sigma_\zeta \subset \operatorname{graph} \beta$, by contradiction. Otherwise, there would exist $(h, z) \in \Sigma_\zeta$ with $z>\beta(h)$. Consequently, there would exist limit measures $\mu_1, \ldots, \mu_k$ of ζ and numbers $\tau_1>0, \ldots, \tau_k>0$ such that $\Sigma \tau_i=1$,

$$\Sigma \tau_i \rho(\mu_i)=h, \quad \text{and} \quad \Sigma \tau_i A(\mu_i)=z.$$

Let $\varepsilon=(z-\beta(h))/10$. Choose $\delta>0$ and $T_1 \geqq 1$ so that if $\zeta^*:[a^*, b^*] \to \tilde{M}$ is a minimizer, then

$$|(b^*-a^*)^{-1} A(\zeta^*)-\beta(h)|<\varepsilon,$$

if $\|\rho(\zeta^*)-h\| \leqq 2\delta$ and $b^*-a^* \geqq T_1$. Such a choice is possible by the corollary to Proposition 1. Let M_0 be a relatively compact fundamental domain of the group H of Deck transformations of $\tilde{M}$. Let $\Delta=\sup\{\|y-x\|: x, y \in M_0\}$. Let $T \geqq \max(T_1, 2\Delta/\delta)$.

For each $i=1, \ldots, k$, we choose an infinite sequence of mutually disjoint intervals $I_{ij}=[a_{ij}, b_{ij}], j=1, 2, \ldots$ such that $b_{ij}-a_{ij}$ is an integral multiple of T, $b_{ij}-a_{ij} \to \infty$, as $j \to \infty$, and $\mu_{ij} \to \mu_i$, as $j \to \infty$, where μ_{ij} denotes the probability measure evenly distributed along $\gamma|I_{ij}$. (As before, we set $\gamma(t)=(d\pi\zeta(t), t \bmod 1)$, where π is the projection of $\tilde{M}$ on M.)

From the lemma, it follows that $\limsup_{b-a\to+\infty} \|\zeta(b)-\zeta(a)\|/(b-a)<\infty$. Then, using the minimality of ζ, we see that $\limsup_{j\to\infty} A(\mu_{ij})<\infty$. Since ρ is continuous on sets where A is bounded, it follows that $\rho(\mu_{ij}) \to \rho(\mu_i)$, as $j \to \infty$. From the lower semi-continuity of A, it follows that $\liminf_{j\to\infty} A(\mu_{ij}) \geqq A(\mu_i)$.

Now we consider the partition $\{I_{ij\alpha}\}_\alpha$ of I_{ij} into intervals of length T. Obviously, the mean value of the $\rho(I_{ij\alpha})$ is $\rho(I_{ij})$. Since $\rho(\mu_{ij}) \to \rho(\mu_i)$, as $j \to \infty$, and h is a convex combination of the $\rho(\mu_i)$, it follows that it is possible to choose a finite subcollection $\{J_\beta\}_{\beta=1,\ldots,N}$ of the family $\{I_{ij\alpha}\}_{i,j,\alpha}$ of intervals such that $\|h'-h\|<\delta$, where h' is the mean value of the $\rho(\zeta|J_\beta)$. In addition, it is possible to make this choice so that the mean value of $A(\zeta|J_\beta)/T$ is $\geqq z-\varepsilon$, since $\liminf A(\mu_{ij}) \geqq A(\mu_i)$ and (h, z) is a convex linear combination of the $(\rho(\mu_i), A(\mu_i))$.

Let $c_\beta<d_\beta$ denote the endpoints of J_β and suppose that the intervals J_β are indexed in increasing order, so that $d_\beta<c_{\beta+1}$. We construct a new curve $\zeta^*: \mathbb{R} \to \tilde{M}$, as follows: We let $\zeta^*|(-\infty, c_1] \cup [d_N, +\infty)=\zeta|(-\infty, c_1] \cup [d_N, +\infty)$. We let $\zeta^*|[d_\beta, c_{\beta+1}]=D_\beta \zeta|[d_\beta, c_{\beta+1}]$, where D_β is a suitably chosen Deck

transformation of $\tilde{M}$ over M. We let $\zeta^*|[c_\beta, d_\beta]$ be a minimizer joining $D_{\beta-1}\zeta(c_\beta)$ to $D_\beta\zeta(d_\beta)$, where we set $D_0 = D_N =$ identity.

We choose the Deck transformations so that

$$\| T^{-1} \sum_{\alpha=1}^{\beta} [D_\alpha \zeta(d_\alpha) - D_{\alpha-1}\zeta(c_\alpha)] - \beta h' \| < \delta/2.$$

It is possible to choose the $D_\beta, 1 \leqq \beta \leqq N-1$, inductively, so that this holds, since $T \geqq 2\Delta/\delta$. It follows that

$$\| T^{-1} [D_\beta \zeta(d_\beta) - D_{\beta-1}\zeta(c_\beta)] - h' \| < \delta.$$

We have $\zeta(d_N) - \zeta(c_1) = \rho^* + T \sum_{\alpha=1}^{N} \rho(\zeta|J_\alpha) = \rho^* + TNh'$, where

$$\rho^* = \sum_{\alpha=1}^{N-1} \zeta(c_{\alpha+1}) - \zeta(d_\alpha) = \sum_{\alpha=1}^{N-1} D_\alpha \zeta(c_{\alpha+1}) - D_\alpha \zeta(d_\alpha).$$

Moreover $D_{N-1}\zeta(c_N) - \zeta(c_1) = \rho^* + \sum_{\alpha=1}^{N-1} [D_\alpha \zeta(d_\alpha) - D_{\alpha-1}\zeta(c_\alpha)]$, so we obtain

$$\| T^{-1} [\zeta(d_N) - D_{N-1}\zeta(c_N)] - h' \| \leqq \delta/2.$$

Since $T \geqq T_1$ and $\|h' - h\| < \delta$, we have $|(d_\alpha - c_\alpha)^{-1} A(\zeta^*|[c_\alpha, d_\alpha]) - \beta(h)| < \varepsilon$, so we obtain $A(\zeta^*|[c_1, d_N]) \leqq A^* + TN(\beta(h) + \varepsilon)$, where $A^* = \sum_{\alpha=1}^{N-1} A(\zeta|[d_\alpha, c_{\alpha+1}])$. But $A(\zeta|[c_1, d_N]) \geqq A^* + TN(z - \varepsilon)$ since the mean value of the $A(\zeta|J_\beta)/T$ is $\geqq z - \varepsilon$. Hence $A(\zeta^*|[c_1, d_N]) < A(\zeta|[c_1, d_N])$ and we have a contradiction to the assumption that ζ is a minimizer. □

Consider $c \in H^1(M, \mathbb{R})$. We will denote by $\mathfrak{M}_c$ the set of invariant probability measures μ which minimize A_c over $\mathfrak{M}_L$. Clearly, $\mathfrak{M}_c$ is a compact, convex set, and its extremal points are ergodic measures. By the *support* $\operatorname{supp}\mathfrak{M}_c$ of $\mathfrak{M}_c$, we mean the set of $x \in P$ such that every neighborhood of x has positive μ-measure for some $\mu \in \mathfrak{M}_c$.

Proposition 3 *For any $c \in H^1(M, \mathbb{R})$, every trajectory of the Euler-Lagrange flow in $\operatorname{supp}\mathfrak{M}_c$ is an $\tilde{M}$-minimizer.*

Proof. Let $\gamma: \mathbb{R} \to P$ be a trajectory in $\operatorname{supp}\mathfrak{M}_c$. If γ is not an $\tilde{M}$-minimizer, then there is a finite interval $[a, b]$ such that $\gamma|[a, b]$ is not an $\tilde{M}$-minimizer. Let N be a small neighborhood of $\gamma(a)$. Since $\gamma(a) \in \operatorname{supp}\mathfrak{M}_c$, there exists $\mu \in \mathfrak{M}_c$ such that $\operatorname{supp}\mu \cap N \neq \emptyset$. In fact, μ can be taken to be an extremal point of $\mathfrak{M}_c$. Then μ is ergodic. Let γ_1 be a trajectory in $\operatorname{supp}\mu$ such that the proportion of time it spends in N is $\mu(N)$ and $\lim_{T\to\infty} (2T^{-1} A(\gamma_1|[-T, T]) = A(\mu)$. Such a trajectory exists by Birkhoff's ergodic theorem. Then there are intervals $[a_i, b_i], i \in \mathbb{Z}$ with $a - a_i \in \mathbb{Z}$, $b - b_i \in \mathbb{Z}$, $b_i - a_i = b - a$, $b_i < a_{i+1}$ and $\limsup_{i\to\pm\infty} i^{-1} a_i < \infty$, such that $\gamma_1(a_i) \in N$. By choosing N small enough we may suppose that

$\gamma_1|[a_i, b_i]$ is as close as we wish to $\gamma|[a, b]$. Since $\gamma|[a, b]$ is not an $\tilde{M}$ minimizer, neither will $\gamma_1|[a_i, b_i]$ be an $\tilde{M}$ minimizer, if the latter is close enough to the former. In fact, there will exist $\varepsilon > 0$ such that for each i, there exists $\zeta_i^*: [a_i, b_i] \to M$ satisfying $A(\zeta_i^*) \leqq A(\pi\gamma_1|[a_i, b_i]) - \varepsilon$ where $\pi: P \to M$ denotes the projection, and $\tilde{\zeta}_i^*(a_i) = \pi\gamma_1(a_i)$, $\tilde{\zeta}_i^*(b_i) = \pi\gamma_1(b_i)$, for appropriate lifts $\tilde{\zeta}_i^*$ and $\pi\gamma_1$ of ζ_i^* and $\pi\gamma_1$ to $\tilde{M}$. We construct a curve $\tilde{\zeta}^*: \mathbb{R} \to \tilde{M}$ by $\tilde{\zeta}^*|[a_i, b_i] = \tilde{\zeta}_i^*$ and $\tilde{\zeta}^*|[b_i, a_{i+1}] = \pi\gamma_1|[b_i, a_{i+1}]$.

Then $\limsup_{T\to\infty} (2T)^{-1} A(\tilde{\zeta}^*|[-T, T]) \leqq A(\mu) - \varepsilon/\eta$, where $\eta = \limsup_{i\to\pm\infty} i^{-1} a_i$. Moreover, $\lim_{T\to\infty} \rho(\tilde{\zeta}^*([-T, T])) = \rho(\mu)$. Let $\tilde{\zeta}_T: [-T, T] \to \tilde{M}$ be a minimizer with the same endpoints as $\tilde{\zeta}^*$. Let ζ_T be the projection of $\tilde{\zeta}_T$ on M. Let γ_T: $[-T, T] \to P$ be defined by $\gamma_T(t) = (d\zeta_T(t), t \bmod 1)$. Let μ_T be the probability measure evenly distributed along γ_T. Let μ^* be an accumulation point of μ_T as $T \to \infty$. Obviously,

$$\rho(\mu^*) = \rho(\mu), \qquad A(\mu^*) \leqq A(\mu) - \varepsilon/\eta.$$

Thus, we obtain a contradiction to the assumption that $\mu \in \mathfrak{M}_c$. □

4 The Lipschitz property

Let $c \in H^1(M, \mathbb{R})$. Recall that $\mathfrak{M}_c$ denotes the set of Φ-invariant probability measures which minimize A_c. In this section, we prove a couple of properties of the subset $\operatorname{supp} \mathfrak{M}_c$ of P. The Lipschitz property stated in Theorem 2 below is the main result of this paper.

Proposition 4 $\operatorname{supp} \mathfrak{M}_c$ *is compact.*

For the proof, we need the following:

Remark. The conclusion of the lemma used in the proof of Proposition 2 is valid for $a \leqq a' < b' \leqq b$, if $b - a \geqq 1$. In other words, we may drop the condition $b' - a' \geqq 1$.

To show this, we have to use the hypothesis of completeness of the Euler-Lagrange flow. Without this hypothesis, the examples of Ball and Mizel [3] would contradict this remark.

To prove the remark, we argue by contradiction. For, otherwise, there would exist a sequence $\zeta_i: [a_i, b_i] \to \tilde{M}$, $i = 1, 2, \ldots$ of minimizers satisfying $\operatorname{dist}(\zeta_i(a_i), \zeta_i(b_i))/(b_i - a_i) \leqq K$ and $c_i \in [a_i, b_i]$ such that $\|d\zeta_i(c_i)\| \to \infty$, as $i \to \infty$. Using the periodicity of L, we may assume that $c_i \in [0, 1)$. Passing to a subsequence, we may suppose that $c_1, c_2, \ldots$ converges to $c \in [0, 1]$. Translating each ζ_i by a Deck transformation and passing to a subsequence, we may suppose that $\zeta_i(c_i)$ converges to a point $x \in \tilde{M}$, as $i \to \infty$. For each i, we choose an interval $[a_i', a_i' + 1]$ in $[a_i, b_i]$ which contains c_i. We may suppose that a_i' converges to $a \in \mathbb{R}$. By the lemma used in the proof of Proposition 2, there exists K' such that $\operatorname{dist}(\zeta_i(a_i'), \zeta_i(a_i' + 1)) \leqq K'$ for all i. Since each ζ_i is a minimizer there exists an upper bound on $A(\zeta_i|[a_i', a_i' + 1])$, independent of i. Let $\pi: \tilde{M} \to M$ denote the projection. It follows from the lemma used to prove Tonelli's theorem that the sequence of curves $\pi\zeta_i|[a_i', a_i' + 1]$ has a subsequence which is C^0 convergent. Let $\pi\zeta$ denote the limit, where ζ maps $[a, a+1]$ into $\tilde{M}$ and $\zeta(c) = x = \lim \zeta_i(c_i)$.

Since ζ is a limit of minimizers, it is a minimizer. By the semi-continuity property of A, we have $A(\zeta) \leqq \liminf A(\zeta_i)$. Moreover, we cannot have $A(\zeta) < \limsup A(\zeta_i)$, since this would contradict the fact that the ζ_i's are minimizers. Hence, $A(\zeta) = \lim A(\zeta_i)$. By the addendum to the lemma used to prove Tonelli's theorem, it therefore follows that γ_i converges to γ in the C^{ac}-topology.

From this, we may deduce that ζ cannot be C^1 at c. For, otherwise, there would be a small interval J containing c and $K>0$ such that $\|d\zeta(t)\| \leqq K$, for all $t \in J$. Since $t \to (d\pi\gamma(t), t \bmod 1)$ is a trajectory of the Euler-Lagrange flow, there would exist $K' > 2K$ and $\delta > 0$ such that $\|d\zeta_i(c_i)\| \geqq K'$ implies $\|d\zeta_i(t)\| \geqq 2K$ when $|t - c_i| < \delta$. We may assume J has length $< \delta$; then $\|d\zeta_i(t)\| \geqq 2K$ but $\|d\zeta(t)\| \leqq K$, for all $t \in J$, contrary to the fact that $\pi\zeta_i$ converges to $\pi\zeta$ with respect to the metric d_{ac}. This contradiction shows that ζ cannot be C^1 at c. But, we have already shown that the hypothesis of completeness implies that any minimizer is C^1. This contradiction proves the remark. □

Proof of Proposition 4 Since $\operatorname{supp} \mathfrak{M}_c$ is a closed subset of P by definition, it is enough to show that there exists K' such that $(\xi, \theta) \in \operatorname{supp} \mathfrak{M}_c \subset TM \times (\mathbb{R}/\mathbb{Z})$ implies $\|\xi\| \leqq K'$. Since $\beta: H_1(M, \mathbb{R}) \to \mathbb{R}$ has superlinear growth, there exists K such that $\|\rho(\mu)\| \leqq K$, for all $\mu \in \mathfrak{M}_c$. Let μ be an extremal point of $\mathfrak{M}_c$ (so it is an ergodic measure). Let γ be generic, in the sense of Birkhoff's ergodic theorem, for μ, so if $\zeta: \mathbb{R} \to \tilde{M}$ is a corresponding minimizer, then

$$\lim_{a \to -\infty, b \to \infty} \|\zeta(b) - \zeta(a)\|/(b-a) = \|\rho(\mu)\| \leqq K,$$

on K, so that $\|d\gamma(t)\| \leqq K'$ for all $t \in \mathbb{R}$. But the union of the set of such trajectories is dense in $\operatorname{supp} \mathfrak{M}_c$. Thus, we have $\|\xi\| \leqq K'$, for all $(\xi, \theta) \in \operatorname{supp} \mathfrak{M}_c$. □

Now we come to the main result: We let $\pi: P = TM \times (\mathbb{R}/\mathbb{Z}) \to M \times (\mathbb{R}/\mathbb{Z})$ denote the projection and we denote the restriction of π to $\operatorname{supp} \mathfrak{M}_c$ by the same symbol.

Theorem 2 *$\pi: \operatorname{supp} \mathfrak{M}_c \to M \times (\mathbb{R}/\mathbb{Z})$ is injective. Its inverse (considered as a mapping from $\pi(\operatorname{supp} \mathfrak{M}_c)$ is Lipschitz, i.e. there exists a constant C such that for any $x, y \in \pi(\operatorname{supp} \mathfrak{M}_c)$, we have*

$$\operatorname{dist}(\pi^{-1}(x), \pi^{-1}(y)) \leqq C \operatorname{dist}(x, y).$$

As usual, distance is measured with respect to smooth Riemannian metrics. Since M is compact and $\operatorname{supp} \mathfrak{M}_c$ is a compact subset of P, it doesn't matter which Riemannian metrics we choose to measure distance.

The proof is based on the following:

Lemma. *If $K>0$, then there exist $\varepsilon, \delta, \eta, C > 0$ such that if $\alpha, \beta: [t_0 - \varepsilon, t_0 + \varepsilon] \to M$ are solutions of the Euler-Lagrange equation with $\|d\alpha(t_0)\|$, $\|d\beta(t_0)\| \leqq K$, $\operatorname{dist}(\alpha(t_0), \beta(t_0)) \leqq \delta$, and $\operatorname{dist}(d\alpha(t_0), d\beta(t_0)) \geqq C \operatorname{dist}(\alpha(t_0), \beta(t_0))$, then there exist C^1 curves $a, b: [t_0 - \varepsilon, t_0 + \varepsilon] \to M$ such that $a(t_0 - \varepsilon) = \alpha(t_0 - \varepsilon)$, $a(t_0 + \varepsilon) = \beta(t_0 + \varepsilon)$, $b(t_0 - \varepsilon) = \beta(t_0 - \varepsilon)$, $b(t_0 + \varepsilon) = \alpha(t_0 + \varepsilon)$, and*

$$A(\alpha) + A(\beta) - A(a) - A(b) \geqq \eta \operatorname{dist}(d\alpha(t_0), d\beta(t_0))^2.$$

Proof. We may choose a cover of M by a finite family (U^j, x^j) of smooth coordinate charts and for each j choose a compact subset $\Sigma^j \subset U^j$ such that the family of Σ^j's still covers M. We may make these choices so that for each $j, x^j(U^j)$ is a convex subset of $\mathbb{R}^m$, where $m = \dim M$. We choose positive numbers $\delta_0 > \delta_1$ and ε_0 such that the closed δ_0 neighborhood Σ^{*j} of Σ^j is in U^j and such that if $\alpha: [t_0 - \varepsilon_0, t_0 + \varepsilon_0] \to M$ is a minimizer with $\operatorname{dist}(\alpha(t_0), \Sigma^j) \leqq \delta_1$ and $\|d\alpha(t_0)\| \leqq K$, then $\alpha([t_0 - \varepsilon_0, t_0 + \varepsilon_0]) \subset \Sigma^{*j}$.

We will choose positive numbers $\delta \leqq \delta_1$ and $\varepsilon \leqq \varepsilon_0$. Thus, if α, β are two minimizers which satisfy the hypotheses of the lemma, we may choose j such that $\alpha(t_0) \in \Sigma^j$. Then $\beta(t_0)$ is in the δ_1 neighborhood of Σ^j, so the images of both α and β are in Σ^{*j}.

From now on, we consider only this coordinate neighborhood and drop the index j. Sums and scalar products will be taken with respect to the system of coordinates given in this neighborhood. It will be clear that we can form all the sums introduced below if δ and ε are small enough. How small they have to be depends only on K.

We set $\mu(t) = (\alpha(t) + \beta(t))/2$. We set

$$\begin{aligned} a(t) &= \mu(t) + (2\varepsilon)^{-1}\{(-t + t_0 + \varepsilon) \\ &\quad (\alpha(t_0 - \varepsilon) - \mu(t_0 - \varepsilon)) + (t - t_0 + \varepsilon)(\beta(t_0 + \varepsilon) - \mu(t_0 + \varepsilon))\}, \\ b(t) &= \mu(t) + (2\varepsilon)^{-1}\{(-t + t_0 + \varepsilon)(\beta(t_0 - \varepsilon) - \mu(t_0 - \varepsilon)) + (t - t_0 + \varepsilon) \\ &\quad (\alpha(t_0 + \varepsilon) - \mu(t_0 + \varepsilon))\} \end{aligned}$$

and we set $w = \alpha(t_0) - \beta(t_0)$, $D = (\dot{\alpha}(t_0) - \dot{\beta}(t_0))/2$, where the dot denotes differentiation with respect to t. Then we have

$$\begin{aligned} \dot{a}(t) - \dot{\mu}(t) &= (2\varepsilon)^{-1}\{\beta(t_0 + \varepsilon) - \mu(t_0 + \varepsilon) + \mu(t_0 - \varepsilon) - \alpha(t_0 - \varepsilon)\} \\ &= (4\varepsilon)^{-1}\{\beta(t_0 + \varepsilon) - \alpha(t_0 + \varepsilon) + \beta(t_0 - \varepsilon) - \alpha(t_0 - \varepsilon)\} \\ &= -(2\varepsilon)^{-1} w + O(\varepsilon(\|D\| + \|w\|)), \end{aligned}$$

for $t_0 - \varepsilon \leqq t \leqq t_0 + \varepsilon$. The estimate may be seen by expanding the various terms in Taylor series with remainder, e.g.

$$\beta(t_0 + \varepsilon) = \beta(t_0) + \varepsilon \dot{\beta}(t_0) + \int_0^\varepsilon (\varepsilon - s) \ddot{\beta}(t_0 + s)\, ds$$

and similarly expanding $\alpha(t_0 + \varepsilon)$, $\beta(t_0 - \varepsilon)$, and $\alpha(t_0 - \varepsilon)$. The constant terms from these expansions contribute $-(2\varepsilon)^{-1} w$, the linear terms cancel out, and the remainder terms contribute

$$\begin{aligned} &\leqq (\varepsilon/2) \sup\{\|\ddot{\beta}(s) - \ddot{\alpha}(s)\| : t_0 - \varepsilon \leqq s \leqq t_0 + \varepsilon\} \\ &\leqq \text{const}\, \varepsilon(\|D\| + \|w\|). \end{aligned}$$

This last inequality follows from the fact that α and β both satisfy the Euler-Lagrange equation: it implies that

$$\begin{aligned} \|\ddot{\beta}(s) - \ddot{\alpha}(s)\| &\leqq C(\|\dot{\beta}(s) - \dot{\alpha}(s)\| + \|\beta(s) - \alpha(s)\|) \\ &\leqq C(1 + C_1 \varepsilon)(2\|D\| + \|w\|), \end{aligned}$$

provided ε is small enough, where the constants C and C_1 depend only on K.

This proves the estimate for $\dot{a}(t)-\dot{\mu}(t)$. Similarly, we have

$$\begin{aligned}
&\dot{b}(t)-\dot{\mu}(t)=(2\varepsilon)^{-1}w+O(\varepsilon(\|D\|+\|w\|)),\\
&\dot{\alpha}(t)-\dot{\mu}(t)=D+O(\varepsilon(\|D\|+\|w\|)),\\
&\dot{\beta}(t)-\dot{\mu}(t)=-D+O(\varepsilon(\|D\|+\|w\|),
\end{aligned}$$

for $t_0-\varepsilon\leqq t\leqq t_0+\varepsilon$.

Let $\Lambda_t(\dot{x})=L(\mu(t),\dot{x},t)$. Clearly,

$$\begin{aligned}
&\|L(x,\dot{x},t)-\Lambda_t(\dot{x})-L_x(\mu(t),\dot{\mu}(t),t)\cdot(x-\mu(t))\|\\
&\quad\leqq\|L(x,\dot{x},t)-\Lambda_t(\dot{x})-L_x(\mu(t),\dot{x},t)\cdot(x-\mu(t))\|\\
&\qquad+\|(L_x(\mu(t),\dot{x},t)-L_x(\mu(t),\dot{\mu}(t),t))\cdot(x-\mu(t))\|\\
&\quad\leqq C_1(\|x-\mu(t)\|+\|\dot{x}-\dot{\mu}(t)\|)\|x-\mu(t)\|,
\end{aligned}$$

for $t_0-\varepsilon\leqq t\leqq t_0+\varepsilon$, provided $\|\dot{x}\|\leqq K$, where C_1 is a constant, which depends only on K.

Next, we have

$$\begin{aligned}
&L(\alpha(t),\dot{\alpha}(t),t)+L(\beta(t),\dot{\beta}(t),t)-L(a(t),\dot{a}(t),t)\\
&\quad-L(b(t),\dot{b}(t),t)\\
&\geqq\Lambda_t(\dot{\alpha}(t))+\Lambda_t(\dot{\beta}(t))-\Lambda_t(\dot{a}(t))-\Lambda_t(\dot{b}(t))\\
&\quad-C_2\,\varepsilon^{-1}(\varepsilon\|D\|+\|w\|)^2\\
&\geqq C_3\|\dot{\alpha}(t)-\dot{\beta}(t)\|^2-C_4\|\dot{a}(t)-\dot{b}(t)\|^2\\
&\quad-C_2\,\varepsilon^{-1}(\varepsilon\|D\|+\|w\|)^2\\
&\geqq C_5\|D\|^2-C_6\|\varepsilon^{-1}w\|^2-C_7\,\varepsilon^{-1}(\varepsilon\|D\|+\|w\|)^2.
\end{aligned}$$

Here, the C's are constants which depend only on K. This estimate is valid for $t_0-\varepsilon\leqq t\leqq t_0+\varepsilon$, provided ε is small enough. How small ε has to be depends only on K.

The first inequality follows from the bound on $\|L(x,\dot{x},t)-\Lambda_t(\dot{x})-L_x(\mu(t),\dot{\mu}(t),t)\cdot(x-\mu(t))\|$ which we noted above. Notice that since $\mu(t)=(\alpha(t)+\beta(t))/2=(a(t)+b(t))/2$, the contributions of $L_x(\mu(t),\dot{\mu}(t),t)\cdot(x-\mu(t))$ cancel in pairs, i.e. the contribution from α cancels that from β and the contribution from a cancels that from b. In each of the cases $(x,\dot{x})=(\alpha(t),\dot{\alpha}(t))$, $(\beta(t),\dot{\beta}(t))$, $(a(t),\dot{a}(t))$, or $(b(t),\dot{b}(t))$, we have that $(\|\dot{x}-\dot{\mu}(t)\|+\|x-\mu(t)\|)\|x-\mu(t)\|$ is bounded by const $\varepsilon^{-1}(\varepsilon\|D\|+\|w\|)^2$, as may be seen from the estimates we obtained above on $\dot{a}(t)-\dot{\mu}(t)$, etc.

The second inequality above follows from two inequalities:

$$\Lambda_t(\dot{\alpha}(t))+\Lambda_t(\dot{\beta}(t))\geqq 2\Lambda_t(\dot{\mu}(t))+C_3\|\dot{\alpha}(t)-\dot{\beta}(t)\|^2,$$

which is a consequence of the positive definiteness of L and the fact that $\dot{\mu}(t)=(\dot{\alpha}(t)+\dot{\beta}(t))/2$; and

$$\Lambda_t(\dot{a}(t))+\Lambda_t(\dot{b}(t))\leqq 2\Lambda_t(\dot{\mu}(t))+C_4\|\dot{a}(t)-\dot{b}(t)\|^2,$$

which is a consequence of the fact that L is twice continuously differentiable and the fact that $\dot{\mu}(t)=(\dot{a}(t)+\dot{b}(t))/2$. Here, C_3 and C_4 are constants, which depend only on K.

The last inequality is a consequence of our estimates for $\dot{a}(t)-\dot{\mu}(t)$, $\dot{b}(t)-\dot{\mu}(t)$, $\dot{\alpha}(t)-\dot{\mu}(t)$, and $\dot{\beta}(t)-\dot{\mu}(t)$.

Integrating from $t_0-\varepsilon$ to $t_0+\varepsilon$ and absorbing the last term on the right side of the above inequality into the two previous ones, we obtain

$$A(\alpha)+A(\beta)-A(a)-A(b)$$
$$\geqq (2\varepsilon)(C_8\|D\|^2-C_9\|\varepsilon^{-1}w\|^2),$$

where ε, C_8, and C_9 are contants which depend only on K and not on α or β.

The conclusions of the lemma follow: We have $a(t_0-\varepsilon)=\alpha(t_0-\varepsilon)$, $a(t_0+\varepsilon)=\beta(t_0+\varepsilon)$, $b(t_0-\varepsilon)=\beta(t_0-\varepsilon)$, $b(t_0+\varepsilon)=\alpha(t_0+\varepsilon)$ by the formulas defining a and b. Taking $C^2=2C_9/C_8\,\varepsilon^2$ and $\eta=\varepsilon C_8$, we have

$$A(\alpha)+A(\beta)-A(a)-A(b)\geqq\eta\|D\|^2,$$

whenever $\|D\|\geqq C\|w\|$. Taking into account that distances measured in any two Riemannian metrics are comparable, we obtain the conclusion of the lemma (after possibly changing C and η). □

Proof of Theorem 2 By proposition 4, we may choose K such that $(\xi,t_0)\in\operatorname{supp}\mathfrak{M}_c$ implies that $\|\xi\|<K$. Let ε, δ, η and C be as in the lemma. We first show that if $(\xi,t_0),(v,t_0)\in\operatorname{supp}\mathfrak{M}_c$ and $\operatorname{dist}(\pi(\xi),\pi(v))<\delta$, then $\operatorname{dist}(\xi,v)\leqq C\operatorname{dist}(\pi(\xi),\pi(v))$. Suppose the contrary, i.e. suppose $\operatorname{dist}(\pi(\xi),\pi(v))<\delta$ but $\operatorname{dist}(\xi,v)>C\operatorname{dist}(\pi(\xi),\pi(v))$. We may choose open neighborhoods N_ξ of ξ in TM and N_v of v in TM and a small positive number δ_1 such that for $\xi'\in N_\xi$, $v'\in N_v$, we have $\operatorname{dist}(\pi(\xi'),\pi(v'))<\delta$ but $\operatorname{dist}(\xi',v')>C\operatorname{dist}(\pi(\xi'),\pi(v'))+\delta_1$ and $\|\xi'\|,\|v'\|<K$.

Since $(\xi,t_0),(v,t_0)\in\operatorname{supp}\mathfrak{M}_c$, there exist extremal points μ_0,μ_1 of $\mathfrak{M}_c$ such that $\operatorname{supp}\mu_0$ has non-void intersection with $N_\xi\times t_0$ and $\operatorname{supp}\mu_1$ has nonvoid intersection with $N_v\times t_0$. Since μ_0 and μ_1 are extremal points, they are ergodic measures. Therefore, we may choose points $\xi'\in N_\xi\times t_0$ and $v'\in N_v\times t_0$ which are generic (in the sense of Birkhoff's ergodic theorem) for μ_0 and μ_1, resp.

Since ξ' is generic for μ_0 and $N_\xi\times(t_0-\varepsilon,t_0+\varepsilon)$ has positive measure with respect to μ_0, the orbit of the Euler-Lagrange flow through $(\xi',t_0 \bmod 1)$ returns to $N_\xi\times t_0 \pmod 1$ with positive frequency, i.e. there exists a strictly increasing bi-infinite sequence $(\ldots,n_i,\ldots)$ of integers such that $\Phi((\xi',t_0 \bmod 1),n_i)\in N_\xi\times t_0 \pmod 1$, where Φ is the Euler-Lagrange flow on P. Moreover, $\lim_{i\to+\infty} i^{-1}n_i$ exists and is finite.

Likewise, there exists a strictly increasing bi-infinite sequence $(\ldots,n_i',\ldots)$ of integers such that $\Phi((v',t_0 \bmod 1),n_i')\in N_v\times t_0$ and $\lim_{i\to\pm\infty} i^{-1}n_i'$ exists and is finite.

Let $\alpha(t)=\pi\Phi((\xi',t_0 \bmod 1),t)$, $\beta(t)=\pi\Phi((v',t_0 \bmod 1),t)$, where $\pi\colon TM\times(\mathbb{R}/\mathbb{Z})\to M$ denotes the projection. Then α and β are curves on M which satisfy the Euler-Lagrange equation. Moreover, for any i, the curves $\alpha_i=\alpha|[t_0+n_i-\varepsilon,t_0+n_i+\varepsilon]$ and $\beta_i=\beta|[t_0+n_i'-\varepsilon,t_0+n_i'+\varepsilon]$ satisfy the hypotheses of the lemma, since $d\alpha(t_0)\in N_\xi$, $d\beta(t_0)\in N_v$. Note that by the periodicity of L, we may still apply the lemma when the domains of α and β differ by an integer, as here.

From the lemma, it follows that, for each integer i, there exist curves a_i, b_i: $[t_0-\varepsilon, t_0+\varepsilon] \to M$ such that $a_i(t_0-\varepsilon)=\alpha(t_0+n_i-\varepsilon)$, $a_i(t_0+\varepsilon)=\beta(t_0+n_i'+\varepsilon)$, $b_i(t_0-\varepsilon)=\beta(t_0+n_i'-\varepsilon)$, $b_i(t+\varepsilon)=\alpha(t_0+n_i+\varepsilon)$, and

$$A(\alpha_i)+A(\beta_i)-A(a_i)+A(b_i) \geqq \eta \operatorname{dist}(d\alpha(t_0+n_i), \\ d\beta(t_0+n_i'))^2 \geqq \eta \delta_1^2.$$

Now we construct two new curves $\alpha^*, \beta^*: \mathbb{R} \to M$, as follows. First, we choose two sequences $(\dots, m_i, \dots)$ and $(\dots m_i', \dots)$ of integers such that

$$m_{2i+1}-m_{2i}=n_{2i+1}-n_{2i}, \quad m'_{2i+1}-m'_{2i}=n'_{2i+1}-n'_{2i},$$
$$m_{2i}-m_{2i-1}=n'_{2i}-n'_{2i-1}, \quad m'_{2i}-m'_{2i-1}=n_{2i}-n_{2i-1},$$

for all integers i. We let

$$\begin{aligned} \alpha^*(t) &= \alpha(t+n_{2i}-m_{2i}), && \text{for } m_{2i}+\varepsilon \leqq t-t_0 \leqq m_{2i+1}-\varepsilon, \\ &= \beta(t+n'_{2i}-m_{2i}), && \text{for } m_{2i-1}+\varepsilon \leqq t-t_0 \leqq m_{2i}-\varepsilon, \\ &= b_{2i}(t-m_{2i}), && \text{for } m_{2i}-\varepsilon \leqq t-t_0 \leqq m_{2i}+\varepsilon, \\ &= a_{2i+1}(t-m_{2i+1}), && \text{for } m_{2i+1}-\varepsilon \leqq t-t_0 \leqq m_{2i+1}+\varepsilon, \end{aligned}$$

and

$$\begin{aligned} \beta^*(t) &= \beta(t+n'_{2i}-m'_{2i}), && \text{for } m'_{2i}+\varepsilon \leqq t-t_0 \leqq m'_{2i+1}-\varepsilon, \\ &= \alpha(t+n_{2i}-m'_{2i}), && \text{for } m'_{2i-1}+\varepsilon \leqq t-t_0 \leqq m'_{2i}-\varepsilon, \\ &= a_{2i}(t-m'_{2i}), && \text{for } m'_{2i}-\varepsilon \leqq t-t_0 \leqq m'_{2i}+\varepsilon, \\ &= b_{2i+1}(t-m'_{2i+1}), && \text{for } m'_{2i+1}-\varepsilon \leqq t-t_0 \leqq m_{2i+1}+\varepsilon. \end{aligned}$$

For each positive integer N, we set

$$\begin{aligned} \alpha_N &= \alpha \mid [t_0+n_{-N}-\varepsilon, t_0+n_N+\varepsilon] \\ \beta_N &= \beta \mid [t_0+n'_{-N}-\varepsilon, t_0+n'_N+\varepsilon] \\ \alpha_N^* &= \alpha^* \mid [t_0+m_{-N}-\varepsilon, t_0+m_N+\varepsilon] \\ \beta_N^* &= \beta^* \mid [t_0+m'_{-N}-\varepsilon, t_0+m'_N+\varepsilon]. \end{aligned}$$

By the previous inequality

$$A(\alpha_N)+A(\beta_N)-A(\alpha_N^*)-A(\beta_N^*) \geqq (2N+1)\,\eta\,\delta_1^2.$$

Let $\alpha_N^{**}, \beta_N^{**}$ be minimizers whose lifts $\tilde{\alpha}_N^{**}$, $\tilde{\beta}_N^{**}$ to $\tilde{M}$ join the endpoints of lifts of α_N^*, β_N^*. Then $A(\alpha_N^{**}) \leqq A(\alpha_N^*)$, $A(\beta_N^{**}) \leqq A(\beta_N^*)$, so

$$A(\alpha_N)+A(\beta_N)-A(\alpha_N^{**})-A(\beta_N^{**}) \geqq (2N+1)\,\eta\,\delta_1^2.$$

To finish the proof of Theorem 2, we introduce the following notation: Let $\lambda_1, \dots, \lambda_l$ be the closed 1-forms on M which were introduced at the beginning of § 2. Since $[\lambda_1], \dots, [\lambda_l]$ is a basis of $H^1(M, \mathbb{R})$, the cohomology class c can be uniquely expressed in the form $c=\sum_{i=1}^{l} a_i[\lambda_i]$, with $a_i \in \mathbb{R}$. Let $\lambda=\sum a_i \lambda_i$, so that $c=[\lambda]$. For a curve $\zeta:[a,b] \to M$, we set $A_c(\zeta)=\int_a^b (L-\lambda)(d\zeta(t), t)\,dt$, so that $A_c(\zeta)=A(\zeta)-(b-a)\langle c, \rho(\zeta)\rangle$.

It is easy to see that the 1-chain $\alpha_N^{**}-\beta_N^{**}-\alpha_N-\beta_N$ is a 1-cycle which represents the homology class 0. Consequently:

$$A_c(\alpha_N)+A_c(\beta_N)-A_c(\alpha_N^{**})-A_c(\beta_N^{**})$$
$$=A(\alpha_N)+A(\beta_N)-A(\alpha_N^{**})-A(\beta_N^{**}).$$

Because ξ' and v' have been chosen to be generic for μ_0 and μ_1, resp., we have

$$\lim_{N\to\infty}(n_N-n_{-N})^{-1}A_c(\alpha_N)=A_c(\mu_0)=\min A_c$$
$$\lim_{N\to\infty}(n'_N-n'_{-N})^{-1}A_c(\beta_N)=A_c(\mu_1)=\min A_c$$

and the limits $L=\lim_{i\to\pm\infty} i^{-1}n_i$, $L'=\lim_{i\to\pm\infty} i^{-1}n'_i$ exist.

Because $m_N-m_{-N}+m'_N-m'_{-N}=n_N-n_{-N}+n'_N-n'_{-N}$, it follows from the inequality

$$A_c(\alpha_N)+A_c(\beta_N)-A_c(\alpha_N^{**})-A_c(\beta_N^{**})\geqq(2N+1)\,\eta\,\delta_1^2$$

and from the two equations above that at least one of the following two inequalities holds:

$$\liminf_{N\to\infty}(m_N-m_{-N})^{-1}A_c(\alpha_N^{**})<\min A_c$$
$$\liminf_{N\to\infty}(m'_N-m'_{-N})^{-1}A_c(\beta_N^{**})<\min A_c,$$

where the "min" is taken over all Φ-invariant probability measures.

But this leads to a contradiction:

By the compactness of the set of probability measures, we may choose a sequence $N_1<N_2<\dots$ of positive integers such that the vague limits (as $i\to\infty$) of the probability measures uniformly distributed along $\alpha_{N_i}^{**}$ and $\beta_{N_i}^{**}$ exist. Call these limits μ_0^* and μ_1^*. Since one of the above inequalities holds, we obtain that one of the following inequalities holds:

$$A_c(\mu_0^*)<\min A_c \quad \text{or} \quad A_c(\mu_1^*)<\min A_c.$$

This is obviously impossible.

This contradiction shows that if $(\zeta,t_0),(v,t_0)\in\operatorname{supp}\mathfrak{M}_c$ and $\operatorname{dist}(\pi(\zeta),\pi(v))<\delta$, then $\operatorname{dist}(\xi,v)\leqq C\operatorname{dist}(\pi(\xi),\pi(v))$. The injectivity of π and the Lipschitz property of π^{-1} follow immediately. □

5 Perturbations of a system with an invariant torus

Let (N,ω) be a symplectic manifold, i.e. let N be a $2n$-manifold and ω a closed non-degenerate 2-form on N. Let $f\colon N\to N$ be a symplectic diffeomorphism of N, i.e. a C^∞ diffeomorphism such that $f^*\omega=\omega$. Let $\mathfrak{L}$ be an n-dimensional submanifold of N. Suppose that $\mathfrak{L}$ is invariant under f, i.e. $f(\mathfrak{L})=\mathfrak{L}$ and that $f|\mathfrak{L}$ is C^∞ conjugate to a translation on the n-torus T^n by a vector ρ

$=(\rho_1, \ldots, \rho_n)\in\mathbb{R}^n$ which satisfies a Diophantine condition, i.e. there exists a C^∞ diffeomorphism $\phi: \mathfrak{L}\to T^n$ such that $\phi f \phi^{-1}(\theta)\equiv\theta+\rho(\operatorname{mod}\mathbb{Z}^n)$ and there exist $C, \beta>0$ such that

$$|k_0+k_1\rho_1+\ldots+k_n\rho_n|\geqq c(|k_1|+\ldots+|k_n|)^{-\beta},$$

for all $k=(k_1, \ldots, k_n)\in\mathbb{Z}^n\setminus 0$.

It is well known that it is possible to introduce a C^∞ system of coordinates $(q, p)=(q_1, \ldots, q_n, p_1, \ldots, p_n)$ in a neighborhood of $\mathfrak{L}$, where q is defined mod $\mathbb{Z}^n$, with the following properties: $\omega=\sum_{i=1}^n dq_i\wedge dp_i$ where (q, p) is defined; $\mathfrak{L}=\{p=0\}$; and $(q', p')\circ f=(q, p)$, where

$$q'=q+\rho+A\cdot p+O(p^2), \qquad p'=p+O(p^2),$$

and A is an $n\times n$ symmetric matrix of real numbers. (The more classical way of writing $(q', p')\circ f=(q, p)$ is $f(q, p)=(q', p')$. However, the way we express this relation is more logical, since coordinates are functions on the manifold.) See e.g., [11], Appendix 2.

Here is a brief sketch of the proof of the existence of such coordinates: By the hypothesis that $f|\mathfrak{L}$ is C^∞ conjugate to translation by ρ on the n-torus, it follows that there exists a C^∞ system $q=(q_1, \ldots, q_n)$ of coordinates on $\mathfrak{L}$ (defined mod $\mathbb{Z}^n$), such that $q'=q+\rho$, where $q=q'\circ f|\mathfrak{L}$. Moreover, $\mathfrak{L}$ is a Lagrangian submanifold of N, by a theorem of Herman [13], i.e. $i^*\omega=0$, where i is the inclusion mapping of $\mathfrak{L}$ into N. Since $\mathfrak{L}$ is Lagrangian, it follows from the global form of Darboux's theorem (cf. Weinstein [27]), that there exists a neighborhood of $\mathfrak{L}$ which is C^∞ symplecticly diffeomorphic to a neighborhood of the zero section of the cotangent bundle $T^*\mathfrak{L}$. It follows that it is possible to extend $q_1, \ldots, q_n$ to functions $Q_1, \ldots, Q_n$ defined in a neighborhood of $\mathfrak{L}$ and find other functions $P_1, \ldots, P_n$ such that (Q, P) is a diffeomorphism of that neighborhood onto $T^n\times U$, where U is an open set in $\mathbb{R}^n$, and $\omega=\sum_{i=1}^n dQ_i\wedge dP_i$ on that neighborhood. In addition, we may choose the P's so that $\mathfrak{L}=\{P=0\}$. Since $\mathfrak{L}$ is invariant, we have (setting $(Q', P')\circ f=(Q, P)$),

$$Q'=Q+\rho+B(Q)\cdot P+O(P^2), \qquad P'=H(Q)\cdot P+O(P^2),$$

where B and H are $n\times n$ matrices whose entries are C^∞ functions on $\mathfrak{L}$. Since the symplectic form ω is preserved by f, i.e. $\sum_{i=1}^n dQ'_i\wedge dP'_i=\sum_{i=1}^n dQ_i\wedge dP_i$, it follows that $H(Q)$ is the identity matrix and $B(Q)$ is symmetric. This would be the form we desire to obtain, except for the fact that B is not constant. To obtain the form we want, we introduce new coordinates $(q, p)=(q_1, \ldots, q_n, p_1, \ldots, p_n)$ defined by the generating function

$$V(Q, p)=Q\cdot p+p\cdot W(Q)\cdot p/2,$$

where $W(Q)$ is an $n\times n$ symmetric matrix of C^∞ functions depending periodically on Q (i.e. $W(Q+k)=W(Q)$ for $k\in\mathbb{Z}^n$). Thus,

$$P=\partial V/\partial Q; \qquad q=\partial V/\partial p$$

and so

$$p=P+O(P^2); \qquad q=Q+W(Q)\cdot P+O(P^2).$$

Defining the new coordinates by a generating function in this way guarantees that $\sum dq_i\wedge dp_i=\sum dQ_i\wedge dP_i$. In this new system of coordinates, f has the expression

$$q'=q+\rho+A(q)\cdot p+O(p^2), \qquad p'=p+O(p^2),$$

where

$$A(q)=B(q)+W(q+\rho)-W(q).$$

Thus, $B(q)$ is a symmetric $n\times n$ matrix depending in a C^∞ and $\mathbb{Z}^n$-periodic way on q and we wish to find a symmetric $n\times n$ matrix $W(q)$ depending in a C^∞ and $\mathbb{Z}^n$-periodic way on q such that $A(q)$ is constant (i.e. independent of q). Of course, it is enough to solve this difference equation separately for each entry. It is well known that it is possible to solve this difference equation when ρ satisfies the Diophantine condition we have imposed above (and only then): one may see this by expanding all the functions which appear in Fourier series in $q=(q_1,\ldots,q_n)$. The Diophantine condition on ρ is then precisely the condition for the resulting Fourier series for W to converge to a C^∞ function, whenever B is C^∞.

This completes our sketch of a proof that f has the normal form (in a neighborhood of $\mathfrak{L}$)

$$q'=q+\rho+A\cdot p+O(p^2), \qquad p'=p+O(p^2),$$

where A is an $n\times n$ symmetric matrix of real numbers, and $(q,p)=(q',p')\circ f$.

Now we consider a tubular neighborhood U of $\mathfrak{L}$ in N and a symplectic diffeomorphism g of U into N. We suppose that g is C^1 close to $f\,|\,U$. In addition, we suppose that g is a Hamiltonian perturbation of f, in the following sense. Since $\mathfrak{L}$ is Lagrangian, i.e. $i^*\omega=0$, where i is the inclusion of $\mathfrak{L}$ in N, it follows that the cohomology class of $\omega|U$ vanishes, since U (being a tubular neighborhood of a Lagrangian torus) is diffeomorphic to the Cartesian product of $\mathfrak{L}$ with an open ball. Therefore, $\omega|U=d\eta$, for a suitable 1-form η on U. Note that since $f(\mathfrak{L})=\mathfrak{L}$, and $f\,|\,\mathfrak{L}$ is homotopic to the identity, $f^*\eta$ is cohomologous to η, i.e. the cohomology class of $f^*\eta-\eta$ in $H^1(U,\mathbb{R})$ vanishes. We will say that g is a *Hamiltonian* perturbation of f if $(g^*\eta-\eta)|\mathfrak{L}$ is exact.

According to *KAM* theory, if r is sufficiently large, A is non-singular, and g is a C^r sufficiently small Hamiltonian perturbation of f, then there exists a compact submanifold $\mathfrak{L}'$ of U such that $g(\mathfrak{L}')=\mathfrak{L}'$ and $g|\mathfrak{L}'$ is C^1 conjugate to a translation of the torus T^n by ρ. In fact, it has been shown that this result is valid for $r>2\beta+2$, where β is the exponent which appears in the Diophantine condition. Cf. Moser [21], Salamon [24], and Salamon and Zehnder [25].

The purpose of this section is to show that when A is positive (or negative) definite, there is still a g-invariant set in U near $\mathfrak{L}$, if g is a C^1 small Hamiltonian perturbation of f. Moreover, if there is a g-invariant torus in U on which g is C^1 conjugate to a translation by ρ, then the set which we will construct is this torus. Thus, the set which we will construct may be regarded as a generalization of the *KAM* torus.

A related result is contained in the paper of Bernstein and Katok [6], where periodic orbits are constructed in a similar circumstance. However, the invariant sets which we construct are, in general, different from the periodic orbits found in [6].

The coordinates $(q, p) = (q_1, \ldots, q_n, p_1, \ldots, p_n)$ which we found above provide a symplectic diffeomorphism of an open neighborhood of $\mathfrak{L}$ in N onto an open neighborhood of $T^n \times O$ in $T^n \times \mathbb{R}^n = T^* T^n$. Thus, without loss of generality, we may suppose that $U = T^n \times V$, where V is an open ball about O in $\mathbb{R}^n$ and f and g map U into $T^n \times \mathbb{R}^n$. We may also suppose that $f(g, p) = (q', p')$ where

$$q' = q + \rho + A \cdot p + O(p^2), \qquad p' = p + O(p^2),$$

and A is positive definite. For, we may reduce the case when A is negative definite to the case when A is positive definite by replacing p by $-p$.

Let $f_t \colon T^n \times \mathbb{R}^n \to T^n \times \mathbb{R}^n$ be defined by $f_t(q, p) = (q + tp + tA \cdot p)$, for $t \in \mathbb{R}$, so $f(q, p) = f_1(q, p) + O(p^2)$. We need a generating function for $g \circ f_1^{-1}$ or, more precisely, a function $G(q', p) = q' \cdot p + G_1(q', p)$, where $G_1 \colon T^n \times \mathbb{R}^n \to \mathbb{R}$ is C^2, lies in a pre-assigned C^2 neighborhood of O, and vanishes outside of $T^n \times B^n$, where B^n is the unit ball in $\mathbb{R}$. We require that this satisfy the condition to be a generating function for $g \circ f_1^{-1}$, i.e. for $q \in T^n$, $p \in V$, we should have

$$g \cdot f_1^{-1}(q, p) = (q', p')$$

if and only if

$$q = \partial G / \partial p \quad \text{and} \quad p' = \partial G / \partial q'.$$

For the case $g = f$, we may find such a generating function, after possibly replacing V with a smaller ball containing the origin. Since g is a C^1 small Hamiltonian perturbation of f, we may still find such a generating function in general. The proof, both for the case $g = f$ and in general, (using the case $g = f$) is the usual calculus exercise combined with standard extension lemmas. We leave it to the reader. The reason for doing the argument in two steps this way is to show that the amount of shrinking of V which is required depends only on f, not on g.

Let $u \colon [0, 1] \to [0, 1]$ be such that u is C^∞, u vanishes identically in a neighborhood of 0 and u is identically 1 in a neighborhood of 1. Let $\phi_t \colon T^n \times \mathbb{R}^n \to T^n \times \mathbb{R}^n$ be the Hamiltonian diffeomorphism whose generating function is $V_t(q', p) = q' \cdot p + u(t)\, G_1(q', p)$, i.e. the diffeomorphism which satisfies the condition that

$$\phi_t(q, p) = (q', p')$$

if and only if

$$q = \partial V_t / \partial p, \qquad p' = \partial V_t / \partial q'.$$

We assume that such a diffeomorphism ϕ_t exists. This will be the case if G_1 is sufficiently close to 0 in the C^2 topology.

We let $g_t = \phi_t \circ f_t$. By construction, $\{g_t\}$ is a 1-parameter family of Hamiltonian diffeomorphisms of $T^n \times \mathbb{R}^n$ with $g_0 =$ identity. Thus, g_t is the Hamiltonian flow associated to a time dependent Hamiltonian h_t, i.e., we may find a function $h_t: T^n \times \mathbb{R}^n \to \mathbb{R}$ such that

$$\frac{d(q \circ g_t)}{dt} = \frac{\partial h_t}{\partial p} \quad \text{and} \quad \frac{d(p \circ g_t)}{dt} = -\frac{\partial h_t}{\partial q}.$$

Moreover, h_t is a C^2-small perturbation of $H(q,p) = \rho \cdot p^T + p \cdot A \cdot p^T/2$ (where p^T denotes the column vector which is the transpose of the row vector p) in the sense that if $G_1 = 0$, then $h_t = R$, and h_t depends continuously on G_1 in the C^2 topology, provided that G_1 is in a sufficiently C^2-small neighborhood of the identity. The verification of this may be done in two stages: First, $G_1 \to dg_t/dt$ is continuous, with respect to the C^2 topology on the functions G_1 and the C^1 topology on the functions dg_t/dt. Second, $dg_t/dt \to h_t$ is continuous, with respect to the C^1 topology on the functions dg_t/dt and the C^2 topology on the functions h_t. This is because h_t is obtained from dg_t/dt by an integration.

Since H has positive definite Hessian second derivative along the fibers of $T^* T^n = T^n \times \mathbb{R}^n$, so does h_t if the latter is close enough to H in the C^2 topology. As we observed above, this will be the case if we shrink V enough and choose g close enough in the C^1 norm to f, so as to be able to choose G_1 in a suitable C^2-small neighborhood of 0.

The Hamiltonian h_t, in addition to having positive definite Hessian second derivative along the fibers of $T^* T^n$, equals H outside of a compact set. Consequently it has superlinear growth along the fibers. By applying the Legendre transformation, we get a Lagrangian system, which is equivalent to the original system. Recall that the Lagrangian of this system is $L(q, \dot{q}, t) = \dot{q} \cdot p^T - h_t(q, p)$, where $\dot{q}$ is defined by the Legendre transformation $\dot{q} = \partial h_t / \partial p$. Since h_t has positive definite Hessian second derivative and superlinear growth in the p variables and is C^2, the Legendre transformation is a C^1 diffeomorphism of $T^* T^n \times \mathbb{R}/\mathbb{Z}$ onto $TT^n \times \mathbb{R}/\mathbb{Z}$, which commutes with the projections onto $T^n \times \mathbb{R}/\mathbb{Z}$. (As usual, $T^* T^n$ and TT^n denote the cotangent bundle and the tangent bundle of the torus.) The inverse transformation is given by $p = \partial L / \partial \dot{q}$. Notice also that $\partial L/\partial q = \partial h_t / \partial q$ and $\partial L / \partial t = -\partial h_t / \partial t$, so all first partial derivatives of L are C^1, i.e. L is C^2. Notice that $\partial^2 L / \partial \dot{q}^2 = \partial p / \partial \dot{q} = (\partial \dot{q} / \partial p)^{-1} = (\partial^2 h_t / \partial p^2)^{-1}$. Consequently, $\partial^2 L / \partial \dot{q}^2$ is positive definite. Since $h_t(q,p) = H(q,p) = \rho \cdot p^T + p \cdot A \cdot p^T/2$ outside a compact set, $L(q, \dot{q}, t) = (\dot{q} - \rho) \cdot A^{-1} (\dot{q} - \rho)^T 2$ outside a compact set, and so L has superlinear growth. Furthermore, the flow defined by the Euler-Lagrange equation is complete in this case, because it is integrable outside a compact subset of $TT^n \times (\mathbb{R}/\mathbb{Z})$.

Thus, we have verified all the conditions imposed on L in § 1. It follows that the results stated in §§ 1–4 apply to this L. In view of the definition of this L, they translate to results about g_1. For, $T^* T^n = T^n \times \mathbb{R}^n$ is a global Poincaré surface of section of the Hamiltonian flow with Hamiltonian h_t and this flow is C^1 conjugate to the flow Φ_L defined by the Euler-Lagrange equations associated to L. The induced mapping on this Poincaré surface of section is g_1. To put this in another way, Φ_L is C^1 conjugate to the suspension of g_1, i.e. the flow $\partial/\partial t$ on the quotient manifold of $T^* T^n \times \mathbb{R}$ obtained by identifying (ξ, t) with $(g_1(\xi), t+1)$.

Thus, there is a one-one correspondence between invariant probability measures of g_1 and invariant probability measures of Φ_L. We may define the *average action* of a g_1-invariant probability measure as the average action of the corresponding Φ_L-invariant measure. The function which assigns to a g_1-invariant probability measure μ its average action is actually a symplectic invariant of g_1 modulo addition of affine functions of the rotation vector. A special case of this was proved in [18] (by arguments which go back to [8]), and the same argument carries over to the situation we are considering here, without change. We will use the same symbol for an invariant probability measure of g_1 and the corresponding invariant probability measure of Φ_L.

Of course, what we want are results about g, not about g_1. By construction of g_1, we have $g_1|U'=g$, for an appropriate neighborhood U' of $\mathfrak{L}$ in U, so results about invariant measures μ of g_1 apply to g, as long as $\operatorname{supp}\mu\subset U'$.

First consider the case when $g=f$. In this case, we have the *KAM* torus $\mathfrak{L}=T^n\times O$ which supports a unique invariant measure μ_0, which is minimal (Appendix 2). In other words, $\beta(\rho)=A(\mu_0)$, where β is the function which appears Theorem 1. (We identify $H_1(T^n,\mathbb{R})$ with $\mathbb{R}^n$.) It follows from Birkhoff normal form ([11], Appendix 2) that β is differentiable at ρ and that the unique supporting hyperplane of the epigraph of β at $(\rho,\beta(\rho))$ meets epigraph β only at that one point. Let $c_0\in H^1(T^n,\mathbb{R})=\mathbb{R}^n$ be the derivative of β at ρ.

For $c\in H^1(T^n,\mathbb{R})=\mathbb{R}^n$, let $\operatorname{supp}\mathfrak{M}_c\subset TT^n$ denote the support of the set of g_1-invariant probability measures μ which minimize $A_c(\mu)=A(\mu)-\langle c,\rho(\mu)\rangle$. It is easy to see that for every neighborhood $\mathfrak{N}$ of μ_0 in the vague topology, there are neighborhoods $\mathfrak{N}_1$ of f in the C^1 topology on Hamiltonian perturbations of f, and $\mathfrak{N}_2$ of c_0 in $H_1(T^n,\mathbb{R})=\mathbb{R}^n$ such that if g is in $\mathfrak{N}_1$ and c is in $\mathfrak{N}_2$, then $\mathfrak{M}_c\subset N$. However, by Theorem 2, $\operatorname{supp}\mathfrak{M}_c$ is the graph of a Lipschitz function from a subset of T^n to $\mathbb{R}^n$. Moreover, the proof of Theorem 2 gives an *a priori* bound on the Lipschitz constant. Choosing $\mathfrak{N}$ appropriately, using the fact that $\mathfrak{M}_c\subset\mathfrak{N}$ and the *a priori* bound on the Lipschitz constant, we obtain $\operatorname{supp}\mathfrak{M}_c\subset U'$.

Supp $\mathfrak{M}_c$ is the g-invariant set in U' which we sought. As c tends to c_0 and g tends to f in the C^1 topology, $\operatorname{supp}\mathfrak{M}$ converges to the *KAM* torus in the Hausdorff topology, as may be seen by the argument above.

We may summarize what we have proved as follows:

Proposition 5 *Let f be a C^∞ symplectic diffeomorphism of a $2n$-dimensional symplectic manifold N and let $\mathfrak{L}$ be a KAM torus of f, i.e. suppose that $\mathfrak{L}$ satisfies the conditions listed at the beginning of this section. Let q (defined* $\operatorname{mod}\mathbb{Z}^n$*) and p be symplectic coordinates, defined in a neighborhood of $\mathfrak{L}$, such that $\mathfrak{L}=\{p=0\}$ and f has the form*

$$q'=q+p+A\cdot p+O(p^2),\qquad p'=p$$

where $(q',p')\cdot f=(q,p)$. Suppose that the symmetric matrix A of real numbers is positive definite.

Let $U=T^n\times V$ be an open neighborhood of $\mathfrak{L}$ in N. Let c_0 be the derivative of β at ρ, where ρ is the rotation vector of $f|\mathfrak{L}$. Then we have the following:

If c is close enough to c_0 in $H^1(T^n,\mathbb{R})=\mathbb{R}^n$ and g is close enough to f in the C^1 topology on Hamiltonian perturbations of f, then $\operatorname{supp}\mathfrak{M}_c\subset U'$ *and consequently is g-invariant. Moreover,* $\operatorname{supp}\mathfrak{M}_c$ *is the graph of a function from a subset of T^n to V and it converges (in the Hausdorff topology) to $\mathfrak{L}$ as c tends to c_0 and g tends to f in the C^1 topology. Moreover, if g has a KAM torus C^1-sufficiently close to $\mathfrak{L}$, then that torus is one of the sets $\mathfrak{M}_c$.*

The last sentence of this proposition follows from Appendix 2. The rest has been proved above. Note that the restriction of β to a sufficiently small neighborhood of c_0 depends only on f, not on its extension to $T^n \times \mathbb{R}^n$. This is a consequence of the fact that supp $\mathfrak{M}_c$ lies in U', for c close enough to c_0.

6 Twist maps

In this section, we apply the theory developped in §§ 1–4 to the case when $M = S^1$. In this case, TM is the cylinder. Moser has shown [20] that any finite composition of exact area preserving twist maps of the cylinder may be represented as the time one map associated to a Lagrangian satisfying our conditions for the case $M = S^1$. Thus, the results we prove in this section apply to finite compositions of twist maps. However, the results we prove in this section have already been prove by related methods in previous articles. The purpose of this section is to show that the results of this article generalize earlier results about twist maps. Of the many articles which discuss these results about twist maps, that of Denzler [10] is closest to the approach which we adopt here. See also Mather [18], which expresses results about twist diffeomorphisms in terms of minimal measures.

Proposition 6 *In the case that $M = S^1$, the function $\beta: H_1(M, \mathbb{R}) \to \mathbb{R}$ is strictly convex, i.e. every point on graph β is an extremal point of the epigraph of β.*

Proof. Suppose the contrary. Then there exists a supporting hyperplane $l \subset H_1(M, \mathbb{R}) \times \mathbb{R} = \mathbb{R}^2$ of epigraph β which meets graph β in more than one point. Let $c \in H^1(M, \mathbb{R})$ be the slope of l. According to Proposition 4 and Theorem 2, supp $\mathfrak{M}_c$ is a compact subset of $TM \times (\mathbb{R}/\mathbb{Z}) = S^1 \times \mathbb{R} \times (\mathbb{R}/\mathbb{Z})$, whose projection on $M \times (\mathbb{R}/\mathbb{Z}) = S^1 \times (\mathbb{R}/\mathbb{Z})$ is injective. Since l meets graph β in more than one point, $l \cap$ graph β is a closed line segment, and its endpoints are extremal points of epigraph β.

From the discussion at the end of § 2, it follows that there exist ergodic invariant measures μ_0, μ_1 such that $(\rho(\mu_0), A(\mu_0))$ and $(\rho(\mu_1), A(\mu_1))$ are the endpoints of this line segment. Let $\zeta_0: \mathbb{R} \to TM \times \mathbb{R}/\mathbb{Z}$ and $\zeta_1: \mathbb{R} \to TM \times \mathbb{R}/\mathbb{Z}$ be Birkhoff generic trajectories for μ_0 and μ_1, resp. Let $\gamma_0, \gamma_1: \mathbb{R} \to M \times (\mathbb{R}/\mathbb{Z}) = S^1 \times \mathbb{R}/\mathbb{Z}$ denote the projections of ζ_0, ζ_1. Let $\tilde{\gamma}_0, \tilde{\gamma}_1: \mathbb{R} \to \mathbb{R}^2$ denote the lifts to the universal cover.

Since $\mu_0 \neq \mu_1$, we have $\zeta_0 \neq \zeta_1$. Since these are trajectories, their images in $TM \times \mathbb{R}/\mathbb{Z}$ are disjoint. These images lie in supp $\mathfrak{M}_c$. Since the projection of supp $\mathfrak{M}_c$ on $M \times \mathbb{R}/\mathbb{Z}$ is injective, it follows that the images of γ_0 and γ_1 are also disjoint.

On the other hand, the asymptotic slopes of $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ are $\rho(\mu_0)$ and $\rho(\mu_1)$, resp., since ζ_0 and ζ_1 are Birkhoff generic trajectories for μ_0 and μ_1, resp. This implies that the curves $\tilde{\gamma}_0, \tilde{\gamma}_1$ cross, contradicting the fact that the images of γ_0 and γ_1 are disjoint. This contradiction proves the proposition. □

The idea of this proof and of Proposition 8 is similar to the idea of Moser, explained in Denzler [10].

Let $h \in H_1(S^1, \mathbb{R})$, let $l \subset H^1(S^1, \mathbb{R}) \times \mathbb{R}$ be a supporting hyperplane of epigraph β which touches epigraph β at h, and let c be the slope of l. Let $M_h = (TS^1$

$\times O) \cap \operatorname{supp} \mathfrak{M}_c$. Note that M_h is independent of the choice of l by the strict convexity of β, since $\mathfrak{M}_c$ is the set of invariant measures which minimize A, subject to the condition of having rotation number h.

Proposition 7 *The projection* π_1 *of* $M_h(\subset TS^1)$ *on* S^1 *is injective and the inverse* $\pi_1^{-1}: \pi_1(M_h) \to M_h \subset TS^1$ *is Lipschitz.*

Proof. Immediate from Theorem 2. □

Let f be the section mapping of $TS^1 = TS^1 \times O$ into itself, corresponding to the Euler-Lagrange flow associated to L. By definition, $f(M_h) = M_h$. Let $\pi: \mathbb{R}^2 \to S^1 \times \mathbb{R} = TS^1$ denote the projection and let $\tilde{M}_h = \pi^{-1}(M_h) \subset \mathbb{R}^2$. Let $\pi_1: \mathbb{R}^2 \to \mathbb{R}$ denote the projection on the first factor and let $\tilde{f}$ denote a lift of f to the universal cover. By Proposition 7, $\pi_1: \tilde{M}_h \to \mathbb{R}$ is injective, so that $\tilde{M}_h$ inherits an order from that on $\mathbb{R}$.

Proposition 8 $\tilde{f}: \tilde{M}_h \to \tilde{M}_h$ *is order preserving.*

Proof. If not, the projection of $\operatorname{supp} \mathfrak{M}_c$ on $S^1 \times (\mathbb{R}/\mathbb{Z})$ would not be injective, contradicting Theorem 2. □

Corollary. *If* h *is irrational,* M_h *supports a unique* f*-invariant measure* μ_h*, which is the unique minimal measure of rotation number* h*.*

Proof. To show that M_h supports a unique f-invariant measure, use Proposition 8 and copy the well known proof that an orientation preserving homeomorphism of the circle of irrational rotation number has a unique invariant measure.

Since all minimal measures of rotation number h have support in M_h, it follows that μ_h is the only one. □

Appendix 1

Tonelli's theorem

We stated a version of Tonelli's Theorem in § 2. This is slightly different from any version we have found in the published literature. However, it may be proved by modification of the proof found in a standard text [2]. (For a thorough discussion of Tonelli's theorem, especially in more variables, see [9]). For completeness sake, we prove our version of Tonelli's theorem here. As we observed in § 2, it is enough to prove:

Lemma. *Let* $K \in \mathbb{R}$*. The set* $\{A \leqq K\}$*, consisting of all* $\gamma \in C^{ac}([a, b], M)$ *for which* $A(\gamma) \leqq K$*, is compact in the* C^0*-topology.*

The rest of this appendix is devoted to the proof of this lemma and its addendum.

The first step is the observation that the family $\{A \leqq K\}$ of curves satisfies the condition of absolute equicontinuity: For every $\varepsilon > 0$, there exists $\delta > 0$ such that if $a \leqq a_0 < b_0 \leqq a_1 < b_1 \leqq \ldots \leqq a_n < b_n \leqq b$ and $\sum_{i=0}^{n} b_i - a_i < \delta$, then $\sum_{i=0}^{n} \operatorname{dist}(\gamma(a_i), \gamma(b_i)) < \varepsilon$. For this, we use the superlinear growth of L: Choose

C so that $K/C < \varepsilon/2$ and B so that $\|\xi\| \geqq B$ implies $L(\xi, t) \geqq C\|\xi\|$. Let $\delta = \varepsilon/2B$. For $\gamma \in \{A \leqq K\}$, let $E = \{t \in [a, b]: \|d\gamma(t)\| \geqq B\}$. Then

$$\int_E \|d\gamma(t)\|\, dt \leqq C^{-1} \int_E L(d\gamma(t), t)\, dt \leqq K/C < \varepsilon/2.$$

Let $J = [a_1, b_1] \cup \ldots \cup [a_n, b_n]$. It follows that

$$\sum_{i=0}^{n} \operatorname{dist}(\gamma(a_i), \gamma(b_i)) \leqq \int_J \|d\gamma(t)\|\, dt < \sum_{i=0}^{n} (b_i - a_i) B + \varepsilon/2 < \varepsilon.$$

Note that δ is independent of γ, as long as $A(\gamma) \leqq K$.

In particular, the family $\{A \leqq K\}$ of curves is equicontinuous. Since these curves lie in the compact metric space M, it follows from the Ascoli-Arzela theorem that every sequence $\gamma_1, \gamma_2, \ldots$, in $\{A \leqq K\}$ has a subsequence which is convergent with respect to the C^0 topology. It follows immediately from the absolute *equi*continuity of the sequence that the limit γ of any convergent subsequence is absolutely continuous.

So far, we have shown that any sequence in $\{A \leqq K\}$ has a subsequence $\gamma_1, \gamma_2, \ldots$ which converges in the C^0 topology to an absolutely continuous curve γ. To complete the proof of the lemma, we will show that $A(\gamma) \leqq K$.

Consider $t \in [a, b]$ where γ is differentiable. Let (U, x) be a C^∞ coordinate chart about $\gamma(t)$. Here, $x = (x_1, \ldots, x_n)$ is a local system of coordinates. We let $(x, \dot{x}) = (x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n)$ denote the system of coordinates on TU canonically associated to it, and we express $d\gamma(t)$ in these coordinates as $(\gamma(t), \dot{\gamma}(t))$. For $\varepsilon > 0$, we have

$$L(x, \dot{x}, s) \geqq L(\gamma(t), \dot{\gamma}(t), t) + dL_{(\gamma(t), \dot{\gamma}(t))}(0, \dot{x} - \dot{\gamma}(t)) - \varepsilon,$$

if x is close enough to $\gamma(t)$ and s is close enough to t. For $s = t$ and $x = \gamma(t)$, this inequality (with $\varepsilon = 0$) follows immediately from the fiberwise convexity of L. There exists $C > 0$ such that for $\|\dot{x}\| > C$, this inequality follows from the superlinear growth condition on L. For $\|\dot{x}\| \leqq C$, this inequality follows from the continuity of L and the fact that it holds for $\varepsilon = 0$ when $s = t$ and $x = \gamma(t)$, provided x is close enough to $\gamma(t)$ and s is close enough to t, although how close these must be taken depends on C and ε.

We will apply this inequality with $x = \gamma_i(s)$, $\dot{x} = \dot{\gamma}_i(s)$. Note that

$$\begin{aligned}(\delta + \delta')^{-1} &\int_{t-\delta'}^{t+\delta} dL_{(\gamma(t), \dot{\gamma}(t))}(0, \dot{\gamma}_i(s) - \dot{\gamma}(t))\, ds \\ &= (\delta + \delta')^{-1} dL_{(\gamma(t), \dot{\gamma}(t))}(0, \gamma_i(t + \delta) - \gamma_i(t - \delta') \\ &\quad - (\delta + \delta')\, \dot{\gamma}(t)).\end{aligned}$$

Taking $\lim\limits_{\delta, \delta' \downarrow 0} \lim\limits_{i \to \infty}$ of this quantity, we obtain zero, since γ_i converges C^0 to γ. Thus, the inequality above implies

$$\lim_{\delta, \delta' \downarrow 0} \inf \liminf_{i \to \infty} (\delta + \delta')^{-1} A(\gamma_i | [t - \delta', t + \delta]) \geqq L(d\gamma(t), t).$$

This is valid at any point where γ is differentiable.

This inequality implies that for every $\varepsilon>0$, there exists $\delta_0>0$ such that if $0<\delta, \delta' \leqq \delta_0$, then

$$\liminf_{i\to\infty}(\delta+\delta')^{-1} A(\gamma_i|[t-\delta', t+\delta]) \geqq L(d\gamma(t), t) - \varepsilon/2.$$

Until now, we have not proved that $L(d\gamma(t), t)$ is an integrable function of t. For this reason, it is convenient to introduce the functions $u_C(t) = \min(L(d\gamma(t), t), C)$ and $U_C(t) = \int_a^t u_C(s)\, ds$. We let E_C denote the set of points $t\in[a, b]$ where γ and U_C are differentiable and $u_C(t) = dU_C(t)/dt$. (I am indebted to Odet Schramm for a considerable simplification at this point of the proof which I presented in my graduate course in spring 1988. I spent an hour proving that E_C or some similarly defined set has full measure. At the end of the hour, he remarked that one of the conditions in my definition amounted to $u_C(t) = dU_C(t)/dt$, and formulating the condition this way showed that my result about full measure was an immediate consequence of well known results in function theory.) If $t\in E_C$, then for every $\varepsilon>0$, we may choose $\delta_0>0$ such that if $0<\delta, \delta' \leqq \delta_0$, then

$$L(d\gamma(t), t) - \varepsilon/2 \geqq (\delta+\delta')^{-1}(U_C(t+\delta) - U_C(t-\delta')) - \varepsilon,$$

if $t\in E_C$. Combining this with the previous inequality, we obtain

$$\liminf_{i\to\infty}(\delta+\delta')^{-1} A(\gamma_i|[t-\delta', t+\delta])$$
$$\geqq (\delta+\delta')^{-1}(U_C(t+\delta) - U_C(t-\delta')) - \varepsilon,$$

for $t\in E_C$ and $0<\delta$, $\delta' \leqq \delta_0$.

The set E_C has full measure in $[a, b]$, since u_C is a bounded measurable function. We have shown that for each $t\in E_C$, there exists δ_0 such that the inequality above holds for $0<\delta, \delta' \leqq \delta_0$. We construct a countable sequence $[a_1, b_1], [a_2, b_2], \ldots, [a_j, b_j], \ldots$ of closed intervals which cover E_C, which are mutually disjoint, and for which

$$\liminf_{i\to\infty}(b_j-a_j)^{-1} A(\gamma_i|[a_j, b_j])$$
$$\geqq (b_j-a_j)^{-1}(U_C(b_j) - U_C(a_j)) - \varepsilon,$$

as follows. Let $t_1, t_2, \ldots$ be a countable dense sequence in E_C. Let $[a_1, b_1]$ be such an interval containing t_1 for which $\min(b_1-t_1, t_1-a_1)$ is as large as possible. Assuming $[a_1, b_1], \ldots, [a_{i-1}, b_{i-1}]$ have been constructed, we construct $[a_i, b_i]$, as follows. Let t_j be the first element of the sequence which is not in $[a_1, b_1] \cup \ldots \cup [a_{i-1}, b_{i-1}]$. (If there is none then $[a_1, b_1], \ldots, [a_{i-1}, b_{i-1}]$ already covers E_C.) Let $I=[a', b']$ be the closure of the component of the complement of $[a_1, b_1] \cup \ldots \cup [a_{i-1}, b_{i-1}]$ in $\mathbb{R}$ which contains t_j. For $t_j\in(a, b) \subset [a', b']$, define $c(a, b) = +\infty$ if $a=a'$, $b=b'$, $c(a, b) = b-t_j$ if $a=a'$, $b<b'$, $c(a, b) = t_j - a$, if $a'<a$, $b=b'$, and $c(a, b) = \min(b-t_j, t_j-a)$, if $a'<a<b<b'$. Choose $[a_i, b_i]$ with $t_j\in(a_i, b_i) \subset [a', b']$, satisfying the above inequality, and so that $c(a_i, b_i)$ is as large as possible. In view of the fact that for each $t\in E_C$, there exists δ_0 such that the previous inequality holds for $0<\delta, \delta' \leqq \delta_0$, it is easily seen that

$[a_1, b_1], \ldots, [a_i, b_i], \ldots$ cover E_C. Since E_C has full measure in $[a, b]$, L is bounded below, and the $[a_i, b_i]$'s are mutually disjoint, it follows that

$$\liminf_{i\to\infty} A(\gamma_i) \geqq U_C(b) - U_C(a) - \varepsilon(b-a).$$

Since this is true for every $\varepsilon > 0$ and $C \in \mathbb{R}$ and since $\gamma_i \in \{A \leqq K\}$, we obtain

$$K \geqq \liminf_{i\to\infty} A(\gamma_i) \geqq \lim_{C\uparrow\infty} U_C(b) - U_C(a) = A(\gamma). \quad \square$$

Proof of the addendum. We must show that if γ_i converges C^0 to γ and $A(\gamma_i) \to A(\gamma) < \infty$, then γ_i converges C^{ac} to γ, i.e.

$$\lim_{i\to\infty} \int_a^b \operatorname{dist}(d\gamma_i(t), d\gamma(t))\, dt = 0.$$

Let $u(t) = L(d\gamma(t), t)$. Since $A(\gamma) < \infty$, the function u is integrable. Let $U(t) = \int_a^t u(s)\, ds$. Let E denote the set of $t \in [a, b]$ at which U is differentiable and $u(t) = dU(t)/dt$.

Consider $t \in [a, b]$ where γ is differentiable and let (U, x) be a C^∞ coordinate chart about $\gamma(t)$. Using the same notation as in the proof of the lemma which we have just given, we have

$$\lim_{\delta, \delta' \downarrow 0} \lim_{i\to\infty} (\delta + \delta')^{-1} \int_{t-\delta'}^{t+\delta} (\dot{\gamma}_i(s) - \dot{\gamma}(t))\, ds = 0.$$

This follows immediately from the assumption that γ_i converges C^0 to γ and the assumption that γ is differentiable at t.

Consider $\delta, \delta' > 0$. By the lemma we have just proved,

$$\liminf_{i\to\infty} A(\gamma_i | [a, t-\delta'] \cup [t+\delta, b]) \geqq A(\gamma | [a, t-\delta'] \cup [t+\delta, b]).$$

From our assumption that $A(\gamma_i)$ converges to $A(\gamma)$, we therefore obtain

$$\limsup_{i\to\infty} A(\gamma_i | [t-\delta', t+\delta]) \leqq A(\gamma | [t-\delta', t+\delta]).$$

Now suppose $t \in E$, so that

$$\lim_{\delta, \delta' \downarrow 0} (\delta + \delta')^{-1} A(\gamma | [t-\delta', t+\delta]) = L(d\gamma(t), t).$$

Combining the above estimate on lim sup with the argument in the proof of the lemma which shows that

$$\liminf_{\delta, \delta' \downarrow 0} \liminf_{i\to\infty} (\delta + \delta')^{-1} A(\gamma_i | [t-\delta', t+\delta])$$
$$\geqq L(d\gamma(t), t),$$

and using the fact that

$$\dot\gamma(t)=\lim_{\delta,\delta'\downarrow 0}\lim_{i\to\infty}(\delta+\delta')^{-1}\int_{t-\delta'}^{t+\delta}\dot\gamma_i(s)\,ds,$$

we obtain

$$\lim_{\delta,\delta'\downarrow 0}\limsup_{i\to\infty}(\delta+\delta')^{-1}\int_{t-\delta'}^{t+\delta}\operatorname{dist}(\dot\gamma_i(s),\dot\gamma(t))\,ds=0,$$

for $t\in E$.

For, if $\Delta>0$, there exists $\eta>0$ such that

$$L(x,\dot x,s)\geqq L(\gamma(t),\dot\gamma(t),t)+dL_{(\gamma(t),\dot\gamma(t))}(0,\dot x-\dot\gamma(t))\\ +\eta\operatorname{dist}(\dot x,\dot\gamma(t))$$

provided that $\operatorname{dist}(\dot x,\dot\gamma(t))\geqq\Delta$, x is close enough to $\gamma(t)$ and s is close enough to t. There exists $C>0$ such that if $\|\dot x\|\geqq C$, this inequality follows from the superlinear growth condition on L. Moreover, by the positive definiteness condition on L, this inequality holds for $x=\gamma(t)$ and $s=t$. For $\|\dot x\|\leqq C$, this inequality, with η replaced by a smaller positive number, holds for x close enough to $\gamma(t)$ and s close enough to t, by the continuity of L. Thus, our previous argument shows that

$$\limsup_{\delta,\delta'\downarrow 0}\limsup_{i\to\infty}(\delta+\delta')^{-1}A(\gamma_i|[t-\delta',t+\delta])\\ \geqq L(d\gamma(t),t)+\limsup_{\delta,\delta'\downarrow 0}\limsup_{i\to\infty}(\delta+\delta')^{-1}\int_{t-\delta'}^{t+\delta}\phi_\Delta(s)\,\eta\operatorname{dist}(\dot\gamma_i(s),\dot\gamma(t))\,ds,$$

where $\phi_\Delta(s)=1$ when $\operatorname{dist}(\dot\gamma(s),\dot\gamma(t))\geqq\Delta$ and $\phi_\Delta(s)=0$ otherwise. Therefore,

$$\limsup_{\delta,\delta'\downarrow 0}\limsup_{i\to\infty}(\delta+\delta')^{-1}\int\phi_\Delta(s)\operatorname{dist}(\dot\gamma_i(s),\dot\gamma(t))\,ds=0.$$

Since this holds for every $\Delta>0$, we obtain

$$\lim_{\delta,\delta'\downarrow 0}\limsup_{i\to\infty}(\delta+\delta')^{-1}\int\operatorname{dist}(\dot\gamma_i(s),\dot\gamma(t))\,ds=0,$$

for $t\in E$, as asserted.

We may, in particular, apply this for the constant sequence $\gamma_i=\gamma$ and obtain

$$\lim_{\delta,\delta'\downarrow 0}(\delta+\delta')^{-1}\int_{t-\delta'}^{t+\delta}\operatorname{dist}(\dot\gamma(s),\dot\gamma(t))\,ds=0.$$

Combining these two inequalities, we obtain

$$\lim_{\delta,\delta'\downarrow 0}\limsup_{i\to\infty}(\delta+\delta')^{-1}\int_{t-\delta'}^{t+\delta}\operatorname{dist}(\dot\gamma_i(s),\dot\gamma(s))\,ds=0.$$

Let $F_i(t)=\int_a^t \operatorname{dist}(d\gamma_i(s), d\gamma(s))\,ds$ and $f_i(t)=dF_i(t)/dt$ (so that $f_i(t) =\operatorname{dist}(d\gamma_i(t), d\gamma(t))$ almost everywhere). By what we have just proved $f_i(t)\to 0$ as $i\to\infty$ if $t\in E$ and $f_i(t)$ is defined for every i.

Thus, f_i converges pointwise almost everywhere to 0. What we wish to prove is equivalent to $\int_a^b f_i(t)\,dt\to 0$, as $i\to\infty$. For any $C>0$, we have that $\int_a^b \min(f_i(t), C)\,dt\to 0$, as $i\to\infty$ by the bounded convergence theorem. Moreover, for any $\varepsilon>0$, there exists $C>0$ and $i_0>0$ such that

$$\int_a^b [f_i(t)-\min(f_i(t), C)]\,dt<\varepsilon$$

for all $i\geqq i_0$. For, at least one of $\|d\gamma_i(s)\|$ or $\|d\gamma(s)\|$ is $\geqq C/3$, so we can use the superlinear growth condition on L, together with the assumption that there is a uniform bound on $A(\gamma_i)$, to obtain this estimate. Thus, $\int_a^b f_i(t)\,dt\to 0$, as was required to be proved. □

Appendix 2

In this appendix, we prove the theorem of Weierstrass which was stated in § 2. This result is only slightly different from results stated in [7], but we give the proof for completeness sake. We follow the method of [7], which is due to Weierstrass. We also use Weierstrass's method to show that the unique invariant measure on a KAM torus which is a graph (§ 5) is minimal.

From classical mechanics, it is known that there is a 2-form Ω on $TM\times\mathbb{R}$ which may be expressed in terms of C^∞ local coordinates $x=(x_1, \ldots, x_n)$ defined on an open set U in M as $\Omega=\Sigma\, dp_i\wedge dx_i-dH\wedge dt$, where $p_i=\partial L/\partial \dot{x}_i$ and $H =\Sigma\,\dot{x}_i\,p_i-L$. As usual, t denotes the $\mathbb{R}$ coordinate, and $(x,\dot{x}) =(x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n)$ defines the system of local coordinates on TU, canonically associated to $(x_1, \ldots, x_n)$.

Lemma 1 *Let V be a connected, smooth m-manifold ($m=\dim M$) and let $a<b\in\mathbb{R}$ be real numbers. Let $\Phi\colon V\times[a,b]\to TM\times\mathbb{R}$ be a C^1 mapping with the following properties:*

1) *$\Phi(v,t)=(\Phi_1(v,t), t)$ with $\Phi_1(v,t)\in TM$, for all $v\in V$, $t\in[a,b]$,*

2) *For each $v\in V$, the mapping $t\to\Phi(v,t)$ is a trajectory of the Euler-Lagrange flow,*

3) *$\Phi^*\Omega=0$, and*

4) *$\pi\Phi$ is a diffeomorphism of $V\times[a,b]$ onto an open subset of $M\times[a,b]$, where $\pi\colon TM\times[a,b]\to M\times[a,b]$ denotes the projection.*

Then for any compact subset V_1 of V there exist $C_0, C_1>0$, such that if $v\in V_1$ and $\gamma(t)=\Phi_1(v,t)$, for $a\leqq t\leqq b$, then $A(\gamma_1)\geqq A(\gamma)+F(d_{ac}(\gamma,\gamma_1))$ for any absolutely continuous curve $\gamma_1\colon[a,b]\to M$ such that $\gamma_1(a)=\gamma(a)$, $\gamma_1(b)=\gamma(b)$,

$\gamma_1(t) \in \Phi_1(V_1 \times t)$, *for* $a \leq t \leq b$, *and* γ_1 *is homologous* (*in* $\Phi_1(V \times t)$) *to* γ *rel. endpoints. Here*

$$F(t) = \min(C_0 t^2, C_1 t).$$

This lemma is the basis of Weierstrass's method, as explained in [7].

Proof. From classical mechanics, it is know that there is a 1-form η on $TM \times \mathbb{R}$ which may be expressed in local coordinates as $\eta = \Sigma p_i\, dx_i - H\, dt$, so $\Omega = d\eta$ (where x_i, p_i, t, and H are as above). Since $\Phi^* \Omega = 0$, we have that $\Phi^* \eta$ is closed. Since η is C^∞ and Φ is C^1, we have that $\Phi^* \eta$ is C^0. (Note that we may still say that $\Phi^* \eta$ is closed, since we may define $d\Phi^* \eta$ in the sense of distributions.) Let V be the covering space of V defined by $\pi_1(\tilde{V}) = \ker(\pi_1(V) \to H_1(V, \mathbb{R}))$, and let $p: \tilde{V} \times [a, b] \to V \times [a, b]$ denote the projection. Since $\Phi^* \eta$ is closed, there is a C^1 function W on $\tilde{V} \times [a, b]$ such that $dW = p^* \Phi^* \eta$. Let $L^*: T\tilde{V} \times [a, b] \to \mathbb{R}$ be defined by

$$L^* = L \circ T(\pi \Phi p) - d_V W - \partial W / \partial t.$$

Here, $T(\pi \Phi p): T\tilde{V} \times [a, b] \to TM \times [a, b]$ is the tangent mapping associated to $\pi \Phi p$, and $d_V W$ denotes the differential of W taken with respect to the V-variables (with the $\mathbb{R}$ variable omitted.) The function $\partial W / \partial t$ is defined on $\tilde{V} \times [a, b]$, but we denote its pull back to $T\tilde{V} \times [a, b]$ by the same symbol.

Since $\pi \Phi p$ is a local diffeomorphism we may express functions on $T\tilde{V} \times [a, b]$ in terms of local coordinates $x = (x_1, \ldots, x_n)$ on M. In such local coordinates,

$$L^* = L - \Sigma \dot{x}_i (\partial W / \partial x_i) - \partial W / \partial t.$$

For $v \in \tilde{V}$ and $t \in [a, b]$, let $\Psi(v, t) \in T\tilde{V} \times [a, b]$ be defined by $T(\pi \Phi p)(\Psi(v, t)) = \Phi p(v, t)$. Then $\Psi: \tilde{V} \times [a, b] \to T\tilde{V} \times [a, b]$ is a section of this vector bundle. The equation $dW = p^* \Phi^* \eta$ amounts to

$$\frac{\partial W}{\partial x_i} = \frac{\partial L}{\partial \dot{x}_i}, \qquad \frac{\partial W}{\partial t} = -H$$

on the image of Ψ.

It follows that $L^*_{\dot{x}} = 0$ on the image of Ψ, and consequently, the restriction of L^* to the fiber over (v, t) takes its minimum at $\Psi(v, t)$. Moreover, the Euler-Lagrange flow associated to L^* is related by $T(\pi \Phi p)$ to the Euler-Lagrange flow associated to L, as may be verified by checking that the variational problems $\delta \int L^*(d\gamma(t), t)\, dt = 0$ and $\delta \int L(d\gamma(t), t)\, dt = 0$ for the fixed endpoint problem are the same. Therefore, the image of Ψ is a union of trajectories of the Euler-Lagrange flow of L^*. Since $L^*_{\dot{x}} = 0$ on the image of Ψ, it follows from the Euler-Lagrange equation $dL^*_{\dot{x}}/dt = L^*_x$ that $L^*_x = 0$ on the image of Ψ. Consequently, $L^* |\text{image}\, \Psi$ is a function of t alone.

From the fact that the restriction of L^* to the fiber over (v, t) takes its minimum at $\Psi(v, t)$, the fact that $L^* |\text{image}\, \Psi$ is a function of t alone, and the fact that L^* satisfies the positive definitness and the superlinear growth conditions, it follows that

$$A^*(\gamma_1) \geq A^*(\gamma) + F(d_{ac}(\gamma, \gamma_1)),$$

where by abuse of terminology we continue to denote suitable lifts of γ and γ_1 to $\tilde{V}$ by the same symbols. Here, we use the fact that $\pi \Phi: V \times [a, b] \to M$

$\times [a, b]$ is a diffeomorphism onto an open subset and the fact that p is a covering map. Moreover, we use the fact that γ and γ_1 are homologous (in $\pi\Phi(V \times [a, b])$) rel. endpoints to quarantee that they can be lifted to curves in $\tilde{V}$ having the same endpoints. We set

$$A^*(\gamma_1) = \int_a^b L^*(d\gamma_1(t), t)\, dt.$$

It is easily verified that $A^*(\gamma_1) - A(\gamma_1) = W(\gamma_1(a)) - W(\gamma_1(b))$. Consequently, $A(\gamma_1) - A(\gamma) = A^*(\gamma_1) - A^*(\gamma)$ and hence we obtain the conclusion of Lemma 1. □

For $c \in [a, b]$, we let i_c: $TM \to TM \times \mathbb{R}$ be defined by $i_c(\xi) = (\xi, c)$ and set $i_c^* \Omega = \Omega_c$. We let Φ_c: $V \to TM$ be defined by $\Phi_c(v) = \Phi_1(v, c)$, where Φ and Φ_1 are as in Lemma 1.

Lemma 2 *Let* Φ: $V \times [a, b] \to TM \times \mathbb{R}$ *be a* C^1 *mapping satisfying properties* 1) *and* 2) *of Lemma* 1. *Let* $c \in [a, b]$. *Then for* $\Phi^* \Omega = 0$ *to hold, it is sufficient that* $\Phi_c^* \Omega_c = 0$.

Proof. Let ξ be the vector field on $\Phi(V \times [a, b])$ whose trajectories are the curves $t \to \Phi(v, t)$. The Euler-Lagrange equation is equivalent to Hamilton's equation, which is equivalent to the assertion that ξ is in the kernel of Ω, i.e. $\Omega(\xi(\Phi(v, t)), \eta) = 0$ for all tangent vectors η to $T\tilde{M} \times \mathbb{R}$ at $\Phi(v, t)$. Since ξ is in the kernel of Ω, it is enough to show that $\Phi_t^* \Omega_t = 0$, for all t. But this is a consequence of the fact that it is true for $t = c$, together with the fact that Hamilton's flow is symplectic. □

Lemma 2 permits us to construct lots of examples of Φ which satisfy the hypotheses of Lemma 1. For once Φ_c: $V \to TM$ satisfying $\Phi_c^* \Omega = 0$ is given, there is a unique way to extend Φ_c to Φ: $V \times [a, b] \to TM \times \mathbb{R}$ satisfying 1) and 2) of Lemma 1, in view of the fact that for each $(\xi_0, t_0) \in TM \times \mathbb{R}$, there is a unique integral curve of the Euler-Lagrange vector field through (ξ_0, t_0). In view of condition 4) in Lemma 1, we wish to find Φ_c with the additional property that $\pi\Phi_c$ is a diffeomorphism of V onto an open subset of M. To put this in another way, we are looking for sections s of TM over open subsets of M with the property that $s^* \Omega = 0$. Using the Legendre transformation, we see that this is the same as finding sections of $T^* M$ over open subsets of M which pull back the canonical 2-form on $T^* M$ to zero. These sections are precisely the closed 1-forms, and the differential of any function is a closed 1-form.

The rest of the proof of our formulation in § 2 of Weierstrass's theorem is elementary. For example, we may proceed as follows.

Let $\tilde{\xi}_0 \in T\tilde{M}$, $c \in \mathbb{R}$ with $\|\tilde{\xi}_0\| \leq K$. Let $\xi_0, \tilde{x}_0, x_0$. Denote the projections of $\tilde{\xi}_0$ on TM, $\tilde{M}$, M, resp. Let $\mathfrak{L}_c$: $TM \to T^* M$ denote the Legendre transformation corresponding to the Lagrangian $L | TM \times c$. Recall that if $x = (x_1, \ldots, x_n)$ is a C^∞ chart defined on an open set U in M, $(x, \dot{x})$ is the canonically associated chart on TU and (x, p) is the canonically associated chart on $T^* U$, then the Legendre transformation is defined in these local coordinates by $p = L_{\dot{x}}$. We let Ξ be a C^∞ function on M such that $d\Xi(x_0) = \mathfrak{L}_C(\xi_0)$ and let $\Phi_c = \mathfrak{L}_c^{-1} \circ d\Xi$. Thus, Φ_c is a C^1 section of TM. We use the Euler-Lagrange flow to extend Φ_c to a mapping Φ: $M \times [a, b] \to TM$, where $a = c - \varepsilon$, $b = c + \varepsilon$, satisfying conditions 1)–3) in Lemma 1. Since $\pi\Phi_c$ = identity, it is clear that condition 4) in Lemma 1 will also be satisfied, provided that ε is small enough. There is a positive uniform lower bound on how small ε must be taken, depending only on K, provided that Ξ is chosen carefully. It is obviously possible to lift this construction to $\tilde{M}$. In this way, the inequality in our formulation of Weierstrass's theorem is seen to be a special case of the inequality in Lemma 1.

The last assertion in our formulation of Weierstrass's theorem is a consequence of the fact that the flow generated by the Euler-Lagrange vector field is C^1 and the form of the Euler-Lagrange equation, i.e. in local coordinates $d\dot{x}/dt = G(x, \dot{x}, t)$, $dx/dt = \dot{x}$. Let B_K denote the ball in TM_m of radius K, about the zero vector. Let $\Psi_{b-a}: B_K \to M$ be the mapping which assigns to $\zeta \in B_K$ the value $\gamma(b)$ where $\gamma: [a, b] \to M$ is the unique solution of the Euler-Lagrange equation with $d\gamma(a) = \xi$. This is a diffeomorphism of B_K onto a subset of M which contains the ball of radius $(b-a)K/2$ about m, provided ε is small enough, since $dx/dt = \dot{x}$. This proves our formulation of Weierstrass's theorem. □

In § 5, we asserted that the unique invariant measure supported by a *KAM* torus which is a graph is minimal. For this, we use the remark of Herman [13] that such a torus is a Lagrangian submanifold and proceed just as before. I am indebted to J. Moser for pointing out to me that it is possible to apply Weierstrass's theorem to prove that the trajectories which lie in a *KAM* torus are minimal.

Acknowledgements. I would like to thank M. Herman and J. Moser for listening to oral expositions of these results and for encouraging me to write them up. I would like to thank J. Moser for useful references and historical remarks. I would like to thank A. Katok for telling me about his results [14] when I visited Cal. Tech. in January 1988. At that time I knew a version of the results presented here for the geodesic flow on a Riemannian manifold, and his interest stimulated me to generalize these results to the situation described in this paper. I would like to thank Jochen Denzler for proofreading this article and Odet Schramm for pointing out a simplification of the proof which appears in Appendix 1. I would like to thank Rahel Boller for the care she took typing this article. I wrote this article during two visits to the Forschungsinstitut für Mathematik, ETH, Zürich and I would like to thank ETH for its hospitality. During the second visit, I was partially supported by a Fellowship of the John Simon Guggenheim Foundation, and I would like to thank it for its support.

References

1. Aubry, S., Le Daeron, P.Y.: The discrete Frenkel-Kontorova model and its extensions I. Physica D **8**, 381–422 (1983)
2. Akheizer: The calculus of variations. New York: Blaisdell 1962
3. Ball, J., Mizel, V.: One dimensional variational problems whose minimizers do not satisfy the Euler-Lagrange equation. Arch. Ration. Mech. Anal. **90**, 325–388 (1985)
4. Bangert, V.: Mather sets for twist maps and geodesics on tori. Dyn. Rep. **1**, 1–56 (1988)
5. Bangert, V.: Minimal geodesics. Preprint (1987)
6. Bernstein, D., Katok, A.: Birkhoff periodic orbits for small perturbations of completely integrable systems with convex Hamiltonians. Invent. Math. **88**, 225–241 (1987)
7. Caratheodory, C.: Variationsrechnung und partielle Differentialgleichung erster Ordnung. Leipzig-Berlin: B.G. Teubner 1935
8. Cartan, E.: Leçons sur les invariants integrals. Paris: Herman 1922
9. Dacorogna, B.: Direct methods in the calculus of variations. Appl. Math. Sci. vol. 78, Berlin Heidelberg New York: Springer 1989

10. Denzler, J.: Mather sets for plane Hamiltonian systems. J. Appl. Math. Phys. (ZAMP) **38**, 791–812 (1987)
11. Douady, R.: Stabilité ou instabilité des points fixes elliptiques. Ann. Sci. Éc. Norm. Supér., IV. Sér. **21**, 1–46 (1988)
12. Hedlund, G.: Geodesics on a two-dimensional Riemannian manifold with periodic coefficients. Ann. Math., II. Ser. **33**, 719–739 (1932)
13. Herman, M.R.: Existence et non existence de tores invariants par des diffeomorphismes symplectiques. Preprint (1988)
14. Katok, A.: Minimal orbits for small perturbations of completely integrable Hamiltonian systems. Preprint (1988)
15. Kryloff, N.M., Bogoliuboff, N.N.: La théorie générale de la mesure et son application à l'étude des systèmes dynamiques de la mécanique non linéaire. Ann. Math., II. Sér. **38**, 65–113 (1937)
16. Lanford, O.: Selected Topics in Functional Analysis. In: DeWitt, Stora (eds.): Statistical Mechanics and Quantum Field Theory. Proceedings of the Summer School of Theoretical Physics, Les Houches 1970, pp. 109–214. New York: Gordon and Breach 1971
17. Mather, J.: Existence of quasiperiodic orbits for twist homeomorphisms of the annulus. Topology **21**, 457–467 (1982)
18. Mather, J.: Minimal Measures. Comment. Math. Helv. **64**, 375–394 (1989)
19. Mather, J.: Minimal action measures for positive definite Lagrangian systems. In: Simon, B. et al. (eds.): IX International Congress on Mathematical Physics. Bristol-New York: Adam Hilger 466–468 (1989)
20. Moser, J.: Monotone Twist Mappings and the Calculus of Variations. Ergodic Theory Dyn. Syst. **6**, 401–413 (1986)
21. Moser, J.: On the construction of almost periodic solutions of ordinary differential equations. Proc. Int. Conf. Funct. Anal. and Rel. Top., Tokyo 60–67 (1969)
22. Nemytskii, V.V., Stepanov, V.V.: Qualitative Theory of Differential Equations. Princeton, N.J.: Princeton University Press 1960
23. Rockafellar, R.T.: Convex Analysis. Princeton Math. Ser., vol. 28. Princeton: Princeton University Press 1970
24. Salamon, D.: The Kolmogorov-Arnold-Moser theorem. Forschungsinstitut für Mathematik, ETH-Zentrum, Preprint (1986)
25. Salamon, D., Zehnder, E.: *KAM* theory in configuration space. Comment. Math. Helv. **64**, 84–132 (1989)
26. Schwartzman, S.: Asymptotic cycles, Ann. Math. II. Ser., **66**, 270–284 (1957)
27. Weinstein, A.: Lagrangian submanifolds and Hamiltonian systems. Ann. Math. II. Ser. **98**, 377–410 (1973)