CHAPTER 9 or AMG

*GENERALIZATIONS OF THE AUBRY-MATHER THEOREM

January 16, 2000

Complete

There are, strictly speaking, no *full* generalizations of the Aubry-Mather Theorem in higher dimensions: we will see in this chapter examples of fiber convex Lagrangian systems whose set of minimizers achieves only very few rotation directions. However some attempts of generalizations in higher dimensions are quite successful in what they try to achieve. In Section 47, we survey some results by de la Llave and his collaborators. Their setting is explicitly non dynamical but generalizes naturally the Frenkel-Kontorova model to functions on lattices of any dimension. They are entirely successful in proving an Aubry-Mather type theorem in this setting, as well as in some PDE cases. In Section 48, we review the work MacKay & Meiss (1992) who construct higher dimensional analogs of Aubry-Mather sets in symplectic twist maps that are close to the anti-integrable limit: one where the potential term in the generating function of a standard type map dominates. In Section 49, we survey the work of Mather on minimal measures in convex Lagrangian systems. This is the closest to a generalization of the Aubry-Mather theory as one can get in the setting of general convex Lagrangian systems (as well as symplectic twist maps). We start in Subsection A with an introduction to such minimizers and their relation to hyperbolic orbits. In Subsection B we give a quick review of some notions of ergodic theory that are needed in Subsection C, where we introduce minimal measures in Lagrangian systems. Subsection D explores, through examples, the intrinsic limitations of this theory. Section 50 shows that some of these limitations can be alleviated if one considers systems on cotangent bundles of hyperbolic manifolds.

47.* Aubry-Mather Theory for Functions on Lattices and PDE's.

A*. Functions on Lattices

Remember from Chapter 1 that the Frenkel–Kontorova model describes configurations of interacting particles in a periodic potential. For simplicity, these configurations are assumed to be one dimensional, and the interactions only involve nearest neighbors. The resulting energy function is the familiar:

$$W(\boldsymbol{x}) = \frac{1}{2}\sum_{k\in\mathbb{Z}}(x_k - x_{k+1})^2 - \sum_{k\in\mathbb{Z}} V(x_k)$$

where the potential function V has period 1. W coincides with the energy function for the standard map with generating function $S(x, X) = \frac{1}{2}(X - x)^2 - V(x)$. The variational equation $\nabla W = 0$ for this energy function is

$$(-\Delta \boldsymbol{x})_k - V'(x_k) = 0$$

where $\Delta(\boldsymbol{x})_k = 2x_k - x_{k-1} - x_{k+1}$ is the discretized Laplacian. Note that the configuration $\boldsymbol{x}$ can be seen as a function $\mathbb{Z} \to \mathbb{R}$ which to the integer k makes correspond the real x_k. One obtains (see Koch & al. (1994) , Candel & de la Llave (1997) , de la Llave (1999)) a natural generalization of this model, relevant to Statistical Mechanics, by asking that $\boldsymbol{x} : \mathbb{Z}^d \to \mathbb{R}$ be a function on a lattice of dimension d. We assume nearest neighbor interaction here. The energy becomes:

$$W(\boldsymbol{x}) = \frac{1}{2} \sum_{\{(k,j) \in \mathbb{Z}^d |\ |k-j|=1\}} (x_k - x_j)^2 - \sum_{k \in \mathbb{Z}^d} V(x_k).$$

Again V is of period 1 and the corresponding variational equation is still of the form:

$$(-\Delta \boldsymbol{x})_k - V'(x_k) = 0 \tag{47.1}$$

where $(\Delta \boldsymbol{x})_k = \sum_{|k-j|=1} x_j - 2dx_k$ is the d–dimensional discrete Laplacian. In fact, the theory in Candel & de la Llave (1997) applies to substantially more general settings, where k can belong to a set Λ on which a certain type of groups acts in a mildly prescribed way, and where the interactions involves not just nearest neighbors, but all possible pairs of particles (with some decay condition at infinity).

Remember that the solutions $\boldsymbol{x} : \mathbb{Z} \to \mathbb{R}$ found by Aubry and Mather for the Frenkel–Kontorova model are such that $|x_k - k\omega| \leq \infty$. One way to express this is by saying that the graph of $\boldsymbol{x} : \mathbb{Z} \to \mathbb{R}$ is at bounded distance from a line of slope ω in $\mathbb{R} \times \mathbb{R}$. Likewise, the following generalization of the Aubry-Mather Theorem finds configurations whose graphs are at bounded distances from planes of "slopes" $\omega \in \mathbb{R}^d$:

Theorem 47.1 (de la Llave et. al.) *For every $\omega \in \mathbb{R}^d$, there exists a solution of (47.1) such that*

$$\sup_{k \in \mathbb{Z}^d} |x_k - \omega \cdot k| < \infty.$$

The method of proof is very similar to the proof of the Aubry-Mather Theorem presented in GCchapter. One considers the analog of CO sequences, called Birkhoff configurations by these authors. In complete analogy to the CO sequences, they satisfy:

$$x_{k+j} + l \geq x_k,\ \forall k \in \mathbb{Z}^d \quad \text{or} \quad x_{k+j} + l \leq x_k,\ \forall k \in \mathbb{Z}^d$$

The analog to the set of CO sequences of rotation number ω, which we denoted by CO_ω in GCchapter is:

$$\mathcal{B}_\omega = \{\boldsymbol{x} \mid \ \boldsymbol{x} \text{ is Birkhoff and } \sup_{k \in \mathbb{Z}^d} |x_k - k \cdot \omega| < \infty\}$$

In a way analogous to the proof of Theorem 15.1, one shows that the gradient flow of W (that these authors, justifiably, call the heat flow) preserves order among configurations and is suitably periodic, so that the set $\mathcal{B}_\omega$ is invariant under the flow. The same argument as in the proof of Theorem 15.1 is then used to show that W must have a critical point inside $\mathcal{B}_\omega$. So, as in the classical Aubry-Mather Theorem, one not only finds solutions that have asymptotic slope ω, but these solutions have strong order properties, expressed here in terms of nonintersection: they are Birkhoff.

B*. PDE's

As Equation (47.1) suggests, the above theory smells of discretized PDE's. It is therefore not too surprising that the same kind of methods can be applied to certain PDE problems. The main ingredients necessary are some translation invariance and a heat flow that satisfies a comparison principle $u > v \Rightarrow \phi^t u > \phi^t v$, which occurs in parabolic PDE's. The method can be applied (see de la Llave (1999)) to the following PDE situations, to obtain solutions whose graphs are at bounded distance from planes with prescribed slopes, and have nonintersection properties:

$$\Delta u + V'(x,u) = 0 \tag{47.2}$$

where $V(x+e, u+\ell) = V(x,u)\ \forall x \in \mathbb{R}^d, u \in \mathbb{R}, e \in \mathbb{Z}^d, \ell \in \mathbb{Z}$.

$$\sum_{i=1}^{k} L_i^2 + V'(x,u) = 0 \tag{47.3}$$

where L_i are $\mathbb{Z}^d$ periodic vector fields satisfying Hörmander's hypoellipticity conditions and V is as in the previous case.

$$(-\Delta)^{1/2} u + V'(x,u) = 0 \tag{47.4}$$

with V as above. de la Llave (1999) also looks at the following PDE:

$$\begin{aligned} \square\, u = u_{tt} - u_{xx} &= -V(u) + f(x,t) \\ u(x+1,t) = u(x,t+T) &= u(x,t) \end{aligned} \tag{47.5}$$

where the function f also has the periodicity:

$$f(x+1,t) = f(x,t+T) = f(x,t). \tag{47.6}$$

We say that the real number T is of constant type if its continued fraction expansion is bounded. For instance, noble numbers are of constant type.

Theorem 47.2 (de la Llave) *Let T be a number of constant type, let $f \in L^2$ satisfy (47.6) and let $V : \mathbb{R} \to \mathbb{R}$ satisfy*
(i) $0 < \alpha \le V' \le \beta$ where α is any positive number and β only depends on T (in an explicit manner)
(ii) $|V''(x)| \le K$
Then there exists a weak solution $u \in L^2$ to Equation (47.5) . Moreover, if $f \in H^r$ and $V \in C^{r+2}$ has small enough C^{r+2} norm, then there is a solution $u \in H^r$ of (47.5) which is unique in a ball in H^r around the origin.

The method of proof is different from that of the above PDE's, but still involves a variational approach.

48.* Monotone Recurrence Relations

Angenent (1990) proposes a generalization of twist maps of the annulus to maps of $\mathbb{S}^1 \times \mathbb{R}^N$ which are defined by solving a recurrence relation:

$$\Delta(x_{k-l}, \ldots, x_{k+m}) = 0 \tag{49.1}$$

which generalizes $\partial_2 S(x_{k-1}, x_k) + \partial_1 S(x_k, x_{k+1}) = 0$ in twist maps, where $k = l = 1$. The function Δ is required to satisfy the conditions:

a) *monotonicity* $\Delta(x_{-l}, \ldots, x_{+m})$ is a non decreasing function of all the x_k except possibly for $k = 0$. Moreover, it is strictly increasing in the variables x_{-l} and x_m.

b) *periodicity* $\Delta(x_{k-l}, \ldots, x_{k+m}) = \Delta(x_{k-l} + 1, \ldots, x_{k+m} + 1)$

c) *coerciveness* $\lim_{x_l \to \pm\infty} \Delta(x_{-l}, \ldots, x_m) = \lim_{x_m \to \pm\infty} \Delta(x_{-l}, \ldots, x_m) = \pm\infty$

Under these conditions, Angenent calls (49.1) a *monotone recurrence relation*. Conditions a) and c) imply that one can solve for x_{k+m} in terms of a given $(x_{k-l}, \ldots, x_{k+m-1})$. Hence this defines a map $F_\Delta : (x_{k-l}, \ldots, x_{k+m-1}) \mapsto (x_{k-l+1}, \ldots, x_{k+m})$ from $\mathbb{R}^{l+m}$ to itself. Condition b) implies that this maps descends to a map on $\mathbb{S}^1 \times \mathbb{R}^{l+m-1}$. Hence the N above is $N = l + m - 1$.

The notion of CO configurations, rotation number and partial order on sequences etc... of Chapter AM and GCchapter are still entirely valid here, since the variables x_k are 1 dimensional (Angenent also calls CO sequences Birkhoff). An interesting notion that Angenent (1990) introduces, inspired by PDE methods, is that of sub– or supersolution of the monotone recurrence relation (49.1) : $\underline{\boldsymbol{x}}$ is a *subsolution* if $\Delta(\underline{x}_{k-l}, \ldots, \underline{x}_{k+m}) \leq 0,\ \forall k \in \mathbb{Z}$ and a *supersolution* if $\Delta(\underline{x}_{k-l}, \ldots, \underline{x}_{k+m}) \geq 0,\ \forall k \in \mathbb{Z}$.

Theorem 49.1 (Angenent) *Let $\underline{\boldsymbol{x}}, \overline{\boldsymbol{x}}$ be sub– and supersolutions respectively, which are ordered: $\underline{\boldsymbol{x}} \leq \overline{\boldsymbol{x}}$. Then there is at least one solution of (49.1) , say $\boldsymbol{x}$, for which $\underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}$ holds.*

Using this theorem (whose proof is simple), Angenent (1990) is able to generalize a theorem of Hall (1984), itself a generalization of the Aubry-Mather theorem: if a twist map of the annulus, which is *not necessarily* area preserving, has a (m, n)-periodic orbit, then it must have a CO (m, n)-periodic orbit. If the map is also area preserving, this implies, taking limits, the existence of CO orbits of all rotation numbers. Analogously, Angenent proves that if there is an orbit of F_Δ with rotation number $\omega \in \mathbb{R}$, then F_Δ must also have a CO orbit of rotation number ω.

Suppose that two solutions $\boldsymbol{x}$ and $\boldsymbol{w}$ of (49.1) "exchange rotation numbers" in the sense that:

$$\lim_{n\to+\infty} x_k/k \geq \omega_1 \geq \lim_{n\to-\infty} w_k/k$$

and

$$\lim_{n\to+\infty} w_k/k \leq \omega_0 \leq \lim_{n\to-\infty} x_k/k$$

holds for some $\omega_0 \leq \omega_1$. Then Angenent proves that there must be CO orbits of any rotation number $\omega \in [\omega_0, \omega_1]$. Moreover this exchange of rotation numbers condition implies chaos: the topological entropy $h_{top}(F_\Delta) > 0$, in that there is a compact invariant set semi conjugate to a Bernouilli shift. This also generalizes shadowing results of Hall (1989) and Mather (1991a). Angenent proves a few other interesting results for the map F_Δ.

49.* Anti–Integrable Limit

MacKay & Meiss (1992) explore the existence of Aubry-Mather sets (as well as many other possible configurations) close to the anti-integrable limit, where the potential of a standard like map becomes all powerful. Consider a family F_ϵ of symplectic twist maps of $T^*\mathbb{T}^n$ given by the generating functions:

$$S_\epsilon(\boldsymbol{q},\boldsymbol{Q}) = \epsilon T(\boldsymbol{q},\boldsymbol{Q}) + V(\boldsymbol{q})$$

where, for simplicity, we can assume

$$T(\boldsymbol{q},\boldsymbol{Q}) = \frac{1}{2}(\boldsymbol{Q}-\boldsymbol{q})^2,$$

although many more general T's can be considered. As usual, orbits of F_ϵ correspond to solutions of

$$\partial_2 S_\epsilon(\boldsymbol{q}_{k-1},\boldsymbol{q}_k) + \partial_1 S_\epsilon(\boldsymbol{q}_k,\boldsymbol{q}_{k+1}) = 0. \tag{49.2}$$

Even though F_0 is not defined, it is perfectly acceptable to set $\epsilon = 0$ in Formula (49.2) . This is called the *anti-integrable limit* , a notion that seems to have appeared independently in Aubry & Abramovici (1990) and Tangerman & Veerman (1991). The force of this concept is that the solutions of (49.2) at $\epsilon = 0$ are perfectly understood: they are simply allocations of $\boldsymbol{q}_k$ to one of the critical points of V: (49.2) is just $dV(\boldsymbol{q}_k) = 0$ when $\epsilon = 0$. If V is a Morse function, it has finitely many critical points modulo $\mathbb{Z}^n$ and they are all nondegenerate. This has the following consequence:

Theorem 49.2 (MacKay–Meiss) *Any solution $\boldsymbol{q}(0)$ of (49.2) for $\epsilon = 0$ continues to a solution $\boldsymbol{q}(\epsilon)$ when ϵ is small.*

Proof. Rewrite the infinite system of equations (49.2) in the form

$$G(\epsilon,\boldsymbol{q}) = 0$$

where $G : \mathbb{R}\times X \to (\mathbb{R}^n)^{\mathbb{Z}}$ is given by $G(\epsilon,\boldsymbol{q}) = \partial_2 S_\epsilon(\boldsymbol{q}_{k-1},\boldsymbol{q}_k) + \partial_1 S_\epsilon(\boldsymbol{q}_k,\boldsymbol{q}_{k+1})$ and X is the Banach space of sequences such that $\sup_k \|\boldsymbol{q}_k - \boldsymbol{q}_k(0)\| < \infty$. The Implicit Function Theorem on Banach spaces (see Lang (1983)) applies here to find, for small ϵ, a $\boldsymbol{q}(\epsilon)$ such that $G(\epsilon,\boldsymbol{q}(\epsilon)) = 0$ as long as $\frac{\partial G}{\partial \boldsymbol{q}}(0,\boldsymbol{q}(0))$ is invertible. But this is indeed the case since

$$\frac{\partial G}{\partial \boldsymbol{q}}(0,\boldsymbol{q}(0))_k = V''(\boldsymbol{q}_k(0))$$

so that $\frac{\partial G}{\partial \boldsymbol{q}}(0,\boldsymbol{q}(0))$ is an infinite block diagonal matrix with the $n\times n$ diagonal blocks $V''(\boldsymbol{q}_k(0))$ all invertible and uniformly bounded. Indeed these matrices are chosen among a finite set, since $\boldsymbol{q}_k(0)$ is necessarily a critical point of V, of which there are finitely many mod $\mathbb{Z}^n$, by the assumption that V is Morse. □

One can simultaneously continue compact sets of stationary solutions from the anti-integrable limit. Such sets can be quite complicated, since the set of all stationary configurations of the anti-integrable limit can be seen as a shift on as many symbols as there are critical points. In particular, one can find invariant Cantor sets for F_ϵ. One can also get orbits with all rotation vectors $\omega \in \mathbb{R}^n$. To do so, consider the anti–integrable stationary solution $\boldsymbol{q}(0)$ which is such that $\boldsymbol{q}_0(0)$ is at some arbitrarily chosen critical point of V and

$$\boldsymbol{q}_k(0) = k[\omega] + \boldsymbol{q}_0(0),$$

where $[\omega]$ is the integer part of the vector ω. Each $\boldsymbol{q}_k(0)$ is thus on the same critical point as $\boldsymbol{q}_0(0)$, but translated by the integer vector $[\omega]$. Since $|\boldsymbol{q}_k(0) - k\omega| < \sqrt{n}$, $\boldsymbol{q}(0)$ has rotation vector ω. Now use Theorem 49.2 to continue this to an orbit of F_ϵ with rotation vector ω. One can also continue simultaneously all anti–integrable solutions as the above with rotation vectors in a compact set: they themselves form a compact set.
Even though this seems almost too easy, the anti-integrable limit is a very useful concept in order to understand the spectrum of all possible dynamics of symplectic twist maps . It is fair to say that, to this date, the least understood cases are those that are neither close to integrable nor to anti–integrable.

50.* Mather's Theory of Minimal Measures

We now come to Mather's theory of existence and regularity of minimizers. This theory is quite general: it covers a wide class of convex Lagrangian systems on tangent bundles of arbitrary compact manifolds. Note that similar, but less developed theories were created by Bangert (1989) in the setting of minimal geodesics on compact manifolds and Katok (1992) in the setting of perturbations of integrable symplectic twist maps . There is no doubt that Mather's theory could be worked out for general symplectic twist maps . Even now, the correspondences between Lagrangian systems and symplectic twist maps given in Chapter 6 (see in particular the Bialy-Polterovitch suspension theorem 40.1) should allow an ample transfer of Mather's results to the symplectic twist maps case.

The lesson we get from Mather's work is that, yes, minimizers in general manifolds behave very much like those on the circle (the realm of the classical Aubry-Mather theory), in that they satisfy a graph property. The bad news is that minimizers may be much scarcer than in the circle case: Hedlund (1932) had already constructed a Riemannian metric on $\mathbb{T}^3$ (a setting encompassed by Mather's) which is very small along 3 non intersecting geodesics which generate $H_1(\mathbb{T}^3)$. All other minimizers of a certain length are then bound to spend a good portion of their time close to these geodesics. In particular, these three geodesics are the only possible recurrent minimizers. This limits the possible rotation vectors of minimizers to these three directions only. Bangert (1989) (geodesic setting) and Mather (1991b) (Lagrangian setting) show that, in a precise sense, this is the worst case scenario: there should be at least as many rotation directions represented by minimizers as there are dimensions in $H_1(M, \mathbb{R})$. And, to end on an optimistic note, Levi (1997) construct, in this worst case scenario of Hedlund's example, "shadowing" locally minimizing orbits that spend any prescribed proportion of time close to each of the minimizers. In particular, he constructs *locally minimizing* orbits of all rotation vectors.

A*. Lagrangian Minimizers

Throughout this section and next, we consider time-periodic Lagrangian systems determined by a C^2-Lagrangian function $L : TM \times \mathbb{S}^1 \to \mathbb{R}$, where M is a compact manifold given a Riemannian metric g. Remember (see Appendix 1 or SG and Chapter 6) that extremals of the action

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

satisfies the Euler-Lagrange equations $\frac{d}{dt}\frac{\partial L}{\partial \dot x} - \frac{\partial L}{\partial x} = 0$. Using local coordinates these equations yield a first order time-periodic differential equation on TM, and thus in the standard way, a vector field on $TM \times \mathbb{S}^1$. This can be viewed as the Hamiltonian vector field corresponding to the Lagrangian system, pulled back to TM by the Legendre transformation. Since $TM \times \mathbb{S}^1$ is not compact it is possible that trajectories of this vector field are not defined for all time in $\mathbb{R}$ and thus do not fit together to give a global flow ($i.e.$ an $\mathbb{R}$-action). When the flow does exist, it is called the *Euler-Lagrange (or E-L) flow* . The following quite general hypotheses are the setting of Mather (1991b) .

Mather's Hypotheses

L is a C^2 function $L : TM \times \mathbb{S}^1 \to \mathbb{R}$ that satisfies:

(a) *Convexity:* $\frac{\partial^2 L}{\partial v^2}$ is positive definite.

(b) *Completeness:* The Euler-Lagrange flow determined by L exists.

(c) *Superlinear:* $\frac{L(x,v,t)}{\|v\|} \to \infty$ *when* $\|v\| \to +\infty$.

Mather's Hypotheses are satisfied by mechanical Lagrangians, $i.e.$ those of the form

$$L(x, v, t) = \frac{1}{2} \|v\|^2 - V(x, t),$$

where the norm is taken with respect to any Riemannian metric on the manifold. (In fact, one may allow the norm to vary with time, under some conditions, see Mañẽ (1991) , page 44).

Minimizers. We know that, for twist maps, orbits on Aubry-Mather sets are minimizers in the sense of Aubry. We have also seen in INVchapter that orbits on KAM tori are minimizers for symplectic twist maps . These are natural reasons to look for minimizers in convex Lagrangian systems. Lagrangian minimizers are defined in a way analogous to the discrete case. If $\tilde{M}$ is a covering space of M (see Appendix 2 or TOPO), L lifts to a real valued function (also called L) defined on $T\tilde{M} \times \mathbb{S}^1$. A curve segment $\gamma : [a, b] \to \tilde{M}$ is called a *$\tilde{M}$-minimizing segment* or an *$\tilde{M}$-minimizer* if it minimizes the action among all absolutely continuous curves $\beta : [a, b] \to \tilde{M}$ which have the same endpoints as γ. A curve $\gamma : \mathbb{R} \to \tilde{M}$ is also called a *minimizer* if $\gamma_{|_{[a,b]}}$ is a minimizer for all $[a, b] \subset \mathbb{R}$. When the domain of definition of a curve is not explicitly given it is assumed to be $\mathbb{R}$. In practice, the two main covering spaces that we will consider are the universal cover (in next section) and the universal abelian cover (in this section, see Appendix 2 or TOPO for the definitions of these covering spaces).

A fundamental theorem of *Tonelli* (see Mather (1991b) or Mañẽ (1991)) implies that if L satisfies Mather's Hypotheses, then given $a < b$ and two distinct points $x_a, x_b \in \tilde{M}$ there is always a minimizer γ with $\gamma(a) = x_a$ and $\gamma(b) = x_b$. Moreover such a γ is automatically C^2 and satisfies the Euler-Lagrange equations (this uses the completeness of the E-L flow). Hence its differential $d\gamma(t) = (\gamma(t), \dot{\gamma}(t))$ yields a solution $(d\gamma(t), t)$ of the E-L flow.

Minimizers vs. hyperbolicity. There is a general principle, first unveiled by Morse in Riemannian geometry, which ties the index of the second derivative of the action of a segment of geodesic to the number of conjugate points this segment has. In terms of more general Lagrangian systems, this number can be formulated as a certain rotation index (the Maslov index) of Lagrangian subspaces under the differential of the flow along an orbit segment (see Duistermaat (1976)). If the orbit is hyperbolic, the Lagrangian tangent subspace can be chosen to be the unstable manifold. A strong illustration of this occurs in the realm of symplectic twist maps where Aubry et. al. (1991) found a striking correspondence between (local) minimizers and hyperbolicity, which we now sketch. We first need some definitions.

Given a stationary configuration $\boldsymbol{q}$ for a symplectic twist map on $T^*\mathbb{T}^n$ ($i.e.$ the $\boldsymbol{q}$ coordinate of an orbit of the map), we can define the Hessian or second derivative of the action as in the proof of 14.2 :

$$\nabla^2 W(\boldsymbol{q}) = \begin{pmatrix} \ddots & \ddots & & & & & \\ \beta_0 & \alpha_1 & \beta_1 & & & & \\ & \beta_1 & \alpha_2 & \beta_2 & & & \\ & & \beta_2 & \alpha_3 & \ddots & & \\ & & & \ddots & \ddots & \beta_{q-1} & \\ & & & & \beta_{q-1} & \alpha_q & \beta_q \\ & & & & & \ddots & \ddots \end{pmatrix}$$

an infinite matrix, tridiagonal by blocks, where

$$\alpha_k = \partial_{22}S(\boldsymbol{q}_{k-1}, \boldsymbol{q}_k) + \partial_{11}S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}), \quad \beta_k = \partial_{12}S(\boldsymbol{q}_{k-1}, \boldsymbol{q}_k).$$

If $\boldsymbol{q}$ is a local minimizer, $i.e.$ minimizes W locally on any finite segment, then the spectrum of $\nabla^2 W(\boldsymbol{x})$ is positive. We say that $\boldsymbol{q}$ has a *phonon gap* if moreover Spec$(\nabla^2 W(\boldsymbol{q})) \in [a, \infty),\ a > 0$. An invariant set has a phonon gap a if each of the orbits it contains does, and if their phonon gaps are all greater or than a.

Theorem 50.1 (Aubry–Baesens–MacKay) *Let Λ be a closed invariant set for a symplectic twist map of $\mathbb{R}^{2n}$ and Λ' be the associated set of critical sequences for W. Suppose that $\partial_{12}S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$, $(\partial_{12}S)^{-1}(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$, $\partial_{11}S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ and $\partial_{22}S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ are all bounded for $\boldsymbol{q} \in \Lambda', k \in \mathbb{Z}$. Then Λ is uniformly hyperbolic if and only if Λ has a phonon gap.*

B*. Ergodic Theory

Most of the material surveyed in this subsection can be found in Hasselblat & Katok (1995) . We start by motivating this theory by the following trivial remark: if F is a map of $T^*\mathbb{T}^n$ and $\phi(\boldsymbol{z}) = \pi(F(\boldsymbol{z})) - \pi(\boldsymbol{z})$ ($\pi : T^*\mathbb{T}^n \to \mathbb{T}^n$ is the canonical projection) then, *when it exists*:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \phi(F^k(\boldsymbol{z})) = \lim_{n\to\infty} \frac{\pi(F^n(\boldsymbol{z})) - \pi(F(\boldsymbol{z}))}{n} = \rho_F(\boldsymbol{z}),$$

the rotation vector of $\boldsymbol{z}$ under F. The expression $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \phi(F^k(\boldsymbol{z}))$ for a general continuous function ϕ is called the *time average* of ϕ. Hence the rotation vector of a point, when it exists, is the time average of a specific ϕ. The relevance of this is the following:

Theorem 50.2 (**Birkhoff's Ergodic Theorem**) *Let $F : (X, \mu) \to (X, \mu)$ be a measure preserving transformation for a Borel measure μ on a space X, and $\phi \in L^1(X, \mu)$. Then the time average $\phi_T(\boldsymbol{z})$ of ϕ:*

$$\phi_T(\boldsymbol{z}) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \phi(F^k(\boldsymbol{z}))$$

exists for $\mu - a.e.$ $\boldsymbol{z}$. *Moreover, if* $\mu(X) < \infty$, $\int_X \phi_T d\mu = \int_X \phi d\mu$.

Remember that a Borel measure on a topological space is one whose sigma-algebra of measurable sets is generated by the open sets. That F is measure preserving means $\mu(F^{-1}(A)) = \mu(A)$ for any Borel subset of X. An immediate corollary of Birkhoff's theorem is (one needs to compactify $T^*\mathbb{T}^n$ with a point at ∞):

Corollary 50.3 *Let F be a volume preserving map (eg. symplectic) of $\mathbb{T}^n$. The rotation vector ρ_F is defined on a subset of full Lebesgue measure of $T^*\mathbb{T}^n$.*

It turns out that the Lebesgue measure is only one of the many measures that a symplectic map F preserves. Take $\boldsymbol{z} \in T^*\mathbb{T}^n$ to be a N-periodic point of F, for instance, and let :

$$\eta = \frac{1}{N}\sum_{k=1}^{N} \delta_{F^k(\boldsymbol{z})}$$

where the Dirac measure $\delta_{\boldsymbol{w}}$ is the (Borel) probability measure concentrated at the point $\boldsymbol{w}$ ($\delta_{\boldsymbol{w}}(A)$ is 1 if $\boldsymbol{w} \in A$ and it is 0 if not). Since $\delta_{F^k(\boldsymbol{z})}(F^{-1}(A)) = \delta_{F^{k+1}(\boldsymbol{z})}(A)$, η is invariant under F. One of the many differences between η and the Lebesgue measure is their supports. In general, the *support* of a Borel measure μ is defined as:

$$\text{Supp }\mu = \{\boldsymbol{z} \in X \mid \mu(U) > 0 \quad \text{whenever} \quad \boldsymbol{z} \in U,\ U \quad \text{open} \quad \}$$

Clearly, the support of the measure constructed above is the orbit of the periodic point $\boldsymbol{z}$, whereas the support of the Lebesgue measure is $T^*\mathbb{T}^n$. Hence, the support of invariant measures is another way to conceptualize invariant sets. Let $F : X \to X$ be continuous. Then the support of any F-invariant Borel measure μ is closed, F-invariant and its complement has zero μ–measure. If $\mu(X) < \infty$, Poincaré's Recurrence Theorem implies that $\text{Supp }\mu$ is contained in the set of $F-$recurrent points. In fact, $\boldsymbol{z} \in \text{Supp }\mu \Rightarrow \boldsymbol{z} \in \omega(\boldsymbol{z}) \in \text{Supp }\mu$. Hence, to find recurrent orbits in a dynamical system, as we have been doing in this book, one can look for invariant measures.

Coming back to rotation vectors, and the measure η supported on a periodic orbit, the rotation vector $\rho_F(\boldsymbol{z})$ not only exist $\eta - a.e.$, but it is constant on $\text{Supp }\eta$. In fact, it can easily be checked that the time average ϕ_T is constant on $\text{Supp }\eta$ for any function $\phi \in L^1(T^*\mathbb{T}^n, \eta)$: the measure η is ergodic.

Definition 50.4 An F-invariant probability measure μ on a space X is *ergodic* if it satisfies one of the following equivalent properties:
1) Every F-invariant set has has μ measure 0 or 1.
2) If $\phi \in L^1(X, \mu)$ is F-invariant then ϕ is constant a.e..
3) The time average ϕ_T equals the *space average.* $\int \phi d\mu$ μ–a.e.

In terms of support, if μ is ergodic then F has an orbit in $\text{Supp }\mu$ which is dense in that support. Hence ergodicity relates to topological transitivity. The Lebesgue measure may never be ergodic for twist maps: whenever we have a chain of elliptic islands, it comprises an invariant set which is not of full Lebesgue measure. On the other hand, twist maps do have plenty of ergodic measures. We have seen above the example of a measure η supported on periodic orbits. More generally, *Aubry-Mather sets can be defined as supports* *of ergodic measures*, pull-back of measures on $\mathbb{S}^1$ invariant under circle diffeomorphisms. Indeed, take the set $\pi(M)$ in Theorem AMthmpropertiesam: it is the omega limit set $\Omega(T)$ for a circle diffeomorphism T. Now, pick $x \in \Omega(T)$ and take the weak* limit of the probability measures $\mu_N = \frac{1}{2N-1}\sum_{-N}^{N} \delta_{T^k(x)}$: it defines an ergodic measure for T, and its pull back by π is ergodic for F with support the Aubry-Mather set M. [the *weak* limit* is defined by $\mu_n \xrightarrow{*} \mu$ iff $\int_X \phi\mu_n \to \int_X \phi\mu$ for all continuous ϕ].

Hence our main objects of study in this book, periodic orbits and Aubry-Mather sets, are all supports of ergodic probability measures, part of the larger set $\mathcal{M}_F$ of all F-invariant Borel probability measures.

Remark 50.5 The existence of an ergodic measure with rotation vector (as defined by the space average) ω *does* guarantee the existence of at least one orbit with that rotation vector (the support of the measure is not empty, and the time average is constant on it). This is *not* the case if the measure is not ergodic.

If X is a compact metric space, it turns out that the set $\mathcal{M}$ of all Borel probability measures is convex and compact under the weak* topology. Moreover $\mathcal{M}_F$ itself is a compact and convex subset of $\mathcal{M}$ for this topology. A theorem of convex analysis (Krei-Millman) says that $\mathcal{M}_F$ is then in the convex hull of its *extreme points* : those $\mu \in \mathcal{M}_F$ which cannot be written as $t\mu_1 + (1-t)\mu_2$ for two distinct $\mu_1, \mu_2 \in \mathcal{M}_F$. Finally, the extreme points are all *ergodic* measures. We will see in the next subsection that there is a strong correspondence between the (strict) convexity of a certain projection of $\mathcal{M}_F$ and the Aubry-Mather theorem.

As we will see in next section, Mather (1991b) , (1993) considers measures that are invariant under the Euler-Lagrange (E-L) flow instead of a symplectic twist map . In the light of the suspension theorem of Bialy-Polterovitch (Chapter ham), his setting encompasses a large class of symplectic twist maps . All the statements that we made above are valid for E-L flows on $T^*\mathbb{T}^n$ provided one compactifies $T^*\mathbb{T}^n$ (as Mather does) in order to use the compactness of the space of E-L-invariant probability measures.

C*. Minimal Measures

For a more detailed exposition the reader is urged to consult Mather (1991b) or Manẽ (1991) . There is also a very nice survey of this theory in the beginning of Mather (1993) . Given a E-L invariant probability measure with compact support μ on $TM \times \mathbb{S}^1$, one can define its *rotation vector* $\rho(\mu)$ as follows: let $\beta_1, \beta_2, \dots, \beta_n$ be a basis of $H^1(M)$ and let $\lambda_1, \dots, \lambda_n$ be closed one-forms with $[\lambda_i] = \beta_i$ in DeRham cohomology.(11) We refer the reader uncomfortable with (co)homology to Appendix 2 or TOPO and urge her/him to read through this section thinking of the case $M = \mathbb{T}^n$, taking $[\lambda_i] = [dx_i]$, as a basis for $H^1(\mathbb{T}^n) \simeq \mathbb{R}^n$, where $(x_1, \dots, x_n)$ are angular coordinates on $T^*\mathbb{T}^n$. Define the i^{th} component of the rotation vector $\rho(\mu)$ as

$$\rho_i(\mu) = \int \lambda_i d\mu.$$

Note that this integral makes sense when one looks at λ_i as inducing a function from $TM \times \mathbb{S}^1$ to $\mathbb{R}$ by first projecting $TM \times \mathbb{S}^1$ onto TM, and then treating the form as a function on TM that is linear on fibers. The rotation vector does depend on the choice of basis β_i, but because these 1–forms are closed, $\rho_i(\mu)$ does not depend on the choice of representative λ_i with $[\lambda_i] = \beta_i$. Since the rotation vector is dual to forms, it can be

[11] When homology and cohomology coefficients are unspecified they are assumed to be $\mathbb{R}$, so the notation $H_1(M)$ means $H_1(M;\mathbb{R})$, etc.

viewed as an element of $H_1(M)$. In the case $M = \mathbb{T}^n$, one can check that, if $\gamma(0)$ is a generic point of an *ergodic* measure μ, the natural definition of *rotation vector of a curve* γ coincides with $\rho(\mu)$:

$$\rho_i(\gamma) = \lim_{b-a\to\infty} \frac{\tilde\gamma_i(b) - \tilde\gamma_i(a)}{b-a} = \lim_{b-a\to\infty} \frac{1}{b-a} \int_{d\gamma|_{[a,b]}} dx_i = \int dx_i d\mu = \rho_i(\mu)$$

where $\tilde\gamma$ is a lift of γ to $\mathbb{R}^n$ and the second equality uses the Ergodic Theorem (again, dx_i is seen as a function $TM \times \mathbb{S}^1 \to \mathbb{R}$). This prompts the following formula for the (i^{th} coordinate of the) rotation vector of a curve $\gamma : \mathbb{R} \to M$ for a general manifold M:

$$\rho_i(\gamma) = \lim_{b-a\to\infty} \frac{1}{b-a} \int_{d\gamma|_{[a,b]}} \lambda_i,$$

if the limit exists. As before, if $\gamma(0)$ is a generic point for an ergodic measure μ, $\rho(\gamma)$ exists and coincides with $\rho(\mu)$. Next we define the *average action* of a E-L invariant probability on $TM \times \mathbb{S}^1$:

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

i.e. the space average of L, which equals, when μ is ergodic, to its time average along μ-a.e. orbit γ:

$$A(\mu) = \lim_{b-a\to\infty} \frac{1}{b-a} \int_a^b L(\gamma, \dot\gamma) dt.$$

The set of E–L invariant probability measures, denoted by $\mathcal{M}_L$, is a convex set in the vector space of all measures, as we have seen in the previous subsection (It is also compact for the weak* topology if, as Mather does, one compactifies TM). and the extreme points of $\mathcal{M}_L$ are the ergodic measures (see Mañe (1987)). Now consider the map $\mathcal{M}_L \to H_1(M) \times \mathbb{R}$ given by:

$$\mu \mapsto (\rho(\mu), A(\mu)).$$

This map is trivially linear and hence maps $\mathcal{M}_L$ to a convex set U_L whose extreme points are images of extreme points of $\mathcal{M}_L$, *i.e.* images of ergodic measures. Mather shows, by taking limits of measures supported on long minimizers representing rational homology classes, that for each ω, there exists an invariant (but not necessarily ergodic) measure μ such that $\rho(\mu) = \omega$ and $A(\mu) < \infty$.(12) Since L is bounded below, the action coordinate is bounded below on U_L. Hence we can define a map $\beta : H_1(M) \to \mathbb{R}$ by

$$\beta(\omega) = \inf\{A(\mu) \mid \mu \in \mathcal{M}_L, \rho(\mu) = \omega\},$$

which is bounded below and convex: the graph of β is the boundary of U_L. We say that a probability measure $\mu \in \mathcal{M}_L$ is a *minimal measure* if the point $(\rho(\mu), A(\mu))$ is on the graph of β. Hence, an extreme point $(\omega, \beta(\omega))$ of $graph(\beta)$ corresponds to at least one minimal ergodic measure of rotation vector ω. It turns out that if μ is minimal, μ-a.e. orbit lifts to a E-L minimizer in the universal abelian cover $\overline{M}$ of M (whose deck transformation group is $H_1(M;\mathbb{Z})/torsion$, see Appendix 2 or TOPO). Conversely, if μ is an ergodic probability measure whose support consists of $\overline{M}$-minimizers, then μ is a minimal measure.

Hence, each time we prove the existence of an extreme point $(\omega, \beta(\omega))$, we find at least one recurrent orbit of rotation vector ω which is a $\overline{M}$-minimizer.

[12] The impatient reader may be tempted to proclaim, from this fact, the existence of orbits of all rotation vectors. Alas, as we noted in Remark 50.5, we can guarantee that the rotation vector of orbits in the support of a measure μ are equal to $\rho(\mu)$ only when μ is ergodic

Another important property of β is that it is *superlinear*, i.e $\frac{\beta(x)}{\|x\|} \to \infty$ when $\|x\| \to \infty$. We motivate this in the simple case where $L = \frac{1}{2}\|\dot{x}\|^2 - V(x)$ and $\|\cdot\|$ comes from the Euclidean metric on the torus. If μ is any invariant probability measure, then

$$(50.1)\qquad \begin{aligned} A(\mu) = \int L\, d\mu &\geq \int \left(\frac{\|\dot{x}\|^2}{2} - V_{max}\right) d\mu \\ &\geq \frac{1}{2}\left|\int \dot{x}\, d\mu\right|^2 - V_{max} \\ &= \frac{1}{2}|\rho(\mu)|^2 - V_{max} \end{aligned}$$

where we used the Cauchy-Schwarz inequality for the second inequality. So we see that in this particular, but important, case β grows at least quadratically with the rotation vector. The superlinearity of β implies the existence of many extreme points for $graph(\beta)$ (although in most cases still too few, as we will see in the next subsection). Indeed, this growth condition implies that β's graph cannot have flat, or linear domains going to infinity. Any point $(\omega, \beta(\omega))$ is part of at least one linear domain of $graph(\beta)$, which we call S_c, where the index c denotes the "slope" (normal vector) of the supporting hyperplane whose intersection with U_L is exactly the convex and flat domain S_c. [Since c acts linearly on homology classes ω to give the equation $c \cdot \omega = a$ of S_c, it can be seen as an element of first cohomology.] Let X_c be the projection on $H_1(M)$ of S_c. The sets X_c are compact and convex domains which "tile" the space $H_1(M)$. Extreme points of X_c are projections of extreme points of S_c. Hence there are infinitely many such extreme points, and infinitely many outside any compact set. Their convex hull is $H_1(M)$, and in particular, they must span $H_1(M)$ as a vector space. Since these extreme points are the rotation vectors of minimal ergodic measures, we have found that

Theorem 50.6 *There exist at least countably many minimal ergodic measures and at least* $n = \dim H_1(M)$ *of them with distinct rotation directions.*

In particular there are at least n rotation directions represented by minimal measures for a E-L flow on $T^*\mathbb{T}^n$. We will see in Hedlund's example that this lower bound *is* attained by some systems. Finally, the *generalized Mather sets* are defined as

$$M_c = Support(\mathcal{M}_c),$$

where $\mathcal{M}_c$ is the set of all minimal measures whose rotation vectors lies in X_c. Let $\pi : TM \times \mathbb{S}^1 \to M \times \mathbb{S}^1$ denote the projection. Mather's main result in Mather (1991b) is the following theorem.

Theorem 50.7 **(Mather's Lipschitz Graph Theorem)** *For all* $c \in H^1(M)$, M_c *is a compact, non-empty subset of* $TM \times \mathbb{S}^1$. *The restriction of* π *to* M_c *is injective. The inverse mapping* $\pi^{-1} : \pi(M_c) \to M_c$ *is Lipschitz.*

In the case $M = \mathbb{T}^n$, Mather proves that, when they exist, KAM tori coincide with the sets M_c (see also Katok (1992) for some related results in the symplectic twist maps context). The proof of the Lipschitz Graph Theorem (see Mather (1991b) or Mañẽ (1991)), which is quite involved, uses a curve shortening argument: if curves in $\pi(M_c)$ were too close to crossing transversally, one could "cut corners" and, because of recurrence, construct a closed curve with lesser action than A_{min}. This argument by surgery is reminiscent of the proof of Aubry's Fundamental Lemma in Chapter AM.

Remark 50.8 An important special case is that of *autonomous* systems (*i.e.* with time independent L). In this case, one can discard the time component and view M_c as a compact subset of TM. In this case, Mather's theorem implies that M_c is a Lipschitz graph for the projection $\pi : TM \to M$. To see this, suppose that two curves $x(t)$ and $y(t)$ in $\pi(M_c)$ have $x(0) = y(s)$ for some s. Mather's theorem rules out immediately the possibility that s is an integer, unless $x = y$ is a periodic orbit. For a general s, consider the curve $z(t) = y(t+s)$. Then, $\dot z(t) = \dot y(t+s)$ and, by time-invariance of the Lagrangian, $(z(t), \dot z(t))$ is a solution of the E-L flow. It has same average action and rotation vector as $(y, \dot y)$ and hence it is also in M_c. But then $z(0) = x(0)$ is impossible, by Mather's theorem, unless $\dot z(0) = \dot y(s) = \dot x(0)$ and thus, by uniqueness of solutions of ODEs, $x(t) = y(t+s)$.
By using Theorem 40.1 , one can translate the results of Mather to the realm of symplectic twist maps (see Exercise 50.9) and deduce the existence of many invariant sets that are graphs over the base and are made of minimizers. As noted before, one could also redo all of Mather's theory in the setting of symplectic twist maps (see Katok (1992), who considered the near integrable case).

D*. Examples and Counterexamples

Recovering past results. When Mather's function β (see previous section) is strictly convex, each point on $graph(\beta)$ is an extreme point and there are ergodic minimal measures (and hence minimal orbits) of all rotation vectors. One can prove that this is true when $M = \mathbb{S}^1$, and Mather (1991b) shows how his Lipschitz Graph Theorem implies the classical Aubry-Mather Theorem, by taking a E-L flow that suspends the twist map. The fact that M_c is a graph nicely translates into the fact that orbits in an Aubry-Mather set are cyclically ordered: as pointed out by Hall (1984), the CO property corresponds to trivial braiding of the suspended orbit, itself guaranteed by the graph property.

The graph of β is also strictly convex when L is a Riemannian metric on $\mathbb{T}^2$, and hence there are minimal geodesics of all rotation vectors for any metric on the torus. This was known by Hedlund (1932), who had basically worked out the same results as Aubry and Mather in that setting, albeit in a different language. [See Bangert (1988) for a unified approach of the two theories.] Hence one could hope, as a generalization of the Aubry-Mather theorem, that β is strictly convex for any Lagrangian systems satisfying Mather's hypotheses. This statement is false as we will see in the following examples.

Examples of gaps in the rotation vector spectrum of minimizers for Lagrangian on $\mathbb{T}^2$. Take $L : T\mathbb{T}^2 \to \mathbb{R}$, given by $L(x, \dot x) = \|\dot x - X\|^2$ where X is a vector field on $\mathbb{T}^2$. The integral curves x of X are automatically E-L minimizers since $L \equiv 0$ on these curves. Manẽ (1991) chooses the vector field X to be a (constant) vector field of irrational slope multiplied by a carefully chosen function on the torus which is zero at exactly one point q. The integral flow of X has the rest point $q(t) = q$, and all the other solutions are dense on the torus. The flow of X (and its lift to $T\mathbb{T}^2$ by the differential) has exactly two ergodic measures: one is the Dirac measure supported on $(\boldsymbol{q}, 0)$, with zero rotation vector, the other is equivalent to the Lebesgue measure on $\mathbb{T}^2$ and has nonzero rotation vector, say ω. Mañe checks that $\beta^{-1}(0)$ (trivially always an X_c) is the interval $\{\lambda\, \omega \mid \lambda \in [0,1]\}$, and that no ergodic measure has a rotation vectors strictly inside this interval. Thus the Mather set M_0 is the union of the supports of the two above measures .

Boyland & Golé (1996a) give an example of an autonomous *mechanical* Lagrangian on $\mathbb{T}^2$ which displays a similar phenomenon, although we also show in that paper that all autonomous Lagrangian systems satisfying Mather's Hypothesis do have minimizers of all rotation *directions*. We also give in this paper a very precise description of the β function for such systems and show that the support of minimal ergodic measures have to be either a point, a suspension of a Cantor set or a torus.

Hedlund-Bangert's counterexamples. Consider in $\mathbb{R}^3$ the three nonintersecting lines given by the x-axis, the y-axis translated by $(0,0,1/2)$ and the z-axis translated by $(1/2,1/2,0)$. Construct a $\mathbb{Z}^3$- lattice of nonintersecting axes by translating each one of these by integer vectors. Take a metric in $\mathbb{R}^3$ which is the Euclidean metric everywhere except in small, nonintersecting tubes around each of the axes in the lattice. In these tubes, multiply the Euclidean metric by a positive function λ which is 1 on the boundary and attains its (arbitrarily small) minimum along the points in the center of the tubes, *i.e.* at the axes of the lattice. Because the construction is $\mathbb{Z}^3$ periodic, this metric induces a Riemannian metric on $\mathbb{T}^3$. One can show (Bangert (1989)), if λ is taken sufficiently small, that a minimal geodesic (which is a E-L minimizer in our context) can make at most three jumps between tubes. In particular, a recurrent E-L minimizer has to be one of the three disjoint periodic orbits which are the projection of the axes of the lattice. Thus there are only three rotation directions that minimizers can take in this example, or six if one counts positive and negative orientations. In terms of Mather's theory, the level sets of the function β are octahedrons with vertices $(\pm a,0,0), (0,\pm a,0), (0,0,\pm a)$ (we assume here that the function λ is the same around each of the tubes). Since we are in the case of a metric, one can check that β is quadratic when restricted to a line through the origin (a minimizer of rotation vector $a\omega$ is a reparameterization of a minimizer of rotation ω). Hence a set S_c is either a face, an edge or a vertex of some level set $\{\beta = b\}$, and the corresponding M_c is, respectively, the union of three, two (parameterized at same speed) or one of the minimal periodic orbits one gets by projecting the disjoint axes. Note that, instead of the function β of Mather, Bangert uses the *stable norm*. Mather's function β is a generalization of that norm.

Levi's counter–counterexample. It is important to note that the nonexistence of minimizers of a certain rotation vector ω does not mean that there are no orbits of the E-L flow that have rotation vector ω. For example, Levi (1997) has shown the existence of orbits of all rotation vectors in the Hedlund example. He construct, using some broken geodesic methods, *local* minimizers shadowing any curve made of segments (of sufficient length) of the minimizing axes and jumps between the axes. This makes for extremely rich, chaotic dynamics.

Exercise 50.9 Find hypotheses on the generating function of an symplectic twist map F which translate to Mather's hypotheses for the Lagrangian that suspends F (*Hint.* You may want to include Bialy and Polterovitch's conditions of Theorem 40.1 for F to have a convex suspension. Note that completeness of the flow is for free: F is defined everywhere.)

51.* The Case Of Hyperbolic Manifolds

We start this section with another counterexample to the strict convexity of Mather's β function. The setting is that of a metric on the two-holed torus, the simplest example of a compact hyperbolic manifold. However, we finish the section on a positive note, by quoting a result of Boyland & Golé (1996b), in which we introduce another definition of rotation vector suited to hyperbolic manifolds and show the existence of minimal orbits of *all* rotation directions for a class of Lagrangian systems on hyperbolic manifolds only slightly smaller than that considered by Mather.

A*. Hyperbolic Counterexample

Take the metric of constant negative curvature on the surface of genus 2 (the two-holed torus) which has a long neck between the two holes (see Figure 51. 1). A minimizer here is a minimizing geodesic for the hyperbolic metric. With a and b as shown, the minimal measure for the homology class $[a] + [b]$ is a linear combination of the ergodic measures supported on Γ_a and Γ_b, where Γ_a and Γ_b are the closed geodesics in the homotopy classes of a and b, respectively. Indeed, Γ_a and Γ_b must "go around" the same holes as a and b, and any closed curve that crosses the neck will be longer than the sum of the lengths of Γ_a and Γ_b. Hence $([a] + [b], \beta([a] + [b]))$ cannot be an extreme point of $graph(\beta)$.

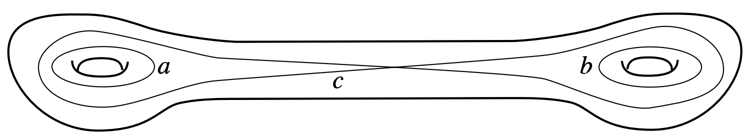

Fig. 51. 1. The surface of genus two and the loops a and b. No minimal measure with rotation vector $[a]+[b]$ can have support passing through the long neck. In particular, a curve in the homotopy class of c cannot yield a minimizer in the abelian cover.

B*. All Rotation Directions in Hyperbolic Manifolds

As the previous example shows, the notion of minimizers in the abelian cover is too restrictive, as it rules out many geodesics. Instead of working on the universal *abelian* cover, we work in the universal cover and define minimizers and rotation vectors with respect to that cover.

All manifolds of dimension n which admit a hyperbolic metric of constant curvature -1 have the Poincaré n–disk as universal covering space $\mathbb{H}^n$. Hence a hyperbolic manifold M is the quotient $\mathbb{H}^n/\pi_1(M)$ where $\pi_1(M)$ acts on $\mathbb{H}^n$ as the group of deck transformations. To visualize $\mathbb{H}^n$, assume $n = 2$, which covers any orientable surface of genus greater or equal to two. $\mathbb{H}^2$ is the usual Euclidean unit disk which is given the hyperbolic metric $\frac{dx^2+dy^2}{1-(x^2+y^2)}$. The ratio between the corresponding hyperbolic distance and the euclidean one tends exponentially to ∞ as points approach the boundary of the disk. Geodesics for the hyperbolic metric are arcs of (euclidean) circles perpendicular to the boundary $\partial\mathbb{H}^2$ of the disk.

The minimizers we consider in this section lift to curves in the *universal cover* which minimize the action between any two of their points. We also assume that the Lagrangian L satisfies Mather's hypotheses (time periodic C^2 function with (a) fiber convexity, (b) completeness of the E–L flow) except that we replace his condition (c) of superlinearity by one of superquadraticity:

(c') *superquadraticity*: There exists a $C > 0$ such that $L(x, v, t) \geq C \|v\|^2$.

(This, again, is satisfied by mechanical systems: if the potential is not positive, one can add a constant to it without changing the solutions to the system).

Theorem 51.1 (Boyland–Golé) *Let (M, g) be a closed hyperbolic manifold. Given a Lagrangian L which satisfies Hypotheses (a), (b), (c'), there are sequences k_i, κ_i, T_i in $\mathbb{R}^+$ depending only on L, with k_i increasing to infinity, such that, for any hyperbolic geodesic $\Gamma \subset \mathbb{H}^n = \tilde{M}$, there are minimizers $\gamma_i : \mathbb{R} \to \tilde{M}$ with $dist(\gamma_i, \Gamma_0) \leq \kappa_i$, $\gamma_i(\pm\infty) = \Gamma_0(\pm\infty)$, and $k_i \leq \frac{1}{d-c} dist(\gamma_i(d), \gamma_i(c)) \leq k_{i+1}$ whenever $d - c \geq T_i$.*

Theorem 51.2 (Boyland–Golé) *Let (M, g) be a closed hyperbolic manifold with geodesic flow g_t. Given a Lagrangian L which satisfies Hypotheses (a), (b), (c') with E-L flow ϕ_t, there exists sequences k_i and T_i with k_i increasing to infinity, and a family of compact, ϕ_t-invariant sets $X_i \subset \mathcal{M}$ so that for all i, (X_i, ϕ_t) is semiconjugate to $(T_1 M, g_t)$ and $k_i \leq \frac{1}{T} dist(\phi_T(\boldsymbol{x}), \phi_0(\boldsymbol{x})) \leq k_{i+1}$, whenever $T \geq T_i$ and $\boldsymbol{x} \in X_i$.*

Hence the geodesic flow and the foliation of invariant ball bundles in T^*M continues to exists, in a weak sense, in any of our general Lagrangian systems.

We now interpret Theorem 51.1 as saying that there exist minimizers of all rotation directions, with a new definition of such a concept valid only for hyperbolic manifolds. Let us first reinterpret the rotation vector on $T^*\mathbb{T}^n$ geometrically: a curve γ on $\mathbb{T}^n$ has rotation vector $\boldsymbol{v} \in \mathbb{R}^n$ if its lift $\tilde{\gamma}$ in the universal cover $\mathbb{R}^n$ is "asymptotically parallel" to the straight line supporting $\boldsymbol{v}$ and if the average of $\|\dot{\gamma}(t)\|$ over all $t \in \mathbb{R}$ is equal to $\|\boldsymbol{v}\|$ (we let the reader make these statement precise and rigorous). Now given two points on $\partial\mathbb{H}^2$, there is exactly one geodesic Γ_0 that goes to the first as $t \to -\infty$, to the other one as $t \to +\infty$. We can declare a curve γ to be asymptotically parallel to Γ_0 iff γ and Γ have same endpoints. This will insure that points of γ are always at a bounded hyperbolic distance from Γ. We also declare that the rotation vector exists iff $\tilde{\gamma}$ has the same endpoints at $\pm\infty$ as a geodesic Γ_0, and if the average $|\rho(\gamma)|$ of $\|\dot{\gamma}\|$ over $t \in \mathbb{R}$ exists, and we define the *rotation vector* to be the pair $\rho(\gamma) = (\Gamma_0, |\rho(\gamma)|)$ (average direction and average speed). In that language, Theorem 51.1 states that, given any geodesic Γ_0, there are infinitely many E-L minimizers with Γ as a rotation direction.

The naive definition of rotation vector that we just outlined has some major flaws:

1. $\rho(\gamma)$ (if it exists) does not belong to a linear space.
2. Two lifts of the same curve γ *will* have different rotation vectors.
3. Rotation direction is not constant $\mu - a.e.$ for many ergodic measures for the geodesic flow.

To remedy that, let $\pi_1(M)$, seen as deck transformation group, act on geodesics in $\mathbb{H}^2$ and declare that two geodesics are parallel iff they belong to the closure of the same $\pi_1(M)$-orbit. Consider the set of tangent vectors at all points of all the geodesics in the closure of a $\pi_1(M)$ orbit. This forms a closed subset of the unit tangent bundle of $\mathbb{H}^2$. The projection by the differential of covering map of this set on the unit tangent bundle of M is the support of a measure μ which is invariant under the geodesic flow. Because of this, Boyland (I996) defines the rotation direction of a curve to be a measure invariant under the geodesic flow, weak* limit of measures supported by geodesics joining two points of the curve. This rotation vector being defined through ergodic theory, it is constant $\mu - a.e.$ for any E-L ergodic μ. Theorem 51.2 implies the existence of minimizer of all rotation directions, in this new sense of the word.

Note that there are many more such "homotopy" directions than there are "homology" directions. For instance the "long neck" metric of Figure 51. 1 has no homology minimizer with rotation direction c, as argued in the previous subsection, but it will have infinitely many homotopy minimizers with that direction.

On the negative side, the counterexamples of Manẽ (1991) and Boyland & Golé (1996a) on $\mathbb{T}^2$ probably have counterparts on hyperbolic manifolds, even with our new definition of rotation vector and we think there is little chance to prove the existence of minimizers of all rotation vectors, even on these manifolds.

52.* Concluding Remark

So what, in the end, are the chances of finding orbits of all rotation vectors for symplectic twist maps or Lagrangian system, in say, $T^*\mathbb{T}^n$? Previous attempts at this problem yielded incomplete results. Bernstein & Katok (1987) "almost" found, for minimizing periodic orbits of symplectic twist maps close to integrable, some uniform modulus of continuity, which they hoped would unable them to take limits and get orbits with the limiting rotation vectors. In my thesis, I hoped that proving some regularity of the ghost tori (invariant set for the gradient flow of the periodic action) might enable one to do the same. This is how ghost circles came about.

One thing is clear: one cannot hope for *global* minimizers to achieve all possible rotation vectors. However, the shadowing methods to construct local minimizers of all rotation vectors of Levi (1997) on the Hedlund counterexamples indicate a possible approach to the general case. The recent work of Mather on existence of unbounded orbits (see Delshams, de la Llave & Seara (1998)and the end of INVchapter), also shows that, for general systems, hyperbolic and variational techniques can combine powerfully to construct orbits shadowing successive minimizers. One possibility to attack this problem would be to try to construct, in a manner analogous to Levi (1997), orbits shadowing the different supports of the ergodic measures which are extreme points of one generalized Mather set $\mathcal{M}_c$. Doing so, one may manage to "fill in" the corresponding set of rotation vectors X_c with rotation vectors of actual orbits, may they be local minimizers.

CHAPTER 10 or CZ

GENERATING PHASES AND SYMPLECTIC TOPOLOGY

July 15 1999

Look up Siburg's work on capacity vs symplectic twist map . What about the title of this Chapter? What about the manifold of points that only move radially? Otherwise, declare this Chapter done, after a last reading.

In Appendix 1 or SG, Section 46, we remark that the differential of a function $W : M \to T^*M$ gives rise to the Lagrangian submanifold $dW(M)$ of T^*M. As a generalization of this fact, one can construct Lagrangian submanifolds of T^*M as symplectic reductions of graphs of differentials of generating phases, which are functions on vector bundles over M.

Generating phases are the common geometric framework to the different discrete variational methods in Hamiltonian systems, including the method developped in this book. Applications of generating phases range from the search for periodic orbits to the Maslov index, symplectic capacities and singularities theory. Generating phases are a viable alternative to the use of heavy functional analytic variational methods in symplectic topology.

This chapter intends to be a basic introduction to generating phases. We first present Chaperon's method, which he used to give an alternate proof of the theorem of Conley & Zehnder (1983) . This theorem, which solved a conjecture by Arnold on the minimum number of periodic points of Hamiltonian maps of $\mathbb{T}^{2n}$, is considered by many as the starting point of symplectic topology$^{(13)}$. We then survey the abstract structure of generating phase, highlighting the common geometric frame for the symplectic twist maps method and that of Chaperon (as well as many others).

53. Chaperon's Method and the Theorem of Conley-Zehnder

Chaperon (1984) introduced a method "du type géodesiques brisées" for finding periodic orbits of Hamiltonians which did not make use of a decomposition by symplectic twist maps. This method has been the basis of later work by Laudenbach, Sikorav and Viterbo.

Until now, we have studied exact symplectic maps that come equipped with a generating function due to the twist condition. The concept of generating function is more general than this, however: we now show how an exact symplectic map of $\mathbb{R}2n$ *which is uniformly* C^1 *close to* Id may have another kind of generating

[13] In the sense that it implies that the C^0 closure of the set of symplectic diffeomorphisms is strictly included in the set of volume preserving diffeomorphisms.

function. The small time t map of a large class of Hamiltonians satisfy this condition. Hence, the time one map of these Hamiltonians can be decomposed into maps that posess this kind of generating function, leading to a new variational setting for periodic orbits. Let

$$F : \mathbb{R}2n \to \mathbb{R}2n$$
$$(\boldsymbol{q}, \boldsymbol{p}) \to (\boldsymbol{Q}, \boldsymbol{P})$$

be an exact symplectic diffeomorphism:

$$\boldsymbol{P}d\boldsymbol{Q} - \boldsymbol{p}d\boldsymbol{q} = F^*\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q} = dS, \tag{53.1}$$

for some $S : \mathbb{R}2n \to \mathbb{R}$ (remember that all symplectic diffeomorphism of $\mathbb{R}2n$ are in fact exact symplectic. We stress *exact* symplectic here in view of our later generalization to T^*M.) The following simple lemma is crucial here.

Lemma 53.1 *Let $F : \mathbb{R}2n \to \mathbb{R}2n$ be an exact symplectic diffeomorphism. Then, if $\|F - Id\|_{C^1}$ is small enough, the map*

$$\phi : (\boldsymbol{q}, \boldsymbol{p}) \to (\boldsymbol{Q}, \boldsymbol{p})$$

is a diffeomorphism of $\mathbb{R}2n$.

Proof. $\boldsymbol{Q}(\boldsymbol{q}, \boldsymbol{p})$ is C^1 close to $\boldsymbol{q}$ and thus ϕ is (uniformly) C^1 close to Id, hence a diffeomorphism. □

We now show how, a way that is slightly different from the twist map case, F can be recovered from S. We define

$$\tilde{S}(\boldsymbol{Q}, \boldsymbol{p}) = \boldsymbol{p}\boldsymbol{q} + S(\boldsymbol{Q}, \boldsymbol{p}), \quad \text{where} \quad \boldsymbol{q} = \boldsymbol{q}(\boldsymbol{Q}, \boldsymbol{p});$$

then

$$d\tilde{S} = \boldsymbol{P}d\boldsymbol{Q} + \boldsymbol{q}d\boldsymbol{p} \tag{53.2}$$

and thus $\tilde{S}$ generates F, in the sense that:

$$\begin{aligned} \boldsymbol{P} &= \frac{\partial \tilde{S}}{\partial \boldsymbol{Q}}(\boldsymbol{Q}, \boldsymbol{p}) \\ \boldsymbol{q} &= \frac{\partial \tilde{S}}{\partial \boldsymbol{p}}(\boldsymbol{Q}, \boldsymbol{p}). \end{aligned} \tag{53.3}$$

Remark 53.2 Note that Id is not a symplectic twist map and thus it cannot be given a generating function in the twist map sense. One of the advantages of the present approach is that Id does have a generating function, which is

$$\tilde{S}(\boldsymbol{Q}, \boldsymbol{p}) = \boldsymbol{p}\boldsymbol{Q}$$

.

As an illustration, fixed points of F are given by the equations:

$$\boldsymbol{p} = \frac{\partial \tilde{S}}{\partial \boldsymbol{Q}} = \boldsymbol{P},$$
$$\boldsymbol{Q} = \frac{\partial \tilde{S}}{\partial \boldsymbol{p}} = \boldsymbol{q},$$

which are equivalent to the following equation:

$$d(\tilde{S} - \boldsymbol{p}\boldsymbol{Q}) = (\boldsymbol{P} - \boldsymbol{p})d\boldsymbol{Q} + (\boldsymbol{q} - \boldsymbol{Q})d\boldsymbol{p} = 0.$$

Hence have reduced the problem of finding fixed point of an exact symplectic diffeomorphism C^1 close to Id on $\mathbb{R}^{2n}$ to the one of finding critical points for a real valued function. We now apply this method to give Hamiltonian maps of $\mathbb{T}^{2n}$ a finite dimensional variational context. It can also be used for time one maps of Hamiltonians with compact support in $\mathbb{R}2n$, or Hamiltonian maps that are C^0 close to Id in a compact symplectic manifold.

Let $H : \mathbb{R}2n \times \mathbb{R}$ be a C^2 function with variables $(\boldsymbol{q}, \boldsymbol{p}, t)$. Assume H is $\mathbb{Z}^{2n}$ periodic in the variables $(\boldsymbol{q}, \boldsymbol{p})$ (i.e., H is a function on $\mathbb{T}^{2n} \times \mathbb{R}$). As in Appendix 1 or SG, we denote by $h^t_{t_0}(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q}(t), \boldsymbol{p}(t))$ the solution of Hamilton's equations with initial conditions $\boldsymbol{q}(t_0) = \boldsymbol{q}, \quad \boldsymbol{p}(t_0) = \boldsymbol{p}$. By assumption, $h^t_{t_0}$ can be seen as a Hamiltonian map on $\mathbb{T}^{2n}$. We know that $h^t_{t_0}$ is exact symplectic (see Theorem 47.7). Furthermore, by compactness of $\mathbb{T}^{2n}$, *when* $|t - t_0|$ *is small,* $h^t_{t_0}$ *is* C^1 *close to* Id (the Hamiltonian vector field of a C^2 function is C^1, hence so is its flow). For $|t - t_0|$ small enough, we can apply Lemma 53.2 to get a generating function for $h^t_{t_0}$. To make this argument global, we decompose h^1 in smaller time maps (see Exercise 47.4):

$$h^1 = h^1_{\frac{N-1}{N}} \circ h^{\frac{N-1}{N}}_{\frac{N-2}{N}} \circ \ldots \circ h^{\frac{1}{N}}_{\frac{2}{N}} \circ h^{\frac{1}{N}}_0 \tag{53.4}$$

and thus, for a large enough N, h^1 can be decomposed into N maps that satisfy Lemma 53.2. [The farther h^1 is from Id, the bigger N must be.] We can then apply the following proposition to h^1:

Proposition 53.3 *Let* $F = F_N \circ \ldots \circ F_1$ *where each* F_k *is exact symplectic in* $T^*\mathbb{R}^n$, C^1 *close to* Id, *and has generating function* $\tilde{S}_k(\boldsymbol{Q}, \boldsymbol{p})$. *The fixed points of* F *are in one to one correspondence with the critical points of :*

$$\tilde{W}(\boldsymbol{Q}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{Q}_N, \boldsymbol{p}_N) = \sum_{k=1}^{N} \tilde{S}_k(\boldsymbol{Q}_k, \boldsymbol{p}_k) - \boldsymbol{p}_k \boldsymbol{Q}_{k-1}$$

where we set $\boldsymbol{Q}_0 = \boldsymbol{Q}_N$.

Proof. We will use the notation

$$(\boldsymbol{P}_k, \boldsymbol{Q}_k) = F_k(\boldsymbol{q}_k, \boldsymbol{p}_k)$$

where we know from (53.3) that $\boldsymbol{P}_k$ and $\boldsymbol{q}_k$ are functions of $\boldsymbol{Q}_k, \boldsymbol{p}_k$. Then, using Equation (53.2) ,

$$\begin{aligned} d\tilde{W}(\overline{\boldsymbol{Q}}, \overline{\boldsymbol{p}}) &= \sum_{k=1}^{N} \boldsymbol{P}_k d\boldsymbol{Q}_k + \boldsymbol{q}_k d\boldsymbol{p}_k - \boldsymbol{p}_k d\boldsymbol{Q}_{k-1} - \boldsymbol{Q}_{k-1} d\boldsymbol{p}_k \\ &= \sum_{k=1}^{N-1} (\boldsymbol{P}_k - \boldsymbol{p}_{k+1}) d\boldsymbol{Q}_k + \sum_{k=2}^{N} (\boldsymbol{q}_k - \boldsymbol{Q}_{k-1}) d\boldsymbol{p}_k \\ &\quad + (\boldsymbol{P}_N - \boldsymbol{p}_1) d\boldsymbol{Q}_N + (\boldsymbol{q}_1 - \boldsymbol{Q}_N) d\boldsymbol{p}_1. \end{aligned} \tag{53.5}$$

This formula proves that $(\overline{\boldsymbol{Q}}, \overline{\boldsymbol{p}})$ is critical exactly when:

$$F_k(\boldsymbol{q}_k, \boldsymbol{p}_k) = (\boldsymbol{q}_{k+1}, \boldsymbol{p}_{k+1}), \forall k \in \{1, \dots, N-1\},$$
$$F_N(\boldsymbol{q}_N, \boldsymbol{p}_N) = (\boldsymbol{q}_1, \boldsymbol{p}_1),$$

that is, exactly when $(\boldsymbol{q}_1, \boldsymbol{p}_1)$ is a fixed point for F. □

As with the function W in Chapter 6, $\tilde{W}$ has the interpretation of the action of a "broken" solution of the Hamiltonian equation. This time, the jumps are both vertical *and* horizontal:

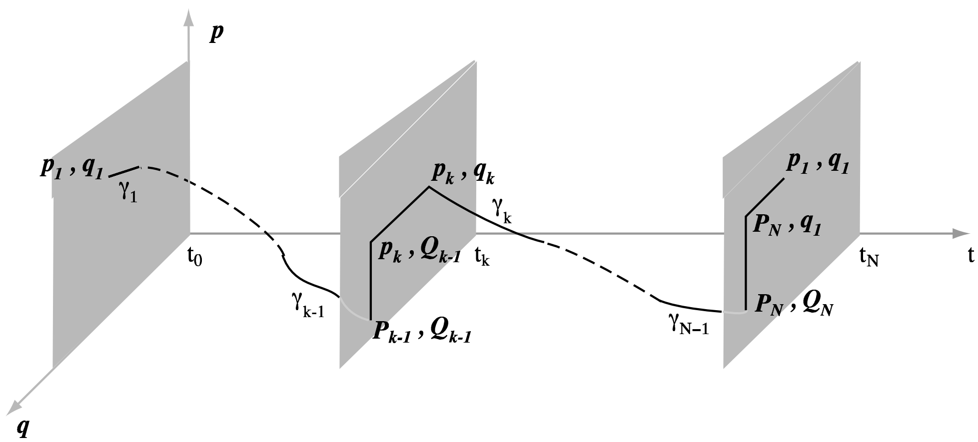

Fig. 53. 1. Interpretation of $\tilde{W}$ as the action of a "broken" solution Γ, concatenation of the solution segments γ_k and "corners" in the $t = t_k$ planes.

Each curve γ_k in Figure 53. 1 is the unique solution of Hamilton's equations starting at $(\boldsymbol{q}_k, \boldsymbol{p}_k, t_k)$ where $t_k = \frac{k-1}{N}$ and flowing for time $1/N$. Since $\tilde{S}_k(\boldsymbol{Q}_k, \boldsymbol{p}_k) = S_k(\boldsymbol{q}_k, \boldsymbol{p}_k) + \boldsymbol{p}_k \boldsymbol{q}_k$ and $S_k(\boldsymbol{q}_k, \boldsymbol{p}_k) = \int_{\gamma_k} \boldsymbol{p} d\boldsymbol{q} - H dt$ (see Theorem 47.7), $\tilde{W}$ measures the action of the broken solution Γ:

$$\begin{aligned} \tilde{W}(\boldsymbol{Q}_1, \boldsymbol{p}_1, \dots, \boldsymbol{Q}_N, \boldsymbol{p}_N) &= \sum_{k=1}^{N} \boldsymbol{p}_k(\boldsymbol{q}_k - \boldsymbol{Q}_{k-1}) + \sum_{k=1}^{N} \int_{\gamma_k} \boldsymbol{p} d\boldsymbol{q} - H dt \\ &= \int_{\Gamma} \boldsymbol{p} d\boldsymbol{q} - H dt. \end{aligned} \tag{53.6}$$

This is the definition given by Chaperon (1984) and (1989).

The following theorem, solved a famous conjecture by Arnold in the case of the Torus. It was hailed as the start of symplectic topology, as it shows that symplectic diffeomorphisms have dynamics necessarily different from that of general diffoemorphisms, or even volume preserving diffeomorphisms. The original proof of Conley and Zehnder also reduces the analysis to finite dimensions, but by truncating Fourier series of periodic orbits. Chaperon's proof avoids the functional analysis altogether.

Theorem 53.4 (Conley-Zehnder) *Let h^1 be a Hamiltonian map of $\mathbb{T}^{2n}$. Then h^1 has at least $2n+1$ distinct fixed points and at least 2^n of them if they all are nondegenerate.*

Proof. Let $\tilde{W}$ be defined as in Proposition 53.3 for the decomposition of h^1 into symplectic maps close to Id given by (53.5) . We will show that $\tilde{W}$ is equivalent to a g.p.q.i. on $\mathbb{T}^{2n}$, and hence it has the prescribed number of critical points, corresponding to fixed points of h^1. We refer the reader to Section TOPOsecgpqi for the definition and properties of generating phases that are relevant here. We first note that $\tilde{W}$ induces a function on $(\mathbb{R}2n)^N/\mathbb{Z}^{2n}$ where $\mathbb{Z}^{2n}$ acts on $(\mathbb{R}2n)^N$ by:

$$(\boldsymbol{m}_q, \boldsymbol{m}_p).(\boldsymbol{Q}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{Q}_N, \boldsymbol{p}_N) = (\boldsymbol{Q}_1 + \boldsymbol{m}_q, \boldsymbol{p} + \boldsymbol{m}_p, \ldots, \boldsymbol{Q}_N + \boldsymbol{m}_q, \boldsymbol{p}_N + \boldsymbol{m}_p)$$

The fact that $\tilde{W}$ is invariant under this action is most easily seen from (53.6) . Indeed, since the Hamiltonian flow is a lift from one on $\mathbb{T}^{2n}$, the curve $\gamma_k + (\boldsymbol{m}_q, \boldsymbol{m}_p, 0)$ is the solution between $(\boldsymbol{q}_k + \boldsymbol{m}_q, \boldsymbol{p}_k + \boldsymbol{m}_p)$ and $(\boldsymbol{Q}_k + \boldsymbol{m}_q, \boldsymbol{P}_k + \boldsymbol{m}_p)$ starting at time $\frac{k-1}{N}$ of that flow. But

$$\int_{\gamma_k + (\boldsymbol{m}_q, \boldsymbol{m}_p, 0)} \boldsymbol{p}d\boldsymbol{q} + Hdt = \int_{\gamma_k} (\boldsymbol{p} + \boldsymbol{m}_p)d\boldsymbol{q} - Hdt = \boldsymbol{m}_p(\boldsymbol{Q}_k - \boldsymbol{q}_k) + \int_{\gamma_k} \boldsymbol{p}d\boldsymbol{q} - Hdt$$

Hence the action of γ_k changes by $\boldsymbol{m}_p(\boldsymbol{Q}_k - \boldsymbol{q}_k)$ under this transformation. On the otherhand, under the same transformation, the sum $\sum_{k=1}^N \boldsymbol{p}_k(\boldsymbol{q}_k - \boldsymbol{Q}_{k-1})$ of Formula (53.6) changes by $\sum_{k=1}^N \boldsymbol{m}_p(\boldsymbol{q}_k - \boldsymbol{Q}_{k-1})$. Summing up the actions of the γ_k, these changes cancel out, from which we deduce that $\tilde{W}$ is invariant under the $\mathbb{Z}^{2n}$ action.

We now show that $\tilde{W}$ is equivalent to a g.p.q.i. over $\mathbb{T}^{2n}$. Let $E = (\mathbb{R}2n)^N \to \mathbb{R}2n$ be the bundle given by the projection map onto $(\boldsymbol{Q}_N, \boldsymbol{p}_N)$ and let $\chi : E \to E$ be the bundle diffeomorphism given by:

$$\chi(\boldsymbol{Q}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{Q}_N, \boldsymbol{p}_N) = (\boldsymbol{a}_1, \boldsymbol{b}_1, \ldots, \boldsymbol{a}_{N-1}, \boldsymbol{b}_{N-1}, \boldsymbol{Q}_N, \boldsymbol{p}_N)$$

where

$$\boldsymbol{a}_k = \boldsymbol{Q}_k - \boldsymbol{Q}_{k-1} \qquad (\boldsymbol{Q}_0 = \boldsymbol{Q}_N)$$
$$\boldsymbol{b}_k = \boldsymbol{p}_k - \boldsymbol{p}_N.$$

In these new coordinates, the $\mathbb{Z}^{2n}$ action only affects $(\boldsymbol{Q}_N, \boldsymbol{p}_N)$, so that $\tilde{W} \circ \chi^{-1}$ induces a function W on $(\mathbb{R}2n)^{N-1} \times \mathbb{T}^{2n}$. We now need to show that W is in fact a g.p.q.i. Define $\tilde{W}_0$ (resp. W_0) to be the functions $\tilde{W}$ (resp. W) obtained when setting the Hamiltonian to zero. Since $\tilde{S}_k(\boldsymbol{Q}_k, \boldsymbol{p}_k) = \boldsymbol{p}_k\boldsymbol{Q}_k$ in this case, $\tilde{W}_0(\overline{\boldsymbol{Q}}, \overline{\boldsymbol{p}}) = \sum_{k=1}^N \boldsymbol{p}_k(\boldsymbol{Q}_k - \boldsymbol{Q}_{k-1})$ and hence a simple computation yields

$$W_0(\overline{\boldsymbol{a}}, \overline{\boldsymbol{b}}, \boldsymbol{Q}_N, \boldsymbol{p}_N) = \sum_{k=1}^{N-1} \boldsymbol{a}_k.\boldsymbol{b}_k$$

which, as easily checked, is quadratic nondegenerate in the fiber.

Finally, we need to check that $\frac{\partial}{\partial \boldsymbol{v}}(W_0 - W)$ is bounded, where $\boldsymbol{v} = (\boldsymbol{a}, \boldsymbol{b})$. It is sufficient for this to check that $d(\tilde{W} - \tilde{W}_0)$ is bounded. Using (53.5) , we obtain:

$$\begin{aligned} d(\tilde{W} - \tilde{W}_0) &= \sum_{k=1}^N (\boldsymbol{P}_k - \boldsymbol{p}_{k+1})d\boldsymbol{Q}_k + \sum_{k=1}^N (\boldsymbol{q}_k - \boldsymbol{Q}_{k-1})d\boldsymbol{p}_k \\ &- \sum_{k=1}^N (\boldsymbol{Q}_k - \boldsymbol{Q}_{k-1})d\boldsymbol{p}_k - \sum_{k=1}^N (\boldsymbol{p}_k - \boldsymbol{p}_{k+1})d\boldsymbol{Q}_k \\ &= \sum_{k=1}^N (\boldsymbol{q}_k - \boldsymbol{Q}_k)d\boldsymbol{p}_k + \sum_{k=1}^N (\boldsymbol{P}_k - \boldsymbol{p}_k)d\boldsymbol{Q}_k \end{aligned}$$

where we have set throughout $\boldsymbol{Q}_0 = \boldsymbol{Q}_N, \boldsymbol{p}_{N+1} = \boldsymbol{p}_1$. Since by definition $(\boldsymbol{Q}_k, \boldsymbol{P}_k) = F_k(\boldsymbol{q}_k, \boldsymbol{p}_k)$ where $F_k = h_{\frac{k-1}{N}}^{\frac{k}{N}}$ lifts a diffeomorphism of $\mathbb{T}^{2n}$, the coefficients of the above differential must be bounded. We can conclude by applying Proposition 52.8. In fact, Proposition TOPOproptrivialgpqi is enough here. □

Remark 53.5 Since the lift of the orbits we find are closed, the orbits in $\mathbb{T}^{2n}$ are contractible. In general, one cannot hope to find periodic orbits of different homotopy classes, as the example $H_0 \equiv 0$ shows. It would be interesting, however, to study the special properties of the set of rotation vectors that orbits of h^1 may have, i.e., to find out if being Hamiltonian implies more properties on this set than those known for general diffeomorphisms of $\mathbb{T}^{2n}$.

54. Generating Phases and Symplectic Geometry

In Section TOPOsecgpqi, we define generating phases as functions $W : E \to \mathbb{R}$, where E is a vector bundle over the manifold M. We then give conditions under which lower estimates on the number of critical points of W can be obtained from the topology of M. In this section, we show how such functions give rise to Lagrangian submanifolds of T^*M, hence the adjective "generating". In particular, we show that the action function obtained either in the symplectic twist map setting or in the Chaperon approach generate a Lagrangian manifold canonically symplectomorphic to the graph of of the map F under consideration. More generally, this construction unifies the different finite, and even infinite, variational approaches in Hamiltonian dynamics.

B. Generating Phases and Lagrangian Manifolds

Let W be a differentiable function $M \to \mathbb{R}$. We have seen in Section 46.C that:

$$dW(M) = \{(\boldsymbol{q}, dW(\boldsymbol{q})) \mid \boldsymbol{q} \in M\}$$

is a Lagrangian submanifold of T^*M. Note that this manifold is a graph over the zero section 0^*_M of T^*M. Heuristicly, we would like to make it possible to similarly "generate" Lagrangian submanifolds that are not graphs. One way to do this is to add auxilliary variables and see our Lagrangian manifold as an appropriate projection in T^*M of a manifold in some bundle over M. This is what is behind the following construction.

Let $\pi : E \to M$ be a fiber bundle over the manifold M. Let $W(\boldsymbol{q}, \boldsymbol{v})$ be a real valued function on an open set $\mathcal{U} \subset E$. The derivative $\frac{\partial W}{\partial \boldsymbol{v}} : E \to E^*$ of W along the fiber of E is well defined, in the sense that if U is a chart on M and $\psi_1, \psi_2 : U \times V \to \pi^{-1}(M)$ are two local trivializations of E, and $W_1 = W \circ \psi_1,\ W_2 = W \circ \psi_2$, then

$$\Phi^* \frac{\partial W_1}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) d\boldsymbol{v} = \frac{\partial W_2}{\partial \boldsymbol{v}}(\Phi(\boldsymbol{q}, \boldsymbol{v})) d\boldsymbol{v}$$

where $\Phi = \psi_2 \circ \psi_1^{-1}$ is the change of trivialization. We assume that the map: $(\boldsymbol{q}, \boldsymbol{v}) \mapsto \frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v})$ is *transverse to* 0. This means that the second derivative (in any coordinates) $(\frac{\partial^2 W}{\partial \boldsymbol{v} \partial \boldsymbol{q}}, \frac{\partial^2 W}{\partial \boldsymbol{v}^2})$ is of maximum rank at points $(\boldsymbol{q}, \boldsymbol{v})$ where $\frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0$. With this assumption, the following set of *fiber critical points* is a manifold of same dimension as M:

$$\Sigma_W = \left\{ (\boldsymbol{q}, \boldsymbol{v}) \in E \mid \frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0 \right\}. \tag{54.1}$$

[For a proof of this general fact about transversality, see *eg.* the theorem p.28 in Guillemin & Pollack (1974)] Define the map:

$$i_W : \Sigma_W \to T^*M$$
$$(\boldsymbol{q},\boldsymbol{v}) \to \left(\boldsymbol{q}, \frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q},\boldsymbol{v})\right)$$

Exercise CZexoimmersion shows that this is an immersion. We now show directly that this immersion is Lagrangian:

$$i_W^* \boldsymbol{p} d\boldsymbol{q} = \frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q},\boldsymbol{v}) d\boldsymbol{q} = dW\big|_{\Sigma_W}(\boldsymbol{q},\boldsymbol{v})$$

and hence:

$$i_W^*(d\boldsymbol{q} \wedge d\boldsymbol{p}) = d^2W\big|_{\Sigma_W} = 0.$$

We will say that W is a *generating phase* for a Lagrangian immersion $j : L \to T^*M$ if $j(L) = i_W(\Sigma_W)$.

Exercise 54.1 Show that $i_W : \Sigma_W \to T^*M$ is an immersion, *i.e.* that $Di_W\big|_{\Sigma_W}$ has full rank (*Hint.* Use the transversality condition to show that $KerDi_W \cap T\Sigma_W = \{0\}$.)

B. Symplectic Properties of Generating Phases

We start with the trivial, but important:

Proposition 54.2 *Suppose the Lagrangian submanifold $L \subset T^*M$ is generated by a function $W : E \to \mathbb{R}$. The points in the intersection of L with the zero section 0_M^* of T^*M are in a one to one correspondance with the critical points of W.*

Proof. $i_W(\boldsymbol{q},\boldsymbol{v})$ is in L if and only if $\frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q},\boldsymbol{v}) = 0$. It is in 0_M^* if and only if $\frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q},\boldsymbol{v}) = 0$. □

In TOPOsecgpqi, we find that critical points persist under elementary operations on generating phases: if $W_1 : E_1 \to \mathbb{R}$, and $W_2 : E_2 \to \mathbb{R}$ are two generating phases such that

$$W_2 \circ \Phi = W_1 + ct, \quad \text{or}$$
$$W_2(\boldsymbol{q},\boldsymbol{v}_1,\boldsymbol{v}_2) = W_1(\boldsymbol{q},\boldsymbol{v}_1) + f(\boldsymbol{q},\boldsymbol{v}_2)$$

where Φ is a fiber preserving diffeomorphism and f is nondegenerate quadratic in $\boldsymbol{v}_2$, then W_1 and W_2 had the same number of critical points. The first operation is called equivalence, the second stabilization. This persistence is now geometrically explained by Proposition 54.2 and the following:

Lemma 54.3 *Two equivalent generating phases generate the same Lagrangian immersion. This is also true under stabilization.*

Proof. Let $W_2 \circ \phi = W_1 + cst$ where Φ is a fiber preserving diffeomorphism between $E_1 \to M$ and $E_2 \to M$. Writing

$$\Phi(\boldsymbol{q},\boldsymbol{v}) = (\boldsymbol{q}, \phi(\boldsymbol{q},\boldsymbol{v})) = (\boldsymbol{q},\boldsymbol{v}'),$$

where $\boldsymbol{v} \to \phi(\boldsymbol{q},\boldsymbol{v})$ is a diffeomorphism for each fixed $\boldsymbol{q}$. we have:

$$W_2(\boldsymbol{q}, \phi(\boldsymbol{q},\boldsymbol{v})) = W_1(\boldsymbol{q},\boldsymbol{v}) + C$$

and thus

$$\frac{\partial W_1}{\partial \boldsymbol{v}} = \left(\frac{\partial W_2{}'}{\partial \boldsymbol{v}} \circ \Phi\right) \cdot \frac{\partial \phi}{\partial \boldsymbol{v}}$$

This implies that $\Sigma_{W_2} = \Phi(\Sigma_{W_1})$, and we conclude the proof of the first assertion by noticing that:

$$\frac{\partial W_1}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}) = \frac{\partial W_2}{\partial \boldsymbol{q}}(\Phi(\boldsymbol{q}, \boldsymbol{v})).$$

Now let

$$W_2(\boldsymbol{q}, \boldsymbol{v}_1, \boldsymbol{v}_2) = W_1(\boldsymbol{q}, \boldsymbol{v}_1) + f(\boldsymbol{q}, \boldsymbol{v}_2)$$

where f is quadratic and nondegenerate. we have:

$$\partial W_2/\partial \boldsymbol{v} = 0 \Leftrightarrow \boldsymbol{v}_2 = 0 \quad \text{and} \quad \partial W_1/\partial \boldsymbol{v}_1 = 0$$

so that $\Sigma_{W_2} = \Sigma_{W_1} \times 0_{E_2}$, where 0_{E_2} is the zero section of E_2. Moreover $\partial f/\partial \boldsymbol{q}\big|_{\{\boldsymbol{v}_2=0\}} = 0$ so that, at points $(\boldsymbol{q}, \boldsymbol{v}_1, 0)$ of Σ_2,

$$\left(\boldsymbol{q}, \frac{\partial W_2}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}_1, 0)\right) = \left(\boldsymbol{q}, \frac{\partial W_1}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}_1)\right).$$

□

C. The Action Function Generates the Graph of F

We examine here the twist map case, and let the reader perform the analysis for the Chaperon case in Exercise 54.4. Let M be an n–dimensional manifold and F be a symplectic twist map on $U \subset T^*M$, where U is of the form $\{(\boldsymbol{q}, \boldsymbol{p}) \in T^*M \mid \|\boldsymbol{p}\| < K\}$. Let $S(\boldsymbol{q}, \boldsymbol{Q})$ be a generating function for a lift $\tilde{F}$ of F. S can be seen as a function on some open set V of $\tilde{M} \times \tilde{M}$, diffeomorphic to $\tilde{U}$. [14] Since $\boldsymbol{P}d\boldsymbol{Q} - \boldsymbol{p}d\boldsymbol{q} = dS(\boldsymbol{q}, \boldsymbol{Q})$, we can describe the graph of $\tilde{F}$ as:

$$Graph(\tilde{F}) = \left\{\left(\boldsymbol{q}, -\frac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{Q}), \boldsymbol{Q}, \frac{\partial S}{\partial \boldsymbol{Q}}(\boldsymbol{q}, \boldsymbol{Q})\right) \mid (\boldsymbol{q}, \boldsymbol{Q}) \in V\right\} \subset (T^*\tilde{M})^2,$$

which is canonically symplectomorphic to:

$$\left\{\left(\boldsymbol{q}, \boldsymbol{Q}, \frac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{Q}), \frac{\partial S}{\partial \boldsymbol{Q}}(\boldsymbol{q}, \boldsymbol{Q})\right) \mid (\boldsymbol{q}, \boldsymbol{Q}) \in V\right\} \subset T^*(\tilde{M} \times \tilde{M}).$$

One can easily check that this manifold has S as a generating phase. In other words *the generating function of a symplectic twist map F is a generating phase for the graph of $\tilde{F}$.*

We expand in more details for the more general case where $F = F_N \circ \ldots \circ F_1$ is a product of symplectic twist maps of $U \subset T^*M$. This time, the candidate for a generating phase is:

$$\tilde{W}(\overline{\boldsymbol{q}}) = \sum_{k=1}^{N} S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}),$$

where we do not identify $\boldsymbol{q}_{N+1}$ and $\boldsymbol{q}_1$ in any way. Then, writing

$$\boldsymbol{v} = (\boldsymbol{q}_2, \ldots, \boldsymbol{q}_N), \quad \boldsymbol{q} = (\boldsymbol{q}_1, \boldsymbol{q}_{N+1}),$$

we will show that $\tilde{W}(\boldsymbol{q}, \boldsymbol{v})$ is a generating phase for $Graph(\tilde{F}) \subset (T^*\tilde{M})^2$. Let

---

[14] in the case where $M = \mathbb{T}^n$, and the map is defined on all of $T^*\mathbb{T}^n$, we have $V \cong \tilde{U} \cong \mathbb{R}2n$.

$$\mathcal{U} = \left\{(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k) \mid (\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) \in \psi_k(\tilde{U})\right\}$$

where ψ_k is the "Legendre transformation" attached to the twist map F_k. Let $\beta : M^{N+1} \to M \times M$ be the map defined by: $(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N+1}) \to (\boldsymbol{q}_1, \boldsymbol{q}_{N+1})$ The bundle that we will consider here is:

$$\mathcal{U} \to \beta(\mathcal{U}) \subset M \times M.$$

Proposition 24.1 states that $\frac{\partial \tilde{W}}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0$ exactly when $\overline{\boldsymbol{q}} = (\boldsymbol{q}, \boldsymbol{v})$ is the $\boldsymbol{q}$ component of the orbit of $(\boldsymbol{q}_1, \boldsymbol{p}_1(\boldsymbol{q}_1, \boldsymbol{q}_2))$ under the successive $\tilde{F}_k$'s. This means that *the set of orbits under the successive $\tilde{F}_k$'s is in bijection with the set $\Sigma_{\tilde{W}} = \{\frac{\partial \tilde{W}}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0\}$ as defined in (54.1)* . Since this set is parametrized by the starting point of an orbit, it is diffeomorphic to U, hence a manifold.

For $\overline{\boldsymbol{q}} \in \Sigma_{\tilde{W}}$, we have:

$$\tilde{F}(\boldsymbol{q}_1, \boldsymbol{p}_1(\boldsymbol{q}_1, \boldsymbol{q}_2)) = (\boldsymbol{q}_{N+1}, \boldsymbol{P}_{N+1}(\boldsymbol{q}_N, \boldsymbol{q}_{N+1}))$$

but:

$$\boldsymbol{p}_1(\boldsymbol{q}_1, \boldsymbol{q}_2) = -\partial_1 S_1(\boldsymbol{q}_1, \boldsymbol{q}_2) = -\frac{\partial \tilde{W}}{\partial \boldsymbol{q}_1}(\boldsymbol{q}_1, \boldsymbol{q}_{N+1}, \boldsymbol{v})$$

$$\boldsymbol{P}_{N+1}(\boldsymbol{q}_N, \boldsymbol{q}_{N+1}) = \partial_2 S_N(\boldsymbol{q}_N, \boldsymbol{q}_{N+1}) = \frac{\partial \tilde{W}}{\partial \boldsymbol{q}_{N+1}}(\boldsymbol{q}_1, \boldsymbol{q}_{N+1}, \boldsymbol{v})$$

In other words, the graph of $\tilde{F}$ in $T^*\tilde{M} \times T^*\tilde{M}$ can be expressed as:

$$Graph(\tilde{F}) = \left\{\left(\boldsymbol{q}_1, -\frac{\partial W}{\partial \boldsymbol{q}_1}(\boldsymbol{q}, \boldsymbol{v}), \boldsymbol{q}_{N+1}, \frac{\partial \tilde{W}}{\partial \boldsymbol{q}_{N+1}}(\boldsymbol{q}, \boldsymbol{v})\right) \;\middle|\; (\boldsymbol{q}, \boldsymbol{v}) \in \Sigma_{\tilde{W}}\right\}.$$

To finish our construction, we define the following symplectic map:

$$j : \left(T^*\tilde{M} \times T^*\tilde{M}, -\Omega_{\tilde{M}} \oplus \Omega_{\tilde{M}}\right) \to \left(T^*(\tilde{M} \times \tilde{M}), \Omega_{\tilde{M} \times \tilde{M}}\right)$$
$$(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{P}) \to (\boldsymbol{q}, \boldsymbol{Q}, -\boldsymbol{p}, \boldsymbol{P}).$$

where Ω_X denotes the canonical symplectic structure on T^*X. Clearly:

$$j(Graph(\tilde{F})) = i_{\tilde{W}}(\Sigma_{\tilde{W}}),$$

that is, $\tilde{W}$ generates the Lagrangian manifold $Graph(\tilde{F})$. Note that the fixed points of F correspond to $Graph(\tilde{F}) \cap \Delta(T^*\tilde{M} \times T^*\tilde{M})$, i.e. to $\overline{\boldsymbol{q}} \in \Sigma_{\tilde{W}}$ such that $\boldsymbol{q}_1 = \boldsymbol{q}_{N+1}$ and $-\partial_1 S_1(\boldsymbol{q}_1, \boldsymbol{q}_2) = \partial_2 S_N(\boldsymbol{q}_N, \boldsymbol{q}_{N+1})$, which are critical points of $W = \tilde{W}\big|_{\{\boldsymbol{q}_1 = \boldsymbol{q}_{N+1}\}}$, as we well know.

Exercise 54.4 Show that the generating function $\tilde{W}$ of Chaperon (see Proposition 53.3) generates the graph of the Hamiltonian map $F : \mathbb{T}^{2n} \to \mathbb{T}^{2n}$. (*Hint.* If you are stuck, consult Laudenbach & Sikorav (1985)

D. Symplectic Reduction

We introduce yet another geometric point of view for the generating phase construction. We will see that if a Lagrangian manifold $L \subset T^*M$ is generated by the phase $W : E \to \mathbb{R}$, than in fact L is the symplectic reduction of the Lagrangian manifold $dW(E) \subset T^*E$. We introduce symplectic reduction in the linear case, and only sketch briefly the manifold case, refering the reader to Weinstein (1979) for more detail.

Consider V, Ω_0, be a symplectic vector space of dimension $2n$. Let C a coisotropic subspace of V. Let $\Lambda(V)$ be the set of Lagrangian subspaces of V (a Grassmanian manifold). The process of symplectic reduction gives a natural map $\Lambda(V) \to \Lambda\left(C/C^{\perp}\right)$ that we now describe. By Theorem 43.1, we know that we can find coordinates for V in which:

$$C = \{(q_1, \ldots, q_n, p_1, \ldots, p_k)\}$$

and we have $C^{\perp} = \{(q_{k+1}, \ldots, q_n)\} \subset C$. Then

$$C/C^{\perp} \simeq \{(q_1, \ldots, q_k, p_1, \ldots, p_k)\}$$

which is obviously symplectic. It is called the *reduced symplectic space* along C. We denote by Red the quotient map $C \to C/C^{\perp}$. The symplectic form Ω_C of $C/C^{\perp}$ is natural in the sense that it makes Red into a symplectic map:

$$\Omega_C(Red(\boldsymbol{v}), Red(\boldsymbol{v}')) = \Omega(\boldsymbol{v}, \boldsymbol{v}'). \tag{54.2}$$

Proposition 54.5 *Let $L \subset V$ be a Lagrangian subspace and $C \subset V$ a coisotropic subspace. Then*

$$L_C = Red(L \cap C) = L \cap C / L \cap C^{\perp}$$

is Lagrangian in $C/C^{\perp}$.

We say that L_C is *the symplectic reduction* of L along the coisotropic space C.

Proof. Formula (54.2) tells us that L_C is isotropic. We need to show that $dim L_C = \frac{1}{2} dim C/C^{\perp}$. Linear algebra tells us that:

$$dim L_C = dim(L \cap C) - dim(L \cap C^{\perp}).$$

As would be the case for any nondegenerate bilinear form, the dimensions of a subspace and that of its orthogonal add up to the dimension of the ambient space. Also, the orthogonal of an intersection is the sum of the orthogonal. Hence:

$$dim V = dim(L \cap C^{\perp}) + dim(L \cap C^{\perp})^{\perp} = dim(L \cap C^{\perp}) + dim L + dim C,$$

since $L^{\perp} = L$. Thus

$$\begin{aligned} dim L_C &= dim(L \cap C) - dim V + dim(L + C) = dim L + dim C - dim V \\ &= dim C - \frac{1}{2} dim V \end{aligned} \tag{54.3}$$

But

$$\begin{aligned} dim(C/C^{\perp}) &= dim C - dim C^{\perp} = dim C - (dim V - dim C) \\ &= 2 dim C - dim V \end{aligned} \tag{54.4}$$

We conclude that $dim L_C = \frac{1}{2} dim(C/C^{\perp})$ by putting (54.3) and (54.4) together. □

We now sketch the reduction construction in the manifold case. Let C be a coisotropic submanifold of a symplectic manifold (M, Ω). Then $TC^{\perp}$ is a subbundle of TC (that is, the fibers are of same dimension and vary smoothly) so we can form the quotient bundle $TC/TC^{\perp}$, with base C and fiber the quotient $T_{\boldsymbol{q}}C/T_{\boldsymbol{q}}C^{\perp}$ at each point $\boldsymbol{q}$ of C. It turns out that this quotient bundle can actually be seen as the tangent bundle of a certain manifold $C/C^{\perp}$, whose points are leaves of the integrable foliation $TC^{\perp}$. Moreover one can show that the naturally induced form Ω_C is indeed symplectic on $C/C^{\perp}$. Finally, we define $red : C \to C/C^{\perp}$ as the propjection. Its derivative is basically the map Red defined above. One can show that, if C intersect a Lagrangian submanifold L transversally, then $L_C = red(L)$ is an immersed symplectic manifold of $C/C^{\perp}$, which is the *reduction of L along C*.

We now apply this new point of view to the generating function construction. Let $E = M \times \mathbb{R}^N$. We show that if $L = i_W(\Sigma_W) \subset T^*M$ is generated by the generating phase $W : E \to \mathbb{R}$, then L is in fact the reduction of $dW(E) \subset T^*E$ along the coisotropic manifold $C = \{\boldsymbol{p_v} = 0\}$, where we have given T^*E the coordinate $(\boldsymbol{q}, \boldsymbol{v}, \boldsymbol{p_q}, \boldsymbol{p_v})$. This is just a matter of checking through the construction. We know that $dW(E)$ is Lagrangian in T^*E. Its intersection with C is the set:

$$dW(E) \cap C = \left\{ (\boldsymbol{q}, \boldsymbol{v}, \boldsymbol{p_q}, \boldsymbol{p_v}) \in T^*E \;\middle|\; \boldsymbol{p_q} = \frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}),\ \boldsymbol{p_v} = \frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0 \right\}$$
$$= dW(\Sigma_W).$$

where Σ_W is the set of fiber critical points. Since by the tranversality condition in our definition of generating phase Σ_W is a manifold, so is $dW(E) \cap C$: for any W, the map $dW : E \to T^*E$ is an embedding. The bundle $TC^{\perp}$ is the one generated by the vector fields $\frac{\partial}{\partial \boldsymbol{v}}$ and thus $C/C^{\perp}$ can be identified with $T^*M = \{(\boldsymbol{q}, \boldsymbol{p_q})\}$. The image of $dW(E) \cap C$ under the projection $red : C \to C/C^{\perp}$ is exactly $i_W(\Sigma_W) = \{(\boldsymbol{q}, \frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v})) \mid \frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0\} = L$. Note that because $E = M \times \mathbb{R}^N$, the above argument is independent of the coordinate chosen (e.g. C is well defined.) With a little care, the argument extends to the case where E is a nontrivial bundle over M.

Exercise 54.6 Show that, in the Darboux coordinate used above, the $\boldsymbol{q}$–plane and the $\boldsymbol{p}$–plane of V both reduce to the $\boldsymbol{q}$ and $\boldsymbol{p}$–plane (resp.) of $C/C^{\perp}$.

E. Further Applications of Generating Phases

The symplectic theory of Generating Phases does not only provide a unifying packaging for the different variational approaches to Hamiltonian systems. It can also serve as the basis of symplectic topology, where invariants called *capacities* play a crucial role. Roughly speaking, capacities are to symplectic geometry what volume is to Riemannian geometry: they provide obstruction for sets to be symplectomorphic, or for sets to be squeezed inside other sets. Viterbo (1992) uses generating phases to define such capacities, in contrast to prior approaches by Gromov (1985)who uses the theory of “pseudo–holomorphic curves”. The basis fo Viterbo’s definition of capacity is a converse statement to Lemma 54.3:

Proposition 54.7 *If W_1 and W_2 both generate $h^t(0^*_M)$, where h^t is a Hamiltonian isotopy, then after stabilization W_1 and W_2 are equivalent.*

In view of this, Viterbo is able to define a capacity for a Lagrangian manifold L Hamiltonian isotopic to 0^*_M by choosing minimax values of a given (and hence any) generating phase for L.

In another work, Viterbo (1987) shows that a certain integer function called Maslov Index on the set of paths in the Lagrangian Grassmannian is invariant under symplectic reduction. It can be shown that the Lagrangian Grassmanian $\Lambda(V)$ has first fundamental group $\pi_1(\Lambda(V)) = \mathbb{Z}$. Very roughly, we can interpret this by saying that $\Lambda(V)$ has a "hole" and the *Maslov index* measures the number of turns a curve makes around that hole. Now let W_t be generating phases for a Hamiltonian isotopy h^t. The set $dW_T(E)$ is Lagrangian in T^*E and its reduction is graph of h^t. The Maslov Index in $\Lambda(T^*E)$ detects the change in Morse Index of the second derivative of W_t, whereas on the graph of h^t, it detects a non transverse intersection with the plane $\{(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P})\}$. This can be used to give a neat generalization to Lemma 31.2and to explain the classical relationship discovered by Morse between the index of the second variation of the action function and the number of "conjugate points"(see Milnor (1969) for the classical, Riemannian geometry case, and Duistermaat (1976) for the more general convex Lagrangian case.) Finally, we refer to Weinstein (1979) , Lecture 6, for further survey on generating phases (called Morse families there) .

Proposition CZproplagrim is 41.5, Exercise CZexoimmersion is CZexoimmersion