Lectures on Symplectic Geometry

Fraydoun Rezakhanlou
Departmet of Mathematics, UC Berkeley

August 17, 2009

1 Introduction

Hamiltonian systems appear in conservative problems of mechanics as in celestial mechanics but also in statistical mechanics governing the motion of particles and molecules in fluid. A mechanical system of N planets (particles) is modeled by a Hamiltonian function $H(x)$ where $x = (q, p)$, $q = (q_1, \dots, q_N)$, $p = (p_1, \dots, p_N)$ with $(q_i, p_i) \in \mathbb{R}^d \times \mathbb{R}^d$ being the position and the momentum of the i-th particle. The Hamiltonian's equations of motion are

$$\dot{q} = H_p(q, p),\ \dot{p} = -H_q(q, p) \tag{1.1}$$

which is of the form

$$\dot{x} = J\nabla_x H(x), \quad J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} \tag{1.2}$$

where $n = dN$ and I_n denotes the $n \times n$ identity matrix. The equation (1.2) is an ODE that possesses a unique solution for every initial data x_0 provided that we make some standard assumptions on H. If we denote such a solution by $\phi_t(x_0) = \phi(t, x_0)$, then ϕ enjoys the group property

$$\phi_t \circ \phi_s = \phi_{t+s}, \quad t, s \in \mathbb{R}.$$

The ODE (1.2) is a system of $2n = 2dN$ unknowns. Such a typically large system can not be solved explicitly. A reduction of such a system is desirable and this can be achieved if we can find some conservation laws associated with our system. To find such conservation laws systematically, let us look at a general ODE of the form

$$\frac{dx}{dt} = b(x), \tag{1.3}$$

with the corresponding flow denoted by ϕ_t, and study $u(x, t) = T_t f(x) = f(\phi_t(x))$. A celebrated theorem of Liouville asserts that the function u satisfies

$$\frac{\partial u}{\partial t} = \mathcal{L}u = b \cdot u_x. \tag{1.4}$$

Recall that a function $f(x)$ is *conserved* if $\frac{d}{dt} f(\phi_t(x)) = 0$. From (1.4) we learn that a function f is conserved if and only if

$$b \cdot f_x = 0. \tag{1.5}$$

In the case of a Hamiltonian system, $b = (H_p, -H_q)$ and the equation (1.5) becomes

$$\{H, f\} := f_q \cdot H_p - f_p \cdot H_q = 0. \tag{1.6}$$

As an obvious choice, we may take $f = H$ in (1.6). In general, we may have other conservation laws that are not so obvious to be found. *Nother's principle* tells us how to find a conservation law using a symmetry of the ODE (1.3). With the aid of the symmetries, we may reduce our system to a simpler one that happens to be another Hamiltonian-type system.

Liouville discovered that for a Hamiltonian system of Nd-degrees of freedom ($2Nd$ unknowns) we only need Nd conserved functions in order to solve the system completely by means of quadratures. Such a system is called *completely integrable* and unfortunately hard to come by. Recently there has been a revival of the theory of completely integrable systems because of several infinite dimensional examples (Korteweg–deVries equation, nonlinear Schrödinger equation, etc.).

As we mentioned before, the conservation laws can be used to simplify a Hamiltonian system by reducing its size. To get more information about the solution trajectories, we may search for other conserved quantities. For example, imagine that we have a flow ϕ_t associated with (1.3) and we may wonder how the volume of $\phi_t(A)$ changes with time for a given measurable set A. For this, imagine that there exists a density function $\rho(x,t)$ such that

$$\int J(\phi_t(x))\rho^0(x)dx = \int J(x)\rho(x,t)dx \tag{1.7}$$

for every bounded continuous function J. This is equivalent to saying that for every nice set A,

$$\int_{\phi_{-t}(A)} \rho^0(x)dx = \int_A \rho(x,t)dx. \tag{1.8}$$

In words, the ρ^0-weighted volume of $\phi_{-t}(A)$ is given by the $\rho(\cdot,t)$-weighted volume of A. Using (1.4), it is not hard to see that in fact ρ satisfies the (dual) Liouville's equation

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho b) = 0. \tag{1.9}$$

As a result, the measure $\rho^0(x)dx$ is invariant for the flow ϕ_t if and only if

$$\operatorname{div}(\rho^0 b) = 0.$$

In particular, if $\operatorname{div} b = 0$, then the Lebesgue measure is invariant. In the case of a Hamiltonian system $b = J\nabla H$, we do have $\operatorname{div} b = 0$, and as a consequence,

$$\operatorname{vol}(\phi_t(A)) = \operatorname{vol}(A), \tag{1.10}$$

for every measurable set A.

In our search for other invariance properties, let us now look for vector fields $F : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\frac{d}{dt}\int_{\phi_t(\gamma)} F \cdot dx \equiv 0 \tag{1.11}$$

for every closed curve γ. Such an invariance property is of interest in (for example) fluid mechanics because $\int_\gamma F \cdot dx$ measures the *circulation* of the velocity field F around γ. To calculate the left-hand side of (1.11), observe

$$\int_{\phi_t(\gamma)} F \cdot dx = \int_\gamma F(\phi_t(x)) D\phi_t(x) \cdot dx$$

where $D\phi_t$ denotes the derivative of ϕ_t in x and we regard F as a row vector. Set $F = (F^1, \ldots, F^k)$, $\phi = (\phi^1, \ldots, \phi^k)$, $u = (u^1, \ldots, u^k)$, $u(x,t) = T_t F(x) = F \circ \phi_t(x) D\phi_t(x)$, so that

$$u^j(x,t) = \sum_i F^i(\phi_t(x)) \frac{\partial \phi_t^i}{\partial x_j}(x).$$

To calculate the time derivative, we write

$$\begin{aligned} u(x, t+h) &= F(\phi_{t+h}) D\phi_{t+h} = F(\phi_t \circ \phi_h)\ D\phi_t \circ \phi_h D\ \phi_h \\ &= u(\phi_h(x), t) D\phi_h(x), \end{aligned}$$

so that

$$\begin{aligned} \frac{d}{dh} u^j(x, t+h)\bigg|_{h=0} &= \sum_i \left[(\nabla u^i \cdot b)\delta_{ij} + u^i \frac{\partial b^i}{\partial x_j} \right] \\ &= \nabla u^j \cdot b + \sum_i u^i \frac{\partial b^i}{\partial x_j} \\ &= (u \cdot b)_{x_j} + \sum_i (u^j_{x_i} - u^i_{x_j}) b^i. \end{aligned}$$

In summary,

$$\frac{\partial u}{\partial t} = \nabla(u \cdot b) + \mathcal{C}(u) b \tag{1.12}$$

where $\mathcal{C}(u)$ is the matrix $[u^i_{x_j} - u^j_{x_i}]$. In particular,

$$\frac{d}{dt}\int_{\phi_t(\gamma)} F \cdot dx = \int_\gamma \mathcal{C}(u) b \cdot dx \tag{1.13}$$

for every closed curve γ. Recall that we would like to find vector fields F for which (1.11) is valid. For this it suffices to have $\mathcal{C}(F)b$ a gradient. Indeed if $\mathcal{C}(F)b$ is a gradient, then

$$\begin{aligned}\frac{d}{dt}\int_{\phi_t(\gamma)} F\cdot dx &= \frac{d}{dh}\int_{\phi_{t+h}(\gamma)} F\cdot dx\bigg|_{h=0}\\ &= \frac{d}{dh}\int_{\phi_h(\phi_t(\gamma))} F\cdot dx\bigg|_{h=0}\\ &= \int_{\phi_t(\gamma)} \mathcal{C}(F)b\cdot dx = 0.\end{aligned}$$

Let us examine this for some examples.

Example 1.1. (i) Assume that $k = 2n$ with $x = (q,p) = (q_1,\dots,q_n,p_1,\dots,p_n)$. Let $b(q,p) = (H_p,-H_q)^t = J\nabla H$ for a Hamiltonian $H(q,p)$. Choose $F(q,p) = (p,0)$. We then have $\mathcal{C}(F) = \begin{bmatrix} 0 & I_n \\ -I_n & 0\end{bmatrix} = J$, and $\mathcal{C}(F)b = JJ\nabla H = -\nabla H$. This and (1.13) imply that for a Hamiltonian flow ϕ_t and closed γ,

$$\frac{d}{dt}\int_{\phi_t(\gamma)} p\cdot dq = 0, \tag{1.14}$$

which was discovered by Poincaré originally.

(ii) Assume $k = 2n+1$ with $x = (q,p,t)$ and $b(q,p,t) = (H_p,-H_q,1)^t$ where H is now a time-dependent Hamiltonian function. Define

$$F(q,p,t) = (p,0,-H(q,p,t)).$$

We then have

$$\mathcal{C}(F) = \left[\begin{array}{c|c|c} 0 & I_n & H_q^t \\ \hline -I_n & 0 & H_p^t \\ \hline -H_q & -H_p & 0\end{array}\right] = \left[\begin{array}{c|c} J & \nabla H^t \\ \hline \nabla H & 0\end{array}\right]. \tag{1.15}$$

Since $\mathcal{C}(F)b = 0$, we deduce that for any closed (q,p,t)-curve γ,

$$\frac{d}{ds}\int_{\phi_s(\gamma)} (q\cdot dp - H(q,p,t)dt) = 0, \tag{1.16}$$

proving a result of *Poincaré* and *Cartan*. Note that if γ has no t-component in (1.16), then (1.16) becomes (1.14).

(iii) Assume $n = 3$. Then $\mathcal{C}(F) = \begin{bmatrix} 0 & -\alpha_3 & \alpha_2 \\ \alpha_3 & 0 & -\alpha_1 \\ -\alpha_2 & \alpha_1 & 0 \end{bmatrix}$ with $(\alpha_1, \alpha_2, \alpha_3) = \nabla \times F$. Now if $b = \nabla \times F$, then $\mathcal{C}(F)b = (\nabla \times F) \times b$ and $\frac{d}{dt} \int_{\phi_t(\gamma)} F \cdot dx = 0$. In words, the F-circulation of a curve moving with velocity field $\nabla \times F$ is preserved with time. $\square$

Example 1.2. A Hamiltonian system (1.2) simplifies if we can find a function $w(q, t)$ such that $p(t) = w(q(t), t)$. If such a function w exists, then $q(t)$ solves

$$\frac{dq}{dt} = H_p(q, w(q, t), t). \tag{1.17}$$

The equation $\dot{p}$ gives us the necessary condition for the function w:

$$\dot{p} = w_q \dot{q} + w_t = w_q \cdot H_p(q, w, t) + w_t,$$
$$\dot{p} = -H_q(q, w, t).$$

Hence $w(q, t)$ must solve,

$$w_t + w_q \cdot H_p(q, w, t) + H_q(q, w, t) = 0. \tag{1.18}$$

For example, if $H(q, p, t) = \frac{1}{2}|p|^2 + V(q, t)$, then (1.18) becomes

$$w_t + w_q w + V_q(q, t) = 0. \tag{1.19}$$

The equation (1.17) simplifies to

$$\frac{dq}{dt} = w(q, t) \tag{1.20}$$

in this case. If the flow of (1.20) is denoted by ψ_t, then $\phi_t(q, p) = (\psi_t(q), w(\psi_t(q), t))$. Now (1.14) means that for any closed q-curve η,

$$\frac{d}{dt} \int_{\psi_t(\eta)} w(q, t) \cdot dq = 0.$$

This is the celebrated *Kelvin's circulation theorem.* $\square$

We may use Stokes' theorem to rewrite (1.14) as

$$\frac{d}{dt} \int_{\phi_t(\Gamma)} \bar{\omega} := \frac{d}{dt} \int_{\phi_t(\Gamma)} dp \wedge dq = 0 \tag{1.21}$$

for every two-diimensional surface Γ. In words, the 2-form $\bar{\omega}$ is invariant under the Hamiltonian flow ϕ_t. In summary, we have found various invariance principles for Hamiltonian flows:

- The conserved functions f satisfying (1.6) is an example of an invariance principle for 0-forms.
- The Liouville's theorem (1.10) is an example of an invariance principle of an n-form.
- Poincaré's theorem (1.21) is an instance of an invariance principle involving a 2-form.

In fact (1.21) implies (1.10) because the invariance of $\bar{\omega}$ implies the invariance of the $k = 2n$ form $\bar{\omega}^n = \bar{\omega} \wedge \cdots \wedge \bar{\omega}$ which is a constant multiple of the volume form. More generally, we may take an arbitrary l-form ω and evolve it by the flow ϕ_t of a velocity field b. If we write $\omega(t)$ for $\phi_t^*\omega$:

$$\int_\Gamma \omega(t) = \int_\Gamma \phi_t^*\omega = \int_{\phi_t(\Gamma)} \omega,$$

then by a formula of Cartan,

$$\frac{d\omega}{dt} = \mathcal{L}_b\omega := d(i_b\omega) + i_b(d\omega)$$

where $\mathcal{L}_b\omega$ denotes the *Lie derivative.*

The configuration space of a system with constraints is a manifold. Also, when we use conservation laws to reduce our Hamiltonian system, we obtain a Hamiltonian system on a manifold. If the configuration space is a n-dimensional differentiable manifold N, and $L : TN \to \mathbb{R}$ is a differentiable *Lagrangian function*, then $p = \frac{\partial L}{\partial \dot{q}}$ is a cotangent vector. The cotangent bundle $M = T^*N$ is an example of a *symplectic manifold* because it possesses a natural closed non-degenerate form $\bar{\omega}$ which is simply $\sum_1^n dp_i \wedge dq_i$, in local coordinates. More generally we may study an even dimensional manifold M, equipped with a non-degenerate closed 2-form ω, and construct vector fields X_H associated with scalar functions H such that $i_{X_H}(\omega) = -dH$. The vector field X_H is the analog of $J\nabla H$ in the Euclidean case $M = \mathbb{R}^{2n}$. By the non-degeneracy of ω, such X_H exists for every differentiable Hamiltonian function H.

A celebrated theorem of Darboux asserts that any symplectic manifold is locally equivalent to a Euclidean space with its standard symplectic structure. As a result, the most important questions in symplectic geometry are the global ones.

Consider the Euclidean space $(\mathbb{R}^{2n}, \bar{\omega})$. If the hypersurface $\Gamma = H^{-1}(c)$ is a compact energy level set with $\nabla H \neq 0$ on Γ, then the unparametrized orbits on Γ of the Hamiltonian vector field $X_H = J\nabla H$ are independent of the choice of H. One can therefore wonder what hypersurfaces carry a periodic orbit. P. Rabinowitz showed that every star-like hypersurface carries a periodic orbit. Later, Viterbo showed that the same holds more generally for hypersurfaces of *contact type*, establishing affirmatively a conjecture of A. Weinstein.

Consider two compact connected domains U_1 and U_2 in $\mathbb{R}^n$ with smooth boundaries. U_1 and U_2 are diffeomorphic and volume(U_1) = volume(U_2), then we can find a diffeomorphism between U_1 and U_2 that is also volume preserving (DaCorogna–Moser). We may wonder whether or not there exists a symplectic diffeomorphism between U_1 and U_2. *Gromov's squeezing theorem* shows that the symplectic transformations are more rigid; if there exists a symplectic embedding from the ball

$$B_R(0) = \{(q,p) : |q|^2 + |p|^2 < R^2\}$$

into the cylinder

$$Z_r(0) = \{(q,p) : q_1^2 + p_1^2 < r^2\},$$

then we must have $r \geq R$! Motivated by this, Gromov defines the *symplectic radius* $r(M)$ of a symplectic manifold (M, ω) as the largest r for which there exists a symplectic embedding from $B_r(0)$ into M. The Gromov's radius is an example of a *symplectic capacity* that is a symplectic invariant. Since the discovery of the Gromov radius, new capacities have been discovered. The existence of some of these capacities can be used to prove various global properties of Hamiltonian systems such as Viterbo's existence of periodic orbits.

Another rigidity of symplectic transformation is illustrated in an important result of Eliashberg and Gromov: If $\{f_m\}$ is a sequence of symplectic transformation that converges uniformly to a differentiable function f, then f is also symplectic. The striking aspect of this result is that our definition of a symplectic function f involves the first derivative of f. As a result, we should expect to have a definition of symplicity that does not involve any derivative. This should be compared to the definition of a volume preserving transformation that can be formulated with or without using derivative.

2 Quadratic Hamiltonian

Quadratic functions are the simplest non-trivial examples of Hamiltonian functions. In this section we define the *symplectic spectrum* and explore its monotonicity. Before embarking on this, let us first review some well-known facts for symmetric matrices, which is the symmetric counterpart of what we will discuss for symplectic matrices,

To begin, let us use the dot product for the standard Euclidean inner product and this inner product is preserved by a matrix A if A is orthogonal. That is

$$Aa \cdot Ab = a \cdot b \text{ for all } a, b \in \mathbb{R}^k \Leftrightarrow A^{-1} = A^t.$$

Let us write $O(k)$ for the space of $k \times k$ orthogonal matrices. We also write $S(k)$ for the space of symmetric matrices. A quadratic function $H : \mathbb{R}^k \to \mathbb{R}$ is defined by $H(x) = \frac{1}{2}Bx \cdot x$ with $B \in S(k)$. This function induces a gradient vector field which is simply $\nabla H(x) = Bx$. By an *ellipsoid* we mean a set E of the form

$$E = \{x : H(x) \leq 1\}$$

where $H(x) = \frac{1}{2}Bx \cdot x$ with $B > 0$. By Spectral Theory we have

Proposition 2.1. *Let H_1 and H_2 be two quadratic functions associated with the symmetric matrices B_1 and B_2. Then there exists $A \in O(k)$ such that $H_1 \circ A = H_2$ if and only if B_1 and B_2 have the same spectrum.*

As a consequence, if $H(x) = \frac{1}{2}Bx \cdot x$ and B has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then there exists $A \in O(k)$ such that $H(Ax) = \frac{1}{2}\sum_{j=1}^k \lambda_j x_j^2$. In particular, if $B > 0$, then λ_j's are positive and we may define *radii* $\mathbf{R}(H) = (R_1(H), \ldots, R_k(H))$ by $R_i^2 = R_i^2(H) = \frac{2}{\lambda_j}$ so that $0 < R_1(H) \leq R_2(H) \leq \cdots \leq R_k(H)$ and

$$H(A(x)) = \sum_{j=1}^k \frac{x_j^2}{R_j^2}.$$

If E is the corresponding ellipsoid:

$$E = \{x : H(x) \leq 1\},$$

then we write

$$\mathbf{R}(E) = (R_1(E), \ldots, R_k(E))$$

for $\mathbf{R}(H)$ and refer to them as the radii of E. We now rephrase Proposition 2.1 as

Corollary 2.2. *Let E_1 and E_2 be two ellipsoids. Then there exists $A \in O(k)$ such that $A(E_1) = E_2$ if and only if $\mathbf{R}(E_1) = \mathbf{R}(E_2)$.*

We next discuss the monotonicity of $\mathbf{R}$.

Proposition 2.3. (i) *Let H_1 and H_2 be two quadratic functions with eigenvalues $\boldsymbol{\lambda}^1$ and $\boldsymbol{\lambda}^2$. Then $H_1 \circ A \leq H_2$ for some $A \in O(k)$ if and only if $\boldsymbol{\lambda}^1 \leq \boldsymbol{\lambda}^2$.*

(ii) *Let E_1 and E_2 be two ellipsoids. Then $A(E_2) \subseteq E_1$ for some $A \in O(k)$ if and only if $\mathbf{R}(E_2) \leq \mathbf{R}(E_1)$.*

Proof. We note that (i) implies (ii) because if $E_r = \{x : H_r(x) \leq 1\}$ for $r = 1$ and 2, then

$$E_2 \subseteq A^{-1}E_1 \Leftrightarrow H_1 \circ A \leq H_2.$$

The proof of (i) is an immediate consequence of Courant–Hilbert Minimax Principle. □

Lemma 2.4 (Courant–Hilbert). *Let $B \in S(k)$ with eigenvalues $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$. Then*

$$\mu_j = \inf_{\dim V = j} \sup_{x \in V - \{0\}} \frac{Bx \cdot x}{|x|^2}, \tag{2.1}$$

$$\mu_j = \sup_{\dim V = j-1} \inf_{x \in V^{\perp} - \{0\}} \frac{Bx \cdot x}{|x|^2}, \tag{2.2}$$

where V denotes a linear subspace of $\mathbb{R}^k$.

Proof. Let us write X for the right-hand side of (2.1). Let $u_1, u_2, \ldots, u_k$ be an orthonormal basis with $Bu_j = \mu_j u_j$, $j = 1, \ldots, k$. Note that

$$\sup\left\{\frac{Bx \cdot x}{|x|^2} : x \in \operatorname{span}\{u_1, \ldots, u_j\}, x \neq 0\right\} = \sup_{c_1, \ldots, c_j} \frac{\sum_1^j \mu_l c_l^2}{\sum_1^j c_l^2} \leq \mu_j,$$

proving $X \leq \mu_j$. For $X \geq \mu_j$, pick a linear subspace V of dimension j and choose non-zero $x \in V$ such that $x \perp u_1, \ldots, u_{j-1}$. Such x exists because $\dim V = j$ and we are imposing $j - 1$ many conditions. Since we can write $x = \sum_{l=j}^k c_l u_l$, we have

$$\frac{Bx \cdot x}{|x|^2} = \frac{\sum_j^k \mu_l c_l^2}{\sum_j^k c_l^2} \geq \mu_j. \tag{2.3}$$

As a result, $X \geq \mu_j$ and this completes the proof of (2.1).

As for (2.2), note that if $x \perp u_1, \ldots, u_{j-1}$, $x \neq 0$, then $\frac{Bx \cdot x}{|x|^2} \geq \mu_j$ by (2.3). Hence, if Y denotes the right-hand side of (2.2), then $Y \geq \mu_j$ by choosing $V = \operatorname{span}\{u_1, \ldots, u_{j-1}\}$. For $\mu_j \geq Y$, let V be any linear space of dimension $j-1$ and pick a non-zero $x \in \operatorname{span}\{u_1, \ldots, u_j\} \cap V^{\perp}$.

For such a vector x we have $x = \sum_1^j c_l u_l$, $(c_1, \dots, c_j) \neq 0$, and $\frac{Bx \cdot x}{|x|^2} \leq \mu_j$. This implies that $\mu_j \geq Y$. □

We would like to develop a theory similar to what we have seen in this section but now for the bilinear form $\bar{\omega}$. The following table summarizes our main results:

	Symmetric	Antisymmetric
Form	$a \cdot b$	$\bar{\omega}(a,b) = Ja \cdot b$
Invariant matrix	$A \in O(k) : A^{-1} = A^t$	$T \in Sp(n) : T^{-1} = -JT^tJ$
Vector field	$\nabla H(x) = Bx,\ B \in S(k)$	$J\nabla H(x) = JBx;\ JB \in \mathrm{Ham}(n)$
Spectral theorem	Proposition 2.1	Weirstrass Theorem (Th. 2.6)
Monotonicity	Courant–Hilbert Minimax (Lemma 2.4)	Theorem 2.11, Lemma 2.12

We say a matrix T is *symplectic* if $\bar{\omega}(Ta, Tb) = \bar{\omega}(a, b)$. Equivalently $T^tJT = J$ or $T^{-1} = -JT^tJ$. The space of $2n \times 2n$ symplectic matrices is denoted by $Sp(n)$. We say a matrix C is Hamiltonian if $C = JB$ for a symmetric matrix B. The space of $2n \times 2n$ Hamiltonian matrices is denoted by $\mathrm{Ham}(n)$. We have

$$C \in \mathrm{Ham}(n) \Leftrightarrow JB + B^tJ = 0 \Leftrightarrow B^t = JBJ.$$

We note that if $H(x) = \frac{1}{2}Bx \cdot x$ with $B \in S(2n)$, then $J\nabla H(x) = JBx$ with $JB \in \mathrm{Ham}(n)$. We continue with some elementary properties of symplectic and Hamiltonian matrices.

Proposition 2.5. (i) *If $C_1, C_2 \in Ham(n)$ and $r \in \mathbb{R}$, then $C_1 + C_2, [C_1, C_2] = C_1C_2 - C_2C_1, C_1^t, rC_1 \in Ham(n)$.*

(ii) *If $T_1, T_2 \in Sp(n)$, then $T_1^{-1}, T_1^t, T_1T_2 \in Sp(n)$.*

(iii) *If $Z(t_0, t)$ denotes the fundamental solution of $\dot{x} = JB(t)x$ with $B : [t_0, \infty) \to S(2n)$ a C^1-function, then $Z(t_0, t) \in Sp(n)$ for every $t \geq t_0$.*

(iv) *If $e^{tC} \in Sp(n)$ for every t, then $C \in Ham(n)$.*

(v) *If $B > 0$, then all eigenvalues of $C = JB$ are purely imaginary.*

(vi) *If $C \in Ham(n)$, and $p_C(\lambda) = \det(\lambda I - C)$, then $p_C(\lambda) = p_C(-\lambda)$.*

(vii) *If $T \in Sp(n)$, then $p_T(\lambda^{-1}) = \lambda^{-2n} p_T(\lambda)$.*

Proof. (i) and (ii) are straightforward.

For (iii), set $X(t) = Z(t_0, t)^t J Z(t_0, t)$. Note that $X(t_0) = J$. We wish to show that $X(t) = J$ for $t \geq t_0$. This is achieved if we can show that $\dot{X} \equiv 0$. Indeed

$$\dot{X} = \dot{Z}^t JZ + Z^t J\dot{Z} = -Z^t BJJZ + Z^t JJBZ = 0.$$

For (iv), assume that $e^{tC} \in Sp(n)$. As a result, $e^{tC^t} J e^{tC} = J$. Differentiating this with respect to t and evaluating at $t = 0$ yields $C^tJ + JC = 0$.

For (v), assume that $C = JB$ with $B > 0$ and let $\lambda_1 + i\lambda_2$ be an eigenvalue of C associated with the (non-zero) eigenvector $a + ib$. As a result,

$$Ca = \lambda_1 a - \lambda_2 b, \quad Cb = \lambda_2 a + \lambda_1 b. \tag{2.4}$$

Hence

$$\begin{aligned} Ba \cdot b &= -JJBa \cdot b = -JCa \cdot b = -\lambda_1 Ja \cdot b, \\ Bb \cdot a &= -JJBb \cdot a = -JCb \cdot a = -\lambda_1 Jb \cdot a = \lambda_1 Ja \cdot b. \end{aligned}$$

Since B is symmetric, either $\lambda_1 = 0$ or $Ja \cdot b = 0$. To rule out the latter, observe that by (2.4)

$$\begin{aligned} Ba \cdot a &= -JJBa \cdot a = -JCa \cdot a = \lambda_2 Jb \cdot a, \\ Bb \cdot b &= -JJBb \cdot b = -JCb \cdot b = -\lambda_2 Ja \cdot b. \end{aligned}$$

Hence $Ja \cdot b = 0$ implies that $Ba \cdot a = Bb \cdot b = 0$. This in turn implies that $a = b = 0$ because $B > 0$. But this is impossible because $a + ib \neq 0$.

For (vi), take $C \in \mathrm{Ham}(n)$. Then $C = JC^t J$, and

$$\begin{aligned} p_C(\lambda) &= \det(\lambda - C) = \det(\lambda - JC^t J) \\ &= \det(-\lambda JJ - JC^t J) = \det(-\lambda - C^t) \\ &= \det(-\lambda - C) = p_C(-\lambda). \end{aligned}$$

Finally for (vii), pick $T \in Sp(n)$. We have

$$\begin{aligned} p_T(\lambda) &= \det(\lambda - T^t) = \det(\lambda + JT^{-1} J) \\ &= \det(-\lambda JJ + JT^{-1} J) = \det(-\lambda + T^{-1}) \\ &= \det(T^{-1}) \det(-\lambda T + I) = \lambda^{2n} p_T(\lambda^{-1}), \end{aligned}$$

using $\det(T) = 1$. (This will be discussed in Exercise 2.17 below.) □

We are now ready to state Weirstrass Theorem which allows us to diagonalize a Hamiltonian matrix using a symplectic change of variable. Note that if $H(x) = \frac{1}{2} Bx \cdot x$ with $B > 0$, then the corresponding Hamiltonian $C = JB$ has purely imaginary eigenvalues. By Proposition 2.5(vi), these eigenvalues come in conjugate pairs. Let us write these eigenvalues as $\pm i\lambda_j$, $j = 1, \dots, n$, with

$$0 < \lambda_n \leq \lambda_{n-1} \leq \dots \leq \lambda_1.$$

Let us write $\frac{1}{2}\lambda_j = \frac{1}{r_j^2}$ so that $r_j = r_j(H)$ satisfy

$$0 < r_1(H) \leq r_2(H) \leq \dots \leq r_n(H).$$

We also write $\mathbf{r}(H) = (r_1(H), \dots, r_n(H))$ and if E is the corresponding ellipsoid, we write $\mathbf{r}(E)$ for $\mathbf{r}(H)$. The following theorem is due to Weirstrass:

Theorem 2.6. *Let $H(x) = \frac{1}{2}Bx \cdot x$ with $B > 0$. Then there exists $T \in Sp(n)$ such that*

$$H \circ T(x) = \sum_{j=1}^{n} \frac{q_j^2 + p_j^2}{r_j^2}$$

where $x = (q_1, p_1, \dots, q_n, p_n)$ and $r_j = r_j(H)$, $j = 1, \dots, n$.

Corollary 2.7. (i) *If H_1 and H_2 are two positive definite quadratic forms, then $\mathbf{r}(H_1) = \mathbf{r}(H_2)$ if and only if $H_2 = H_1 \circ T$ for some $T \in Sp(n)$.*

(ii) *Let E_1 and E_2 be two ellipsoid. Then $T(E_2) = E_1$ for $T \in Sp(n)$ if and only if $\mathbf{r}(E_1) = \mathbf{r}(E_2)$.*

Proof of Corollary 2.7. Suppose that $H_2 = H_1 \circ T$ for some $T \in Sp(n)$. If $H_2(x) = \frac{1}{2}B_2x \cdot x$ and $H_1(x) = \frac{1}{2}B_1x \cdot x$, then $B_2 = T^tB_1T$. So, for the corresponding Hamiltonian,

$$C_2 = JB_2 = JT^tB_1T = -JT^tJJB_1T = T^{-1}C_1T.$$

As a result, C_1 and C_2 have the same spectrum which means that $\mathbf{r}(H_2) = \mathbf{r}(H_1)$.

For the converse, assume that $\mathbf{r}(H_1) = \mathbf{r}(H_2) = (r_1, \dots, r_n)$. By Weirstrass Theorem, we can find T_1 and $T_2 \in Sp(n)$ such that $H_l \circ T_l(x) = \sum_1^n \frac{q_j^2+p_j^2}{r_j^2}$, $l = 1, 2$. As a result, $H_2 \circ T_2 = H_1 \circ T_1$, or $H_2 = H_1 \circ T$ with $T = T_1 \circ T_2^{-1}$. □

Example 2.8. Let $n = 1$ and $H(q_1, p_1) = \frac{q_1^2}{R_1^2} + \frac{p_1^2}{R_2^2}$ so that $\mathbf{R}(H) = (R_1, R_2)$. Here $H(x) = \frac{1}{2}Bx \cdot x$ with $B = \begin{bmatrix} \frac{2}{R_1^2} & 0 \\ 0 & \frac{2}{R_2^2} \end{bmatrix}$. We have $C = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} B = \begin{bmatrix} 0 & \frac{2}{R_2^2} \\ -\frac{2}{R_1^2} & 0 \end{bmatrix}$. The matrix C has eigenvalues $\pm i\frac{2}{R_1R_2}$. Hence $\mathbf{r}(H) = (r_1(H))$ with $r_1(H) = \sqrt{R_1R_2}$ and there exists $T \in Sp(n)$ such that $H \circ T(q_1, p_1) = \frac{q_1^2+p_1^2}{R_1R_2}$. □

It remains to establish Theorem 2.6. As a preparation, let us develop some understanding about *symplectic vector spaces.* A symplective vector space (V, ω) is a pair of finite dimensional real vector space V and a bilinear form $\omega : V \times V \to \mathbb{R}$ which is antisymmetric and non-degenerate. That is, $\omega(a, b) = -\omega(b, a)$ for all l$a, b \in V$, and that $\forall a \in V$ with $a \neq 0$, $\exists b \in V$ such that $\omega(a, b) \neq 0$. The non-degeneracy is equivalent to saying that the transformation $a \mapsto \omega(a, \cdot)$ is a linear isomorphism between V and its dual V^*. Clearly $(\mathbb{R}^{2n}, \bar{\omega})$ is an example of a symplectic vector space. Given a symplectic (V, ω), then we say a and b are *ω-orthogonal* and write $a \amalg b$ if $\omega(a, b) = 0$. If W is a linear subspace of V, then

$$W^{\amalg} = \{a \in V : a \amalg W\}.$$

Proposition 2.9. (i) $\dim W + \dim W^{\rm II} = \dim V$.

(ii) (W, ω) *is symplectic iff* $W \oplus W^{\rm II} = V$.

(iii) *If* W *is a symplectic subspace, then* $W^{\rm II}$ *is also symplectic.*

Proof. (i) Assume that $\dim V = n$ and $\dim W = m$. Choose a basis $\{a_1, \dots, a_m\}$ for W. Then by non-degeneracy $a_1^*, \dots, a_m^*$ are independent where $a_j^*(b) = \omega(a_j, b)$. Since $W^{\rm II} = \{a : a_j^*(a) = 0 \text{ for } j = 1, \dots, m\}$, with a_j^* independent, we have that $\dim W^{\rm II} = n - m$.

(ii) By definition, (W, ω) is symplectic iff $W \cap W^{\rm II} = \{0\}$. Since $\dim W + \dim W^{\rm II} = \dim V$, we have that $W \oplus W^{\rm II} = V$. (iii) Evidently $W \subseteq (W^{\rm II})^{\rm II}$. Since $\dim W + \dim W^{\rm II} = \dim W^{\rm II} + \dim (W^{\rm II})^{\rm II}$, we deduce that $W = (W^{\rm II})^{\rm II}$. If W is symplectic, then $V = W \oplus W^{\rm II} = (W^{\rm II})^{\rm II} \oplus W^{\rm II}$, which implies that in fact $W^{\rm II}$ is symplectic. □

The following elementary fact is the linear version of the celebrated Darboux Theorem.

Proposition 2.10. *Let* (V, ω) *be a symplectic vector space. Then* $\dim V = 2n$ *is even and there exists a basis* $e_1, \dots, e_n, f_1, \dots, f_n$ *such that* $\omega(e_i, e_j) = \omega(f_i, f_j) = 0$ *and* $\omega(f_i, e_j) = \delta_{ij}$. *Equivalently, if* $x = (q_1, p_1, \dots, q_n, p_n)$, $y = (q_1', p_1', \dots, q_n', p_n')$, $a = \sum_1^n q_j e_j + p_j f_j$, $a' = \sum_1^n q_j' e_j + p_j' f_j$, *then* $\omega(a, a') = \bar\omega(x, y)$.

Proof. Evidently $\dim V \geq 2$. Let e_1 be a non-zero vector of V. Since ω is non-degenerate, we can find $f_1 \in V$ such that $\omega(f_1, e_1) = 1$. Clearly f_1, e_1 are linearly independent. Let $V_1 = \mathrm{span}\{e_1, f_1\}$. If $V = V_1$, then we are done. Otherwise $V = V_1 \oplus V_1^{\rm II}$ with both (V_1, ω), $(V_1^{\rm II}, \omega)$ symplectic. Now we repeat the previous argument to find f_2, e_2 etc. □

With the experience of Proposition 2.10, we are ready to prove Theorem 2.6.

Proof of Theorem 2.6. Write $\lambda_1 = \frac{2}{r_1^2}$ so that $\pm i\lambda_1$ is an eigenvalue for $C = JB$. Then we can find $a_1 + ib_1 \neq 0$ so that $C(a_1 + ib_1) = -i\lambda_1(a_1 + ib_1)$. This means

$$Ca_1 = \lambda_1 b_1, \quad Cb_1 = -\lambda_1 a_1.$$

As a result,

$$H(a_1) = \frac{1}{2} Ba_1 \cdot a_1 = -\frac{1}{2}\bar\omega(Ca_1, a_1) = \frac{\lambda_1}{2}\bar\omega(a_1, b_1),$$
$$H(b_1) = \frac{1}{2} Bb_1 \cdot b_1 = -\frac{1}{2}\bar\omega(Cb_1, b_1) = \frac{\lambda_1}{2}\bar\omega(a_1, b_1).$$

Since $a_1 + ib_1 \neq 0$ and $B > 0$, we deduce that $\bar\omega(a_1, b_1) \neq 0$. We may choose a_1, b_1 so that $\bar\omega(a_1, b_1) = -1$. In this case $H(a_1) = H(b_1) = r_1^{-2}$ and

$$Ba_1 \cdot b_1 = -\bar\omega(Ca_1, b_1) = -\lambda_1 \bar\omega(b_1, b_1) = 0.$$

From all this we learn that

$$H(q_1a_1 + p_1b_1) = q_1^2H(a_1) + p_1^2H(b_1) = \frac{q_1^2 + p_1^2}{r_1^2}.$$

Let $V_1 = \text{span}\{a_1, b_1\}$. Then V_1 and V_1^{II} are symplectic and $\mathbb{R}^{2n} = V_1 \oplus V_1^{\text{II}}$. We now claim

$$CV_1 \subseteq V_1, \quad CV_1^{\text{II}} \subseteq V_1^{\text{II}}, \tag{2.5}$$

$$a \in V_1,\ b \in V_1^{\text{II}} \Rightarrow Ba \cdot b = 0. \tag{2.6}$$

The proof of $CV_1 \subseteq V_1$ is obvious. For the second part of (2.5), we need to show that if $a \in V_1^{\text{II}}$, $b \in V_1$, then $\bar{\omega}(Ca, b) = 0$. Indeed

$$\begin{aligned}\bar{\omega}(Ca, b) &= JCa \cdot b = -Ba \cdot b = -a \cdot Bb = -Ja \cdot Cb \\ &= -\bar{\omega}(a, Cb) = 0,\end{aligned}$$

because $Cb \in V_1$ and $a \in V_1^{\text{II}}$. The proof of (2.6) is similar: if $a \in V_1$, $b \in V_1^{\text{II}}$,

$$Ba \cdot b = \bar{\omega}(Ca, b) = 0,$$

because $Ca \in V_1$.

Note that (2.6) implies that if $a \in V_1$ and $b \in V_1^{\text{II}}$, then

$$H(a + b) = H(a) + H(b).$$

Let us look at the restriction of H to the symplectic vector space $(V_1^{\text{II}}, \bar{\omega})$. A repetition of the above argument yields a pair of vectors $a_2, b_2 \in V_1^{\text{II}}$ with $\bar{\omega}(a_2, b_2) = -1$, $H(a_2) = H(b_2) = r_2^{-2}$ and $Ba_2 \cdot b_2 = 0$.

Continuing this process would yield a basis $(a_1, b_1, a_2, b_2, \ldots, a_n, b_n)$ such that

$$\bar{\omega}(a_i, a_j) = \bar{\omega}(b_i, b_j) = 0,\ \bar{\omega}(a_i, b_j) = -\delta_{ij}, \tag{2.7}$$

$$H(a_j) = H(b_j) = r_j^{-2},\ Ba_j \cdot b_j = 0. \tag{2.8}$$

From (2.7) we learn that $T : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ defined by

$$T(q_1, p_1, \ldots, q_n, p_n) = \sum_1^n q_ja_j + p_jb_j$$

is symplectic. From (2.8) we deduce

$$H(T(x)) = \sum_{1}^{n} \frac{q_j^2 + p_j^{-2}}{r_j^2},$$

as desired. □

We now turn to the question of monotonicity.

Theorem 2.11. (i) *If H_1 and H_2 are two positive definite quadratic forms, then $\mathbf{r}(H_1) \geq \mathbf{r}(H_2)$ if and only if there exists $T \in Sp(n)$ such that $H_1 \circ T \leq H_2$.*

(ii) *Let E_1 and E_2 be two ellipsoids. Then $\mathbf{r}(E_2) \leq \mathbf{r}(E_1)$ if and only if $T(E_2) \subseteq E_1$ for some $T \in Sp(n)$.*

Proof. As before (i) implies (ii) and for (ii), it suffices to find a variational formula for $r_j(H)$ which is non-increasing in H. This formula is given in Lemma 2.12 below. □

Lemma 2.12. *Let H be a positive definite quadratic function of $\mathbb{R}^{2n}$. Then*

$$\frac{1}{2} r_j^2(H) = \inf_{\dim V = 2n+2j} \sup_{[x,y]^t \in V - \{0\}} \frac{\bar{\omega}(x,y)}{H(x) + H(y)}. \tag{2.9}$$

Here V is for linear subspace of $\mathbb{R}^{4n}$.

Remark 2.13. In (2.9) we can replace $\bar{\omega}$ with its positive part $\bar{\omega}^+$ because the left-hand side is positive. After this replacement, we get an expression which is non-increasing in H.

Proof of Lemma 2.12. Recall that $\pm i2r_j^{-2}$ are the eigenvalues of $C = JB$ where $H(x) = \frac{1}{2} Bx \cdot x$. Hence $\pm \frac{i}{2} r_j^2$ are the eigenvalues of C^{-1}. If $a_j + ib_j$ denotes the corresponding eigenvector, then $C^{-1}(a_j + ib_j) = \frac{i}{2} r_j^2 (a_j + ib_j)$. This means

$$C^{-1} a_j = -\frac{r_j^2}{2} b_j, \; C^{-1} b_j = \frac{r_j^2}{2} a_j. \tag{2.10}$$

This suggests looking at the $4n \times 4n$ matrix

$$D = \begin{bmatrix} 0 & C^{-1} \\ -C^{-1} & 0 \end{bmatrix}.$$

From (2.10) we readily deduce

$$D \begin{bmatrix} a_j \\ b_j \end{bmatrix} = \frac{r_j^2}{2} \begin{bmatrix} a_j \\ b_j \end{bmatrix}, \; D \begin{bmatrix} b_j \\ -a_j \end{bmatrix} = \frac{r_j^2}{2} \begin{bmatrix} b_j \\ -a_j \end{bmatrix},$$
$$D \begin{bmatrix} b_j \\ a_j \end{bmatrix} = -\frac{r_j^2}{2} \begin{bmatrix} b_j \\ a_j \end{bmatrix}, \; D \begin{bmatrix} -a_j \\ b_j \end{bmatrix} = -\frac{r_j^2}{2} \begin{bmatrix} -a_j \\ b_j \end{bmatrix}.$$

Note that since $a_j + ib_j \neq 0$, the vectors $\begin{bmatrix} a_j \\ b_j \end{bmatrix}$, $\begin{bmatrix} b_j \\ -a_j \end{bmatrix}$ are linearly independent. Hence $\pm\frac{i}{2}r_j^2$ produces eigenvalue $\pm\frac{r_j^2}{2}$ of multiplicity 2 for D. We would like to apply Courant–Hilbert minimax principle to D, except that D is not symmetric with respect to the dot product of $\mathbb{R}^{4n}$. However if we define an inner product

$$\langle [x,y]^t, [x',y']^t \rangle = Bx \cdot x' + By \cdot y',$$

with corresponding norm

$$\| [x,y]^t \| = 2H(x) + 2H(y),$$

then D is $\langle \cdot, \cdot \rangle$-symmetric. Indeed,

$$\begin{aligned} \left\langle D \begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} x' \\ y' \end{bmatrix} \right\rangle &= \left\langle \begin{bmatrix} C^{-1}y \\ -C^{-1}x \end{bmatrix}, \begin{bmatrix} x' \\ y' \end{bmatrix} \right\rangle \\ &= BC^{-1}y \cdot x' - BC^{-1}x \cdot y' \\ &= Jx' \cdot y + Jx \cdot y', \end{aligned}$$

which is symmetric. Moreover,

$$\left\langle D \begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} x \\ y \end{bmatrix} \right\rangle = 2\bar{\omega}(x,y).$$

We are now in a position to apply Lemma 2.4 to obtain (2.9) because $\frac{1}{2}r_j^2(H)$ is the $2n+2j$-th eigenvalue of D. □

So far we have learned that a symplectic change of variables does not change symplectic radii. The converse to this is "almost" true. In fact what we have said so far is also valid for *anti-symplectic* matrices. We say a matrix T is anti-symplectic if $\bar{\omega}(Tx,Ty) = -\bar{\omega}(x,y)$, or equivalently $T^tJT = -J$. The proof of the following is very similar to the proof of Corollary 2.7 and Theorem 2.6.

Proposition 2.14. (i) *Let H be a positive definite quadratic function. Then there exists an anti-symplectic T such that $H \circ T(x) = \sum_1^n \frac{q_j^2+p_j^2}{r_j^2}$, where $r_j = r_j(H)$.*

(ii) *Let E_1 and E_2 be two ellipsoids. Then $T(E_2) = E_1$ for some anti-symplectic T if and only if $\mathbf{r}(E_2) = \mathbf{r}(E_1)$.*

We are now ready for a converse to Corollary 2.7(ii) and Proposition 2.14(ii), which will be used for the proof of Eliashberg's theorem in Section 6.

Theorem 2.15. *Let T be an invertible $2n \times 2n$ matrix. The following statements are equivalent:*

(i) $r_1(E) = r_1(T(E))$ *for every ellipsoid* E.

(ii) T *is either symplectic or anti-symplectic.*

Proof. We already know that (ii) implies (i) by Corollary 2.7(ii) and Proposition 2.14(ii).

For the converse, assume that T is neither symplectic nor anti-symplectic. Hence the sets

$$G_1 = \{(a, b) : \bar{\omega}(T^t a, T^t b) \neq \bar{\omega}(a, b)\},$$
$$G_2 = \{(a, b) : \bar{\omega}(T^t a, T^t b) \neq -\bar{\omega}(a, b)\}$$

are both non-empty open subsets of $\mathbb{R}^{4n}$. We now argue that $G_1 \cap G_2 \neq \emptyset$. If this were not the case, then $|\bar{\omega}(T^t a, T^t b)| = |\bar{\omega}(a, b)|$ for all a, b. Both sets $V^+ = \{\bar{\omega}(a, b) > 0\}$ and $V^- = \{\bar{\omega}(a, b) < 0\}$ are connected and by continuity, either $\bar{\omega}(T^t a, T^t b) = \bar{\omega}(a, b)$ for all $(a, b) \in V^+$ or $\bar{\omega}(T^t a, T^t b) = -\bar{\omega}(a, b)$ for all $(a, b) \in V^+$. If the former occurs, then by replacing b with $-b$ we learn that the same is true for $(a, b) \in V^-$. Hence $\bar{\omega}(T^t a, T^t b) = \bar{\omega}(a, b)$ everywhere which contradicts our assumption. A similar reasoning applies if $\bar{\omega}(T^t a, T^t b) = -\bar{\omega}(a, b)$ for all $(a, b) \in V^+$. In summary $G_1 \cap G_2 \neq \emptyset$.

We now pick $(a, b) \in G_1 \cap G_2$. We have that $|\bar{\omega}(T^t a, T^t b)| \neq |\bar{\omega}(a, b)|$. By perturbing (a, b) if necessary, we may assume that $\bar{\omega}(a, b) \neq 0$, $\bar{\omega}(T^t a, T^t b) \neq 0$. By switching to T^{-1} if necessary, we may assume that

$$0 < |\bar{\omega}(T^t a, T^t b)| < |\bar{\omega}(a, b)|.$$

After a scaling, we may assume that $\bar{\omega}(a, b) = -1$. Set $\lambda^2 = |\bar{\omega}(T^t a, T^t b)|$ so that $0 < \lambda^2 < 1$. We now construct symplectic bases $\{a_1, b_1, \dots, a_n, b_n\}$ and $\{a'_1, b'_1, \dots, a'_n, b'_n\}$ such that

$$a_1 = a,\ b_1 = b,\ \lambda a'_1 = T^t a,\ \lambda b'_1 = \pm T^t b.$$

Here $\pm$ is chosen so that $\pm\bar{\omega}(T^t a, T^t b) = -\lambda^2$. Now let U_1 and U_2 be two symplectic matrices which map the standard Euclidean basis $\{e_1, f_1, \dots, e_n, f_n\}$ to $\{a_1, b_1, \dots, a_n, b_n\}$ and $\{a'_1, b'_1, \dots, a'_n, b'_n\}$ respectively. We now have that if $\hat{T} = U_2^{-1} T^t U_1$, then

$$\hat{T} e_1 = \lambda e_1,\ \hat{T} f_1 = \pm\lambda f_1.$$

From this we learn that if $|x| \leq 1$, then

$$(\hat{T}^t x \cdot e_1)^2 + (\hat{T}^t x \cdot f_1)^2 = \lambda^2 (x \cdot e_1)^2 + (x \cdot f_1)^2 \leq \lambda^2.$$

Pick $\lambda' \in (\lambda, 1)$ and choose R large enough so that for $|x| \leq 1$,

$$(\hat{T}^t x \cdot e_1)^2 + (\hat{T}^t x \cdot f_1)^2 + \sum_{j=2}^{n} \frac{(\hat{T}^t x \cdot e_j)^2 + (\hat{T}^t x \cdot f_j)^2}{R^2} \leq (\lambda')^2.$$

This means that $\hat{T}^t = U_1^t T (U_2^t)^{-1}$ maps the ball $B = \{x : |x| \leq 1\}$ into the ellipsoid

$$E = \left\{ x : \frac{q_1^2 + p_1^2}{(\lambda')^2} + \sum_{j=2}^{n} \frac{q_j^2 + p_j^2}{R^2} \leq 1 \right\}.$$

Since T satisfies (i) and both U_1 and U_2 are symplectic, $\hat{T}^t$ satisfies (i) as well. But

$$1 = r_1(B) = r_1(\hat{T}^t(B)) \leq r_1(E) = \lambda' < 1$$

which is impossible. This completes the proof of (i) $\Rightarrow$ (ii). □

So far we have only considered ellipsoids $E = \left\{x : \frac{1}{2} Ax \cdot x \leq 1\right\}$ for $A > 0$. Of course we may also consider the degenerate case $A \geq 0$ which means that the corresponding symplectic radii may take values in $(0, \infty]$. By a limiting procedure, we may include such ellipsoids as well. To this end let us define

$$B_R = \{x : |x| \leq R\}, \quad Z_R = \{x : q_1^2 + p_1^2 \leq R^2\}$$

which correspond to ellipsoids of radii $(R, R, \ldots, R)$ and $(R, \infty, \infty, \ldots, \infty)$ respectively. In particular, we may apply Theorem 2.11(ii) to assert that if for some $T \in Sp(n)$, we have $T(B_r) \subseteq Z_R$, then $r \leq R$. This is the linear version of Gromov's non-squeezing theorem. We now slightly improve this and give a direct proof of it.

Theorem 2.16. *Suppose that for some $T \in Sp(n)$ and $z^0 \in \mathbb{R}^{2n}$, $T(B_r) \subseteq z^0 + Z_R$. Then $r \leq R$.*

Proof. Write $z^0 = (q_1^0, p_1^0, \ldots, q_n^0, p_n^0)$ and let $(s_1, t_1, \ldots, s_n, t_n)$ denote the rows of T. By assumption

$$(x \cdot s_1 - q_1^0)^2 + (x \cdot t_1 - p_1^0)^2 \leq R^2$$

for x satisfying $|x| \leq r$. Hence

$$(x \cdot s_1)^2 + (x \cdot t_1)^2 - 2x \cdot (q_1^0 s_1 + p_1^0 t_1) \leq R^2. \tag{2.11}$$

On the other hand, since T^t is symplectic,

$$\bar{\omega}(s_1, t_1) = \bar{\omega}(T^t e_1, T^t f_1) = \bar{\omega}(e_1, f_1) = -1 \tag{2.12}$$

where $\{e_1, f_1, \ldots, e_n, f_n\}$ denote the standard symplectic basis for $\mathbb{R}^{2n}$, i.e., $e_j \cdot x = q_j$ and $f_j \cdot x = p_j$ for $x = (q_1, p_1, \ldots, q_n, p_n)$. From (2.12) we learn that

$$1 = |\bar{\omega}(s_1, t_1)| = |Js_1 \cdot t_1| \leq |s_1||t_1|.$$

So either $|s_1| \geq 1$ or $|t_1| \geq 1$. Both cases can be treated similarly, so let us assume that for example $|t_1| \geq 1$. We then choose $x = \pm r\frac{t_1}{|t_1|}$ in (2.11). We select + or − for x so that $x \cdot (q_1^0 s_1 + p_1^0 t_1) \leq 0$. This would allow us to deduce $r^2 \leq R^2$ from (2.11), and this completes the proof. □

Exercises 2.17.

(i) (1) Let V be a vector space with $\dim V = 2n$. Then a 2-form $\omega : V \times V \to \mathbb{R}$ is non-degenerate if and only if $\omega^n = \underbrace{\omega \wedge \cdots \wedge \omega}_{n \text{ times}} \neq 0$.

(ii) Use part (1) to deduce that if $T \in Sp(n)$, then $\det T = 1$.

(iii) Recall that an invertible matrix T can be written as $T = PO$ with $p > 0$ and O orthogonal, and that this decomposition is unique (polar decomposition). Show that if $T \in Sp(n)$, then P and $O \in Sp(n)$.

(iv) Prove Proposition 2.14(i).

(v) Show that if $T \in Sp(n) \cap O(2n)$, then $T = \begin{bmatrix} x & -y \\ y & x \end{bmatrix}$ with x, y two $n \times n$ matrices such that $x + iy$ is a unitary matrix.

(vi) Let V be a vector space with $\dim V = 2n + 1$. Assume β is an antisymmetric 2-form on V with
$$\dim\{v \in V : \beta(v, a) = 0 \text{ for all } a \in V) = 1.$$
Then there exists a basis $\{e_1, \ldots, e_n, f_1, \ldots, f_n, \bar{a}\}$ such that $\beta(e_i, e_j) = \beta(f_i, f_j) = 0$, $\beta(f_j, e_i) = \delta_{ij}$, and $\beta(\bar{a}, f_j) = \beta(\bar{a}, e_j) = 0$.

(vii) An invertible T maps the flows of $\frac{dx}{dt} = Ax$ to the flow of $\frac{dx}{dt} = Bx$ iff $B = TAT^{-1}$.

(viii) If $T \in Sp(n)$ and $A \in Ham(n)$, then $T^{-1}AT \in Ham(n)$.

□

3 Symplectic Manifolds and Darboux's Theorem

Before discussing symplectic manifolds, let us review some useful facts about our basic example $(\mathbb{R}^{2n}, \bar{\omega})$ with $\bar{\omega}(a,b) = Ja \cdot b$. We write $Sp(\mathbb{R}^{2n})$ for the space of differentiable functions φ such that $\varphi^*\bar{\omega} = \omega$. This means

$$\varphi^*\bar{\omega}(a,b) = \bar{\omega}(\varphi'(x)a, \varphi'(x)b) = \bar{\omega}(a,b),$$

for every $a, b, x \in \mathbb{R}^{2n}$. Here $\varphi'(x)$ denotes the derivative of φ. A function $\varphi \in Sp(\mathbb{R}^{2n})$ is called *symplectic*. Note that $\varphi \in Sp(\mathbb{R}^{2n})$ iff $\varphi'(x) \in Sp(n)$ for every x. Hence for a symplectic transformation φ, we have

$$\varphi'(x)^t J \varphi'(x) = J. \tag{3.1}$$

Evidently $\bar{\omega} = \sum_{i=1}^n dp_i \wedge dq_i = d\bar{\lambda}$ where $\bar{\lambda} = \sum_1^n p_i dq_i = p \cdot dq$. Let us define

$$A(\gamma) = \int_\gamma \bar{\lambda}. \tag{3.2}$$

Clearly $\varphi \in Sp(\mathbb{R}^{2n})$ iff $d(p^*\bar{\lambda} - \bar{\lambda}) = 0$. Hence $\varphi \in Sp(\mathbb{R}^{2n})$ is equivalent to saying

$$A(\varphi \circ \gamma) = A(\gamma) \tag{3.3}$$

for every closed curve γ. It is worth mentioning that if γ is parametrized by $\theta \mapsto x(\theta)$, $\theta \in [0,1]$, then

$$\begin{aligned} A(\gamma) &= \int_0^1 p \cdot \dot{q} d\theta = \frac{1}{2}\int_0^1 (p \cdot \dot{q} - q \cdot \dot{p}) d\theta \\ &= -\frac{1}{2}\int_0^1 \bar{\omega}(\dot{x}, x) d\theta = \frac{1}{2}\int_0^1 (Jx \cdot \dot{x}) d\theta. \end{aligned} \tag{3.4}$$

Given a scalar-valued (0-form) function H, we may use non-degeneracy of $\bar{\omega}$ to define a vector-field X_H such that

$$\bar{\omega}(X_H(x), a) = -dH(x)a = -\nabla H(x) \cdot a,$$

which means that $X_H = J\nabla H$. We write $\phi_t = \phi_t^H$ for the corresponding flow:

$$\begin{cases} \dfrac{d}{dt}\phi_t(x) = X_H(\phi_t(x)), \\ \phi_0(x) = x. \end{cases} \tag{3.5}$$

Our interest in symplectic transformation stems from two important examples. As the first example, recall that $\phi_t \in Sp(\mathbb{R}^{2n})$ if ϕ_t is a Hamiltonian flow. We have seen this in section 1 and will be proved later in this section for general symplectic manifolds. The second example has to do with the fact that symplectic change of coordinates preserve Hamiltonian structure. We have seen this in the linear case in the proof of Corollary 2.7.

Proposition 3.1. *Let $\varphi \in Sp(\mathbb{R}^{2n})$ be a diffeomorphism and assume that ϕ_t is the flow of X_H. Then $\psi_t = \varphi^{-1} \circ \phi_t \circ \varphi$ is the flow of $X_{H\circ\varphi}$.*

Proof. From $\psi_t = \varphi^{-1} \circ \phi_t \circ \varphi$, we learn

$$\begin{aligned}\frac{d}{dt}\psi_t\Big|_{t=0} &= (\varphi^{-1})' \circ \varphi\ X_H \circ \varphi = (\varphi')^{-1}\ X_H \circ \varphi \\ &= -(\varphi')^{-1}\ J\nabla H \circ \varphi = J(\varphi')^t\ JJ\nabla H \circ \varphi \\ &= -J(\varphi')^t\ \nabla H \circ \varphi = -J\nabla(H \circ \varphi).\end{aligned}$$

□

A pair (M, ω) is called a *symplectic manifold* if M is an even dimensional manifold and ω is a closed non-degenerate 2-form on M. This implies that for each $x \in M$, the pair (T_xM, ω_x) is a symplectic vector space. Also, by Exercise 2.17(i) we know that if $dim(M) = 2n$, then the form ω^n is a volume form. Hence M is an orientable manifold. In fact if M is a compact symplectic manifold without boundary, then ω is never exact. This is because if $\omega = d\lambda$, then $\Omega := \omega^n = d(\lambda \wedge \omega^{n-1})$. But by Stokes' theorem $\int_M \Omega = \int_M d(\lambda \wedge \omega^{n-1}) = 0$, which contradicts the non-degeneracy of Ω. Note however that $(\mathbb{R}^{2n}, \bar{\omega})$ is an example of a non-compact symplectic manifold with $\bar{\omega} = d\bar{\lambda}$. This example has a natural generalization: Every cotangent bundle T^*N can be equipped with a non-degenerate $\omega = d\lambda$ with λ a 1-form. To see this, let us assume that N is a n-dimensional C^2 manifold and let $\pi : T^*N \to N$ denote the projection onto the base point, i.e., $\pi(q, p) = p$ with $q \in N$ and $p \in T_q^*N$.

The derivative $d\pi : T(T^*N) \to TN$ can be written as $d\pi(q, p)b = (q, v)$ with $v \in T_qN$. In words, v is the T_qN-component of b. Of course $p \in T_q^*N$ or $p : T_qN \to \mathbb{R}$. We now define the 1-form λ on $M = T^*N$ by

$$\lambda_{(q,p)}(b) = p(v),$$

and $\omega = d\lambda$. The pair (M, ω) is a symplectic manifold. We still need to verify the non-degeneracy of ω. This would follow once we find a natural Darboux chart for M. More precisely, we find an open cover for N and a family of transformations $\bar{h} : T^*U \to \mathbb{R}^{2n}$ such that $\bar{h}^*\bar{\lambda} = \lambda$ and $\bar{h}^*\bar{\omega} = \omega$. To construct $\bar{h}$, let us take a local description of N by taking open subset U of N and a diffeomorphism $h : U \to h(U) \subseteq \mathbb{R}^n$. In other words (U, h) is a chart of N. This induces a transformation $h : TU \to \mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$. Indeed, we may regard T_qN as the set of equivalent classes $[\gamma]_q$ of curves where two curves $\gamma_1, \gamma_2 : [-\delta, \delta] \to U$, with $\gamma_1(0) = \gamma_2(0) = q$ are equivalent if $(h \circ \gamma_1)'(0) = (h \circ \gamma_2)'(0)$. We now define $dh(q, [\gamma]_q) = (h(q), (h \circ \gamma)'(0))$ and $(dh)^{-1}(a, v) = (h^{-1}(a), [\gamma_{a,v}])$ with $\gamma_{a,v}(\theta) = h^{-1}(a + \theta v)$. We next define a natural $\bar{h} : T^*U \to \mathbb{R}^{2n} = T^*\mathbb{R}^n$. For this, take the standard basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$ and write $\hat{e}_j(q)$ for the equivalent class of $\gamma^j(\theta) = h^{-1}(h(q) + \theta e_j)$. Certainly $\{\hat{e}_1(q), \dots, \hat{e}_n(q)\}$ defines a basis for T_qN. We now define a basis for T_q^*N by taking dual vectors: $\bar{e}_1, \dots, \bar{e}_n$ are defined so that $\bar{e}_i(q)\left(\sum_{j=1}^n v_j \hat{e}_j(q)\right) = v_j$. We finally define $\bar{h}$ by

$$\bar{h}\left(q, \sum_{j=1}^n p_j \bar{e}_j(q)\right) = (h(q), (p_1, \dots, p_n)).$$

We now assert that $\bar{h}^*\bar{\lambda} = \lambda$. To see this, define $\hat{\pi} : T^*\mathbb{R}^n \to \mathbb{R}^n$ by $\pi^*(a, b) = a$ and

$$d\hat{\pi} : T(T^*\mathbb{R}^n) \to T\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$$

is simply given by

$$d\hat{\pi}(a, b)(\alpha, \beta) = (a, \beta).$$

Clearly $\bar{\lambda}_{(a,b)}(\alpha, \beta) = b \cdot \alpha$. Since the following diagram commutes

$$\begin{array}{ccc} T^*U & \xrightarrow{\pi} & U \\ {\scriptstyle \bar{h}}\downarrow & & \downarrow{\scriptstyle h} \\ \mathbb{R}^{2n} & \xrightarrow[\hat{\pi}]{} & \mathbb{R}^n \end{array}$$

we also have that their derivatives

$$\begin{array}{ccc} T(T^*U) & \xrightarrow{d\pi} & TU \\ {\scriptstyle d\bar{h}}\downarrow & & \downarrow{\scriptstyle dh} \\ T\mathbb{R}^{2n} & \xrightarrow[d\hat{\pi}]{} & \mathbb{R}^{2n} \end{array}$$

commute. From this, we immediately deduce that $\bar{h}^*\bar{\lambda} = \lambda$. □

Example 3.2. The sphere S^{2n} with $n > 1$ is not symplectic because any closed 2-form is exact. $\square$

Example 3.3. Any orientable 2-dimensional manifold is symplectic where ω is chosen to be any volume form. $\square$

A differentiable map $f : (M_1, \omega^1) \to (M_2, \omega^2)$ between two symplectic manifolds is called *symplectic* if $f^*\omega^2 = \omega^1$. This means

$$\omega^2_{f(x)}(df(x)a, df(x)b) = \omega^1_x(a, b) \tag{3.6}$$

for $x \in M_1$ and $a, b \in T_xM_1$. We write $Sp(M_1, M_2)$ for the space of symplectic transformations. When $M_1 = M_2 = M$ and $\omega^1 = \omega^2 = \omega$, we simply write $Sp(M)$ for $Sp(M^1, M^2)$. We note that if $f \in Sp(M^1, M^2)$, then $df(x)$ is injective by (3.6) and non-degeneracy of ω^1_x. Hence, if such f exists, then $\dim M^1 \leq \dim M^2$.

Using the symplectic structure on a manifold, we can define an isomorphism between vector fields and 1-forms given by $X \mapsto \omega(X, \cdot)$. In particular, for any $H : M \to \mathbb{R}$, we can find a unique vector field X_H such that

$$(i_{X_H}\omega)(x) = \omega(X_H(x), \cdot) = -dH(x).$$

The flow associated with the vector field X_H will be denoted by ϕ^H_t.

Proposition 3.4. *Let (M, ω) be a symplectic manifold.*

(i) $\phi^H_t \in Sp(M)$ *for every H.*

(ii) *If $\varphi \in Sp(M)$, then $\varphi_* X_H = X_{H \circ \varphi}$ where $\varphi_* X = (d\varphi)^{-1} X \circ \varphi$.*

As a preparation let us state a useful result of Cartan:

Theorem 3.5. *Let X be a vector field with flow ϕ_t and let α be a k-form. Then*

$$\frac{d}{dt}\phi^*_t\alpha = \mathcal{L}_X\phi^*_t\alpha = \phi^*_t\mathcal{L}_X\alpha \tag{3.7}$$

with $\mathcal{L}_X = i_X \circ d + d \circ i_X$.

Proof of Proposition 3.4. (i) We need to show that $\frac{d}{dt}\phi^*_t\omega = 0$. By Theorem 3.5, it suffices to verify that $\mathcal{L}_X\omega = 0$. But

$$\mathcal{L}_{X_H}\omega = i_{X_H}(d\omega) + di_{X_H}\omega = -ddH = 0.$$

(ii) We have

$$\begin{aligned} i_{\varphi_* X_H}\omega &= i_{\varphi_* X_H}\varphi^*\omega = \varphi^* i_{X_H}(\omega) = -\varphi^* dH \\ &= -d(H\circ\varphi). \end{aligned}$$

□

We now turn to the proof of (3.7).

Proof of Theorem 3.5. Let us define

$$\mathcal{L}\beta = \lim_{h\to 0}\frac{1}{h}(\phi_h^*\beta - \beta)$$

whenever the limit exists. Since

$$(\phi_{t+h}^* - \phi_t^*)\alpha = \phi_t^*(\phi_h^*\alpha - \alpha) = \phi_h^*(\phi_t^*\alpha) - \phi_t^*\alpha,$$

it suffices to show that $\mathcal{L} = \mathcal{L}_X$. Let us study some properties of $\mathcal{L}$. From $\phi_t^*(\alpha\wedge\beta) = \phi_t^*\alpha\wedge\phi_t^*\beta$, we learn

$$\phi_t^*(\alpha\wedge\beta) - \alpha\wedge\beta = (\phi_t^*\alpha - \alpha)\wedge\phi_t^*\beta + \alpha\wedge(\phi_t^*\beta - \beta).$$

From this we deduce

$$\mathcal{L}(\alpha\wedge\beta) = \alpha\wedge\mathcal{L}\beta + \mathcal{L}\alpha\wedge\beta. \tag{3.8}$$

From $\phi_t^*\circ d = d\circ\phi_t^*$, we deduce

$$\mathcal{L}\circ d = d\circ\mathcal{L}. \tag{3.9}$$

We can readily show that $\mathcal{L}_X$ satisfy (3.8) and (3.9) as well. Since locally every form can be built from 0-th forms using the operations $\wedge$ and d, we only need to check that $\mathcal{L} = \mathcal{L}_X$ on 0-forms. That is, if $f : M \to \mathbb{R}$, then $\mathcal{L}f = i_X\circ df = df(X)$. This is trivially verified because $\phi_t(x) = x + tX(x) + o(t)$. □

As our next result, we show that a symplectic manifold is locally equivalent to $(\mathbb{R}^{2n}, \bar{\omega})$.

Theorem 3.6. *(Darboux) Let (M, ω) be a symplectic manifold of dimension $2n$ and take $x^0 \in M$. Then there exists an open set $U \subseteq \mathbb{R}^{2n}$ with $0 \in U$ and a diffeomorphism $\varphi : U \to M$ such that $\varphi(0) = x^0$ and $\varphi^*\omega = \bar{\omega}$.*

Proof. Since M is locally diffeomorphic to an open subset of $\mathbb{R}^{2n}$, we may assume that $M = U \subseteq \mathbb{R}^{2n}$ and $x^0 = 0 \in U$. In view of Proposition 2.10, we may also assume that

$\omega_0 = \bar{\omega}$. Our goal is finding a diffeomorphism $\varphi : U \to U$ with $\varphi(0) = 0$ and $\varphi^*\omega = \bar{\omega} = \omega_0$. For this we may have to go to a smaller neighborhood of 0. Following an idea of Moser, it is more convenient to consider a family of forms $\omega(t) = \bar{\omega} + t(\omega - \bar{\omega})$, $t \in [0, 1]$ connecting $\bar{\omega}$ to ω, and search for a family of diffeomorphisms φ_t such that

$$\varphi_0 = id, \ \varphi_t^*\omega(t) = \bar{\omega} \text{ for } t \in [0, 1]. \tag{3.10}$$

In order to find φ_t, let us search for a time-dependent vector field X_t for which the associated flow satisfies (3.10). Using Cartan's formula,

$$\begin{aligned} 0 &= \frac{d}{dt}(\varphi_t^*\omega(t)) = \varphi_t^*\mathcal{L}_{X_t}\omega(t) + \varphi_t^*(\omega - \bar{\omega}) \\ &= \varphi_t^*[d \circ i_{X_t}\omega(t) + \omega - \bar{\omega}]. \end{aligned}$$

Hence (3.10) would follow if we can find a time-dependent vector field X_t such that

$$d(i_{X_t}\omega(t)) = \bar{\omega} - \omega.$$

Without loss of generality, we may assume that U is simply connected. By Poincaré's lemma, there exists a 1-form α such that $d\alpha = \bar{\omega} - \omega$. Without loss of generality, we may assume that $\alpha_0 = 0$. Now for (3.10), we need to find X_t with $i_{X_t}\omega(t) = \alpha$. The form $\omega(t)_0 = \bar{\omega} + t(\omega_0 - \bar{\omega}) = \bar{\omega}$ is non-degenerate for every $t \in [0, 1]$. Hence in a small neighborhood U of 0, $\omega(t)$ is non-degenerate for every $t \in [0, 1]$. This allows us to solve $i_{X_t}\omega(t) = \alpha$ uniquely for X_t in U. The condition $\alpha_0 = 0$ means that $X_t(0) = 0$ for every $t \in [0, 1]$. In a small neighborhood of the origin of flow φ_t, $t \in [0, 1]$ exists with $\varphi_t(0) = 0$. We are done. □

Remark 3.7. If we write $\omega = F \cdot dx$ and $\bar{\omega} = \bar{F} \cdot dx$ near the origin, then X_t constructed in the proof has the form

$$X_t = \mathcal{C}(tF + (1 - t)\bar{F})^{-1}(\bar{F} - F)$$

where the matrix $\mathcal{C}(A) = \left[\frac{\partial A^i}{\partial x_j} - \frac{\partial A^j}{\partial x_i}\right]$ with $A = (A^1, \dots, A^k)$. □

As an immediate consequence of Darboux's theorem we learn that any symplectic manifold M of dimension $2n$ has a covering $\{U_j\}$ with $\varphi_j : U_j \to \mathbb{R}^{2n}$, such that

$$\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$$

is symplectic for every i and j. The family $\{(U_j, \varphi_j)\}$ is an example of a symplectic coordinate that always exists by Darboux's theorem.

The idea of the proof of Theorem 3.6 is due to Moser and is known as *deformation method*. This method can be used to show that the total volume is the only invariant of volume preserving diffeomorphisms. More precisely, if

$$\mathcal{V}(M) = \{\alpha : \alpha \text{ is a volume form on } M\}$$

with M compact with no boundary, and

$$c : \mathcal{V}(M) \to \mathbb{R}$$

is a function such that $c(\alpha) = c(\beta)$ whenever $\varphi^*\alpha = \beta$ for some diffeomorphism φ, then c must be a function of $\int_M \alpha$. Here is Moser's theorem:

Theorem 3.7. *Let M be a connected oriented compact manifold with no boundary. Assume that α and β are two volume forms with $\int_M \alpha = \int_M \beta$. Then there exists a diffeomorphism φ such that $\varphi^*\alpha = \beta$.*

Proof. Define $\omega(t) = t\alpha + (1-t)\beta = \beta + t(\alpha - \beta)$. Since α and β are volume forms, locally $\alpha = adx_1 \wedge \cdots \wedge dx_k$ and $\beta = bdx_1 \cdots \wedge \dots dx_k$ with a and b non-zero and continuous. Since $\int_M \alpha = \int_M \beta$, a and b must have the same sign. Hence $ta + (1-t)b$ is never zero, and as a result, $\omega(t)$ is never degenerate. We search for a vector field X_t with flow φ_t so that $\varphi_0 = id$ and $\frac{d}{dt}\varphi_t^*\omega(t) = 0$. If such X_t can be found, then $\varphi_1^*\alpha = \varphi_1^*\omega(1) = \varphi_0^*\omega(0) = \beta$ would be the desired conclusion. By Cartan's formula, it suffices to have

$$d(i_{X_t}\omega(t)) = \beta - \alpha.$$

But $\int_M(\beta - \alpha) = 0$ implies that $\beta - \alpha = d\lambda$ for some $k-1$-form λ. The existence of X_t with $i_{X_t}\omega(t) = \lambda$ follows from the non-degeneracy of $\omega(t)$. □

Remark 3.8. If $M = \mathbb{R}^k$ and $\alpha = \rho^1(x)e^{-V(x)}dx_1 \wedge \cdots \wedge dx_k$, $\beta = \rho^2(x)e^{-V(x)}dx_1 \wedge \cdots \wedge dx_k$, then the vector field in Moser's proof can be constructed by the following recipe: First find a vector field Y such that

$$\text{div } Y - \nabla V \cdot Y = \rho^2 - \rho^1, \tag{3.11}$$

so that if $\lambda = \sum_1^k (-1)^j Y^j e^{-V} dx_1 \wedge \cdots \wedge \widehat{dx_j} \wedge \cdots \wedge dx_k$, then $d\lambda = \beta - \alpha$. Such a vector field Y exists because $\int(\beta - \alpha) = 0$. We then set

$$X_t = \frac{Y}{t\rho^1 + (1-t)\rho^2}.$$

In fact for (3.11), we may first find a scalar-valued function u such that

$$\Delta u - \nabla V \cdot \nabla u = \rho^2 - \rho^1, \tag{3.12}$$

and then set $Y = \nabla u$. Hence, if ρ^1 and ρ^2 are C^l, then X_t is C^l as well. Again (3.12) has a solution because $\int (\rho^2 - \rho^1)e^{-V} dx = 0$. □

For the sake of completeness, we provide a proof of De Rham's lemma which was used in the proof of Theorem 3.7.

Lemma 3.9. *Let M be a compact connected orientable k-dimensional manifold and let α be a k-form with $\int_M \alpha = 0$. Then α is exact.*

Proof. Let $\{U_1, \dots, U_r\}$ be a finite cover of M with each U_i diffeomorphic to a simply connected subset of $\mathbb{R}^k$. To ease the notation, we assume that $r = 2$. Choose φ_1 and φ_2 with $\varphi_1 + \varphi_2 = 1$, $\varphi_1, \varphi_2 \geq 0$, $\varphi_1, \varphi_2 \in C^1$ and supp $\varphi_i \subseteq U_i$ for $i = 1, 2$. We can readily construct a k-form β such that supp $\beta \subseteq U_1 \cap U_2$ and $\int_M \beta = 1$. Define $\gamma = \varphi_1 \alpha - c\beta$ with $c = \int_M \varphi_1 \alpha$. Then supp $\gamma \subseteq U_1$ and $\int_{U_1} \gamma = 0$. By Poincaré's lemma, there exists a form $\hat{\gamma}$ such that $d\hat{\gamma} = \gamma_1$. Similarly, if $\tau = \varphi_2 \alpha + c\beta$, then supp $\tau \subseteq U_2$ and $\int_{U_2} \tau = \int_K \tau = \int_M (1 - \varphi_1)\alpha + c\beta = \int_M \alpha = 0$. Hence, we can apply Poincaré's lemma to find $\hat{\tau}$ with $d\hat{\tau} = \tau$. We now have

$$d(\hat{\gamma} + \hat{\tau}) = \varphi_1 \alpha - c\beta + \varphi_2 \alpha + c\beta = \alpha.$$

□

Remark 3.10. As a consequence of Theorem 3.8, if (M_1, ω^1) and (M_2, ω^2) are two compact connected symplectic manifolds of dimension 2, then they are isomorphic by a volume preserving diffeomorphism, if they are diffeomorphic and $\int_{M_1} \omega^1 = \int_{M_2} \omega^2$. Indeed if $\varphi : M_1 \to M_2$ is a diffeomorphism, then there exists a diffeomorphism $\psi : M_1 \to M_2$ with $\psi^* \omega_2 = \omega_1$. This follows from Theorem 3.8 with $\alpha = \omega^1$ and $\beta = \varphi^* \omega^2$ two volume forms on $M = M_1$. □

There has been new developments for Moser's theorem. In fact the transformation φ in Theorem 3.8 is by no means unique. However, one may wonder whether or not a "nice" φ exists. More precisely, let (M, g) be a Riemannian manifold. Assume that M is compact, connected with no boundary. Using g we can talk about Riemannian distance. More precisely, let $d(x, y)$ be the length of the geodesic distance between two points x and y. We have a natural volume form Ω that is expressed by $(\det[g_{ij}])^{1/2} dx_1 \wedge \cdots \wedge dx_k$ in local coordinates. Consider two forms $\alpha = a\Omega$ and $\beta = b\Omega$ with $a, b > 0$ and $\int_M \alpha = \int_M \beta = 1$. Set

$$S(\alpha, \beta) = \{f : M \to M \text{ with } f^* \alpha = \beta\}.$$

By Moser's theorem, this set is non-empty. Monge's problem searches for a function $f \in S(\alpha, \beta)$ which minimizes the cost function

$$\mathcal{I}(f) = \int_M c(x, f(x))\beta,$$

with $c(x, y)$ a suitable function of $M \times M$. If we choose $c(x, y) = \frac{1}{2}(d(x, y))^2$, then the minimizer f is a *monotone* function. This was shown by Brenier (1987, 1991) in the Euclidean case. By monotonicity we mean that for every x, y,

$$(f(x) - f(y)) \cdot (x - y) \geq 0.$$

The existence of a unique minimizer in the case of a Riemannian manifold was established by McCann (2001). There is an analogous notion of monotonicity in the Riemannian case that we don't discuss here. Brenier observed that such a minimizer can be used to find a *non-linear polar decomposition*. To explain this, let $f : U \to \mathbb{R}^k$ be an invertible integrable function with $\alpha = (F^{-1})^*\beta$ where $\beta = dx_1 \wedge \cdots \wedge dx_k$ and α is a volume form. According to Brenier's theorem, there exists a convex function ψ such that $(\nabla\psi)^*\alpha = \beta$. If we write $\rho = (\nabla\psi)^{-1} \circ F$, then $\rho^*\beta = F^*(\nabla\psi)^{-1*}\beta = F^*\alpha = \beta$. As a consequence,

Theorem 3.1.1 (Brenier). *Any function F can be decomposed as $F = \nabla\psi \circ \rho$ with ψ convex and ρ volume preserving.*

Remark 3.12. It turns out that polar decomposition implies the Hodge decomposition. To see this assume that $F^\epsilon(x) = x + \epsilon f(x)$ with ϵ small. We then expect to have $\psi^\epsilon(x) = \frac{1}{2}|x|^2 + \epsilon\varphi(x) + o(\epsilon)$ and $\rho^\epsilon(x) = x + \epsilon m(x) + o(\epsilon)$. Hence

$$\begin{aligned} x + \epsilon f(x) &= x + \epsilon\nabla\varphi(x + \epsilon m(x)) + \epsilon m(x) + o(\epsilon) \\ &= x + \epsilon(\nabla\varphi(x) + m(x)) + o(\epsilon). \end{aligned} \tag{3.13}$$

On the other hand, since ρ^ϵ is volume preserving,

$$1 = \det(1 + \epsilon m') = 1 + \epsilon \ \mathrm{div}(m) + o(\epsilon).$$

From this and (3.13) we learn

$$f(x) = \nabla\varphi(x) + m(x) \text{ with div } m = 0.$$

□

We end this section with a discussion regarding complex structures on manifolds. Let M be a C^1 manifold. By an *almost complex structure* on M, we mean a continuous $x \mapsto J_x$ with $J_x : T_xM \to T_xM$ linear function satisfying $J_x^2 = -id$. The pair (M, J) is called an *almost complex manifold*. If (M, ω) is symplectic, then we say J and ω are compatible if

$$\hat{g}_x(a, b) = \omega_x(a, J_x b), \ \ a, b \in T_xM \tag{3.14}$$

is a Riemannian metric on M. If this happens, then we write (M, ω, J) for the triplet.

Proposition 3.13. *Let (M,ω) be symplectic with a Riemannian metric g. Then there exists a compatible almost complex structure J on (M,ω).*

Proof. Fix x. Both $a \mapsto \omega_x(a,\cdot)$ and $a \mapsto g_x(a,\cdot)$ are linear isomorphisms between T_xM and $(T_xM)^*$. Hence there exists a linear invertible $A_x : T_xM \to T_xM$ such that $\omega_x(a,b) = g_x(A_xa,b)$. Note that A_x is skew symmetric because

$$\begin{aligned} g(A^ta,b) &= g(a,Ab) = g(Ab,a) = \omega(b,a) \\ &= -\omega(a,b) = -g(Aa,b). \end{aligned}$$

The transformation A_x has a polar decomposition $A_x = S_xJ_x$ with $S_x = (A_xA_x^t)^{1/2}$ symmetric with respect to g_x, and J_x orthogonal. Note that since $A^t = -A$, the matrices A and S commute. We have $JJ^t = id$ and $J^t = A^tS^{-1} = -S^{-1}A = -J$ because A and S commute. In summary

$$J^2 = -id, \;\; J^t = -J.$$

We also have

$$\begin{aligned} \hat{g}(a,b) &:= \omega(a,Jb) = g(Aa,Jb) \\ &= g(Sa,b). \end{aligned}$$

We are done because $S > 0$. □

Exercises 3.14. (i) Let (U,ω) be symplectic with U open subset of $\mathbb{R}^{2n}$. Show that $\omega_x(a,b) = A_xa\cdot b$ with A_x skew symmetric and invertible. In this case $X_H(x) = -A_x^{-1}\nabla H(x)$. Also, if $\omega = d\lambda$ with $\lambda = F\cdot dx$, then $A = \mathcal{C}(F)$ with $\mathcal{C}(F) = [F^i_{x_j}]$. Here $F = (F^1,\dots,F^{2n})$.

(ii) Verify Remark 3.7.

(iii) Let (M,ω) be a $2n$-dimensional symplectic manifold. Assume that M is compact with no boundary. Then for every $j = \{1,\dots,n\}$, there exists a closed $2j$-form which is not exact.

(iv) Let (M,ω^1) and (M,ω^2) be two symplectic manifolds. Define $(M_1\times M_2, \omega^1\times\omega^2)$ with

$$(\omega^1\times\omega^2)_{(x,y)}((a_1,a_2),(b_1,b_2)) = \omega^1_x(a_1,b_1) + \omega^2_y(a_2,b_2).$$

Show that $(M_1\times M_2,\omega^1\times\omega^2)$ is symplectic.

(v) Let $\alpha = a dx_1\wedge\cdots\wedge dx_k$, $\beta = b dx_1\wedge\cdots\wedge dx_k$ be two volume forms. Then $f^*\alpha = \beta$ means that

$$a(f(x))\det Df(x) = b(x).$$

If $f = \nabla w$ for some $w : \mathbb{R}^k \to \mathbb{R}$, then

$$a(\nabla w)\det(D^2w) = b.$$

This is the celebrated Monge–Ampére's equation.

4 Contact Manifolds

Consider a Hamiltonian system $\dot{x} = J\nabla H(x)$. We know that the level sets of the energy

$$S = \{x : H(x) = c\} \subseteq \mathbb{R}^{2n} = \mathbb{R}^k \tag{4.1}$$

are invariant for the flow of the Hamiltonian system. We also know that the 2-form $\bar{\omega}$ and the volume form $\bar{\Omega} = \bar{\omega}\wedge\cdots\wedge\bar{\omega} = n!\, dx_1\wedge dx_2\wedge\cdots\wedge dx_k$ with $(x_1 \dots x_k) = (q_1, p_1, q_2, p_2, \dots, q_n, p_n)$ are invariant. We assume that the surface S is compact and regular. The latter means that $\nabla H(x) \neq 0$ for $x \in S$ (or $dH \neq 0$ on S) and guarantees $\dim S = 2n - 1$. Evidently $(S, \bar{\omega})$ is not a symplectic manifold because the dimension of S is odd. This means that $\bar{\omega}$, restricted to T_xS cannot be non-degenerate. As a classical problem in celestial mechanics, we are interested in the existence of a periodic solution on S. It is quite natural to expect the existence of periodic orbits because the Hamiltonian flow has an invariant measure on S and therefore enjoys the recurrence properties predicted by a theorem of Poincaré. Indeed since $dH \neq 0$ on S, we can find a $(2n-1)$-form α such that

$$\bar{\Omega} = dH \wedge \alpha \tag{4.2}$$

in a neighborhood of S. If we write $j : S \to \mathbb{R}^{2n}$ for the inclusion map, then $\beta = j^*\alpha$ defines a volume form on S. Note that for $x \in S$, the form β_x is simply the restriction of α_x to $(T_xS)^{2n-1}$.

Proposition 4.1. (i) *If* $\bar{\Omega} = dH \wedge \alpha = dH \wedge \alpha'$ *then* $j^*\alpha = j^*\alpha'$, *so* (4.2) *determines* β *uniquely.*

(ii) $\Phi_t^*\beta = \beta$, *i.e.,* β *is invariant for* Φ_t *restricted to* S.

Proof. (i) We certainly have $dH \wedge (\alpha - \alpha') = 0$. From Exercise 4.2 below we deduce that for some $2n-2$ form γ, $\alpha - \alpha' = dH \wedge \gamma$. As a result, $j^*(\alpha - \alpha') = j^*(dH \wedge \gamma) = (j^*dH) \wedge j^*\gamma = 0$ because $j^*dH = d(H \circ j) = 0$.

(ii) By (4.2)

$$\begin{aligned}\bar{\Omega} = \Phi_t^*\bar{\Omega} &= (\Phi_t * dH) \wedge \Phi_t^*\alpha \\ &= d(H \circ \Phi_t) \wedge \Phi_t^*\alpha \\ &= dH \wedge \Phi_t^*\alpha.\end{aligned}$$

By uniqueness of (i) we deduce $j^*\Phi_t^*\alpha = j^*\alpha$, or $(\Phi_t \circ j)^*\alpha = j^*\alpha$. We may write $\Phi_t \circ j = j \circ \Phi_t$ where the second $\Phi_t : S \to S$ is the restriction of Φ_t to S. Hence $\Phi_t^* j^*\alpha = j^*\alpha$, as desired. □

Exercise 4.2. Let η be $2n - 1$ form and τ a 1-form with $\tau \wedge \eta = 0$. Show that $\eta = \tau \wedge \gamma$ for some $n - 2$-form γ. □

The form α of (4.2) induces an $n-1$-dimensional measure on S that is easy to describe. Note that from (4.2) we learn that if $\{g_j : j = 1, \dots, n\}$ is a collection of smooth functions with $\sum_{j=1}^{k} g_j H_{x_j} (-1)^{j-1} \equiv 1$, then we may choose for α,

$$\alpha = n! \sum_{j=1}^{k} g_j dx_1 \wedge \cdots \wedge \widehat{dx_j} \wedge \cdots \wedge dx_k.$$

We now claim that if Γ is a subset of S, then $\int_\Gamma \alpha$ is

$$c \int_\Gamma |\nabla H|^{-1} d\sigma \tag{4.3}$$

where c is a constant (independent of Γ and H) and $d\sigma$ is the surface measure on Γ. To see this, let us assume $H_{x_k} \neq 0$ so that locally we can find a function $f(x_1, \dots, x_{k-1})$ with

$$H(x_1, \dots, x_{k-1}, f(x_1, \dots, x_{k-1})) = c.$$

This means that the graph of f gives a parametrization of S and that $f_{x_j} = -\frac{H_{x_j}}{H_{x_k}}$ for $j = 1, \dots, k-1$. On the other we can use the function f to express the volume measure on S that is induced by the standard volume of $\mathbb{R}^k$. More precisely, consider the vector $a = (a_1, \dots, a_k)$ with

$$a_j = \frac{\partial(x_1, \dots, \hat{x}_j, \dots, x_k)}{\partial(x_1, \dots, x_{k-1})}.$$

A straightforward calculation yields

$$a_j = (-1)^{k-j} \frac{H_{x_j}}{H_{x_k}} \; 1 \leq j \leq k.$$

If $\Gamma = \{(x_1, \dots, x_{k-1}, f(x_1, \dots, x_{n-1})) : (x_1, \dots, x_k) \in U\}$, then

$$\int_\Gamma \alpha = n! \int_U \left(\sum_{j=1}^{k} a_j g_j \right) dx_1, \dots, dx_{k-1}.$$

On the other hand

$$\bar{a} = \left(\sum_{j=1}^{k} a_j^2 \right)^{1/2} = \frac{|\nabla H|}{|H_{x_k}|}.$$

Hence locally either $H_{x_k} > 0$ or $H_{x_k} < 0$, and

$$\begin{aligned} \int_\Gamma \alpha &= n! \int_\Gamma (-1)^{k-1} \frac{1}{H_{x_k}} \; dx_1, \dots, dx_{k-1} \\ &= \pm n! \int_\Gamma \frac{1}{|\nabla H|} \bar{a} \; dx_1, \dots, dx_{k-1} \\ &= \pm n! \int_\Gamma \frac{1}{|\nabla H|} \; d\sigma, \end{aligned}$$

proving (4.3). (Note that this is consistent with the coarea formula

$$\int J\ dx_1, \dots, dx_k = \int_{-\infty}^{\infty} \left[\int_{H=c} J \frac{d\sigma}{|\nabla H|} \right] dc.)$$

As we mentioned earlier, the existence of a finite invariant measure on S implies that an orbit of X_H comes arbitrarily close to its starting point by Poincaré's recurrence theorem. As a result of this, there are chances that in fact some orbits close up in finite time and we have periodic orbits. This is the so-called Closing Problem and we would like to address this problem for the level set S. We now assert that the existence of a periodic orbit on S is a property of S and $\bar{\omega}$ and does not depend on H. In other words, if S given by (4.1) possesses a periodic orbit of X_H and of $S = \{x : H'(x) = c'\}$ for another regular H', then $X_{H'}$ possesses a periodic orbit as well. In fact the orbits of X_H and $X_{H'}$ differ only on their parametrizations. To see this, let us find an expression for the orbits as subsets of S that is purely stated in terms of S and $\bar{\omega}$. For this, let us observe that if $a \in T_xS$, then

$$\bar{\omega}(X_H(x), a) = -dH(x)a = 0,$$

because $T_xS = \{a : dH(x)a = 0\}$. Hence, if we define

$$l_x = \{v : \bar{\omega}(v, a) = 0 \text{ for all } a \in T_xS\}, \tag{4.4}$$

then

$$X_H(x) \in l_x, \tag{4.5}$$

for every $x \in S$. Using our notation from Section 2,

$$l_x = (T_xS)^{\amalg}.$$

By (2.4)

$$\dim(T_xS)^{\amalg} + \dim T_xS = 2n,$$

and since $\dim T_xS = 2n - 1$, we deduce that $\dim(T_xS)^{\amalg}$ is 1. Hence l_x is a line and $X_H(x)$ is parallel to it. Evidently $l_x = (T_xS)^{\amalg}$ offers an H-independent candidate for the tangent lines to the orbit. As a result, we can express our problem purely in terms of the line bundle $L_S = \bigsqcup_{x \in S} l_x$; our goal is finding a closed characteristic of the line bundle L_S. It turns out that there are hypersurfaces S and symplectic structures ω such that the corresponding line bundles have no closed characteristics. Our hope is that perhaps if S has some practically checkable properties, then such a periodic characteristic can be found.

In our search for some natural structure for certain class of hypersurfaces, let us ask the following question that is meaningful for $(\mathbb{R}^{2n}, \bar{\omega})$, the cotangent bundle (T^*N, ω) or any symplectic manifold (M, ω) for each $\omega = d\lambda$ for a 1-form λ: Does a Hamiltonian flow on M

preserve the form λ? We have the following straightforward fact. Let us assume that N is a connected manifold.

Proposition 4.2. *The Hamiltonian vector field X_H on T^*N preserves the* 1*-form λ if and only if $H(q,p)$ is positively homogeneous of degree* 1 *in p, i.e., $H(q,cp) = cH(q,p)$ for any $c > 0$.*

Proof. By Theorem 3.5, λ is invariant with respect to the flow of X_H iff $\mathcal{L}_{X_H}\lambda = 0$. But

$$\mathcal{L}_{X_H}\lambda = d \circ i_{X_H}\lambda + i_{X_H} \circ d\lambda = d(i_{X_H}\lambda) - dH.$$

The term $i_{X_H}\lambda - H$ is a scalar function and this function has zero exterior derivative iff $i_{X_H}\lambda - H$ is a constant. Since adding a constant does not change X_H, we may assume that the constant is zero so that the condition now is $i_{X_H}\lambda = H$. Let's examine this condition when $N = \mathbb{R}^n$. In this case $\lambda = \bar{\lambda} = p \cdot dq$ and

$$\begin{aligned}(i_{X_H}\bar{\lambda})(x) &= \bar{\lambda}_x(X_H(x)) = \bar{\lambda}_{(q,p)}(X_H(q,p)) \\ &= \bar{\lambda}_{(q,p)}(H_p, -H_q) = p \cdot H_p.\end{aligned}$$

This equals H iff H is homogeneous of degree 1 in p.

The proof of general N is similar. Use the Darboux charts we defined in Section 3, namely take $h : U \to \mathbb{R}^n$ as a chart of N and use this to construct $\bar{h} : T^*U \to \mathbb{R}^{2n}$. Recall $\bar{h}\left(q, \sum_1^n p_i dq^i\right) = (h(q), p_1, \ldots, p_n)$ and that $\bar{h}^*\bar{\lambda} = \lambda$. Let us write $V = \bar{h}(T^*U)$ and $\psi = \bar{h}^{-1} : V \to T^*U$. We then use $i_{X_H}\lambda = H$ to deduce that $\psi^* i_{X_H}\lambda = H \circ \psi$. We deduce that $i_{\psi_* X_H}\bar{\lambda} = i_{X_{H\circ\psi}}\bar{\lambda} = H \circ \psi$. As a result, $H \circ \psi$ is homogeneous of degree 1 in p. Using the definition of $\bar{h}$, we deduce that H is homogeneous of degree 1 in p. □

We know that in a symplectic manifold (M, ω) we can use ω to define a volume form $\Omega = \omega^n$. We have seen that the hypersurface S possesses a form $\beta = \Omega_S$ that serves as a volume form on S and comes from Ω. Under the circumstances of Proposition 4.2, we would like to use λ to define a volume form on S. Indeed if we define

$$\hat{\beta} = j^*(\lambda \wedge \omega^{n-1}), \tag{4.6}$$

then

$$\begin{aligned}\Phi_t^* j^*(\lambda \wedge \omega^{n-1}) &= (j \circ \Phi_t)^*(\lambda \wedge \omega^{n-1}) = (\Phi_t \circ j)^*(\lambda \wedge \omega^{n-1}) \\ &= j^*\Phi_t^*(\lambda \wedge \omega^{n-1}) = j^*(\lambda \wedge \omega^{n-1}).\end{aligned}$$

As a result, if $\mu = j^*\lambda$, then μ is a 1-form on S such that

$$\hat{\Omega} = \mu \wedge (d\mu)^{n-1}, \tag{4.7}$$

is a $2n-1$-form on S and $\Phi_t^* \hat{\Omega} = \hat{\Omega}$, i.e., Φ_t preserves $\hat{\Omega}$ on S. Note that $\mu = j^*\lambda$ is nothing other than the restriction of λ to T_xS. Also, we learned in the process of proving Proposition 4.2 that indeed $(j^*\lambda)_x(X_H(x)) = \lambda_x(X_H(x)) = H(x)$ for $x \in S$. But $H(x) = c$ is constant and $c \neq 0$ because H is homogeneous and $\{(q,p) : H(q,p) = 0\}$ is not compact. As a result,

$$\mu_x(v) \neq 0 \text{ if } x \in S \text{ and } v \in (T_xS)^{\amalg} = l_x. \tag{4.8}$$

Motivated by this we make the following definition. Let (M,ω) b a symplectic manifold. Let $S \subseteq M$ be an orientable compact hypersurface. We then say S is of *contact type* if there exists a 1-form μ on S such that $d\mu = j^*\omega$ and (4.8) is true. Note that for a hypersurface of contact type we can still talk about the line bundle L_S. This line bundle is orientable because we can uniquely define a vector field $X = X^\mu$ by the requirement $X_x = X(x) \in l_x$ and $\mu_x(X_x) = 1$. The vercor field X^μ is called the *Reeb's vecor field.*

Lemma 4.3. *Let (S,μ) be a submanifold of contact type. Then $\mu \wedge (d\mu)^{n-1}$ is a volume form.*

Proof. Let $\gamma = (d\mu)^{n-1}$. Since $d\mu_x = (j^*\omega)_x$ is non-degenerate on $V = \{v : \mu_x(v) = 0\}$ and $\dim V = 2(n-1)$, we may use Exercise 2.12(i) to assert that $(d\mu)_x^{n-1}$ is a volume form on V. In other words γ_x is non-degenerate on V. From this we would like to deduce that $\mu_x \wedge \gamma_x$ is non-degenerate. Indeed if $\{a_1, a_2, \dots, a_{2n-1}\}$ is a basis with $a_1 \in l_x$ and $\text{span}\{a_2, \dots, a_{2n-1}\} = V$, then with respect to this basis, $\mu_x = c_1 dx_1$, $\gamma_x = c_2 dx_2 \wedge \cdots \wedge dx_{2n-1}$ with $c_1, c_2 \neq 0$. As a result, $\mu_x \wedge \gamma_x$ is a volume form. □

Exercise 4.4. Let S be a hypersurface in (M,ω) that is compact and orientable. Suppose there exists a 1-form μ such that $\mu \wedge (d\mu)^{n-1}$ is a volume form and $d\mu = j^*\omega$. Show that (S,μ) is of contact type. □

Recall that $M = T^*N$ is equipped with a natural 1-form λ such that $d\lambda = \omega$. In view of Proposition 4.2, if S is a level set of a homogeneous Hamiltonian, then S is a submanifold of contact type with $\mu = j^*\lambda$. In general, λ may not exist or if such a 1-form λ exists, the Hamiltonian flow may not preserve it. However near S we have a similar structure if S is of contact type.

Theorem 4.5. *If (S,μ) is a submanifold of contact type, then there exists a 1-form τ on a neighborhood U of S such that $d\tau = \omega$ and $\mu = j^*\tau$ on S.*

To prepare for this, let us start with a lemma.

Lemma 4.6. *Let S be a codimension 1 compact submanifold of (M,ω) with an orientable line bundle $\{l_x\}$. Then there exists a diffeomorphism $\psi : S \times (-1,1) \to U$ with U an open* *neighborhood of S, such that $\psi(x,0)=x$. Moreover there exist a Hamiltonian H such that $S=H^{-1}(0)$ and $\nabla H \neq 0$ on S.*

Proof. By Proposition 3.13, we may find a Riemannian metric $\hat{g}$ and a complex structure J such that $J^2=-id$ and

$$\omega_x(a, J_x b)=\hat{g}_x(a,b).$$

Since l is orientable, we can find a verto field X such that $X(x)\in l_x$ for every $x\in S$. We then set $n(x)=J_x X(x)$. Note that $\hat{g}_x(n(x),b)=\hat{g}_x(J_xX(x),b)=\omega_x(J_xX(x),J_xb)=\omega_x(X(x),b)=0$, for $b\in T_xS$. Using the non-zero vector field n we may define $\psi: S\times(-1,1)\to U\subseteq M$ by $\psi(x,t)=\exp_x(\epsilon t n(x))$ for some small $\epsilon>0$ and a neighborhood $U\subseteq M$ of S where exp is the exponential map ($\psi(x,t)$ is the location of a geodesic starting at x and traveling for t seconds).

Finally we define $H=F\circ\psi^{-1}$ where $F(x,t)=t$. □

Proof of Theorem 4.5. Let ψ be as in Lemma 4.4. Define $f: S\times(-1,1)\to S$ by $F(x,t)=x$ and let $f: U\to S$ be defined by $f=F\circ\psi^{-1}$. We define a 1-form γ by $\gamma=f^*\mu$. Note $d(\omega-d\gamma)=d\omega=0$. Moreover

$$\begin{aligned} j^*(\omega-d\gamma) &= d\mu - dj^*\gamma = d\mu - dj^*f^*\mu \\ &= d\mu - d(f\circ j)^*\mu = d\mu - d\mu = 0. \end{aligned}$$

It suffices to find a form ν such that

$$j^*\nu=0,\ \omega-d\gamma=d\nu. \tag{4.9}$$

Indeed if we define $\tau=\gamma+\nu$, we then have that $j^*\tau=j^*\gamma=j^*f^*\mu=(f\circ j)^*\mu=\mu$ and $d\tau=d\gamma+d\nu=\omega$.

For the existence of ν as in (4.9), we need a Poincaré-type lemma. This is carried out in Lemma 4.7 below. □

Lemma 4.7. *Let $E=S\times\mathbb{R}$ and let α be a closed k-form on E with $j^*\alpha=0$ where $j: S\to E$ is the inclusion. Then there exists a $k-1$-form β on E such that $d\beta=\alpha$ and $\beta_{|S}\equiv 0$.*

Proof. Define $\Phi_\theta(y)=\theta\cdot(x,s)=(x,\theta s)$ for $y=(x,s)\in E$ and $\theta\in(0,1)$. This defines a flow with $\Phi_1=id$. We define a vector field X_t so that $\frac{d}{d\theta}\Phi_\theta(y)=X_\theta(\Phi_\theta(y))$. Indeed $X_\theta(y)=\frac{1}{\theta}X(y)$ with $Y(x,s)=(0,s)$. We identify S with $S\times\{0\}$. Note that $X_{|S}\equiv 0$. Also note $\Phi_0(y)=\pi(y)$ where $\pi(x,s)=(x,0)$ is the projection onto the base S. As a result,

$\Phi_0^*\alpha = \pi^*\alpha = 0$ by our assumption. We now write

$$\begin{aligned}\alpha = \Phi_1^*\alpha - \Phi_0^*\alpha &= \int_0^1 \frac{d}{d\theta}\Phi_\theta^*\alpha d\theta \\ &= \int_0^1 \Phi_\theta^*\mathcal{L}_{X_\theta}\alpha d\theta = d\int_0^1 \Phi_\theta^* i_{X_\theta}\alpha d\theta \\ &=: d\beta.\end{aligned}$$

Note that $\Phi_\theta(S) = S$ and $X_\theta(x) = 0$ for $x \in S$. From this we can readily deduce that $\beta_{|S} \equiv 0$. □

We now state an equivalent definition of contact type surfaces S that is of great practical use.

Proposition 4.8. *Let (M, ω) be symplectic. Then a hypersurface $S \subseteq M$ is of contact type iff there exists a vector field X, defined on a neighborhood U of S such that $\mathcal{L}_X\omega = \omega$ on U and $X(x) \notin T_xS$ for every $x \in S$.*

Proof. Assume that (S, μ) is of contact type. Let τ be as in Theorem 4.5. By non-degeneracy, we can find a vector field X such that $i_X\omega = \tau$ on U. We have

$$\omega = d\tau = di_X\omega = di_X\omega + i_Xd\omega = \mathcal{L}_X\omega.$$

On the other hand, if $v \in l_x - \{0\}$, then

$$\omega_x(X(x), v) = \tau_x(v) = \mu_x(v) \neq 0.$$

As a result, $X(x) \notin T_xS$ because $v \in (T_xS)^{\mathrm{II}}$.

Conversely, if a vector field X meets the conditions of our theorem, then we may define $\tau = i_X\omega$. We then have $d\tau = di_X\omega = \mathcal{L}_X\omega = \omega$ and that if $v \in l_x - \{0\}$, then

$$0 \neq \omega_x(X(x), v) = \tau_x(v),\ x \in S$$

simply because $\dim l_x = 1$, $l_x = (T_xS)^{\mathrm{II}}$ and by assumption $X(x) \notin T_xS$. As a result, if $\mu = j^*\tau$ then $\mu_x(v) \neq 0$ for every $v \in l_x - \{0\}$. Hence S is of contact type. □

Note that if $\mathcal{L}_X\omega = \omega$ then we can readily calculate $\Phi_t^*\omega$ where Φ_t is the flow associated with X. Indeed

$$\frac{d}{dt}\Phi_t^*\omega = \Phi_t^*\mathcal{L}_X\omega = \Phi_t^*\omega$$

implies

$$\mathcal{L}_X\omega = \omega \Leftrightarrow \Phi_t^*\omega = e^t\omega \tag{4.10}$$

We now study some examples.

Exercise 4.9. Consider $(\mathbb{R}^{2n}, \bar{\omega})$ and let S be the boundary of a set A that is star-shaped with respect to the origin. Assume that each ray emenating from the orgin intersect S at exactly one point. Show that S is of contact type.

Hint: Set $X(x) = \frac{1}{2}x$. Show that X meets the conditions of Theorem 4.8.

As our next example, we consider classical Hamiltonian systems. The positions of a conservative system are points in an n-dimensional manifold N that is known as the configuration space. The phase space in Lagrangian formulation is the tangent bundle TN. The motion is determined by a Lagrangian $L : TN \to \mathbb{R}$. In the Hamiltonian formulation we use a Hamiltonian function $H : M = T^*N \to \mathbb{R}$ to determine the motion of the system. In Section 3 we defined a 1-form λ that was simply given by $p \cdot dq$ in local coordinates. The 2-form $\omega = d\lambda$ is a symplectic construction on M. Given a Riemannian metric g on M we can define a bundle isomorphism $F : TN \to T^*N$ that is defined by

$$F(q, v) = (q, F_q(v)), F_q(v)a = g_q(v, a).$$

Using F we can define a metric on T^*N by

$$G_q(p, p') = g_q(F_q^{-1}(p), F_q^{-1}(p')).$$

Given a smooth potential energy $V : N \to \mathbb{R}$, we define

$$H(q, p) = \frac{1}{2}G_q(p, p) + V(q). \tag{4.11}$$

Lemma 4.10. *Let H be as in* (4.11)*. Then $\lambda(X_H) = G_q(p, p)$.*

Proof. A chart $h : U \to \mathbb{R}^n$, $U \subseteq N$ can be used to find a local coordinate $\bar{h} : T^*U \to \mathbb{R}^{2n}$ with $\bar{h}\left(q, \sum_{j=1}^n p_j dq_j\right)$ given by $(h(q), p_1, \dots, p_n)$. In these coordinates H is given by

$$\bar{H}(q_1, \dots, q_n, p_1 \dots p_n) = \frac{1}{2}S(q_1 \dots q_n)(p_1 \dots p_n) \cdot (p_1 \dots p_n) + V(h^{-1}(q_1 \dots q_n)),$$

with $S > 0$. Now it is clear that in these coordinates X_H is given by $\bar{X} = (S(q)p, -\nabla(V \circ h^{-1}) - \frac{1}{2}DS(q)p, p)$. As a result, if we write K for $\frac{1}{2}S(q_1 \dots q_n)(p_1 \dots p_n) \cdot (p_1 \dots p_n)$, then $\bar{\lambda}(\bar{X}) = 2K$. This is $G_q(p, p)$ if we use the point $(q, p) \in U$. $\square$

Using the elementary Lemma 4.10, we can now readily show that if

$$S_E = \{(q, p) : H(q, p) = E\} \tag{4.12}$$

is compact and regular, and if $E > \max V$, then S_E is of contact type. This simply follows from the fact that $\mu = j^*\lambda$ enjoys the property $\mu_x(X_H(x)) = G_q(p,p) >$ for $x = (q,p) \in S$. Indeed if $(q,p) \in S_E$ then $G_q(p,p) > 0$ always because $E > \max V$.

A. Weinstein conjectured that any surface of contact type would have a periodic orbit provided that $H^1(S) = 0$. Viterbo in 1987 proved that this conjecture is true and there is no need for the assumption $H^1(S) = 0$. In view of Viterbo's theorem we must have a periodic orbit on S_E of (4.12) provided that $E > \max V$.

If $E > \max V$, then we can define a new Riemannian metric

$$\hat{G}_q(p,p) = \frac{G_q(p,p)}{E - V(q)}, \tag{4.13}$$

that is known as the *Jacobi metric*.

We then have

$$S_E = \{(q,p) = H(q,p) = E\} = \left\{(q,p) = \frac{1}{2}\hat{G}_q(p,p) = 1\right\}.$$

The Hamiltonian $\hat{H}(q,p) = \frac{1}{2}\hat{G}_q(p,p)$ induces a Hamiltonian vector field $X_{\hat{H}}$. It is simply related to X_H by

$$X_H(q,p) = G_q(p,p)X_{\hat{H}}(q,p), \ (q,p) \in S. \tag{4.14}$$

Hence a periodic orbit exists on S_E for X_H if the same is true for $X_{\hat{H}}$. Geometrically, $X_{\hat{H}}$ generates the geodesic flow defined by the Jacobi metric $\hat{G}$ on $M = T^*N$.

Exercise 4.11. Verify (4.14).

In general we have this:

Theorem 4.12. *Suppose S_E of* (4.12) *is compact and regular. Then S_E is of contact type.*

We already gave a proof of Theorem 4.12 when $E > \max V$ and we omit the proof for general E and refer the reader to [HZ].

Recall that (S,μ) is a hypersurface of contact type iff $\mu \wedge (d\mu)^{n-1}$ is a volume on S and $d\mu = j^*\omega$. We may define contact manifold without referring to a symplectic form. More precisely, we say that an odd-dimensional manifold S with a 1-form μ is a *contact manifold* if $\mu \wedge (d\mu)^{n-1}$ is a volume form on S. If $\dim(S) = 2n-1$, then

$$\xi_x = \{v \in T_xS : \mu_x(v) = 0\} = \ker \mu_x \subseteq T_xS \tag{4.15}$$

defines a field of hyperspaces of dimension $2n-2$. If S is a submanifold of M of contact type, we can then write $T_xS = l_x \oplus \xi_x$. If X_H is a Hamiltonian vector field of (M,ω), then we have

$$i_{j^*X_H} d\mu = j^* i_{X_H}\omega = 0, \ i_{j^*X_H}\mu = \mu(j^*X_H) \neq 0$$

because $\mu_x(v) \neq 0$ for $v \in l_x - \{0\}$. We may normalize to have a vector field X on S with $i_X\mu \equiv 1$. Recall that such a vector field is called a *Reeb vector field* and can be constructed on every contact manifold.

Proposition 4.13. *Let (S, μ) be a contact manifold.*

(i) *The form $\beta = d\omega$ is non-degenerate on ξ.*

(ii) *There exists a unique vector field X^μ on S such that*

$$i_{X^\mu} d\mu \equiv 0,\ i_{X^\mu}\mu \equiv 1. \tag{4.16}$$

Proof. (i) We argue by contradiction. If $\beta = d\mu$ and at a point x and for a vector $v_1 \in \xi_x$ we have $\beta_x(v_1, w) = 0$ for all $w \in \xi_x$, then find a basis $\{v_1, \ldots, v_{2n-1}\}$ of T_xS such that $\xi_x = \text{span}\{v_1, \ldots, v_{2n-2}\}$. We then have

$$(\mu \wedge \beta^{n-1})_x(v_1, \ldots, v_{2n-1}) = \mu_x(v_{2n-1})\beta^{(n-1)}(v_1, \ldots, v_{2n-2}) = 0,$$

contradiction.

(ii) We construct $X = X^\mu$ locally and once we established the local uniqueness, we can deduce that X is globally defined.

Take $x^0 \in S$ and choose a neighborhood $U \subseteq S$ of x^0 for which we can find linearly independent vector fields $v_1(x), \ldots, v_{2n-1}(x)$ for $x \in U$. If necessary, switch to a smaller neighborhood $U' \ni x^0$ so that for some $v, w \in \{v_1, \ldots, v_{2n-1}\}$ we have $(d\mu)_x(v(x), w(x)) \neq 0$ everywhere in U'. This is possible because $\mu \wedge (d\mu)^{n-1}$ is a volume form. If necessary multiply v by a scalar so that $d\mu(v, w) \equiv 1$ in U'. Express $TU' = E^1 \oplus E^2$ with E^1 a bundle with fibers $\text{span}(v(x), w(x))$ and that E^2 consists of points $y \in TU'$ such that $d\mu(y, z) = 0$ for all $z \in E^1$. We now find v', w' in E^2 with $d\mu(v', w') = 1$ etc. By induction we find a family of independent vector field $\{e_1 \ldots e_{n-1} f_1 \ldots f_{n-1}\bar{a}\}$ in a small neighborhood of x^0 so that

$$\begin{cases} d\mu(e_i, e_j) = & d\mu(f_i, f_j) = d\mu(\bar{a}, e_i) = d\mu(\bar{a}, f_j) = 0, \\ & d\mu(f_i, e_j) = \delta_{ij}. \end{cases}$$

We now choose $X(x) = \frac{\bar{a}(x)}{\mu_x(\bar{a}(x))}$. Note that since $\mu \wedge (d\mu)^{n-1}$ is a volume form, we have

$$\begin{aligned} 0 &\neq \mu \wedge (d\mu)^{n-1}(e_1, f_1, \ldots, e_n, f_n, \bar{a}) \\ &= (d\mu)(e_1, f_1) \ldots (d\mu)(e_{n-1}, f_{n-1})\mu(\bar{a}) = \pm\mu(\bar{a}), \end{aligned}$$

so $\mu(\bar{a}) \neq 0$. Also note that by definition $i_X\mu \equiv 1$. Moreover, $i_X d\mu \equiv 0$ because $i_{\bar{a}} d\mu \equiv 0$.

It remains to verify the uniqueness. Suppose we also have $i_Y\mu \equiv 1$, $i_Y d\mu \equiv 0$. We then have $i_{X-Y}\mu \equiv 0$, $i_{X-Y}d\mu \equiv 0$. If $(X-Y)(x) \neq 0$ then we get $(\mu \wedge (d\mu)^{n-1})_x(a_1, \dots, a_{2n-1}) = 0$ for $a_1 = (X-Y)(x)$ and $a_2 \dots a_{2n-1}$ arbitrary. This would contradict the fact that $\mu \wedge (d\mu)^{n-1}$ is a volume form. Hence $X \equiv Y$ and this completes the proof. □

In summary,, if we have a contact manifold (S, μ), then we can write $TS = l \oplus \xi$ with l a line bundle with a distinguished and preferred section X^μ and ξ a bundle of hyperplane of dimension $2n-2$ on which ω is a non-degenerate 2-form. We may say that the pair (ξ, ω) is a *symplectic bundle.*

Evidently $(\mathbb{R}^{2n-1}, \bar{\mu})$ with $\bar{\mu} = \sum_1^{n-1} p_j dq_j + dz$ is a contact manifold. Here we denote the coordinates by

$$(q_1, p_1, \dots, q_{n-1}, p_{n-1}, z).$$

Exercise 4.14. Let (S, μ) be a contact manifold and set $M = S \times \mathbb{R}$, $\lambda_{(x,t)}(a, s) = e^t \mu_x(a)$, and $\omega = d\lambda$. Show that (M, ω) is a symplectic manifold and that S is a submanifold of contact type with $\mu = j^*\lambda$.

Note that if $X = X^\mu$ is the Reeb vector field associated with μ, then $\mathcal{L}_X\mu \equiv 0$ by (4.16). As a result, the flow of X^μ preserves μ and hence the volume form $\mu \wedge (d\mu)^{n-1}$. More generally, we may consider vector fields Y such that $\mathcal{L}_Y\mu = g\mu$ for some $g : M \to \mathbb{R}$. We call such a vector field a *contact vector field.*

Proposition 4.15. (i) *If Y is a contact vector field and Φ_t is its flow, then $\Phi_t^*\xi = \xi$.*

(ii) *If $H = i_Y\mu$ for a contact Y,, then*

$$i_Y d\mu = -dH + (i_{X^\mu} dH)\mu. \tag{4.17}$$

(iii) *If $H : S \to \mathbb{R}$ is a smooth function, then there exists a contact Y such that $H = i_Y\mu$.*

Proof. (i) We have $\frac{d}{dt}\Phi_t^*\mu = \Phi_t^*\mathcal{L}_Y\mu = \Phi_t^* g\mu$. Hence $\frac{d}{dt}\Phi_t^*\mu = g \circ \Phi_t \Phi_t^*\mu$. As a result $\Phi_t^*\mu = \exp\left(\int_0^t g \circ \phi_s ds\right)\mu$. From this we learn that $d\Phi_t$ maps the kernel of μ at x to the kernel of μ at $\Phi_t(x)$.

(ii) See $X = X^\mu$. We have

$$g\mu = \mathcal{L}_Y\mu = di_Y\mu + i_Y d\mu = dH + i_Y d\mu.$$

Hence

$$i_Y d\mu = -dH + g\mu. \tag{4.18}$$

On the other hand evaluating both sides of (4.18) at $X = X^\mu$ yields

$$\begin{aligned} g = g\mu(X) &= dH(X) + d\mu(Y, X) \\ &= dH(X) - i_X d\mu(Y) = dH(X). \end{aligned}$$

Hence $g = dH(X)$ and this completes the proof.

(iii) Given H, find a unique $Z \in \xi$ such that $i_Z d\mu_{|\xi} = -dH_{|\xi}$. Such a vector field Z exists by non-degeneracy of $d\mu_{|\xi}$. We then set $Y = Y_H = Z + HX^\mu$. We have

$$\begin{aligned} i_Y d\mu &= i_Z d\mu + i_{HZ^\mu} d\mu = i_Z d\mu + H i_{X^\mu} d\mu \\ &= i_Z d\mu. \end{aligned}$$

We also know that for some g, $i_Z d\mu + dH = g\mu$. Since $Z \in \xi$, we have $i_Z \mu = 0$. Hence $i_Y \mu = H$ and

$$\mathcal{L}_Y \mu = d i_Y \mu + i_Y d\mu = dH + i_Z d\mu = g\mu.$$

□

Exercise 4.16. Consider the contact manifold $(\mathbb{R}^{2n-1}, \bar{\mu})$. Let $H(q, p, z)$ be a smooth function. Show that the corresponding vector field Y_H is given by

$$(H_p, -H_q + pH_z, -p \cdot H_p + H).$$

Hint: Use $i_Y \bar{\mu} = H$, $i_Y d\bar{\mu} = -dH + (i_X dH)\mu$.

Exercise 4.17. Let S^{2n-1} be a unit sphere in $\mathbb{R}^{2n}$. The form $\bar{\lambda}$ induces a 1-form $\bar{\mu}$ on S^{2n-1}. Show that $(S^{2n-1}, \bar{\mu})$ is not a contact manifold. Define $\mu = \frac{1}{2}(p \cdot dq - q \cdot dp)$. Show that (S^{2n-1}, μ) is a contact manifold. Find its Reeb flow.

Exercise 4.18. Let μ be a 1-form on S with the following property: If $i_X \mu = i_Y \mu \equiv 0$ then $i_{[X,Y]} \mu \equiv 0$ where $[X, Y]$ is the Lie bracket of X and Y. Deduce that for such a 1-form μ, we have $\mu \wedge d\mu = 0$, hence (S, μ) is not a contact manifold.

Hint: Use $i_{[X,Y]} = i_X d i_Y - i_Y d i_X$, to deduce

$$2d\mu(X, Y) = i_Y \mathcal{L}_X \mu - i_X \mathcal{L}_Y \mu + i_{[X,Y]} \mu.$$

5 Variational Principle and Convex Hamiltonian

In this section, we use variational techniques to prove the following result of A. Weinstein:

Theorem 5.1. *Assume that the hypersurface $S \subseteq \mathbb{R}^{2n}$ is the smooth boundary of a compact strictly convex region. Then the Reeb's vector field on S has a periodic orbit.*

Let us first discuss Lagrangian formulation of Hamiltonian systems. To study Newton's equation with constraints, Lagrange initiated variational formulation of conservative mechanical problems. Let $L : TN \to \mathbb{R}$ be a C^1-function. Given $q^0, q^1 \in N$, we define $\mathcal{B} : \Gamma_T(q^0, q^1) \to \mathbb{R}$ with

$$\Gamma_T(q^0, q^1) = \{\gamma : [0,T] \to N \text{ is } C^1 \text{ and } \gamma(0) = q^0,\ \gamma(T) = q^1\},$$
$$\mathcal{B}(\gamma) = \int_0^T L(\gamma(t), \dot\gamma(t))dt.$$

Let us denote the argument of L by (q, v) with $v \in T_qN$ and write L_q and L_v for the partial derivatives of L.

We now claim that if $q(\cdot)$ is a critical point of $\mathcal{B}$, then q solves the Euler–Lagrange–Newton's equation

$$\frac{d}{dt}L_v(q, \dot q) = L_q(q, \dot q). \tag{5.1}$$

Indeed for every $\gamma \in \Gamma_T(q^0, q^1)$, the (Gâteau) derivative of $\mathcal{B}$ is given by

$$\partial\mathcal{B}(\gamma) = L_q(\gamma, \dot\gamma) - \frac{d}{dt}L_v(\gamma, \dot\gamma). \tag{5.2}$$

To see this, let $\eta : (-\delta, \delta) \to \Gamma_T(q^0, q^1)$ be a path with $\eta(0) = \gamma$ and $\eta'(0) = \tau$. The latter means that if $\eta = \eta(\theta, t)$ with $\theta \in (-\delta, \delta)$ and $t \in [0, T]$, then $\eta_\theta(0, t) = \tau(t)$. Now

$$\begin{aligned}
\langle \partial\mathcal{B}(\gamma), \tau\rangle &:= \frac{d}{d\theta}\mathcal{B}(\eta(\theta, \cdot))\Big|_{\theta=0} \\
&= \int_0^T \left[L_q(\gamma, \dot\gamma) \cdot \tau + L_v(\gamma, \dot\gamma) \cdot \dot\tau\right] dt \\
&= \int_0^T \left[L_q(\gamma, \dot\gamma) - \frac{d}{dt}L_v(\gamma, \dot\gamma)\right] \cdot \tau\ dt.
\end{aligned}$$

Here we used the fact that $\tau(0) = \tau(T) = 0$ which follows from $\eta(\theta, 0) = q^0$, $\eta(\theta, T) = q^1$ for all θ. As an example, let $L(q, v) = \frac{m}{2}|v|^2 - V(q)$ in $\mathbb{R}^n$. Then the equation (5.1) reads as $m\ddot q = -\nabla V(q)$.

Imagine that we can solve the relationship

$$p = L_v(q, v)$$

for v. Denoting the solution by $v(q,p)$, and setting

$$H(q,p) = p \cdot v(p,q) - L(q, v(q,p)),$$

we learn

$$\begin{aligned} H_p &= v + p \cdot v_p - L_v \cdot v_p = v + p \cdot v_p - p \cdot v_p = v, \\ H_q &= p v_q - L_q(q,v) - L_v(q,v) v_q = -L_q(q,v). \end{aligned}$$

Hence

$$p = L_v(q,v) \Leftrightarrow H_p(q,p) = v, \tag{5.3}$$

and

$$H_q(q,p) = -L_q(q,v). \tag{5.4}$$

If we set $p(t) = L_v(q(t), \dot{q}(t))$, then by (5.1), (5.3) and (5.4)

$$\dot{q} = H_p(q,p), \quad \dot{p} = -H_q(q,p).$$

The inversion (5.3) is possible if we assume that L is convex. In this case H can be constructed from L by *Legendre transform*:

$$H(q,p) = \sup_v (p \cdot v - L(q,v)).$$

If we assume that L has superlinear growth in variable v as $|v| \to \infty$, then H is finite and the supremum is attained at $v = v(q,p)$ so that

$$L_v(q, v(q,p)) = p,$$

$$H(q,p) = p \cdot v(q,p) - L(q, v(q,p)).$$

Upon differentiation we obtain

$$H_p = v + p v_p - L_v v_p = v.$$

For a v-convex L, we may find solutions to (5.1) by minimizing $\mathcal{B}$. For example, if for some constants $c_1, c_2 > 0$ and $\alpha > 1$,

$$L(q,v) \geq c_1 |v|^\alpha - c_2, \tag{5.5}$$

for every $v \in T_qN$ and $q \in N$, and if γ_l is a sequence in $\Gamma_T(q^0, q^1)$ such that $\lim_{l\to\infty} \mathcal{B}(\gamma_l) = A = \inf \mathcal{B}$, then by (5.5) we have the bound

$$\sup_l \int_0^T |\dot{\gamma}_l(t)|^\alpha dt < \infty.$$

This bound allows us to extract a subsequence of γ_l which converges weakly with respect to the topology of Sobolev space $W^{1,\alpha}$. It turns out that $\mathcal{B}$ is lower semicontinuous because L is convex in v. This allows us to deduce that for any limit point $q(\cdot)$ of the sequence $\gamma_l(\cdot)$, we have that $\mathcal{B}(q) \leq A$. Since $A = \inf \mathcal{B}$, we learn that $\mathcal{B}(q) = A$ and that the infimum is achieved.

In spite of the appeal of the above argument, it is not clear how we can use it to prove Theorem 5.1. Recall that we are searching for a periodic solution on a given energy surface. Of course we could have defined $\mathcal{B}$ on the space of T-periodic paths, and under some additional assumptions, find some T-periodic orbit of (5.1). The point is that such an orbit may not lie on the surface S. To this end let us define another functional $\mathcal{A}$ of which critical points solve the Hamiltonian system. Set $\mathcal{A} : \Gamma_T \to \mathbb{R}$ to be

$$\begin{aligned} \mathcal{A}(x(\cdot)) &= \int_{x(\cdot)} [p \cdot dq - H(x)]\ dt \\ &= \int_0^T \left[\frac{1}{2} Jx \cdot \dot{x} - H(x)\right] dt, \end{aligned}$$

where

$$\Gamma_T = \{x : \mathbb{R} \to \mathbb{R}^{2n} \text{ is } C^1 \text{ and } T\text{-periodic}\}.$$

We have

$$\partial\mathcal{A}(x(\cdot)) = -J\dot{x} - \nabla H(x), \tag{5.6}$$

because

$$\begin{aligned} \langle \partial\mathcal{A}(x(\cdot)), \tau(\cdot)\rangle &= \frac{d}{d\delta}\mathcal{A}(x + \delta\tau)\Big|_{\delta=0} \\ &= \int_0^T \left[\frac{1}{2}J\tau \cdot \dot{x} + \frac{1}{2}Jx \cdot \dot{\tau} - \nabla H(x) \cdot \tau\right]\ dt \\ &= -\int_0^T (J\dot{x} + \nabla H(x)) \cdot \tau\ dt \end{aligned}$$

for every $x, \tau \in \Gamma_T$. From (5.6) we learn that $\partial\mathcal{A}(x(\cdot)) \equiv 0$ iff x solves

$$\dot{x} = J\nabla H(x). \tag{5.7}$$

Note that $\mathcal{A}$ involves H explicitly and no additional assumption such as convexity is needed. However typically the critical points of $\mathcal{A}$ are saddle points and it is helpless to search for (local) maximizers or minimizers. Because of this, finding critical points for $\mathcal{A}$ is far more challenging. Before discovering a remedy for this, let us show that a hypersurface S as in Theorem 5.1 can be realized as a level set of a convex homogeneous Hamiltonian function.

Lemma 5.2. *Let S be as in Theorem 5.1 and assume that the origin is inside S. Then there exists a strictly convex C^1 function $H : \mathbb{R}^{2n} \to \mathbb{R}$ such that $H(0) = 0$, $S = \{x : H(x) = 1\}$ and $H(\lambda x) = \lambda^2 H(x)$ for $\lambda \geq 0$.*

Proof. First define $F : \mathbb{R}^{2n} \to \mathbb{R}$ for $F(x) = \rho$ when $x \in \rho S$ and $x \neq 0$. Set $F(0) = 0$. Then F is a norm and S is the unit sphere for this norm. Evidently F is convex and homogeneous of degree 1. However F is not differentiable at 0 and is strictly convex only tangent to S. That is,

$$D^2F(x)|_{T_xS} > 0, \tag{1}$$

for $x \in S$, but $D^2F(x)x = 0$. The latter follows from the homogeneity of F; $x \cdot \nabla F = F$ and $D^2F(x)x = 0$. To see (5.8), let us look at the curvatures of S. Note that $n(x) = \frac{\nabla F(x)}{|\nabla F(x)|}$ is the unit normal to S and

$$Dn(x) = |\nabla F(x)|^{-1}D^2F(x) + \nabla F(x) \otimes \nabla(|\nabla F(x)|^{-1}).$$

As a result,

$$Dn(x)a \cdot a = |\nabla F(x)|^{-1}D^2F(x)a \cdot a + (\nabla F(x) \cdot a)(\nabla|\nabla F(x)|^{-1} \cdot a).$$

As a result, if $a \in T_xS$, then

$$Dn(x)a \cdot a = |\nabla F(x)|^{-1}D^2F(x)a \cdot a. \tag{5.9}$$

Since S is strictly convex, we have $Dn(x) > 0$. This and (5.9) imply (5.8).

We now define $H(x) = (F(x))^2$. Evidently H is C^1 and C^2 off the origin. Since H is homogeneous of degree 2,

$$\begin{aligned} \nabla H &= 2F\nabla F, \\ D^2H &= 2\nabla F \otimes \nabla F + 2FD^2F. \end{aligned}$$

Hence,

$$D^2H(x)a \cdot a = 2(\nabla F(x) \cdot a)^2 + 2F(x)D^2F(x)a \cdot a.$$

Note that if $a = b + c$ with $b \| \nabla F(x)$ and $c \in T_xS$, then $(\nabla F(x) \cdot a)^2 > 0$ if $b \neq 0$ and $D^2F(x)a \cdot a > 0$ if $b = 0$ and $c \neq 0$. Hence $D^2H(x) > 0$ for all $x \neq 0$. □

Let us use H of Lemma 5.2 to study the corresponding function $\mathcal{A}$. Note that $\mathcal{A} = \mathcal{A}_T$ is defined for T-periodic functions whereas for Theorem 5.1 we need a periodic orbit on S of some period. Of course if $x(\cdot)$ is such a periodic orbit of period T, then $y(t) = x\left(\frac{t}{T}\right)$ is 1-periodic and

$$\dot{y} = \rho J \nabla H(y) \tag{5.10}$$

for $\rho = T^{-1}$. In view of the form of functional $\mathcal{A}_T$, perhaps we should fix the period to be 1 always and now insist that $y(\cdot)$ solves (5.10) for some ρ. As a result, we now want to find the critical point of

$$\int_0^1 \left[\frac{1}{2} Jx \cdot \dot{x} - \rho H(x)\right] dt = \int_0^1 \frac{1}{2} Jx \cdot \dot{x}\ dt - \rho \int_0^1 H(x)\ dt$$

with x a 1-periodic path and some $\rho \in \mathbb{R}$. The scalar ρ resembles a Lagrange multiplier that now is employed for a functional defined on an infinite dimensional space. Motivated by this, define $\mathcal{C} : \Lambda \to \mathbb{R}$ by

$$\mathcal{C}(x(\cdot)) = \int_0^1 H(x(t))\ dt$$

with

$$\Lambda = \left\{ x : \mathbb{R} \to \mathbb{R}^{2n} \text{ is } C^1, \text{ 1-periodic and } \int_0^1 \frac{1}{2} Jx \cdot \dot{x}\ dt = 1 \right\}.$$

Lemma 5.3. *Let $x(\cdot)$ be a critical point of $\mathcal{C} : \Lambda \to \mathbb{R}$. Then there exists a constant η such that $\eta\dot{x} = J\nabla H(x)$. Moreover, if H is as in Lemma 5.2, then $\eta \neq 0$ and $y(t) = \eta^{-1/2} x(\eta t)$ solves $\dot{y} = J\nabla H(y)$ and $H(y) \equiv 1$.*

Proof. First, let us determine the space $T_{x(\cdot)}\Lambda$. Take a path $z : (-\delta, \delta) \to \Lambda$ with $z(0) = x(\cdot)$ and $\frac{dz}{d\theta}(0) = \tau(\cdot)$. Regarding z as $z : (-\delta, \delta) \times S^1 \to \mathbb{R}^{2n}$, we have that $z(0, t) = x(t)$, and $\frac{\partial z}{d\theta}(0, t) = \tau(t)$. The condition

$$\int_0^1 \frac{1}{2} Jz \cdot \dot{z}\ dt = 1$$

can be differentiated with respect to θ to yield

$$0 = \int_0^1 \left[\frac{1}{2} Jx \cdot \dot{\tau} + \frac{1}{2} J\tau \cdot \dot{x}\right] dt = -\int_0^1 J\dot{x} \cdot \tau\ dt.$$

So, $\tau \in T_{x(\cdot)}\Lambda$ iff $\int_0^1 J\dot{x} \cdot \tau\ dt = 0$. Now, if x is critical for $\mathcal{C}$, then $\frac{d}{d\theta}\mathcal{C}(z(\cdot, \theta))\big|_{\theta=0} = 0$. As a result, for a critical $x(\cdot)$,

$$\int_0^1 J\dot{x} \cdot \tau\ dt = 0 \Rightarrow \int_0^1 \nabla H(x) \cdot \tau\ dt = 0. \tag{5.11}$$

From this, it is not hard to deduce that for some $\eta \in \mathbb{R}$, we have that $\eta J\dot{x} + \nabla H(x) \equiv 0$. Indeed, we choose $\tau = \eta J\dot{x} + \nabla H(x)$ which satisfies $\int_0^1 J\dot{x} \cdot \tau \, dt = 0$ if

$$\eta \int_0^1 |\dot{x}|^2 + \int_0^1 J\dot{x} \cdot \nabla H(x) = 0.$$

Such η exists because $\dot{x} \neq 0$. Since $\int_0^1 \nabla H(x) \cdot \tau \, dt = 0$ for such a choice, we learn

$$\int_0^1 |\tau|^2 = \eta \int_0^1 J\dot{x} \cdot \tau \, dt + \int_0^1 \nabla H(x) \cdot \tau \, dt = 0.$$

Hence $\tau \equiv 0$.

Note that if H is homogeneous, then $\eta \neq 0$ because if $\eta = 0$, then $\nabla H(x) = 0$ and this means that $H(x) = x \cdot \nabla H(x) = 0$. But for $x \in \Lambda$ we never have $x = 0$. Now if $y(t) = \eta^{-1/2} x(\eta t)$, then

$$\begin{aligned}
\dot{y}(t) &= \eta^{1/2} \dot{x}(\eta t) = \eta^{-1/2} J \nabla H(x(\eta t)) \\
&= J\nabla H(\eta^{-1/2} x(\eta t)) = J \nabla H(y(t)),
\end{aligned}$$

and

$$\begin{aligned}
\int_0^{\eta^{-1}} H(y(t)) \, dt &= \int_0^{\eta^{-1}} H(\eta^{-1/2} x(\eta t)) \, dt \\
&= \eta^{-2} \int_0^1 H(x(t)) \, dt \\
&= 2\eta^{-2} \int_0^1 x \cdot \nabla H(x) \, dt \\
&= -2\eta^{-2} \int_0^1 \eta J\dot{x} \cdot x \, dt = \eta^{-1}.
\end{aligned}$$

Since $H(y(t))$ is constant, we deduce that $H(y(\cdot)) \equiv 1$. □

On account of Lemma 5.3, we only need to find critical points for $\mathcal{C}$ on Λ. Since H is convex, we may wonder whether or not a minimum or maximum provides us with a critical point. It turns out that $\sup_\Lambda \mathcal{C} = +\infty$ and $\inf_\Lambda \mathcal{C} = 0$. Note that for H as in Lemma 5.2 infimum is not achieved because if $\mathcal{C}(x(\cdot)) = 0$, then $x(\cdot) \equiv 0$ and $\frac{1}{2}\int_0^1 Jx \cdot \dot{x} \neq 1$ for such $x(\cdot)$. Let us study an example to see why the infimum is not achieved.

Example 5.4. Assume that $n = 1$ and $H : \mathbb{R}^2 \to \mathbb{R}$ is given by $H(x) = \pi |x|^2$. We already know how to solve the corresponding system $\dot{x} = J\nabla H(x)$. The solutions are given by $x(t) = (q(t), p(t)) = r(\sin 2\pi t, \cos 2\pi t)$, $t \geq 0$. The set $\{x : H(x) = 1\}$ carries the periodic orbit $x(t) = \frac{1}{\sqrt{\pi}}(\sin 2\pi t, \cos 2\pi t)$. The set Λ consists of 1-periodic y with $\frac{1}{2}\int_0^1 Jy\cdot\dot{y}\,dt = 1$. If y is a simple curve with coordinates q and p, then $\frac{1}{2}\int_0^1 Jy\cdot\dot{y}\,dt = \int_0^1 \frac{1}{2}(p\cdot\dot{q} - q\cdot\dot{p})\,dt$ is the area enclosed by y. We can easily construct a sequence $y_l(t)$ in Λ with $\mathcal{C}(y_l)\to\infty$ as $l\to\infty$. For example, set $y_l(t) = \left(\sqrt{l}\sin 2\pi t, \frac{1}{\pi\sqrt{l}}\cos 2\pi t\right)$, so that $y_l\in\Lambda$ and $\mathcal{C}(y_l) = \frac{\pi l}{2} + \frac{1}{2\pi l}$. This confirms $\sup_\Lambda \mathcal{C} = +\infty$. As for the infimum, let us choose a sequence $z_l\in\Lambda$ of high oscillation, say $z_l(t) = \frac{1}{\pi l}(\sin 2\pi l t, \cos 2\pi l t)$. Since $H(z_l) = \pi l^{-2}$, we learn that $\inf_\Lambda \mathcal{C} = 0$.

□

From Example 5.4 it is clear that a control on $\mathcal{C}(z_l)$ for $z_l\in\Lambda$ guarantees no control on $\dot{z}_l$ and this results in $\inf_\Lambda\mathcal{C} = 0$. We now follow an idea of Clarke to switch to a new functional $\mathcal{D}$ which involves the derivative. To motivate the definition, observe that if

$$\dot{x} = \rho J\nabla H(x) \tag{5.12}$$

for some $\rho\in\mathbb{R}$, then for any constant $c\in\mathbb{R}^{2n}$,

$$\frac{d}{dt}J(c-x) = \rho\nabla H(x). \tag{5.13}$$

If H is as in Lemma 5.2, then $\rho\neq 0$ because $\dot{x}\neq 0$ for $x\in\Lambda$. On the other hand, we can always assume that $\rho>0$ because when $\rho<0$, we may switch to $-x$. Now by homogeneity, $\rho\nabla H(x) = \nabla H(\rho x)$ if $\rho>0$, and (5.13) can be rewritten as

$$\nabla G(\dot{y}) = \rho(Jy+c), \tag{5.14}$$

where y is defined by $Jy+c = x$ and G denotes the Legendre transform of H. The condition $\int_0^1 \frac{1}{2}Jx\cdot\dot{x}\,dt = 1$ becomes

$$1 = \int_0^1 \frac{1}{2}J(Jy+c)\cdot J\dot{y}\,dt = \int_0^1 \frac{1}{2}Jy\cdot\dot{y}\,dt.$$

Motivated by this and (5.13), let us define $\mathcal{D}:\Lambda\to\mathbb{R}$ by

$$\mathcal{D}(y) = \int_0^1 G(\dot{y}(t))\,dt.$$

Lemma 5.5. *If $y(\cdot)$ is a critical point of $\mathcal{D}:\Lambda\to\mathbb{R}$, then y solves* (5.14) *for suitable constants $\rho\in\mathbb{R}$ and $c\in\mathbb{R}^{2n}$. Moreover, $x(t) = Jy(t)+c$ satisfies* (5.12) *and $H(x(\cdot))\equiv\rho^{-1}$.*

Proof. As in the proof of Lemma 5.3, we have that if $\int_0^1 J\dot{y}\cdot\tau\,dt = 0$, then $\int_0^1 \nabla G(\dot{y})\cdot\dot{\tau}\,dt = 0$. If $y\in C^2$, then the latter means that $\int_0^1 \frac{d}{dt}\nabla G(\dot{y})\cdot\tau\,dt = 0$. From this we can readily deduce that for some $\rho\in\mathbb{R}$,

$$\rho J\dot{y} - \frac{d}{dt}\nabla G(\dot{y}) \equiv 0$$

and as a result, we have (5.14) for some constant c. But we only know that $y \in C^1$. To treat the problem under the mere assumption $y \in C^1$, let us search for τ such that

$$\dot{\tau} = \nabla G(\dot{y}) - \rho J y - c, \tag{5.14}$$

for some $\rho \in \mathbb{R}$ and $c \in \mathbb{R}^{2n}$. This has a solution if

$$c = \int_0^1 (\nabla G(\dot{y}) - \rho J y)\ dt. \tag{5.15}$$

We then require that $\int_0^1 J\dot{y} \cdot \tau\ dt = 0$. This and (5.15) uniquely determine c and ρ. Once this is done, we deduce that $\int_0^1 \nabla G(\dot{y}) \cdot \dot{\tau}\ dt = 0$ and

$$\int_0^1 |\dot{\tau}|^2\ dt = \int_0^1 \nabla G(\dot{y}) \cdot \dot{\tau}\ dt - \rho \int_0^1 Jy \cdot \dot{\tau} = 0.$$

Hence $\dot{\tau} \equiv 0$ and this completes the proof. □

Let us remark that even though we have worked with the space C^1, all of our results are valid for a larger space $\mathcal{H}^1$. The space $\mathcal{H}^1$ consists of 1-periodic functions $y : \mathbb{R} \to \mathbb{R}^{2n}$ which are weakly differentiable with the weak derivative $\dot{y} \in L^2$. That is, there exists $z \in L^2$ such that for every $J \in C^1$,

$$\int_0^1 y \cdot \dot{J}\ dt = -\int_0^1 z \cdot J.$$

We simply write $\dot{y}$ for the weak derivative z. Note that we can define $\mathcal{D} : \bar{\Lambda} \to \mathbb{R}$ where

$$\bar{\Lambda} = \left\{ y \in \mathcal{H}^1 : \int_0^1 \frac{1}{2} Jy \cdot \dot{y}\ dt = 1 \right\}.$$

In Lemma 5.5, we may replace Λ with $\bar{\Lambda}$. We are now ready for the proof of Theorem 5.1.

Proof of Theorem 5.1. On account of Lemma 5.5, it suffices to find a critical point of $\mathcal{D} : \bar{\Lambda} \to \mathbb{R}$. This can be achieved by showing the existence of a minimizer of $\mathcal{D}$. First we obtain some useful properties of G. Note that by homogeneity of H,

$$H(x) = |x|^2 H\left(\frac{x}{|x|}\right).$$

As a result, there are positive constants c_1 and c_2, such that

$$c_1 \frac{|x|^2}{2} \le H(x) \le c_2 \frac{|x|^2}{2}$$

for all $x \in \mathbb{R}^{2n}$. Since $G(a) = \sup_x(a \cdot x - H(x))$, we obtain

$$\frac{|a|^2}{2c_2} \leq G(a) \leq \frac{|a|^2}{2c_1}. \tag{5.16}$$

On the other hand G is also homogeneous because

$$\lambda^{-2}G(\lambda a) = \sup_x(a \cdot \lambda^{-1}x - \lambda^{-2}H(x)) = \sup_z(a \cdot z - H(z)) = G(a).$$

From this and $\nabla G(a) = |a|\nabla G\left(\frac{a}{|a|}\right)$, we learn that for some constant c_3,

$$|\nabla G(a)| \leq c_3|a|. \tag{5.17}$$

Set $A = \inf_{\bar{\Lambda}} \mathcal{D}$ and choose a sequence $y_l \in \bar{\Lambda}$ such that $\mathcal{D}(y_l) \to A$. In view of (5.16), we certainly have

$$\sup_l \int_0^1 |\dot{y}_l|^2 \, dt < \infty. \tag{5.18}$$

We need to control y_l as well. For this, observe that if $z \in \bar{\Lambda}$, then $\hat{z} = z + \text{const.} \in \bar{\Lambda}$ as well and $\mathcal{D}(\hat{z}) = \mathcal{D}(z)$. Hence we may also assume that $\int_0^1 y_l(t)\, dt = 0$ for every l. From this and (5.18), it is not hard to deduce

$$\sup_l |y_l(0)| < \infty. \tag{5.19}$$

By Exercise 5.6 below we know that if $y \in \mathcal{H}^1$, then $y(t) = \int_0^1 \dot{y}(\theta)d\theta$ for almost all t. Hence

$$|y_l(t) - y_l(s)| = \left|\int_s^t \dot{y}(\theta)d\theta\right| \leq |t-s|^{1/2}\left(\int_s^t |\dot{y}_l(\theta)|^2 d\theta\right)^{1/2}.$$

From this and (5.18) we deduce

$$\sup_l \sup_{t \neq s} \frac{|y_l(t) - y_l(s)|}{|t-s|^{1/2}} < \infty.$$

This and (5.19) imply that y_l has a convergent subsequence with respect to uniform topology. In view of (5.18) we may choose a subsequence such that

$$y_l \to y \text{ uniformly,}$$
$$\dot{y}_l \to z \text{ weakly,}$$

for some $z \in L^2$ and continuous y. We now assert that in fact y is weakly differentiable and z is the weak derivative of y. Indeed if $J \in C^1$, then

$$\int_0^1 y \cdot \dot{J}\, dt = \lim_{l\to\infty} \int_0^1 y_l \cdot \dot{J}\, dt = -\lim_{l\to\infty} \int_0^1 \dot{y}_l \cdot J\, dt = -\int_0^1 z \cdot J\, dt.$$

It remains to show that $y \in \bar{\Lambda}$ and

$$A = \mathcal{D}(y). \tag{5.20}$$

For the former observe

$$\begin{aligned}\int_0^1 \frac{1}{2} Jy \cdot \dot{y}\, dt &= \int_0^1 \frac{1}{2} Jy \cdot (\dot{y} - \dot{y}_l)\, dt + \int_0^1 \frac{1}{2} J(y - y_l) \cdot \dot{y}_l\, dt + \int_0^1 \frac{1}{2} Jy_l \cdot \dot{y}_l\, dt \\ &=: \Omega_1 + \Omega_2 + \Omega_3.\end{aligned}$$

We certainly have that $\lim_{l\to\infty} \Omega_1 = 0$ and $\Omega_3 = 1$. Moreover

$$|\Omega_2| \le \frac{1}{2} \|y - y_l\|_{L^2} \|\dot{y}_l\|_{L^2} \to 0$$

as $l \to \infty$ by (5.18) and uniform convergence of y_l to y. This shows that $y \in \bar{\Lambda}$.

It remains to establish (5.20). Since $A = \inf \mathcal{D}$, it suffices to show that $\mathcal{D}(y) \le A$. This follows from lower semi-continuity of the functional $\mathcal{D}$ which is a consequence of the convexity of G. Indeed by convexity of G,

$$G(\dot{y}) + \nabla G(\dot{y}) \cdot (\dot{y}_l - \dot{y}) \le G(\dot{y}_l).$$

Hence,

$$\int_0^1 G(\dot{y})\, dt + \int_0^1 \nabla G(\dot{y}) \cdot (\dot{y}_l - \dot{y})\, dt \le \int_0^1 \nabla G(\dot{y}_l)\, dt.$$

We now send $l \to \infty$. Since $\nabla G(\dot{y}) \in L^2$ by (5.17) and $\dot{y} \in L^2$, we know that the second term on the left-hand side goes to 0. As a result,

$$\int_0^1 G(\dot{y})\, dt \le \liminf_{l\to\infty} \int_0^1 G(\dot{y}_l) = A.$$

So far we know that for some $x \in \mathcal{H}^1$, and $\rho \in \mathbb{R}$, $\dot{x} = \rho J \nabla H(x)$. Since the right-hand side is continuous, we use Exercise 5.6(i) to deduce that for almost all t,

$$x(t) = x(0) + \int_0^1 \rho J \nabla H(x(s))\, ds.$$

From this we deduce that in fact $x \in C^1$. □

Exercise 5.6.

- (i) Let $y \in \mathcal{H}^1$. Show that $y(t) = \int_0^t \dot{y}(\theta)d\theta$ for almost all t. (Hint: Use a test function J that approximates $1\!\!1_{[s,t]}$.)
- (ii) Prove that if $y \in \mathcal{H}^1$ and $\int_0^1 y\ dt = 0$, then

$$\|y\|_{L^2} \le \frac{1}{2\pi}\|\dot{y}\|_{L^2}.$$

□

With the aid of the functional $\mathcal{D} : \Lambda \to \mathbb{R}$, we were able to prove the existence of a periodic orbit on a hypersurface $S = H^{-1}(\{1\})$. We now would like to understand the relationship between the minimizer $y(\cdot)$, the minimum value $\mathcal{D}(y(\cdot))$ and the periodic orbit we have constructed on S. Recall that on the account of Lemma 5.5, the minimizer satisfies

$$\nabla G(\dot{y}) = \rho(Jy - c) =: \rho x,$$

$$H(x(\cdot)) \equiv \rho^{-1}.$$

Note that $G(\dot{y}) = a \cdot \dot{y} - H(a)$ for a satisfying $\nabla H(a) = \dot{y}$. As a result, $a = \rho x$ and

$$\begin{aligned} \int_0^1 G(\dot{y})\ dt &= \int_0^1 [\rho x \cdot \dot{y} - H(\rho x)]\ dt \\ &= \rho \int_0^1 Jy \cdot \dot{y}\ dt - \rho^2 \int_0^1 H(x)\ dt = 2\rho - \rho = \rho. \end{aligned} \tag{5.21}$$

On the other hand, if $z(t) = \sqrt{\rho}x(t/\rho)$, then

$$\dot{z} = J\nabla H(z), \quad H(z(\cdot)) \equiv 1.$$

Hence $z(\cdot)$ is a periodic orbit of X_H that lies on S and the corresponding $\int_0^1 G(\dot{y})\ dt = \rho$ is its period. Also note

$$\begin{aligned} \int_0^\rho \frac{1}{2}Jz \cdot \dot{z}\ dt &= \int_0^\rho \frac{1}{2}Jz \cdot J\nabla H(z)\ dt = \int_0^\rho \frac{1}{2}z \cdot \nabla H(z)\ dt \\ &= \int_0^\rho H(z)\ dt = \rho. \end{aligned} \tag{5.22}$$

In fact we have the following:

Proposition 5.7. *Let S and $\mathcal{D}$ be as in Theorem* 5.1 *and Lemma* 5.5. *Then*

$$\inf_{y \in \Lambda} \mathcal{D}(y) = \inf\{\rho : \text{there exists a } \rho\text{-periodic orbit of } X_H \text{ on } S\}. \tag{5.23}$$

Proof. By the preceding discussion, we know that the left-hand side of (5.22) is less than the right-hand side. For the reverse inequality, start with a ρ-periodic orbit Z of X_H and define $x(t) = \frac{1}{\sqrt{\rho}} z(t\rho)$. We certainly have that $\dot{x} = \rho J \nabla H(x)$ and $H(x(\cdot)) \equiv \rho^{-1}$. We then set $y = J(c - x)$ with $c = \int_0^1 x \, dt$. We have

$$\begin{aligned}\int_0^1 \frac{1}{2} Jy \cdot \dot{y} \, dt &= \int_0^1 \frac{1}{2} JJ(c - x) \cdot J(-\dot{x}) \, dt \\ &= \int_0^1 \frac{1}{2} Jx \cdot \dot{x} \, dt = \rho \int_0^1 \frac{1}{2} Jx \cdot J\nabla H(x) \, dt \\ &= \rho \int_0^1 H(x) \, dt = 1.\end{aligned}$$

On the other hand,

$$\dot{y} = -J\dot{x} = -\rho JJ\nabla H(x) = \rho \nabla H(x),$$

so $\nabla G(\dot{y}) = \rho x$, and $\int_0^1 G(\dot{y}) \, dt = \rho$, as in (5.21). From this we readily deduce that the right-hand side of (5.21) is less than the left-hand side. We are done. □

Remark 5.8. Note that if γ is a periodic characteristic the Reeb's vector field of period ρ, then

$$\int_\gamma \bar{\lambda} = \int_\gamma p \cdot dq = \int_0^\rho \frac{1}{2} Jz \cdot \dot{z} \, dt = \rho$$

by (5.22). Hence the right-hand side of (5.23) can be written as

$$\inf \left\{ \left| \int_\gamma \bar{\lambda} \right| : \gamma \text{ is a periodic characteristic of the Reeb's vector field of } S \right\}.$$

We will see later that this number is the Hofer–Zehnder capacity of the convex set C with $\partial C = S$. Alternatively,

$$c(C) = \inf \left\{ \left| \int_\Gamma \bar{\omega} \right| : \Gamma \text{ is a 2-surface, } \Gamma \subseteq S,\ \partial\Gamma \text{ is a characteristic of Reeb's vector field} \right\}.$$

Remark 5.9. Note that if $\rho = \inf_{y \in \Lambda} \mathcal{D}(y)$, then X_H has no periodic orbit of period $T \in (0, \rho)$. But if $K = \rho H$, then we can say that X_K has no periodic orbit of period $T \in (0, 1)$. In fact K is a Hamiltonian function with $\sup_{C^o} K < \rho$ because $C = \{x : H(x) \leq 1\}$. Defining $\|K\| = \sup K - \inf K$, we learn that $\|K\| < \rho$ on C^o. We learn later that ρ can be realized as

$$\sup_{K \in \hat{\mathcal{H}}(C^0)} \|K\|$$

where $\hat{\mathcal{H}}(C^0)$ denotes the set of Hamiltonian functions of compact support in C^0 for which X_K has no periodic orbit of period $T \in (0, 1]$.

6 Capacity and Its Applications

In this section we assume the existence of a capacity for the symplectic manifolds and deduce several properties of Hamiltonian systems. As we mentioned in the introduction the non-squeezing theorem of Gromov inspired the search for symplectic capacities.

A *symplectic capacity* c assigns a number $c(M,\omega) \in [0,\infty]$ to every symplectic manifold (M,ω) with the following properties:

(6.1)
(i) If $\psi : (M_1,\omega_1) \to (M_1,\omega_2)$ with ψ symplectic and injective, and $dim(M_1) = dim(M_2)$, then, $c(M_1,\omega_1) \le c(M_2,\omega_2)$.
(ii) $c(M,\lambda\omega) = \lambda c(M,\omega)$ for $\lambda > 0$,
(iii) $c(B^{2n}(1),\bar{\omega}) > 0$, $c(Z^{2n}(1),\bar{\omega}) < \infty$ where $B^{2n}(1)$ is the Euclidean ball of radius 1, and $Z^{2n}(1)$ is the cylinder $\{(q,p) : q_1^2 + p_1^2 \le 1\}$.

Note that the condition $c(B^{2n}(1),\bar{\omega}) > 0$, guarantees that $c \equiv 0$ is not a capacity. The requirement $c(Z^{2n}(1)) < \infty$, disqualifies $c(M,\omega) = |\int_M \omega^n|^{1/n}$, $n = dim(M)$, to be a capacity.

We may also consider subsets of $\mathbb{R}^{2n}$ and define capacities for these subsets. In this case a symplectic capacity c assigns a number $c(A)$ to every $A \subseteq \mathbb{R}^{2n}$ such that

(6.2)
(i) If there exists an injective symplectic $\psi : U \to \mathbb{R}^{2n}$ with $U \subseteq \mathbb{R}^{2n}$ open and for $A \subseteq U$ we have $\psi(A) \subseteq B$, then $c(A) \le c(B)$.
(ii) $c(\lambda A) = \lambda^2 c(A)$ for $\lambda > 0$,
(iii) $c(B^{2n}(1)) > 0$ and $c(Z^{2n}(1)) < \infty$.

We may also replace the monotonicity assumption $(6.2(i))$ with a weaker assumption

(6.3)
(i)′ If $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is a symplectomorphism (i.e., a symplectic diffeomorphism with $\psi(\mathbb{R}^{2n}) = \mathbb{R}^{2n}$) and $\psi(A) \subseteq B$ then $c(A) \le c(B)$.

Another possibility is that we require a stronger assumption

$$c(B^{2n}(1)) = c(Z^{2n}(1)) = \pi. \tag{6.4}$$

We now discuss how the existence of a capacity can be used to establish several basis results in symplectic geometry. Here are the main theorems of this section.

Theorem 6.1 (Gromov). *If there exists a symplectic injective* $\psi : B^{2n}(r) \to \mathbb{R}^{2n}$ *with* $\psi(B^{2n}(r)) \subseteq Z^{2n}(R)$, *then* $R \ge r$.

Theorem 6.2 (Gromov, Eliashberg). *If* $\psi_k : (M,\omega) \to (M,\omega)$ *are symplectomorphisms such that* $\psi_k \to \psi$ *uniformly with* ψ *a diffeomorphism, then* ψ *is symplectic.*

Theorem 6.3 (Viterbo). *If S is a hypersurface of $\mathbb{R}^{2n}$ of contact type and $S = \{x : H(x) = 1\}$ for a regular $H : \mathbb{R}^{2n} \to \mathbb{R}$, then flow of X_H has a periodic orbit on S.*

Throughout this section we assume that a capacity exists:

Theorem 6.4. *There exists a capacity $c(M, \omega)$ that satisfies* (6.1(i–iii)) *and* $c(B^{2n}(1), \bar{\omega}) = c(Z^{2n}(1), \bar{\omega}) = \pi$.

We remark that the existence of $c(M, \omega)$ implies the existence of $c(\cdot)$ satisfying (6.2) and (6.4). Indeed we may define

$$c(A) = \inf\{c(U, \bar{\omega}) : A \subseteq U,\ U \text{ open in } \mathbb{R}^{2n}\}. \tag{6.5}$$

Exercise 6.5. Show that c given by (6.5) satisfies (6.2) and (6.4).

As an immediate consequence of Theorem 6.4 we have Theorem 6.1.

Proof of Theorem 6.1. If such a ψ exists, then

$$c(B^{2n}(r)) = c(\psi(B^{2n}(r)) \le c(Z^{2n}(R)).$$

On the other hand $c(B^{2n}(r)) = c(rB^{2n}(1)) = r^2\pi$, $c(Z^{2n}(R)) = R^2\pi$. Thus $\pi r^2 \le \pi R^2$, we are done. □

We now turn to Theorem 6.2. The key idea of the proof is an equivalent definition for symplecticity of a transformation that does not involve any derivative. This should be compared to the notion of a volume preserving transformations. We can say that a transformation is volume preserving if its Jacobian is one or equivalently it preserves the volume. The latter criterion is more useful when we want to show that a limit of a sequence of measure preserving transformations is again measure preserving. To prepare for the proof of Theorem 6.2, let us state a simple consequence of Theorem 6.4.

Lemma 6.7. *Let c be as in* (6.5) *and assume A is an ellipsoid with (symplectic) spectrum $0 < r_1(A) \le \cdots \le r_n(A)$. Then $c(A) = \pi r_1(A)^2$.*

Proof. By Theorem 2.2, A is symplectomorphic to the ellipsoid

$$E = \left\{ x : \sum_1^n r_j^{-2}(q_j^2 + p_j^2) \le 1 \right\}.$$

Hence $c(A) = c(E)$. On the other hand,

$$B^{2n}(r_1) \subseteq E \subseteq Z^{2n}(r_1).$$

As a result, $\pi r_1^2 \le c(E) = c(A) \le \pi r_1^2$. This completes the proof. $\square$

Theorem 6.8. *Let $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be a diffeomorphism. Then the following statements are equivalent:*

(i) *For every ellipsoid E, $c(\psi(E)) = c(E)$.*

(ii) *Either $\psi^*\bar{\omega} = \bar{\omega}$ or $\psi^*\bar{\omega} = -\bar{\omega}$.*

Proof. Evidently if $\psi^*\bar{\omega} = \bar{\omega}$ then (i) is true. If $\psi^*\bar{\omega} = -\bar{\omega}$, then $\hat{\psi}^*\bar{\omega} = \bar{\omega}$ for $\hat{\psi} = \psi \circ \tau$, where τ is given by

$$\tau a = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} a. \tag{6.6}$$

We then have that for an ellipsoid E, $c(\psi(\tau(E))) = c(E)$. On the other hand, if E is the ellipsoid

$$\sum_{i=1}^{n} r_i^{-2}(q_i^2 + p_i^2) \le 1 \tag{6.7}$$

then $\tau(E) = E$. Hence for such ellipsoids, $c(\psi(E)) = c(E)$. Since for any symplectic φ we have $(\psi \circ \varphi)^*\bar{\omega} = -\bar{\omega}$, we also have $c(\varphi(E)) = c(\psi(\varphi(E)))$. Now any ellipsoid can be represented as $\varphi(E)$ for some linear symplectic φ and standard E given by (6.7). This completes the proof of (ii) $\Rightarrow$ (i).

For the converse, assume (i) is true and define

$$\phi_k(x) = k(\psi(a + k^{-1}x) - \psi(a)).$$

We have that $\phi_k(x) \to (D\psi)(a)x$ locally uniformly as $k \to \infty$. Note that if E is an ellipsoid and ψ satisfies (i), then

$$\begin{aligned} c(\phi_k(E)) &= c(k\psi(a + k^{-1}E) - k\psi(a)) \\ &= k^2 c(\psi(a + k^{-1}E)) \\ &= k^2 c(k^{-1}E) = c(E). \end{aligned}$$

On the other hand, it follows from Lemma 6.9 below that $\lim \phi_k(\cdot) = \psi'(a)$ also satisfies (i). We finally our linear algebra Theorem 2.15 to deduce that the linear $A = \psi'(a)$ satisfies $A^*\bar{\omega} = \bar{\omega}$ or $A^*\bar{\omega} = -\bar{\omega}$. By continuity we have $\psi(a)'^*\bar{\omega} = \bar{\omega}$ for all a or $\psi(a)'^*\bar{\omega} = -\bar{\omega}$ for all a. $\square$

To complete the proof of Theorem 6.8, we need a lemma.

Lemma 6.9. *Let ψ_k be a sequence of continuous functions for which (i) of Theorem* 6.8 *is valid. If $\psi_k \to \psi$ locally uniformly and ψ is a homeomorphism, then ψ satisfies* (i) *as well.*

Note that for the proof Theorem 6.8, we may also assume that each ψ_k is a homeomorphism and this assumpsion simplifies the proof. However, for Theorem (6.2) we are not assuming that each ψ_k is a homeorphism.

Proof of Lemma 6.9. Imagine that we can prove this: For every $\lambda \in (0,1)$ and every 0-centered ellipsoid E, there exists k_0 such that for $k > k_0$ we have

$$\psi_k(\lambda E) \subseteq \psi(E) \subseteq \psi_k(\lambda^{-1}E). \tag{6.8}$$

Then we are done because

$$\begin{aligned} c(\psi(E)) &\leq c(\psi_k(\lambda^{-1}E)) = c(\lambda^{-1}E) = \lambda^{-2}c(E), \\ c(\psi(E)) &\geq c(\psi_k(\lambda E)) = c(\lambda E) = \lambda^2 c(E) \end{aligned}$$

and yields property (i) for ψ by sending λ to 1.

To establish (6.8), let us write $\varphi_k = \psi^{-1} \circ \psi_k$. Clearly $\lim_{n\to\infty} \varphi_k =$ identity, locally uniformly. From this it is clear that for large k,

$$\psi^{-1} \circ \psi_k(\lambda E) \subseteq E,$$

establishing the first inclusion in (6.8).

If ψ_k is a homeomorphism for each k, then the second inclusion in (6.8) can be established in the same way. However, not the same reasoning applies for the second inclusion in (6.8) because ψ_k is not a homeomorphism. Our goal is showing

$$E \subseteq \varphi_k(\lambda^{-1}E) \tag{6.9}$$

for large k. If (6.9) fails, then there exists $y_l \in E$ such that $y_l \notin \varphi_{k_l}(\lambda^{-1}E)$. This allows us to define $F_k(x) = \frac{\varphi_{k_l}(x) - y_l}{|\varphi_{k_l}(x) - y_l|}$ for $x \in \lambda^{-1}E =: E_\lambda$ and a sequence $l \to \infty$. It follows from Lemma 6.10 below that $\deg f_k = 0$ where $f_k : \partial E_k \to S^{2n-1}$ is the restriction of F_k to ∂E_k. On the other hand, we may define $g_k : \partial E_\lambda \to \partial S^{2n-1}$ by $g_k(x) = \frac{\varphi_{k_l}(x)}{|\varphi_{k_l}(x)|}$. The function g_k is well-defined for large k because φ_{k_l} is uniformly close to identity over the set ∂E_λ. The function g_k has $\deg 1$ simply because g_k is uniformly close to $x \mapsto \frac{x}{|x|}$ which has degree 1. To arrive at a contradiction, it suffices to show that g_k is homotopic to f_k. (By Lemma A.1 homotopic transformations have the same degree.) For homotopy, define

$$\Phi_k(x,t) = \frac{\varphi_{k_l}(x) - ty_l}{|\varphi_{k_l}(x) - ty_l|},$$

for $x \in \partial E_\lambda$, $t \in [0,1]$. Again, since $\varphi_n \to id$ and $ty_l \in E$ the homotopy is well-defined. $\square$

With Theorem 6.8 at our disposal, we can now give a straightforward proof for Theorem 6.2.

Proof of Theorem 6.2. Since this is a local statement, we may assume that $M = \mathbb{R}^{2n}$. Let ψ_n be a sequence of symplectic transformations such that $\psi_n \to \psi$ locally uniformly. Assume that ψ is a diffeomorphism. By Lemma 6.9, ψ preserves the capacity of ellipsoids and by Theorem 6.8, either $\psi^*\bar{\omega} = \bar{\omega}$ or $\psi^*\bar{\omega} = -\bar{\omega}$. We now need to rule out the second possibility. Indeed if $\psi^*\bar{\omega} = -\bar{\omega}$, we then define $\phi_n = \psi_n \times id : \mathbb{R}^{2n} \times \mathbb{R}^{2n}$. We have that if $\omega = \bar{\omega} \times \bar{\omega}$ then $\phi_n^*\omega = \omega$. On the other hand $\phi_n \to \phi = \psi \times id$ locally uniformly. But $\phi^*\omega = \psi^*\bar{\omega} \times \bar{\omega} = (-\bar{\omega}) \times \bar{\omega} \neq \pm\omega$. As a result, we must have $\psi^*\bar{\omega} = \bar{\omega}$. $\square$

Motivated by the above theorem, we may define the notion of symplecticity for homeomorphism. Note that if $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is a diffeomorphism with $\psi^*\bar{\omega} = -\bar{\omega}$, then $\psi^*\bar{\omega}^n = (-1)^n\bar{\omega}^n$, and if n is odd then ψ can not be orientation preserving. Based on this we have the following definition.

Definition 6.10. *Let $\psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be a homeomorphism. We say ψ is a symplectic homeomorphism if either n is odd and ψ is an orientation preserving transformation for which $c(\psi(E)) = c(E)$ for every ellipsoid E. Or n is even and $\psi \times id : \mathbb{R}^{2n+2} \to \mathbb{R}^{2n+2}$ is a symplectic homeomorphsim.*

Clearly if ψ_k is a sequence of symplectic homeomorphism such that $\psi_k \to \psi$ locally uniformly, then ψ is also a symplectic homeomorphism.

We now turn to Theorem 6.3. For the proof of this theorem however we need a specific capacity that was defined by Hofer and Zehnder. Given a symplectic (M,ω), we first define

$$(6.12)\qquad \begin{aligned} \mathcal{H}(M) &= \{H \in C^\infty(M) : H = maxH \text{ ouside a compact subset of } M \text{ and} \\ &\qquad H = \min H = 0 \text{ on some non-empty open set}\}, \\ \hat{\mathcal{H}}(M,\omega) &= \{H \in \mathcal{H}(M) : X_H \text{ has no periodic orbit of period} \\ &\qquad T \in (0,1]\}, \end{aligned}$$

We then define

$$(6.13)\qquad c_{HZ}(M,\omega) = c(M,\omega) = \sup_{H \in \hat{\mathcal{H}}(M,\omega)} \max H.$$

The following theorem is due to Hofer and Zehnder and evidently implies Theorem 6.4.

Theorem 6.12. *Let c be as in* (6.13). *Then c satisfies* (6.1(i–iii)) *and $c(B^{2n}(1),\bar{\omega}) = c(Z^{2n}(1),\bar{\omega}) = \pi$.*

From the definition of c, it is not hard to guess that the only non-trivial part of Theorem 6.12 is the claim $c(Z^{2n}(1),\bar{\omega}) \leq \pi$. We now state a theorem that is equivalent to this.

Theorem 6.13. *Assume $H \in \hat{\mathcal{H}}(Z^{2n}(1))$ with $\sup H > \pi$. Then the Hamiltonian flow of H has a non-constant periodic orbit of period* 1.

We leave the proof of Theorem 6.16 for Section 7. Assuming this theorem, one can readily establish Theorem 6.12.

Proof of Theorem 6.12. We start with monotonicity. Let $\psi : (M_1,\omega_1) \to (M_2,\omega_2)$ be a symplectic embedding. If $H_1 : M_1 \to \mathbb{R}$ with $H_1 \in \mathcal{H}(M_1)$ and $H_1 = H_2 \circ \psi$ for $H_2 : M_2 \to \mathbb{R}$ then there is a 1-1 correspondence between the flows of H_1 and H_2: If ϕ^1, ϕ^2 are the corresponding flows, then $\psi^{-1} \circ \phi^2 \circ \psi = \phi^1$ by Proposition 3.2. Evidently $\max H_1 = \max H_2$. We now write

$$\begin{aligned} c(M_2,\omega_2) &\geq \sup\{\max H_2 : H_2 \in \hat{\mathcal{H}}(M_2,\omega_2) \text{ support of } H_2 \subseteq \psi(M_1) \\ &= \sup\{\max H_1 : H_1 \in \hat{\mathcal{H}}(M_1,\omega_1)\} = c(M_1,\omega_1). \end{aligned}$$

To show $c(M,\lambda\omega) = c(M,\omega)$, observe

$$i_X(\lambda\omega) = -d(\lambda H) \Leftrightarrow i_X\omega = -dH.$$

As a result,

$$\hat{\mathcal{H}}(M,\lambda\omega) = \{\lambda H : H \in \hat{\mathcal{H}}(M,\omega)\}$$

because (ω, H) and $(\lambda\omega, \lambda H)$ correspond to the same vector field. From this it is clear that $\lambda c(M,\omega) = c(M,\lambda\omega)$.

Next we prove $c(B^{2n}(1),\bar{\omega}) \geq \pi$. For this take any smooth $f : [0,1] \to \mathbb{R}$ with $f(r) = 0$ for r near 0, $f(r) = \pi - \epsilon$ for r near 1, and $0 < f'(r) \leq \pi$. Define $H(z) = f(|z|^2)$ for $z \in B^{2n}(1)$. Then $H \in \mathcal{H}(B^{2n}(1))$ and $\max H = \pi - \epsilon$. We are done if we can show that such H belongs to $\hat{\mathcal{H}}(B^{2n}(1),\bar{\omega})$. For this we need to show that any periodic orbit of X_H has a period $T > 1$.

If $x = (q, p)$ is a solution to $\dot{x} = J\nabla H(x)$,, we then have

$$\dot{q} = 2f'(|x|^2)p, \quad \dot{p} = -2f'(|x|^2)q.$$

For such a solution we have $|x(t)|^2 \equiv r$ by the conservation of energy. Indeed $\frac{d}{dt}|x(t)|^2 = 2f'(|x|^2)(p, -q) \cdot (q, p) = 0$. As a result, in complex notation $z(t) = q(t) + ip(t)$ we get $z(t) = \exp(-2f'(r)it)z(0)$. For given r this is periodic of period $T = \frac{2\pi}{2f'(r)}$ whenever $f'(r) \neq 0$. If for some r we have $f'(r) > 1$, then we have a period $T > 1$ and such f provides us with a Hamiltonian $H \in \hat{\mathcal{H}}(B^{2n}(1), \bar{\omega})$ with $\max H = \pi - \epsilon$. This completes the proof of $c(B^{2n}(1), \bar{\omega}) \geq \pi$.

Finally our Theorem 6.16 simply means that $c(Z^{2n}(1), \bar{\omega}) \leq \pi$. □

If X_H has a periodic orbit x of period T, then $y(t) = x(Tt)$ is 1-periodic and solves $\dot{y} = TJ\nabla H(y)$. By differentiating we obtain $\ddot{y} = TJD^2H(y)\dot{y}$, which in turn yields the bound

$$\|\ddot{y}\|_{L^2} \leq aT\|\dot{y}\|_{L^2} \leq \frac{aT}{2\pi}\|\ddot{y}\|_{L^2},$$

where $a = \max|D^2H|$ and for the last inequality we used Exercise 5.6(ii). From this we learn that if the period T satisfies $T < \frac{2\pi}{a}$, then y must be a constant, or equivalenly x is a constant. In other words, if the vector field X_H has a nonconstant periodic orbit, then its period is at least $\frac{2\pi}{\max|D^2H|}$. However, the capacity $c(M, \omega)$, provides us with an upper-bound on the period that depends on $\max H$ only.

Note that Theorem 6.12 or 6.13 in particlar implies that if U is a bounded open subset of $\mathbb{R}^{2n}$, then $c(U) < \infty$. It is this consequence that will be needed for Theorem 6.3. From the definition of $c = c_{HZ}$, we can readily deduce that if $c(M, \omega) < \infty$, and $H \in \mathcal{H}(M)$, then the corresponding Hamiltonian vector field X_H possesses a periodic orbit of period $T \in \big(0, \frac{c(M,\omega)}{\max H} + \varepsilon\big]$, for every positive ε. To see this, take any nonzero $H \in \mathcal{H}$, and let λ be any positive constant such that $\lambda \max H > c(M, \omega)$. We then have that $\lambda H \notin \hat{\mathcal{H}}$. Hence the vector field $\lambda X_H = X_{\lambda H}$ has a periodic orbit of period $T \in (0, 1]$. But a periodic orbit of period T for λX_H yields a periodic orbit of period $\lambda T \in (0, \lambda]$ for X_H.

Proof of Theorem 6.3. **Step 1.** Let S be a hypersurface of $\mathbb{R}^{2n}$ of contact type and choose a vector field X that is defined on a neighborhood U of S with $X(x) \notin T_xS$ for $x \in S$ and $\mathcal{L}_X\bar{\omega} = \bar{\omega}$. Let ϕ_s be the flow of X and set $S_t = \phi_t(S)$ for $t \in (-\delta, \delta)$. From $\mathcal{L}_X\bar{\omega} = \bar{\omega}$ we learn that $\phi_t^*\bar{\omega} = e^t\bar{\omega}$. On the other hand if $v \in l_x(S) = \{v : \bar{\omega}(v, a) = 0 \text{ for } a \in T_xS\}$, then $d\phi_t(x)v \in l_{\phi_t(x)}(\phi_t(S))$ because $\bar{\omega}(d\phi_t(x)v, d\phi_t(x)a) = \phi_t^*\bar{\omega}(v, a) = e^t\bar{\omega}(v, a) = 0$ for every $a \in T_xS$. As a result,

$$d\phi_t(x)l_x(S) = l_{\phi_t(x)}(\phi_t(S)). \tag{6.15}$$

From this we learn that if one of S_t carries a periodic orbit, then all S_t carry a period orbit.

Step 2. We now set $U = \cup_{t\in(\delta,\delta)} S_t$, and use $c(U) < \infty$ to deduce that any X_H with $H \in \mathcal{H}(U)$, has a periodic orbit in U. If we can construct a suitable $H \in \mathcal{H}(U)$ for which the level sets of H coincide with $(S_t : t \in (-\delta, \delta)$, then we are done. For this let us define $H(x) = g(t)$ for $x \in S_t$, where $g : (\delta, \delta) \to \mathbb{R}$ is a smooth function with the following properties: $g(t) = 0$ for $t \in (-\delta, \delta + \varepsilon]$, $g(t) = \max g$ for $t \in [\delta - \varepsilon, \delta)$, and g is strictly increasing in the interval $(-\delta + \varepsilon, \delta - \varepsilon)$. Note that if $H(x) = 0$ or $H(x) = \max H$, then the flow ϕ_t^H of the vector field X_H satisfies $\phi_t^H(x) = x$. Hence any non-constant periodic orbit has an energy in $(0, \max H)$ and lies on some S_t with $t \in (-d + \varepsilon, \delta - \varepsilon)$. □

The following is a straight forward consequence of the proof of Theorem 6.3.

Corollary 6.14. *Let S be a compact hypersurface in $\mathbb{R}^{2n}$ without boundary. Let $\tau : S \times [0,1] \to \mathbb{R}^{2n}$ be an embedding. Then for a dense set of parameters $t \in [0,1]$, the hypersurface $S \times \{t\}$ carries a periodic orbit.*

Proof. Let I be any open subinterval of $(0,1)$, and set $U_I = \tau(S \times I)$ and $S_t = \tau(S \times \{t\})$. We then repeat the second step in the proof of Theorem 6.3 to deduce that one of $(S_t : t \in I)$, carries a periodic orbit. □

7 Existence of a Capacity

This section is devoted to the proof of Theorem 6.13. Let $H \in \mathcal{H}(Z^{2n}(1))$. As our first step we extend H to a Hamiltonian function $\bar{H} : \mathbb{R}^{2n} \to \mathbb{R}$. For this, we take an ellipsoid

$$E^0 = \left\{ x \in \mathbb{R}^{2n} : Q(x) = q_1^2 + p_1^2 + \frac{1}{l^2} \sum_{j=2}^{n} (q_j^2 + p_j^2) < 1 \right\}.$$

Since $H \in \mathcal{H}(Z^{2n}(1))$, we have that $H = \max H$ for $x \notin K$ where K is a compact subset of $Z := Z^{2n}(1)$. Choose l sufficiently large so that $K \subseteq E^0$. We now pick $\epsilon > 0$ so that $\max H > \pi + \epsilon$ and pick a smooth function $f : \mathbb{R} \to \mathbb{R}$ such that

$f(r) = \max H$ for $r \in [0,1]$, $f(r) = (\pi + \epsilon)r$ for large r, $f(r) \geq (\pi + \epsilon)r$ for all r, and $0 < f'(r) \leq \pi + \epsilon$ for $r > 1$. We now define

$$\bar{H}(x) = \begin{cases} H(x) & \text{if } x \in E^0, \\ f(Q(x)) & \text{if } x \notin E^0. \end{cases} \tag{7.1}$$

We note that $\bar{H}(x) = (\pi + \epsilon)Q(x)$ for large x.

Lemma 7.1. *Assume that $x(\cdot)$ is a 1-periodic solution of $\dot{x} = J\nabla \bar{H}(x)$ with*

$$\mathcal{A}(x(\cdot)) = \int_0^1 \left(\frac{1}{2} Jx \cdot \dot{x} - \bar{H}(x) \right) dt > 0. \tag{7.2}$$

Then $x(t) \in E^0$ for all t and $x(\cdot)$ is non-constant.

Proof. Evidently $X_{\bar{H}} = 0$ on ∂E^0. Hence, all points on ∂E^0 are equilibrium points and if $x(t) \equiv a \in \partial E^0$, then $\mathcal{A}(x(\cdot)) = -\bar{H}(a) \leq 0$ which contradicts (7.2). As a result, either $x(t) \in E^0$ for all t, or $x(t) \in \mathbb{R}^{2n} - E^0$ for all t. It remains to rule out the latter possibility.

If $x(t) \notin E^0$ for all t, then

$$\dot{x} = J\nabla\bar{H}(x) = f'(Q(x))J\nabla Q(x),$$
$$\frac{d}{dt}Q(x) = f'(Q(x))\nabla Q(x)\cdot J\nabla Q(x) = 0.$$

Hence, for such $x(\cdot)$ we have that $Q(x(\cdot)) = Q^0$, and

$$\begin{aligned}\mathcal{A}(x(\cdot)) &= \int_0^1 \left(\frac{1}{2}Jx\cdot\dot{x} - \bar{H}(x)\right)dt \\ &= \int_0^1 \left[\frac{1}{2}f'(Q^0)\nabla Q(x)\cdot x - f(Q^0)\right]dt \\ &= f'(Q^0)Q^0 - f(Q^0) \le (\pi+\epsilon)Q^0 - (\pi+\epsilon)Q^0 \le 0,\end{aligned}$$

which contradicts (7.2). Here we used the fact that $2Q(x) = \nabla Q(x)\cdot x$. □

In view of Lemma 7.1, we only need to find a critical point of $\mathcal{A}$ which satisfies (7.2). Let us first observe that $\mathcal{A}$ is not bounded from below or above. Indeed if $y_k(t) = (\cos 2\pi kt)a + (\sin 2\pi kt)Ja$, for some $a \in \mathbb{R}^{2n}$, then

$$\int_0^1 |y_k(t)|^2 dt = |a|^2, \quad \int_0^1 Jy_k(t)\cdot\dot{y}_k(t)dt = \pi k|a|^2,$$

which in particular implies that $\lim_{k\to\pm\infty}\mathcal{A}(y_k) = \pm\infty$, whenever $a \neq 0$. Because of this, we search saddle-type critical points of $\mathcal{A}$. A standard way of locating such critical points is by using *minimax* principle.

To prepare for this, let us first extend the domain of definition of $\mathcal{A}$ from C^1 to the largest possible Sobolev space which turns out to be the space of function with "half" a derivative.

We begin with $\mathcal{H}^0 = L^2$ which consists of measurable functions

$$x(t) = \sum_{k\in\mathbb{Z}} e^{2\pi ktJ}x_k = \sum_{k\in\mathbb{Z}}[(\cos 2\pi kt)I + (\sin 2\pi kt)J]x_k$$

with $x_k \in \mathbb{R}^{2n}$ and $\|x\|_0 = \sum_k |x_k|^2 < \infty$. Here we are using the Fourier expansion of $x(\cdot)$ where instead of $i = \sqrt{-1}$ we use J. We write

$$\langle x, y\rangle_0 = \int_0^1 x(t)\cdot y(t)dt,$$

for the standard inner product of $\mathcal{H}^0$. Note that if

$$x(t) = \sum_k e^{2\pi ktJ}x_k \text{ and } y(t) = \sum_k e^{2\pi ktJ}y_k,$$

then $\langle x,y\rangle_0=\sum_k x_k\cdot y_k$ and $\|x\|_0^2=\langle x,x\rangle_0$. We note that $\int_0^1 H(x(t))dt$ is well defined for every $x\in H^0$ because $H(x)=Q(x)$ for large x. However, to make sense of $\frac{1}{2}\int_0^1 Jx\cdot\dot{x}dt$, we need to assume that $x(\cdot)$ possesses half a derivative. To see this first observe that if $x\in C^1$, with $x=\sum_k e^{2\pi ktJ}x_k$, then

$$\int_0^1 \frac{1}{2}Jx\cdot\dot{x}dt=\pi\sum_k k|x_k|^2.$$

This suggests defining

$$H^{1/2}=\left\{x\in\mathcal{H}^0:\sum_k |k||x_k|^2<\infty\right\}.$$

We turn $\mathcal{H}^{1/2}$ into a Hilbert space by defining

$$\langle x,y\rangle=\langle x,y\rangle_{1/2}=x_0\cdot y_0+2\pi\sum_{k\in\mathbb{Z}}|k||x_k|^2.$$

More generally, we define

$$\langle x,y\rangle_s=x_0\cdot y_0+(2\pi)^{2s}\sum_{k\in\mathbb{Z}}|k|^{2s}|x_k|^2$$

for every $s>0$ and $\mathcal{H}^s$ consists of function $x\in\mathcal{H}^0$ such that $\|x\|_s^2=\langle x,x\rangle_s<\infty$. Observe that if $x\in C^1$, then $\int_0^1|\dot{x}(t)|^2dt=\sum_k(2\pi k)^2|x_k|^2$ and that in general $x\in\mathcal{H}^1$ iff x has a weak derivative in L^2.

So far we know that our functional $\mathcal{A}$ is defined on the Hilbert space $\mathcal{H}^{1/2}$. Let us take an arbitrary Hilbert space $\mathcal{E}$ and a function $F:\mathcal{E}\to\mathbb{R}$, and explain the idea of minimax principle for such a function. First we say that F is continuously differentiable with a derivative ∇F if $\nabla F:\mathcal{E}\to\mathcal{E}$ is a continuous function such that for all x and a,

$$F(x)=F(a)+\langle\nabla F(x),x-a\rangle+o(\|x-a\|).$$

We say that x is a critical point of F if $\nabla F(x)=0$. We say that $F\in C^1(\mathcal{E};\mathbb{R})$ satisfies *Palais–Smale* condition if the conditions

$$\sup_l|F(x_l)|<\infty,\quad \lim_{l\to\infty}\nabla F(x_l)=0 \tag{7.3}$$

for a sequence $\{x_l\}$ imply that $\{x_l\}$ has a convergent subsequence. Given a family $\mathcal{F}$ of subsets of $\mathcal{E}$, define

$$\alpha(F,\mathcal{F})=\inf_{A\in\mathcal{F}}\sup_{x\in A}F(x)\in[-\infty,+\infty]. \tag{7.4}$$

Theorem 7.2 (Minimax Principle). *Let $F : \mathcal{E} \to \mathbb{R}$ be a Palais–Smale function and assume that the flow ϕ_t of the gradient* ODE $\dot{x} = -\nabla F(x)$ *is well-defined for all $t \in \mathbb{R}^+$. If $\phi_t(A) \in \mathcal{F}$ for $A \in \mathcal{F}$ and that $\alpha = \alpha(F, \mathcal{F}) \in \mathbb{R}$, then there exists $x^* \in \mathcal{E}$ such that*

$$\nabla F(x^*) = 0 \text{ and } F(x^*) = \alpha(F, \mathcal{F}). \tag{7.5}$$

Proof. It suffices to show that for every $\epsilon > 0$, there exists $y = y(\epsilon)$ such that

$$\alpha - \epsilon \leq F(y(\epsilon)) \leq \alpha + \epsilon, \text{ and } \|\nabla F(y(\epsilon))\| \leq \epsilon. \tag{7.6}$$

Indeed if such $y(\epsilon)$ exists, then the sequence $x_l = y\left(\frac{1}{l}\right)$ has a convergence subsequence by Palais–Smale property of F, and if x^* is any limit point of x_l, then (7.5) holds true.

To establish (7.6), we argue by contradiction. Suppose to the contrary, there exists $\epsilon > 0$ such that

$$\alpha - \epsilon \leq F(y) \leq \alpha + \epsilon \Rightarrow \|\nabla F(y)\| > \epsilon. \tag{7.7}$$

By the definition of α, we can find a set $A \in \mathcal{F}$ such that

$$\sup_{x \in A} F(x) \leq \alpha + \epsilon. \tag{7.8}$$

The idea is to use the flow of ϕ_t to come up with another set $\hat{A} = \phi_{t_0}(A) \in \mathcal{F}$ with

$$\sup_{x \in \hat{A}} F(x) \leq \alpha - \epsilon \tag{7.9}$$

which of course contradicts the definition of α.

Observe

$$\frac{d}{dt} F(\phi_t(x)) = -\|\nabla F(\phi_t(x))\|^2, \text{ or } F(\phi_t(x)) = F(x) - \int_0^t \|\nabla F(\phi_s(x))\|^2 ds. \tag{7.10}$$

Now for $x \in A$, we have that $F(x) \leq \alpha + \epsilon$, which in turn implies that $F(\phi_t(x)) \leq \alpha + \epsilon$ for all $t \geq 0$. We now assert that (7.9) holds for $\hat{A} = \phi_{2\epsilon^{-1}}(A)$. We argue by contradiction. Take $x \in A$ and assume to the contrary that $y = \phi_{2\epsilon^{-1}}(x)$ satisfies $F(y) > \alpha - \epsilon$. Then $F(\phi_s(x)) > \alpha - \epsilon$ for every $s \in [0, 2\epsilon^{-1}]$. Then by (7.7), we have that $\|\nabla F(\phi_s(x))\| > \epsilon$ for every $s \in [0, 2\epsilon^{-1}]$. As a result, we can use (7.10) to assert

$$F(y) = F(\phi_{2\epsilon^{-1}}(x)) \leq \alpha + \epsilon - \epsilon^2(2\epsilon) = \alpha - \epsilon,$$

which contradicts our starting assumption $F(y) > \alpha - \epsilon$. This completes the proof. □

Example 7.3. Let $F \in C^1(\mathbb{R}^d; \mathbb{R})$ be a function which satisfies the Palais–Smale condition. Assume that F is bounded from below. Then F has a minimizer x^*, i.e., $F(x^*) = \inf F$. This can be shown using Theorem 7.2 by taking $\mathcal{F} = \{\{x\} : x \in \mathbb{R}^d\}$. Similarly, when F is bounded above, take $\mathcal{F} = \{\mathcal{E}\}$ to deduce that the function F has a maximizer. □

Exercise 7.4. Assume $F \in C^1(\mathbb{R}^d; \mathbb{R})$ satisfies $\lim_{|x|\to\infty} F(x) = \infty$ and that F possesses two distinct relative minima x_1 and x_2. Show that F has a third critical point x_3 which is not a relative minimizer of F.

Hint. Use paths connecting x_2 to x_2 for the member of $\mathcal{F}$ in Theorem 7.2. □

It is useful to think of $F(x)$ as representing the height above a point x in a Landscape. For example, in Exercise 7.4, x_1 and x_2 correspond to two villages which are separated by a mountain range and we wish to find a *mountain pass* connecting the two.

Exercise 7.5. Show that $F(x, y) = e^{-x} - y^2$ does not satisfy Palais–Smale condition. Let

$$E^{\pm} = \{(x, y) : F(x, y) \leq 0, \ \pm y \geq 0\}$$

and set

$$\mathcal{F} = \{\gamma[0, 1] : \gamma : [0, 1] \to \mathbb{R}^2, \ \gamma(0) \in E^-, \ \gamma(1) \in E^+ \text{ and } \gamma \text{ is continuous}\}.$$

Show that $\alpha(F, \mathcal{F}) = 0$ but there is no x^* with $F(x^*) = 0$ and that F has no critical point. □

As an application of Theorem 7.2, we prove the celebrated *Mountain Pass Lemma* of

Ambrosetti and Rabinowitz.

Corollary 7.6. *Let $F \in C^1(\mathcal{E})$ be a Palais–Smale function and assume that $R \subseteq \mathcal{E}$ is a* mountain range *relative to F in the following sense:*

(i) *$\mathcal{E} - R$ is not connected,*

(ii) $\inf_R F =: \beta > -\infty$,

(iii) *if A is a connected component of $\mathcal{E} - R$, then $\inf_A F < \beta$. Then F has a critical value α satisfying $\alpha \geq \beta$.*

Proof. Let $\mathcal{E}^1$ and $\mathcal{E}^2$ be two connected components of $\mathcal{E} - R$ and set $\hat{\mathcal{E}}^i = \{x \in \mathcal{E}^i : F(x) < \beta\}$, for $i = 1$ and 2. We now define

$$\mathcal{F} = \{\gamma[0,1] : \gamma : [0,1] \to \mathcal{E} \text{ is continuous with } \gamma(0) \in \hat{\mathcal{E}}^1 \text{ and } \gamma(1) \in \hat{\mathcal{E}}^2\}.$$

We set $\alpha = \alpha(F, \mathcal{F})$ and would like to apply Theorem 7.2. Note that if $A = \gamma[0,1] \in \mathcal{F}$, then $A \cap R \neq \emptyset$ and $\sup_A F \geq \beta$. Hence $\alpha \geq \beta$.

Evidently $\alpha < \infty$ because $A \in \mathcal{F}$ is compact. On the other hand, if $A = \gamma[0,1] \in \mathcal{F}$ with $\gamma(0) = a_1 \in \hat{\mathcal{E}}^1$ and $\gamma(1) = a_2 \in \hat{\mathcal{E}}^2$, then $\phi_t(a_j) \in \hat{\mathcal{E}}^j$ for $j = 1, 2$ and $t \geq 0$, because

$$F(\phi_t(a_j)) \leq F(a_j) < \beta$$

and $\phi_t(a_j) \notin R$ by $\inf_R F = \beta$. □

The following consequence of Theorem 7.2 is what we really need for the proof of Theorem 6.13.

Corollary 7.7. *Let F be as in Theorem 7.2 and Let Γ and Σ be two bounded subsets of $\mathcal{E}$ such that $\inf_\Gamma F = \beta > -\infty$, $\phi_t(\Sigma) \cap \Gamma \neq \emptyset$ for all $t \geq 0$, and $\sup_\Sigma F < \infty$. Then F has a critical point x^* such that*

$$F(x^*) = \inf_{t \geq 0} \sup_{x \in \phi_t(\Sigma)} F(x) \geq \beta.$$

Proof. We simply take $\mathcal{F} = \{\phi_t(\Sigma) : t \geq 0\}$. Evidently $\alpha(F, \mathcal{F}) \leq \sup_\Sigma F < \infty$. Moreover, since $\phi_t(\Sigma) \cap \Gamma \neq \emptyset$, we have

$$\sup_{t \in \phi_t(\Sigma)} F(x) \geq \beta$$

for every $t \geq 0$. We can now apply Theorem 7.2 to complete the proof. □

Our goal is proving Theorem 6.13 with the aid of Corollary 7.7 for $F = \mathcal{A}$ and $\mathcal{E} = \mathcal{H}^{1/2}$. Let us write $\mathcal{A} = \mathcal{A}_0 - \mathcal{C}$ where

$$\mathcal{A}_0(x) = \frac{1}{2} \int_0^1 Jx \cdot \dot{x} dt, \quad \mathcal{C}(x) = \int_0^1 H(x) dt.$$

Note that $\mathcal{A}_0$ is quadratic and therefore smooth. In fact if we write

$$x^\pm = P^\pm x = \sum_{\pm k > 0} e^{2\pi k t J} x_k, \quad x^0 = P^0 x = x_0,$$

for $x = \sum_{k \in \mathbb{Z}} e^{2\pi k t J} x_k$, then

$$\mathcal{A}_0(x) = \sum_k \pi k |x_k|^2 = \frac{1}{2} \|x^+\|^2 - \frac{1}{2} \|x^-\|^2, \tag{7.11}$$

and

$$\nabla \mathcal{A}_0(x) = x^+ - x^- = (P^+ - P^-) x. \tag{7.12}$$

We now turn to the functional $\mathcal{C}$. Of course $\mathcal{C}$ is differentiable with respect to L^2-inner product and its derivative is given by $\nabla H(x)$. But we want to differentiate with respect to $\mathcal{H}^{1/2}$-inner product. To this end, let us write $j : \mathcal{H}^{1/2} \to \mathcal{H}^0 = L^2$ for the inclusion transformation and j^* for its adjoint; $j^* : \mathcal{H}^0 \to \mathcal{H}^{1/2}$ and for $x \in \mathcal{H}^{1/2}$, and $y \in \mathcal{H}^0$,

$$\langle j^*(y), x \rangle = \langle y, j(x) \rangle_0 = \int_0^1 y \cdot x \; dt. \tag{7.13}$$

Ax a result, if $y = \sum_k e^{2\pi k t J} y_k$, then

$$j^*(y) = y_0 + \sum_{k \neq 0} \frac{1}{2\pi |k|} y_k e^{2\pi k t J}. \tag{7.14}$$

Using (7.13) we can use the L^2-derivative of $\mathcal{C}$ to obtain $\mathcal{H}^{1/2}$-derivative:

$$\nabla \mathcal{C}(x) = j^* \nabla H(x). \tag{7.15}$$

By Lemma 7.8 below, $\nabla\mathcal{A}$ is a compact perturbation of the linear operator $\nabla\mathcal{A}_0$, and the gradient ODE

$$\dot{x} = -\nabla\mathcal{A}(x), \tag{7.16}$$

is well-posed.

Lemma 7.8. (i) *$j^*(L^2) \subseteq \mathcal{H}^1$ and $\|j^*(x)\|_1 = \|x\|_0$.*

(ii) *$\nabla\mathcal{C} : \mathcal{E} \to \mathcal{E}$ is a compact operator with*

$$\|\nabla\mathcal{C}(x) - \nabla\mathcal{C}(y)\| \le c_0\|x - y\|,$$

for a constant c_0.

Proof. (i) By (7.14),

$$\|j^*(x)\|_1^2 = |x_0|^2 + \sum_{k\neq 0} \frac{(2\pi|k|)^2}{(2\pi|k|)^2}|x_k|^2 = \|x\|_0^2.$$

(ii) Since $|\nabla^2 H(x)|$ is uniformly bounded, we learn that there exists a constant c_0 such that

$$\mathcal{C}(x+h) = \mathcal{C}(x) + \langle \nabla H(x), h\rangle_0 + R(x,h)$$

with $|R(x,h)| \le c_0\|h\|_0 \le c_0\|h\|$. Hence $\mathcal{C}$ is differentiable with $\nabla\mathcal{C}(x) = j^*\nabla H(x)$. The compactness of the operator $\nabla\mathcal{C}$ follows from the compactness of j^* and the bound $|\nabla^2 H(x)| \le c_0$ for a constant c_0. Finally,

$$\begin{aligned}\|\nabla\mathcal{C}(x) - \nabla\mathcal{C}(y)\| &= \|j^*(\nabla H(x) - \nabla H(y))\| \\ &= \|\nabla H(x) - \nabla H(y)\|_0 \\ &\le c_0\|x-y\|_0 \le c_0\|x-y\|.\end{aligned}$$

□

As our next step we verify Palais–Smale property for $\mathcal{A}$.

Lemma 7.9. *If $\nabla\mathcal{A}(x_l) \to 0$ as $l \to \infty$, then $\{x_l\}$ has a convergent subsequence.*

Proof. **Step 1.** Take a sequence $\{x_l\}$ such that

$$\nabla\mathcal{A}(x_l) = x_l^+ - x_l^- - \nabla\mathcal{C}(x_l) \to 0, \tag{7.17}$$

as $l \to \infty$. We need to show that $\{x_l\}$ has a convergent subsequence. For this it suffices to show that $\{x_l\}$ is bounded. This is because the boundedness of $\{x_l\}$ implies the precompactness of $\{\nabla\mathcal{C}(x_l)\}$ by the compactness of the operator $\nabla\mathcal{C}$. This and (7.17) imply that

$x_l^+ - x_l^-$ has a convergent subsequence. If $x_{l,k}$ denotes the kth Fourier coefficients of x_l and $\bar{x} = \sum_k \bar{x}_k e^{2\pi ktJ}$ is a limit point of $x_l^+ - x_l^-$, then by definition

$$\sum_{k<0} |k||x_{l,k} + \bar{x}_k|^2 + \sum_{k>0} |k||x_{l,k} - \bar{x}_k|^2 \to 0$$

along a subsequence as $l \to \infty$, and as a result, $\{P^+ x_l + P^- x_l\}$ has a convergent subsequence. From this we deduce that $\{x_l\}$ has a convergent subsequence because $\{P^0 x_l\}$ is also bounded.

Step 2. To prove the boundedness of $\{x_l\}$, we argue by contradiction. Assume to the contrary

$$\lim_{l\to\infty} \|x_l\| = \infty. \tag{7.18}$$

Observe that if $y_l = \frac{x_l}{\|x_l\|}$, then by (7.17),

$$y_l^+ - y_l^- - j^*\left(\frac{1}{\|x_l\|}\nabla H(x_l)\right) \to 0. \tag{7.19}$$

Now we use the boundedness of $\nabla H(x_l)$ in $\mathcal{H}^0$ and repeat Step 1 to deduce that $\{y_l\}$ has a convergent subsequence. Let us continue to use $\{y_l\}$ for such a subsequence and write y for its limit. Recall that for $z \in \mathbb{R}^{2n}$ with large $|z|$,

$$H(z) = (\pi + \epsilon)Q(z) =: \hat{Q}(z)$$

where $Q(z) = q_1^2 + p_1^2 + l^{-2}\sum_{j=2}^n (q_j^2 + p_j^2)$. We now argue

$$\lim_{l\to\infty} \left\| \frac{\nabla H(x_l)}{\|x_l\|} - \nabla\hat{Q}(y) \right\|_0 = 0. \tag{7.20}$$

To see this, observe

$$\left\| \frac{\nabla H(x_l)}{\|x_l\|} - \nabla\hat{Q}(y) \right\|_0 \leq \|\nabla H(x_l) - \nabla\hat{Q}(x_l)\|\|x_l\|^{-1} + \|\nabla\hat{Q}(y_l) - \nabla\hat{Q}(y)\|_0.$$

Now (7.20) follows because $|\nabla H - \nabla\hat{Q}|$ is uniformly bounded and $\|y_l - y\| \to 0$.

From (7.20) and Lemma 7.8(i) we deduce

$$\lim_{l\to\infty} \frac{1}{\|x_l\|} j^*(\nabla H(x_l)) = j^*\nabla Q(y),$$

in $\mathcal{H}^{1/2}$. From this, $\lim_{l\to\infty} y_l = y$, and (7.19) we deduce

$$y^+ - y^- - j^*\nabla\hat{Q}(y) = 0$$

for y satisfying $\|y\| = 1$. This means that y is a critical point of

$$\mathcal{A}_1(y) = \int_0^1 \left[\frac{1}{2} Jy \cdot \dot{y} - \hat{Q}(y)\right]\, dt.$$

From this and Lemma 7.10 below, y is C^1 and $\dot{y} = J\nabla\hat{Q}(y)$. Hence

$$y(t) = \Bigg(a_1 \cos 2(\pi+\epsilon)t, a_1 \sin 2(\pi+\epsilon)t, a_2 \cos 2\frac{(\pi+\epsilon)}{l^2}t,$$
$$a_2 \sin 2\frac{(\pi+\epsilon)}{l^2}t, \dots, a_n \cos 2\frac{(\pi+\epsilon)}{l^2}t, a_n \sin 2\frac{(\pi+\epsilon)}{l^2}t\Bigg).$$

This is 1-periodic only if $y \equiv 0$. This contradicts $\|y\| = 1$. □

Lemma 7.10. *Let K be a C^r-function with $|K(x)| \le c_0|x|^2$ for some c_0. Let $x(\cdot)$ be a critical point of $\mathcal{A}(x) = \int_0^1 \left(\frac{1}{2} Jx \cdot \dot{x} - K(x)\right)\, dt$. Then x is a C^r solution of $\dot{x} = J\nabla K(x)$.*

Proof. Since $\nabla\mathcal{A}(x) = 0$, we have that $x^+ - x^- = j^*\nabla K(x)$. If

$$x = \sum_k e^{2\pi k t J} x_k, \quad \nabla K(x) = \sum_k e^{2\pi k t J} a_k,$$

then we deduce that $a_0 = 0$ and $\mathrm{sgn}(k) x_k = (2\pi|k|)^{-1} a_k$ for $k \ne 0$, or $2\pi k x_k = a_k$ for all k. Since $\nabla K(x) \in \mathcal{H}^0 = L^2$, we learn that $\sum_k |k|^2|x_k|^2 < \infty$, or $x \in \mathcal{H}^1$. From $2\pi k x_k = a_k$, we can readily deduce that $\dot{x} = J\nabla K(x)$ weakly. Since the right-hand side is continuous by $\mathcal{H}^1 \subseteq C(S^1)$, we learn that $x \in C^1(S^1)$. By induction, we can easily show that indeed $x \in C^r$ if $\nabla K \in C^{r-1}$. □

On the account of Lemmas 7.1 and 7.10, we need to find a critical point of $\mathcal{A}$ with $\mathcal{A}(x) > 0$. Note that our Hamiltonian H is supposed to vanish on some non-empty open set. In fact we may assume that H vanishes near the origin. This is because we may replace H with $H \circ \psi$ for a symplectic $\psi : Z(1) \to Z(1)$ which satisfies $\psi(x) = x$ outside a compact subset of $Z(1)$. This is because the flows of X_H and $X_{H\circ\psi}$ are conjugated. Using such ψ, we may shift a minimizer of H to the origin. From now on we assume that H in Theorem 6.13 vanishes near the origin. We establish Theorem 6.13 with the aid of Corollary 7.7. This is carried out in Lemmas 7.11 and 7.12. To this end, let us define

$$\Gamma = \Gamma(r) = \{x \in \mathcal{E}^+ : \|x\| = r\},$$
$$\Sigma = \Sigma(\theta) = \{x : x = x^- + x^0 + s e^+ : \|x^- + x^0\| \le \theta,\ 0 \le s \le \theta\}$$

where $\mathcal{E} = \mathcal{H}^{1/2}$, $e^+ = \{x : x^+ = x\}$ and $\mathcal{C}^+(t) = e^{2\pi t J} e_1$ with $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^{2n}$.

Lemma 7.11. $\inf_{\Gamma(r)} \mathcal{A} = \beta > 0$, *for* $r > 0$ *sufficiently small.*

Lemma 7.12. *If* θ *is sufficiently large, then* $\phi_t(\Sigma(\theta)) \cap \Gamma \neq \emptyset$ *for all* $t > 0$.

A quick way of establishing Lemma 7.11 is with the aid of the celebrated *Sobolev's inequality*:

Lemma 7.13. *For every* $p \geq 1$, *there exists a constant* $c_0 = c_0(p)$ *such that*

$$\left(\int_0^1 |x(t)|^p \, dt \right)^{1/p} \leq c_0(p) \|x\|.$$

We will prove this lemma for $p < 3$. Once Lemma 7.13 is accepted for some $p > 2$, then it is straightforward to prove Lemma 7.11.

Proof of Lemma 7.11. Recall that $|H(z)| \leq c_1 |z|^2$ for a constant c_1. This however doesn't do the job and we need to use the fact that H vanishes near the origin. From this we learn that if $p > 2$, then $|H(z)| \leq c_1(p)|z|^p$ which improves the previous bound for small $|z|$. We now have that if $x \in \mathcal{E}^+$, then

$$\begin{aligned} \mathcal{A}(x) &= \frac{1}{2}\|x^+\|^2 - \int_0^1 H(x(t)) \, dt \\ &\geq \frac{1}{2}\|x^+\|^2 - c_1(p) \int_0^1 |x(t)|^p \, dt \\ &\geq \frac{1}{2}\|x\|^2 - c_1(p) c_0(p)^p \|x\|^p \\ &= \frac{1}{2} r^2 - c_1(p) c_0(p)^p r^p =: \beta. \end{aligned}$$

Evidently $\beta > 0$ if r is sufficiently small. $\square$

We now turn to Lemma 7.12. If fact it is for this lemma that the requirement of $\max_Z H > \pi$ will be used. We note that in the first step of the proof, the condition $H(x) = (\pi + \varepsilon) Q(x)$ with $\pi + \varepsilon > \frac{1}{2}\|e^+\|$ is used in an essential way.

Proof of Lemma 7.12. **Step 1.** We first show that for large θ,

$$\phi_t(\partial\Sigma) \cap \Gamma = \emptyset \tag{7.21}$$

for every $t \geq 0$. Since $\inf_\Gamma \mathcal{A} = \beta > 0$, it suffices to show

$$\sup_t \sup_{\phi_t(\partial\Sigma)} \mathcal{A} \leq 0.$$

Since $\frac{d}{dt}\mathcal{A}(\phi_t(x)) \leq 0$, it suffices to show

$$\sup_{\partial\Sigma} \mathcal{A} \leq 0. \tag{7.22}$$

We write $\partial\Sigma = \partial_1\Sigma \cup \partial_2\Sigma$ where $\partial_1\Sigma = \{x \in \partial\Sigma : x = x^- + x^0\}$. In the case of $x \in \partial_1\Sigma$, we have

$$\mathcal{A}(x) = -\frac{1}{2}\|x^-\|^2 - \int_0^1 H(x)\ dt \leq 0$$

because $H \geq 0$. So for (7.22), we only need to show $\sup_{\partial_2\Sigma} \mathcal{A} \leq 0$ for sufficiently large θ. Recall that there exists a constant c_1 such that $H(x) \geq (\pi+\epsilon)Q(x) - c_1$. Hence, for $x = x^- + x^0 + se^+$,

$$\begin{aligned}
\mathcal{A}(x) &= \frac{1}{2}s^2\|e^+\|^2 - \frac{1}{2}\|x^-\|^2 - \int_0^1 H(x)\ dt \\
&\leq \pi s^2 - \frac{1}{2}\|x^-\|^2 - (\pi+\epsilon)\int_0^1 Q(x^- + x^0 + se^+)\ dt + c_1 \\
&= \pi s^2 - \frac{1}{2}\|x^-\|^2 - (\pi+\epsilon)\left[\int_0^1 Q(x^-)\ dt + \int_0^1 Q(x^0)\ dt + \int_0^1 Q(se^+)\ dt\right] + c_1 \\
&\leq -\frac{1}{2}\|x^-\|^2 - (\pi+\epsilon)\left[\int_0^1 Q(x^-)\ dt + Q(x^0)\right] - \epsilon s^2 + c_1 \\
&\leq c_1 - c_2(\|x^- + x^0\|^2 + \|se^+\|^2)
\end{aligned}$$

for some constant c_2, where for the first inequality we used $H \geq (\pi+\epsilon)Q - c_1$ and $\|e^+\|^2 = 2\pi$, and for the second equality we used the fact that Q is quadratic and that $\mathcal{E}^+, \mathcal{E}^-, \mathcal{E}^0$ are orthogonal with respect to L^2-inner product. It is now clear that if either $\|x^- + x^0\| = \theta$ or $s = \theta$ with θ sufficiently large, then $\mathcal{A}(x) \leq 0$, proving $\sup_{\partial_2\Sigma} \mathcal{A} \leq 0$. This completes the proof of (7.22).

Step 2. To show $\phi_t(\Sigma) \cap \Gamma \neq \emptyset$, we want to say that in some sense $\partial\Sigma$ and Γ *link* with respect to ϕ_t. That is, $\phi_t(\partial\Sigma)$ can not cross the circle Γ as t increases, so Γ must intersect the "frame" $\phi_t(\Sigma)$. To prepare for this, let us analyze $\phi_t(x)$ for $x = x^- + x^0 + se^+$. Let us show that in general

$$\phi_t(x) = e^t x^- + x^0 + e^{-t}x^t + K(x,t), \tag{7.23}$$

where $K : \mathcal{E} \times \mathbb{R} \to \mathcal{E}$ is continuous and maps bounded sets to precompact sets. To see this, observe that the flow of the ODE

$$\dot{x} = -\nabla\mathcal{A}_0(x) = x^- - x^+$$

is simply given by

$$\phi_t(x) = e^t x^- + x^0 + e^{-t} x^+.$$

Since ϕ_t is the flow of

$$\dot{x} = -\nabla \mathcal{A}_0(x) + \nabla \mathcal{C}(x),$$

we may use the variation of constants formula to derive (7.23) with

$$K(x,t) = \int_0^t (e^{t-s} P^- + P^0 + e^{s-t} P^+) \nabla \mathcal{C}(\phi_s(x)) ds.$$

Since $P^{\pm}$ and P^0 commute with j^*,

$$\begin{aligned} K(x,t) &= j^* \int_0^t (e^{t-s} P^- + P^0 + e^{s-t} P^+) \nabla H(\phi_s(x)) ds \\ &=: j^* R(x,t). \end{aligned}$$

Now the compactness of K follows from the facts that the flow ϕ_s maps bounded sets to bounded sets and j^* maps bounded sets to precompact sets.

Step 3. To show that $\phi_t(\Sigma) \cap \Gamma \neq \emptyset$ for $t \geq 0$, we need to find $x \in \Sigma$ such that $\|\phi_t(x)\| = r$ and $(P^- + P^0)\phi_t(x) = 0$. The latter means

$$e^t x^- + x^0 + (P^- + P^0) K(x,t) = 0. \tag{7.24}$$

To combine this with the former condition $\|\phi_t(x)\| = r$, define

$$L(x,t) = (e^{-t} P^- + P^0) K(x,t) + P^+ \{(\|\phi_t(x)\| - r) e^+ - x\}.$$

We now claim that $\phi_t(\Sigma) \cap \Gamma \neq \emptyset$ is equivalent to finding $x \in \Sigma$ such that

$$x + L(x,t) = 0. \tag{7.25}$$

To see this, apply the operators $P^{\pm}$ and P^0 to both sides of (7.25) to deduce that (7.25) is equivalent to the equations

$$\begin{aligned} \|\phi_t(x)\| - r &= 0, \\ x^0 + P^0 K(x,t) &= 0, \\ x^- + e^{-t} P^- K(x,t) &= 0, \end{aligned}$$

which are what we need to solve. Note that to find a solution of (7.25) in Σ^o which is an open bounded subset of

$$\hat{\mathcal{E}} = \mathcal{E}^- \oplus \mathcal{E}^0 \oplus \mathbb{R} e^+.$$

We now assert that $L : \hat{\mathcal{E}} \times \mathbb{R} \to \hat{\mathcal{E}}$ is compact simply because K is compact and $\mathcal{E}^+$ part of $\hat{\mathcal{E}}$ is one-dimensional.

Step 4. We will use degree theory to solve (7.25). Note that by (7.21),

$$0 \notin (I + L(\cdot, t))(\partial\Sigma).$$

Because of this, $\deg_0(I + L(\cdot, t))$ is well-defined. Observe that $(I + L(\cdot, s) : s \in [0, t])$ defines a homotopy, which implies

$$\deg_0(I + L(\cdot, t)) = \deg_0(I + L(\cdot, 0)). \tag{7.26}$$

From the definition of K given in (7.23), we know that $K(\cdot, 0) \equiv 0$. As a result,

$$L(x, 0) = P^+\{(\|x\| - r)e^+ - x\}.$$

To calculate the right-hand side of (7.26), let us define

$$L^\alpha(x) = P^+\{(\alpha\|x\| - r)e^+ - \alpha x\},$$

for $\alpha \in [0, 1]$. We claim

$$0 \notin (I + L^\alpha)(\partial\Sigma). \tag{7.27}$$

Indeed if $(I + L^\alpha)(x) = 0$ for some $x \in \partial\Sigma$, then $x = se^+$ for some $s \in \{0, \theta\}$ and $s + \alpha s\|e^+\| - r - \alpha s = 0$, or $s((1 - \alpha) + \alpha\sqrt{2\pi}) = r$ because $\|e^+\| = \sqrt{2\pi}$. Evidently $s \neq 0$ because $r \neq 0$. To rule out $s = \theta$, observe that we may take θ large enough to have $\theta > r$. But $r = s((1 - \alpha) + 2\alpha\pi) > s$ which implies that $\theta > s$. In summary $x + L^\alpha(x) = 0$ has no solution in $\partial\Sigma$, i.e., (7.27) holds.

By (7.27), $\deg_0(I + L^\alpha)$ is well-defined and by homotopy invariance of degree,

$$\begin{aligned}\deg_0(I + L(\cdot, 0)) &= \deg_0(I + L^1) = \deg_0(I + L^0) \\ &= \deg_0(I - re^+) = \deg_{re^+}(I) = 1,\end{aligned}$$

provided that $re^+ \in \Sigma$, which is true by our assumption $r < \theta$. From this and (7.26) we deduce that (7.25) has a solution in Σ, completing the proof of lemma. □

We end this section with a proof of Sobolev's inequality in the case of $p < 3$.

Proof of Lemma 7.13. Note that $\mathbb{R}^{2n}$ can be identified as $\mathbb{C}^n$ using $(q, p) \mapsto q + ip$. Since $J \begin{bmatrix} q \\ p \end{bmatrix} = \begin{bmatrix} p \\ -1 \end{bmatrix}$, multiplication by J in $\mathbb{R}^{2n}$ is the same as multiplication by $-i$ in $\mathbb{C}^n$. Hence $x(t) = \sum_k e^{2\pi ktJ} x_k$ can be rewritten as $\sum_k e^{-2\pi kti} x_k$ with $x_k \in \mathbb{C}^n$. Since it suffices to establish the inequality for each component, we may assume without loss of generality that $n = 1$.

We now find an expression of $\|x\|$ that involves the function $x(\cdot)$ directly and does not involve its Fourier coefficients. We claim

$$\int_0^{2\pi}\int_0^{2\pi} \frac{|x(e^{i\theta}) - x(e^{i\varphi})|^2}{|e^{i\theta} - e^{i\varphi}|^2} d\theta d\varphi = 4\pi^2 \sum_k |k||x_k|^2. \tag{7.28}$$

This follows from a direct calculation; the left-hand side equals

$$\begin{aligned}
&\int_0^{2\pi}\int_0^{2\pi} |e^{i\theta} - e^{i\varphi}|^{-2} \left|\sum_k x_k e^{-ik\theta} - \sum_k x_k e^{-ik\varphi}\right|^2 d\theta d\varphi \\
&= \int_0^{2\pi}\int_0^{2\pi} |1 - e^{i\tau}|^{-2} \left|\sum_k x_k (1 - e^{-ik\tau}) e^{-ik\theta}\right|^2 d\theta d\tau \\
&= 2\pi \int_0^{2\pi} |1 - e^{i\tau}|^{-2} \sum_k |x_k|^2 |1 - e^{-ik\tau}|^2 d\tau \\
&= (2\pi)^2 \sum_k |k||x_k|^2,
\end{aligned}$$

because

$$\int_0^{2\pi} \frac{|1 - e^{-ik\tau}|^2}{|1 - e^{i\tau}|^2} d\tau = \int_0^{2\pi} (e^{-i(k-1)\tau} + \cdots + 1)(e^{i(k-1)\tau} + \cdots + 1) d\tau = 2\pi.$$

Step 2. Let us write $\Lambda(x)$ for the left-hand side of (7.28). To simplify the notation write $y(\theta) = x(e^{i\theta})$. We have that for constants c_1 and c_2,

$$\int_0^{2\pi}\int_0^{2\pi} |y(\theta) - y(\varphi)|^2 1\!\!1(|\theta - \varphi| < l^{-1}) d\theta d\varphi \le c_1 l^{-2} \Lambda(x) = c_2 l^{-2} \|x\|^2. \tag{7.29}$$

Given $l \ge 1$, define $z^l(t) = l \int_0^{l^{-1}} x(t+\alpha)\, d\alpha$. By (7.29),

$$\begin{aligned}
\int_0^1 |x(t) - z^l(t)|^2 dt &= \int_0^1 \left| l \int_0^{l^{-1}} (x(t) - x(t+\alpha))\, d\alpha \right|^2 dt \\
&\le l \int_0^1 \int_0^{l^{-1}} |x(t) - x(t+\alpha)|^2\, dt d\alpha \\
&\le c_2 l^{-1} \|x\|^2.
\end{aligned}$$

From this we deduce that for every $l \geq 1$,

$$x = z^l + w^l$$

with

$$\|w^l\|_0^2 \leq c_2 l^{-1} \|x\|^2, \ \|z^l\|_{L^\infty} \leq l\|x\|,$$

because

$$|z^l(t)| \leq l \int_0^1 |x(t)| \ dt \leq l\|x\|_0.$$

Hence we apply Chebyshev's inequality to assert

$$|\{t : |x(t)| > 2l\|x\|\}| \leq |\{t : |w^l(t)| > l\|x\|\}| \leq \frac{\|w^l\|_0^2}{l^2\|x\|^2} \leq c_2 l^{-3},$$

whenever $l \geq 1$. On the other hand

$$\begin{aligned}\int_0^1 |x(t)|^p \ dt &\leq 1 + \int_0^\infty p l^{p-1} |\{x > l\}| dl \\ &\leq 1 + c_3\|x\|^3 \int_1^\infty p l^{p-1} l^{-3} dl \\ &= 1 + c_4(p)\|x\|^3\end{aligned}$$

with $c_4(p) < \infty$ whenever $p < 3$. Finally, we replace x with λx, $\lambda > 0$ to deduce

$$\|x\|_{L^p}^p \leq \lambda^{-p} + c_4(p)\lambda^{3-p}\|x\|^3.$$

Minimizing the right-hand side over $\lambda > 0$ yields the desired inequality. □

Exercise 7.14. (i) Let $G : \mathcal{E} \to \mathcal{E}$ be a Lipschitz function and let ϕ_t denote the flow of the ODE $\dot{x} = G(x)$. Show that for every $l > 0$,

$$\sup_{0 \leq t \leq l} \sup_{\|x\| \leq l} \|\phi_t(x)\| < \infty.$$

(ii) Let $\mathcal{H}^s = \{x \in L^2 : \|x\|_s < \infty\}$. Show that if $s < t$, then a bounded subset of H^t is precompact in H^s.

(iii) Show that if $x \in \mathcal{H}^\theta$ and $\theta > 1/2$, then x is Hölder continuous with

$$|x(t) - x(s)| \leq c\|x\||t - s|^\alpha$$

with $\alpha = \min\left(1, \theta - \frac{1}{2}\right)$.

Hint: Given $x = \sum_k e^{2\pi k t J} x_k$, write $x = y + z$ with $z = \sum_{|k| \leq N} e^{2\pi k t J} x_k$. Estimate $\sup_{|t-s|<\delta} |y(t) - y(s)|$ and $\sup_t |z(t)|$ in terms of δ and N. This yields a bound for $|x(t) - x(s)|$ that can be minimizer with respect to N. □

8 Generating Function and Twist Map

A Hamiltonian vector field X_H is generated by a scalar-valued function H. It turns out that a similar phenomenon is true for any symplectic diffeomorphism at least locally. To explain this, let us take a symplectic diffeomorphism $\psi : U \to \mathbb{R}^{2n}$ with U a simply connected open subset of $\mathbb{R}^{2n}$, and write

$$\psi(q,p) = (Q,P)$$

with Q and $P \in \mathbb{R}^n$. Since $\psi^* d\bar{\lambda} - d\bar{\lambda} = d(\psi^*\bar{\lambda} - \bar{\lambda}) = 0$ for $\bar{\lambda} = q \cdot dp$, we have that there exists a scalar-valued function S such that $\psi^*\bar{\lambda} - \bar{\lambda} = dS$. In coordinates,

$$P \cdot dQ - p \cdot dq = dS. \tag{8.1}$$

The form of (8.1) suggests that perhaps we should regard S as a function of q and Q so that (8.1) is equivalent to

$$\frac{\partial S}{\partial Q} = P \text{ and } \frac{\partial S}{\partial q} = -p. \tag{8.2}$$

The scalar-valued function S is an example of a *generating function*. Its existence is guaranteed if we make some non-degeneracy assumptions on ψ.

Proposition 8.1. *Let $\psi : U \to \mathbb{R}^{2n}$ be a symplectic transformation and assume that at $(q^0, p^0) \in U$,*

$$\det \frac{\partial Q}{\partial p}(q^0, p^0) \neq 0. \tag{8.3}$$

Then there exist a neighborhood V of q^0 and $Q^0 = Q(q^0, p^0)$, and a C^1 function $S : V \to \mathbb{R}$ such that (8.2) *holds.*

Proof. From (8.3) and Implicit Function Theorem, the relation $Q = Q(q,p)$ can be solved for $p = p(q,Q)$ for q and Q near q^0 and Q^0. We then set $P(q,Q) = P(q,p(q,Q))$. To solve (8.2) for S, we need to verify the solvability criterion

$$P_q + p_Q = 0,$$

regarding P and p as functions of q and Q. This is exactly $d(P \cdot dQ - p \cdot dq) = 0$. □

Later we will discuss other types of generating functions. But let us first study some examples. As our first example, consider $\psi = \phi_{t^0 t}$ where $\phi_{t^0 s}$ is the flow of the Hamiltonian ODE

$$\dot{x} = J\nabla H(x,s). \tag{8.4}$$

More precisely, $x(s) = \phi_{t^0 s}(a)$ solves (8.4) subject to the condition $\phi(t^0) = a$. Let us write $\alpha(s)$ and $\beta(s)$ for the q and p components of $x(s)$. We assume that for some open set V, the equation (8.4) can be solved if $(q, Q) \in V$ is specified. More precisely, if $x(s) = (\alpha(s), \beta(s))$ with $\alpha, \beta \in \mathbb{R}^n$, then (8.4) has a unique solution subject to the initial and terminal conditions $\alpha(t^0) = q$ and $\alpha(t) = Q$. We then set

$$S(q, Q; t^0, t) = \int_{t^0}^{t} [\beta(s) \cdot \dot{\alpha}(s) - H(x(s), s)]ds \tag{8.5}$$

with $x(s) = x(s; q, Q)$.

Proposition 8.2. *Under the above conditions, the function S is a generating function for $\psi = \phi_{t^0 t}$. Moreover, S satisfies the* Hamilton–Jacobi *PDE*

$$S_t + H(Q, S_Q, t) = 0. \tag{8.6}$$

Proof. Differentiating both sides of (8.5) with respect to q_j yields

$$\begin{aligned} S_{q_j} &= \int_{t^0}^{t} [\beta_{q_j} \cdot \dot{\alpha} + \beta \cdot \dot{\alpha}_{q_j} - \nabla H \cdot x_{q_j}]ds \\ &= \int_{t^0}^{t} [\beta_{q_j} \cdot \dot{\alpha} - \alpha_{q_j} \cdot \dot{\beta} - \nabla H \cdot x_{q_j}]ds + \beta(t) \cdot \alpha_{q_j}(t) - \beta(t^0) \cdot \alpha_{q_j}(t^0) \\ &= \int_{t^0}^{t} [-J\dot{x} - \nabla H(x, s)] \cdot x_{q_j} ds + \beta(t) \cdot \alpha_{q_j}(t) - \beta(t^0) \cdot \alpha_{q_j}(t^0). \end{aligned} \tag{8.7}$$

The first term vanishes because of (8.4). On the other hand, since

$$\alpha(t^0; q, Q) = q, \quad \beta(t^0; q, Q) = p \quad \alpha(t; q, Q) = Q,$$

we learn that $\alpha_{q_j}(t) = 0$, and $\alpha_{q_j}(t^0) = e^j$, where e^j denotes the standard unit j-th vector. As a result, $S_{q_j} = -p_j$. The proof of $S_{Q_j} = P_j$ is similar.

As for (8.6), first observe

$$S(q, \alpha(t); t^0, t) = \int_{t^0}^{t} [\beta(s) \cdot \dot{\alpha}(s) - H(\alpha(s), s)]ds.$$

Differentiating both sides with respect to t yields

$$S_Q \cdot \dot{\alpha} + S_t = \beta(t) \cdot \dot{\alpha}(t) - H(\alpha(t), t).$$

This immediately implies (8.6) because $S_Q = P = \beta$ and $\alpha(t) = (Q, P) = (Q, S_Q)$. $\square$

As our second example, let us study generating functions in the simplest case $n = 1$. For this we consider symplectic $\varphi : A \to A$ with

$$A = \{(q,p) : R_-^2 \le q^2 + p^2 \le R_+^2\}.$$

Such a transformation was encountered by Poincaré as he used a Poincaré's section to study solutions to Hamiltonian systems. Poincaré was interested in fixed points of φ because they correspond to periodic orbits of the corresponding Hamiltonian system. As we will see such fixed points exist if φ is a twist map.

A function $\varphi : A \to A$ is called a *twist map* if the following conditions are met:

(i) φ is a homeomorphism and the restriction of ψ to A^0 is a diffeomorphism with $\det \varphi' \equiv 1$.

(ii) φ maps the circles $C_\pm = \{q^2 + p^2 = R_\pm^2\}$ to theirselves with $\deg \varphi|_{C_\pm} = \pm 1$.

Our main result about twist maps is the following result of Poincaré and Birkhoff.

Theorem 8.3. *Any twist map has at least two fixed points.*

In fact Poincaré established Theorem 8.3 provided that φ has a global generating function. Such a generating function exists if φ is a *monotone twist map*. To prepare for this, let us first observe that any twist map on A yields a twist map on the cylinder $S^1 \times [R^-, R^+]$. Indeed, if $h : S^1 \times [R^-, R^+] \to A$ is given by $h(x,y) = (\sqrt{y}\cos 2\pi x, \sqrt{y}\sin 2\pi x)$, then $\psi = h^{-1} \circ \varphi \circ h$ is again orientation and area preserving because $dq \wedge dp = -\pi dx \wedge dy$. Now $\psi : S^1 \times [R^-, R^+] \to S^1 \times [R^-, R^+]$ has a lift

$$\Psi : \mathbb{R} \times [R^-, R^+] \to \mathbb{R} \times [R^-, R^+]$$

which satisfies the following conditions:

(i) Ψ is a homeomorphism and the restriction of Ψ to $\mathbb{R} \times (R^-, R^+)$ is a diffeomorphism with $\det \Psi' \equiv 1$.

(ii) Ψ maps $\mathbb{R} \times \{R^\pm\}$ onto itself with

$$\Psi(x, R^\pm) = (\pm x + f^\pm(x), R^\pm),$$

where $f^\pm$ is 1-periodic.

(iii) $\Psi(x+1, y) = \Psi(x,y) + (1,0)$.

Such a map Ψ is again called a twist map. We also write $\Psi(x,y) = (X,Y)$ with X and Y functions of (x,y). We note that we may replace (ii) with the condition $Y(x,\pm R) = \pm R$ and $\pm A_\pm(x) := \pm X(x,\pm R) > \pm x$.

We now formulate a condition on Ψ that would guarantee the existence of a global generating function $S(x,X)$ for Ψ. A twist map Ψ is called *monotone* if

$$\frac{\partial X}{\partial y}(x,y) > 0 \tag{8.8}$$

for all $(x,y) \in \mathbb{R} \times (R^-, R^+)$.

Proposition 8.4. *Let Ψ be a monotone twist map. Then there exists a C^2 function $S : U \to \mathbb{R}$ with*

$$U = \{(x,X) : A_-(x) < X < A_+(x)\}$$

such that

$$\Psi(x, -S_x(x,X)) = (X, S_X(x,X)).$$

Moreover

$$S(x+1, X+1) = S(x,X), \tag{8.9}$$

and

$$S_{xX} < 0. \tag{8.10}$$

Proof. The image of the line segment $\{x\} \times [R^-, R^+]$ under Ψ is a curve γ with parametrization $\gamma(y) = (X(x,y), Y(x,y))$. By (8.8), the relation $X(x,y) = X$ can be inverted to yield $y = y(x,X)$ which is increasing in X. The set $\gamma[R^-, R^+]$ can be viewed as a graph of the function

$$X \mapsto Y(x, y(x,X))$$

with $X \in [A^-(x), A^+(x)]$. The antiderivative of this function yields S.

This can be geometrically described as the area of the region Δ between the curve $\gamma([R^-, R^+])$, the line $Y = R^-$ and the vertical line $\{x\} \times [R^-, R^+]$. We now apply Ψ^{-1}

on this region. The line segment $\{X\} \times [R^-, R^+]$ is mapped to a curve $\hat{\gamma}([R^-, R^+])$ which coincides with a graph of a function $x \mapsto y$. Since Ψ is area preserving the area of $\Psi^{-1}(\Delta)$ is $S(x, X)$. From this we deduce that $S_X = -y$. Here we have used the fact that Ψ^{-1} is a (negative) twist map. (Negative because the degree of Ψ^{-1} restricted to the top boundary is -1 whereas the degree of Ψ^{-1} restricted to the bottom boundary is 1.) This is because if we write $\Psi^{-1}(X, Y) = (x(X, Y), y(X, Y))$, then

$$(\Psi^{-1})' = \begin{bmatrix} x_X & x_Y \\ y_X & y_Y \end{bmatrix} = \begin{bmatrix} X_x & X_y \\ Y_x & Y_y \end{bmatrix}^{-1} = \begin{bmatrix} Y_y & -X_y \\ -Y_x & X_x \end{bmatrix}$$

which implies that $\frac{\partial x}{\partial Y} = -\frac{\partial X}{\partial y} < 0$.

The periodicity (8.9) is an immediate consequence of $\Psi(x + 1, y) = \Psi(x, y) + (1, 0)$;

$$\Psi(\{x + 1\} \times [R^-, R^+]) = \Psi(\{x\} \times [R^-, R^+]) + (1, 0).$$

As for (8.10), recall that $y(x, X)$ is increasing in X. Hence

$$S_{xX} = -y_X < 0.$$

□

A partial converse to Proposition 8.4 is true, namely if a function S satisfies (8.9–10), then it generates a map Ψ which is area preserving. We don't address the behavior of Ψ on the boundary lines and for simplicity assume that S is defined on $\mathbb{R}^2$.

Proposition 8.5. *Let S be a C^2 function satisfying* (8.9–10). *Then there exists a C^1-function Ψ such that*

(i) $\Psi(x + 1, y) = \Psi(x, y) + (1, 0)$

(ii) $\Psi(x, -S_x(x, X)) = (X, S_X(x, X))$

(iii) $\det \Psi' \equiv 1$.

Proof. Since $S_{xX} < 0$, the function $X \mapsto -S_x(x, X)$ is increasing. As a result, $y = -S_x(x, X)$ can be inverted to yield $X = X(x, y)$. We then set

$$Y(x, y) = S_X(x, X(x, y)) \text{ and } \Psi(x, y) = (X(x, y), Y(x, y)).$$

Evidently (ii) is true and (i) follows from (ii) and (8.9) because $S_x(x+1, X+1) = S_x(x, X)$, and $S_X(x + 1, X + 1) = S_X(x, X)$. It remains to verify (iii). For this, set $\hat{S}(x, y) = S(x, X(x, y))$. We have

$$\hat{S}_x = S_x + S_X X_x = -y + Y X_x,$$
$$\hat{S}_y = S_X X_y = Y X_y.$$

Differentiating again yields

$$\hat{S}_{xy} = -1 + Y_y X_x + Y X_{xy},$$
$$\hat{S}_{yx} = Y_x X_y + Y X_{yx}.$$

Since $S \in C^2$, we must have $\hat{S}_{xy} = \hat{S}_{yx}$, which yields $Y_y X_x - Y_x X_y = 1$, as desired. □

We now show how the existence of a generating function can be used to prove the existence of fixed points.

Proof of Theorem 8.3 for a monotone twist map. Define $L(x) = S(x, x)$. We first argue that a critical point of L corresponds to a fixed point of Ψ. Indeed, if $L'(x^0) = 0$, then $S_x(x^0, x^0) + S_X(x^0, x^0) = 0$. Since $\Psi(x^0, -S_x(x^0, x^0)) = (x^0, S_X(x^0, x^0))$, we deduce that $\Psi(x^0, y^0) = (x^0, y^0)$ for $y^0 = -S_x(x^0, x^0) = S_X(x^0, x^0)$. On the other hand, by (8.9), we have that $L(x + 1) = L(x)$. Either L is identically constant which yields a continuum of fixed points for Ψ, or L is not constant. In the latter case, L has at least two distinct critical points, namely a maximizer and minimizer. These yield two distinct critical points of Ψ. □

Before we discuss the proof of Theorem 8.3 for general twist maps, let us study an example of a map which is not quite a twist map but still possesses a global generating function.

Example 8.6 (Billiard map in a convex domain). Let C be a strictly bounde dconvex domain in $\mathbb{R}^2$ and denote its boundary by S. Without loss of generality, we assume that the total length of S is 1. First we describe the billiard flow in C. This is the flow associated with the Hamiltonian function $H(q, p) = \frac{1}{2}|p|^2 + V(q)$ where $V(q) = \begin{cases} 0 & \text{if } q \in C \\ \infty & \text{if } q \notin C \end{cases}$. Here is the interpretation of the corresponding flow: A ball of velocity p starts from a point $q \in C$ and is bounced off the boundary S by the law of reflection. This induces a transformation for the hitting location and reflection angle. More precisely, if a trajectory $q + tp$, $t > 0$ hits

the boundary at a point $\gamma(x)$ and a post-reflection angle θ, then we write $\gamma(X)$ and Θ for the location and post-reflection angle of the next reflection. Here x is the length of arc between a reference point $A \in S$ and $\gamma(x)$ on S in positive direction, and θ measures the angle between the tangent at $\gamma(x)$ and the post-reflection velocity vector. We write φ for the map $(x,\theta) \mapsto (X,\Theta)$ with $x, X \in S^1$ and $\theta, \Theta \in [0,\pi]$. It is more convenient to define $y = -\cos\theta$ so that in the (x,y) coordinates, we have a map $\psi : S^1 \times [-1,1] \to S^1 \times [-1,1]$. As before, we write Ψ for its lift. We claim that Ψ is a monotone twist map except that the twist conditions on the boundary lines $y = \pm 1$ are violated. We show this by applying Proposition 8.5. In fact the generating function is simply given by

$$S(x,X) = -|\gamma(x) - \gamma(X)|,$$

because

$$\begin{aligned}
-S_x(x,X) &= -\frac{(\gamma(X)-\gamma(x))}{|\gamma(X)-\gamma(x)|} \cdot \dot{\gamma}(x) = -\cos\theta, \\
S_X(x,X) &= -\frac{(\gamma(X)-\gamma(x))}{|\gamma(X)-\gamma(x)|} \cdot \dot{\gamma}(X) = \cos\Theta, \\
S_{Xx}(x,X) &= \sin\Theta \frac{\partial\Theta}{\partial x}.
\end{aligned}$$

Note that if $\Theta \in (0,\pi)$, then $\sin\Theta > 0$, and Θ is decreasing in x which means that $S_{Xx} < 0$. Here of course we are using the strict convexity. As for the boundary lines, we have $\Psi(x,-1) = (x,-1)$, $\Psi(x,1) = (x+1,1)$. Note that $S(x,X)$ is defined for (x,X) satisfying $X \in [x, x+1]$. Also note that Ψ has no fixed point inside $\mathbb{R} \times (-1,1)$. □

The generating function S can be used to study periodic orbits of monotone twist maps. To explain this, let us observe that the twist condition means that if ρ^+ and ρ^- denote the rotation numbers of the top and bottom boundary circles of the cylinder, then $\rho^- < 0 < \rho^+$. By rotation number we mean

$$\rho^{\pm} = \lim_{n\to\infty} \frac{A_{\pm}^n(x)}{n}. \tag{8.10}$$

In fact it is well-known that the limit in (8.10) exists because $A_\pm$ is a lift of a circle homeomorphism, and that $\pm\rho^\pm > 0$ because $\pm A_\pm(x) > x$. Now we can interpret Theorem 8.3 as saying that since $0 \in (\rho^-, \rho^+)$, the map Ψ has an orbit which projects onto an x-sequence of 0 rotation number, namely a fixed point. The following theorem generalizes this property to assert the existence of an orbit which projects onto an x-sequence of rotation number $\rho \in (\rho^-, \rho^+)$ provided that ρ is rational. To this end, let us formulate a definition concerning periodic orbits. We say that the point (x, y) is (r, s)-*periodic point* with $r, s \in \mathbb{N}$ and r, s relatively prime, if $(x_n, y_n) = \Psi^n(x, y)$ satisfies $x_{n+s} = x_n + r$ for every n. This means that on cylinder, the x-projection of the orbit $(\psi^n(x, y) : n \in \mathbb{Z})$ wraps r times around the cylinder in s iterates.

Theorem 8.7. *Let Ψ be a twist map. If $\rho \in (\rho^-, \rho^+)$ with $\rho = \frac{r}{s}$, r and s coprime, then Ψ has at least two (r, s)-periodic orbits.*

We do not present a full proof of Theorem 8.7. We only indicate that its proof is very similar to the proof of Theorem 8.2 and uses a variational principle. If Ψ is a monotone twist map, then the variational principle is the discrete analog of the Lagrange variational principle, as can be seen in the following proposition.

Proposition 8.8. *Let Ψ be a monotone twist map with generating function S. Then the following statements are true.*

(i) *Given x and $X \in \mathbb{R}$, the sequence $x_1, x_2, \dots, x_{n-1}$ is a critical point of*

$$L(x_1, x_2, \dots, x_{n-1}) = \sum_{j=0}^{n-1} S(x_j, x_{j+1})$$

with $x_0 = x$, and $x_n = X$, if and only if there exist $y_0, y_1, \dots, y_n$ such that $\Psi^j(x_j, y_j) = (x_{j+1}, y_{j+1})$ for $j = 1, 2, \dots, n-1$.

(ii) *The sequence $x_0, x_1, x_2, \dots, x_{s-1}$ is a critical point of*

$$K(x_1, x_2, \dots, x_s) = S(x_{s-1}, x_0 + r) + \sum_{j=0}^{s-2} S(x_j, x_{j+1})$$

if and only if there exist $y_0, y_1, y_2, \dots, y_{s-1}$ such $\Psi^j(x_j, y_j) = (x_{j+1}, y_{j+1})$ for $j = 0, \dots, s-1$, with $x_s = x_0 + r$.

Proof. We only prove (ii) because (i) can be proved by a verbatim argument. Let $(x_0, \dots, x_{s-1})$ be a critical point and set $x_s = x_0 + r$. We also set $y_j = -S_x(x_j, x_{j+1})$. The result follows because if $Y_j = S_X(x_j, x_{j+1})$, then

$$K_{x_j} = y_j - Y_{j-1}$$

for $j = 0, 1, 2, \ldots, s-1$ and $\Psi(x_j, y_j) = (x_{j+1}, Y_j)$. □

Example 8.9 (Billiard map revisited). Let Ψ be the Billiard map as in Example 8.6. We certainly have $\rho^- = 0$ and $\rho^+ = 1$. According to Theorem 8.7, Ψ has at least two periodic orbits of type (r, s) whenever r and s are relatively prime and $r < s$.

Recall that the function K of Proposition 8.8 is defined for $(x_0, x_1, \ldots, x_{s-1})$ provided that $x_{j+1} \in [x_j, x_j + 1]$ for $j = 0, 1, \ldots, s-1$ with $x_s = x_0 + r$. This however does not reflect the ordering of the orbit. For our purposes we define K on a smaller set Λ which consists of $(x_0, x_1, \ldots, x_{s-1})$ such that there exists $z_0 \le z_1 \le \cdots \le z_{rs}$, with $z_{i+s} = z_i + 1$ for $i = 0, 1, \ldots, (r-1)s$, and $x_j = z_{jr}$ for $j = 0, 1, \ldots, s$.

Note that once $(x_0, x_1, \ldots, x_{s-1}) \in \Lambda$ is known, then all z_j can be determined. Of course $\mathbf{x} \in \Lambda$ imposes various inequalities between $x_0, x_1, \ldots, x_{s-1}$. On the other hand, we can regard K as a function of $z_0, z_1, \ldots, z_{rs}$. Also there are only s many independent variables among them, say $z_0, z_1, \ldots, z_{s-1}$. So, we now have a function $\hat{K}(z_0, z_1, \ldots, z_{s-1}) = K(x_0, x_1, \ldots, x_{s-1})$. The advantage of $\hat{K}$ to K is that it has a domain which is much easier to describe, namely

$$\hat{\Lambda} = \{(z_0, z_1, \ldots, z_{s-1}) : z_0 \le z_1 \le \cdots \le z_{s-1} \le z_0 + 1\}.$$

Since $K(x_0 + 1, \ldots, x_{s-1} + 1) = K(x_0, \ldots, x_{s-1})$, we learn that $\hat{K}(z_0 + 1, \ldots, {}_{s-1} + 1) = \hat{K}(z_0, \ldots, z_{s-1})$. Introducing $w_j = z_j - z_{j-1}$, we have that $\mathbf{z} = (z_0, \ldots, z_{s-1}) \in \hat{\Lambda}$ if and only if $(z_0, w_1, \ldots, w_{s-1})$ belongs to the set of points with

$$0 \le w_1, w_2, \ldots, w_{s-1}, \ w_1 + w_2 + \cdots + w_{s-1} \le 1.$$

Writing $\bar{K}$ for $\hat{K}$ as a function of $z_0, w_1, \ldots, w_{s-1}$, then $\bar{K}$ is defined on

$$\bar{\Lambda} = \{(z_0, w_1, \ldots, w_{s-1}) : z_0 \in \mathbb{R}, w_1, \ldots, w_{s-1} \ge 0, \sum_1^{s-1} w_j \le 1\}.$$

Since $\bar{K}(z_0 + 1, w_1, \ldots, w_{s-1}) = K(z_0, w_1, \ldots, w_{s-1})$, $\bar{K}$ is a lift of a function $\bar{k}$ which is defined on the set

$$\lambda = S^1 \times \{(w_1, \ldots, w_{s-1}) : w_1, \ldots, w_{s-1} \ge 0, \sum_1^{s-1} w_j \le 1\} \subseteq S^1 \times [0, 1]^{s-1}.$$

Of course $\bar{k}$ has a maximizer and a minimizer. We now argue that a minimizer yields a critical point which is in the interior of λ. To see this, let us assume that to the contrary the minimizer is a point on the boundary. To explain this in its simplest non-trivial case, let us take a boundary point of the form

$$z_{n-1} < z_n = z_{n+1} < z_{n+2}.$$

Recall that $\Psi(x, y(x, X)) = (X, Y(x, X))$ where $y(x, X) = -S_x(x, X)$ is increasing in X and $Y(x, X) = S_X(x, X)$ is decreasing in x. We now examine several cases:

(i) $$y(z_n, z_{n+r}) < Y(z_{n-r}, z_n)$$

In this case $\frac{\partial \hat{K}}{\partial z_n} = S_Y(z_{n-r}, z_n) + S_y(z_n, z_{n-r}) > 0$. Hence by decreasing z_n a little bit, we decrease $\hat{K}$. This contradicts the fact that $\mathbf{z}$ is a minimizer.

(ii) $$y(z_{n+1}, z_{n+r+1}) > Y(z_{n-r+1}, z_{n+1})$$

In this case $\frac{\partial \hat{K}}{\partial z_{n+1}} = S_Y(z_{n-r+1}, z_{n+1}) + S_y(z_n, z_{n-r}) < 0$. Hence by increasing z_{n+1}, the value $\hat{K}$ decreases, contradicting the fact that $\hat{z} = (z_0, \ldots, z_{s-1})$ is a minimizer.

(iii) If (i) and (ii) do not occur, then

$$\begin{aligned} Y(z_{n-r}, z_n) &\le y(z_n, z_{n+r}) = y(z_{n+1}, z_{n+r}) \\ &\le y(z_{n+1}, z_{n+r+1}) \le Y(z_{n-r+1}, z_{n-1}) \\ &= Y(z_{n-r+1}, z_n) \le Y(z_{n-r}, z_n). \end{aligned}$$

Hence $z_{n-r+1} = z_{n-r}$, $z_{n+r+1} = z_{n+r}$ and

$$\begin{aligned} Y(z_{n-r}, z_n) &= y(z_n, z_{n+r}), \\ Y(z_{n-r+1}, z_{n+1}) &= y(z_{n-1}, z_{n+r+1}) \end{aligned}$$

which means that (z_{n-r}, z_n, z_{n+1}) is the x-coordinate of an orbit. This is what we wanted.

In fact the other critical point is a saddle point in λ. To see this, let us examine the problem when $s = 2$ and $r = 1$ which corresponds to a periodic orbit of period 2. In this case, we simply have

$$K(x_1, x_2) = S(x_1, x_2) + S(x_2, x_1 + 1) = 2S(x_1, x_2)$$

which is defined on the set

$$\Lambda = \{(x_1, x_2) : x_1 \leq x_2 \leq x_1 + 1\}.$$

Note that $K(x_1, x_2) = 0$ if $(x_1, x_2) \in \partial\Lambda$ and we always have $K(x_1, x_2) < 0$ if $(x_1, x_2) \in \Lambda^0$. Writing K in terms of x_1 and $w_2 = x_2 - x_1$ yields $\hat{K}(x_1, w_2) = 2S(x_1, x_1 + w_2)$ which is defined for $(x_1, w_2) \in \mathbb{R} \times [0, 1]$. Since $\hat{K}$ is periodic in x_1, $\hat{K}$ is the lift of $\hat{k} : S^1 \times [0, 1] \to \mathbb{R}$ and evidently its minimum is attained in the interior of $S^1 \times [0, 1]$. Note that $-\min \hat{k}$ is simply the diameter of the convex set C. We now assert that $\hat{k}$ has a saddle critical point which corresponds to the width of C. To see this, for any $x_1 \in S^1$, we can find $\eta(x_1) = x_2 \in S^1$ such that the tangents at x_1 and $x_2 = \eta(x_1)$ are parallel.

Now $-\max_{x_1} S(x_1, \eta(x_1)) = -S(x_1^*, \eta(x_1^*))$ yields the *width* of C. We assert that $(x_1^*, x_2^*) = (x_1^*, \eta(x_1^*))$ is the other critical point of K. Indeed, since $\Theta = \pi - \theta$, we have

$$S_x(x_1, \eta(x_1)) = S_X(x_1, \eta(x_1))$$

for all x_1. On the other hand, at a maximizer x_1^* of $S(x_1, \eta(x_1))$, we must have

$$\begin{aligned} 0 &= S_x(x_1^*, \eta(x_1^*)) + S_X(x_1^*, \eta(x_1^*))\eta'(x_1^*) \\ &= S_x(x_1^*, x_2^*)(1 + \eta'(x_1^*)). \end{aligned}$$

It is not hard to show that in fact $\eta'(x_1^*) > 0$. Hence we must have $S_x = S_X = 0$ at (x_1^*, x_2^*). □

Exercise 8.10. (i) Consider the Billiard map in a circle. Determine the generating function. Find periodic orbits and describe the remaining orbits.

(ii) Consider the Billiard map in an ellipse. Show that the 2-periodic orbits correspond to the reflection along the axes of symmetry. Show that the 2-periodic orbit associated with the shorter axis of symmetry is a saddle point of the generating function.

Hint: (x_1^*, x_2^*) maximizers of $S(x_1, \eta(x_1))$ but is a local minimum for $S(x_1^*, x_2)$. □

So far we have seen that for $\rho \in (\rho^-, \rho^+) \cap \mathbb{Q}$ we can find at least two periodic orbit of rotation number ρ. The variational principle can be used to find orbits corresponding to irrational $\rho \in (\rho^-, \rho^+)$. This is the subject of *Mather Theory*. For any irrational $\rho \in (\rho^-, \rho^+)$, there exists an invariant set on the cylinder which projects onto either a Cantor-like subset of S^1 or the whole S^1. The invariant set lies on a graph of a Lipschitz function defined on S^1. These invariant sets are known as *Aubry–Mather* sets and correspond to the irrational rotations of Exercise 8.10(i).

We now turn to Theorem 8.3. So far we have a proof in the case of a monotone twist map. Following an idea of Chaperon, we try to express a twist map as a composition of monotone twist maps. To this end, let us define some function spaces.

(i) $\mathcal{T}$ denotes a space of homeomorphism ψ from cylinder $C = S^1 \times [0, 1]$ onto itself which is an orientation and area preserving diffeomorphism in the interior of C and preserves the boundary circles $S^1 \times \{0\}$ and $S^1 \times \{1\}$. The rotation numbers of ψ restricted to $S^1 \times \{0\}$ and $S^1 \times \{1\}$ are denoted by $\rho_-(\psi)$ and $\rho_+(\psi)$ respectively.

(ii) $\mathcal{T}^*$ denotes the space of $\psi \in \mathcal{T}$ such that $\rho^-(\psi) \neq \rho^+(\psi)$. $\mathcal{T}^+$ denotes the space of $\psi \in \mathcal{T}$ with $\rho^-(\psi) < \rho^+(\psi)$. $\mathcal{T}^-$ denotes the space of $\psi \in \mathcal{T}$ with $\rho^+(\psi) < \rho^+(\psi)$. $\mathcal{M}^+$ denotes the space of monotone twist maps.

We equip the space of $\mathcal{T}$ with the topology of C^1-convergence in the interior of C and uniform-convergence up to the boundary. Evidently $\mathcal{T}$ is a topological group with multiplication given by composition. As an example, note that the shear map ξ with lift $(x, y) \mapsto (x + y, y)$ belongs to $\mathcal{T}^+$ whereas $\xi^{-1} = \lambda$ belongs to $\mathcal{T}^-$. We have the following straightforward lemma.

Lemma 8.11. *Every element ψ in the connected component of identity in $\mathcal{T}$ can be written as*

$$\psi = \lambda \circ \psi_1 \circ \lambda \circ \psi_2 \circ \cdots \circ \lambda \circ \psi_n \tag{8.11}$$

with $\psi_1, \psi_2, \ldots, \psi_n \in \mathcal{M}^+$.

Proof. Evidently there exists an open set U in $\mathcal{T}$ such that $\xi \in U \subseteq \mathcal{M}^+$. As a result, $\mathrm{id} \in \xi^{-1}U = \lambda U =: V$ is an open neighborhood of identity and each $\psi \in V$ can be written as $\psi = \lambda \circ \psi_1$ with $\psi_1 \in \mathcal{M}^+$. We now write Ω for the set of ψ in $\mathcal{T}$ for which the decomposition (8.11) exists with $\psi_1, \psi_2, \dots, \psi_n \in V$. Clearly Ω is open because V is open. If we can show that V is also closed, then we deduce that Ω is the connected component of id in $\mathcal{T}$. To see the closedness of Ω, let $\{\varphi_m\}$ be a convergent sequence in Ω. If $\lim_{m\to\infty} \varphi_m = \varphi$, then $\lim_{m\to\infty} \varphi \circ \varphi_m^{-1} = \mathrm{id}$ and, as a result, $\varphi \circ \varphi_m^{-1} \in V$, for large m. Hence there exists $\bar{\psi} \in \mathcal{M}^+$, such that $\varphi \circ \varphi_m^{-1} = \lambda \circ \bar{\psi}$ for a sufficiently large m. That is, $\varphi = \lambda \circ \bar{\psi} \circ \varphi_m$. Since $\varphi_m \in \Omega$, we deduce that $\varphi \in \Omega$, completing the proof of closedness of Ω. □

Proof of Theorem 8.3. Let Ψ be a twist map. On account of Lemma 8.11, there exist monotone twist maps $\Psi_1, \Psi_2, \dots, \Psi_n$ such that

$$\Psi = \Lambda \circ \Psi_1 \circ \Lambda \circ \Psi_2 \circ \cdots \circ \Lambda \circ \Psi_n.$$

Each Ψ_j has a generating function $S_j(x, X)$ with $S_j : \Gamma_j \to \mathbb{R}$ with

$$\Gamma_j = \{(x, X) : A_-^j(x) \le X \le A_+^j(x)\}.$$

Note that the generating function for Λ is $T(x, X) = -\frac{1}{2}(X - x)^2$ which is defined on the set

$$\{(x, X) : x - 1 \le X \le x\}.$$

We now define

$$L(x_0, x_1, \dots, x_{2n-1}) = -\sum_{j=0}^{n-1} \frac{1}{2}(x_{ij} - x_{ij+1})^2 + \sum_{j=0}^{n-2} S_j(x_{2j+1}, x_{2j+2}) + S_n(x_{2n-1}, x_0)$$

on the set Γ which consists of points $x_0, x_1, \dots, x_{2n-1}$ such that for $j = 0, 1, \dots, n-2$

$$-1 \le x_{2j} - x_{2j+1} \le 0, \;\; A_-^j(x_{2j+1}) \le x_{2j+2} \le A_+^j(x_{2j+1})$$

with $x_{2j} = x_0$. Now as in Proposition 8.8, we can show that if $x_0, x_1, \dots, x_{2n-1}$ is a critical point of L, then

$$\Lambda(x_{2j}, y_{2j}) = (x_{2j+1}, y_{2j+1}), \;\; \Psi_j(x_{2j+1}, y_{2j+1}) = (x_{2j+2}, y_{2j+2})$$

for $j = 0, 1, \dots, n-1$, where $y_{2j+1} = T_X(x_{2j}, x_{2j+1}) = x_{2j} - x_{2j+1}$, $y_{2j+2} = S_X(x_{2j+1}, x_{2j=2})$, $x_{2j} = x_0$ and $y_{2j} = y_0$. Of course, in particular (x_0, y_0) is a fixed point of Ψ. □

9 Arnold's Conjecture and Conley–Zehnder's Solution

In the previous section we used generating functions to study various properties of twist maps on cylinders. We now discuss possible generalizations of such global properties for other manifolds. As a start, let us take a symplectic $\psi : \mathbb{T}^{2n} \to \mathbb{T}^{2n}$ and wonder how many fixed points it can have. Evidently a rotation or translation on $\mathbb{T}^2$ is a symplectic diffeomorphism with no fixed point. The question is what plays the role of the twist condition to guarantee the existence of fixed points.

We first examine the issue of generating function. In section 8, we showed that a generating function $S(q, Q)$ always exists locally provided that $\frac{\partial Q}{\partial p}(q^0, p^0)$ is non-singular. Note that this condition fails for the identity. We now discuss another type of generating function that exists trivially for identity. Again for $\Psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ with $\Psi(q, p) = (Q, P)$ and $P \cdot dQ - p \cdot dq = dS$, we write

$$P \cdot dQ + q \cdot dp = d(p \cdot q) + dS =: d\hat{S},$$

which suggests a generating function $\hat{S}(Q, p)$. The following proposition can be proved as Proposition 8.1.

Proposition 9.1. *Let $\Psi(q, p) = (Q, P)$ be a symplectic diffeomorphism with*

$$\det \frac{\partial Q}{\partial q}(q^0, p^0) \neq 0. \tag{9.1}$$

Then there exist a neighborhood V of $Q^0 = Q(q^0, p^0)$ and p^0, and a C^1-function $\hat{S} : V \to \mathbb{R}$ such that

$$\frac{\partial \hat{S}}{\partial Q} = P, \quad \frac{\partial \hat{S}}{\partial p} = q. \tag{9.2}$$

We note that identity transformation has such a generating function with $\hat{S}(Q, p) = Q \cdot p$. More generally, we may write $\hat{S}(Q, p) = Q \cdot p - V(Q, p)$ with V now satisfying

$$P - p = -V_Q, \quad Q - q = V_p \tag{9.3}$$

which can be thought of as a discrete analog of a Hamiltonian where $V(Q, p)$ plays the role of the Hamiltonian.

Using the generating function V, it is not hard to come up with a compelling conjecture regarding the fixed points of a symplectic $\psi : \mathbb{T}^{2n} \to \mathbb{T}^{2n}$. Let us write $\Psi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ for the lift of ψ. If we assume that Ψ has a globally defined generating function V, then a point (q, p) is a fixed point of Ψ if and only if the corresponding (Q, p) is a critical point of V. This should be compared to our proof of Poincaré–Birkhoff Theorem in the case of a monotone twist map. To have an analogous result, we need to say something about the critical points of a C^1-function of $\mathbb{T}^{2n}$. In this connection we have the following:

Theorem 9.2 (Ljusternik–Schnirelman). *Let M be a compact manifold. Then any C^1-function $V : M \to \mathbb{R}$ has at least $\mathcal{I}(M)$ many critical points where $\mathcal{I}(M)$ denotes the* cup length *of M.*

Here is the definition of cup length: $\mathcal{I}(M)$ is the smallest number l such that there exist open simply connected sets $U_1, \dots, U_l$ such that $M = U_1 \cup \dots \cup U_l$. Alternatively, for closed forms $\alpha_1, \alpha_2, \dots, \alpha_l$, we have that $\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_l$ is exact. Here each α_j is a k_j-form with $k_j \geq 1$ for $j = 1, \dots, l$.

Example 9.3. $\mathcal{I}(\mathbb{T}^k) = k + 1$.

From Theorem 9.2 and Proposition 9.1 we deduce

Proposition 9.4. *Let $\psi : \mathbb{T}^{2n} \to \mathbb{T}^{2n}$ be a symplectic diffeomorphism which is a small perturbation of identity. Then ψ has at least $2n + 1$ many fixed points.*

We are now in a position to formulate a similar result for more general symplectic diffeomorphism. With our experience from the previous section we search for a condition on ψ that guarantees to a representation $\psi = \psi_1 \circ \psi_2 \circ \dots \circ \psi_N$ where each ψ_j is a small perturbation of identity. To this end, let us assume that $\Psi = \phi_{01}$ where ϕ_{st} is the flow of a Hamiltonian ODE

$$\dot{x} = H(x, t)$$

where H is 1-periodic in x and t, that is

$$H(x + n, t) = H(x, t + 1) = H(x, t)$$

for every $n \in \mathbb{Z}^{2d}$. Such a function Ψ is called a *Hamiltonian symplectomorphism*. Evidently Ψ is a lift of a symplectomorphism on the torus $\mathbb{T}^{2n}$.

Theorem 9.5. *Any symplectomorphism $\psi : \mathbb{T}^{2n} \to \mathbb{T}^{2n}$ has at least $2n + 1$ many fixed points.*

Note that $\psi = \phi_{0,1} = \phi_{0,\frac{1}{N}} \circ \phi_{\frac{1}{N},\frac{2}{N}} \circ \dots \circ \phi_{\frac{N-1}{N},1}$ and if N is sufficiently large, then each $\psi_j = \phi_{\frac{j-1}{N},\frac{j}{N}}$ has a generating function V_j.

Appendix

A. Degree Theory

We first review the classical Brouwer degree theory. Consider triplets (f, U, y) with $U \subseteq \mathbb{R}^d$ open and bounded, $f : \bar{U} \to \mathbb{R}^d$ continuous and $y \notin f(\partial U)$. We now would like to assign an integer $\deg(f, U, y)$ to (f, U, y) that, in some sense, counts the solutions to the equation $f(x) = y$, $x \in U$, with a sign. This degree satisfies the following properties:

(i) If $V \subseteq U$ and $f^{-1}(\{y\}) \subseteq V$, then $\deg(f, V, y) = \deg(f, U, y)$.

(ii) For a constant a, $\deg(f + a, U, y + a) = \deg(f, U, y)$.

(iii) If $U \cap V = \emptyset$ and $y \notin f(\partial U) \cup f(\partial V)$, then $\deg(f, U \cup V, y) = \deg(f, U, y) + \deg(f, V, y)$.

(iv) If $f : \bar{U} \times [0, 1] \to \mathbb{R}^d$ is continuous with $y \notin f(\partial U, t)$ for every $t \in [0, 1]$, then

$$\deg(f(\cdot, 1), U, y) = \deg(f(\cdot, 0), U, y).$$

(v) If $\deg(f, U, y) \neq 0$, then $f(x) = y$ has a solution in U.

(vi) $\deg(id, B, 0) = 1$ where $B = \{x : |x| < 1\}$.

It turns out that the above properties determine "deg" uniquely. Indeed one can show that for any triplet as above, we can find (g, U, y) with g smooth and f and g homotopic. As for $g \in C^1$, deg is defined by

$$\deg(g, U, y) = \sum_{x \in f^{-1}\{y\}} \operatorname{sgn}(\det g'(x)).$$

In the same fashion, we can define the degree of a continuous map between manifolds. Given two compact manifolds M and N, and a C^1 map $f : N \to M$, we say $x \in N$ is *regular* if df_x is invertible. We say $x \in M$ is a *regular value* if $f^{-1}\{x\}$ consists of regular points. By inverse mapping theorem, it is not hard to show that if x is a regular value, then $f^{-1}\{x\}$ is finite. For such a value we may define the degree by

$$\deg_x(f) = \sum_{y \in f^{-1}\{x\}} \epsilon_y \tag{A.1}$$

where $\epsilon_y = \pm 1$ according to whether df_x preserves or reverses orientation. In the Euclidean case $\epsilon_y = \operatorname{sgn} \det D_x f$. The degree of a continuous $f : N \to M$ defined to be the degree of a C^1 function $g : N \to M$ that is sufficiently close to f. As we will see in Lemma $A.1$, this is well-defined.

Lemma A.1. (i) *If $f : N \to M$ is a C^1 function and Ω is a volume form with $\Omega > 0$, $\int_M \Omega = 1$, then for every regular value $x \in M$,*

$$\deg_x f = \int_N f^*\Omega.$$

(ii) *The degree is invariant under homotopies consisting of C^1-maps.*

(iii) *Any two C^1-maps $f, g : N \to M$ that are sufficiently C^0-close are homotopic via C^1-maps.*

(iv) *Let X be an orientable manifold with $\partial X = N$ and let $F : X \to M$ be a continuous map. Then the degree of $f = F|_N$ is zero.*

Proof. (i) Let x be a regular value and assume $f^{-1}\{x\} = \{x_1, \dots, x_k\}$. Find an open neighborhood V of x such that $f^{-1}(V) = U_1 \cup \dots \cup U_k$ with $U_1, \dots, U_k$, open and disjoint, $x_i \in U_i$ for $i = 1, \dots, k$, and $f|_{U_i} : U_i \to V$ a diffeomorphism for every i. We now take an n-form α with support in V such that $\int_V \alpha = 1$. By Lemma 3.10, we may find an $(n-1)$-form β such that $\Omega = \alpha + d\beta$. We now have

$$\begin{aligned} \int_N f^*\Omega &= \int_N f^*\alpha + \int_N df^*\beta = \int_N f^*\alpha \\ &= \sum_{i=1}^{k} \int_{U_i} f^*\alpha = \sum_{i=1}^{k} \epsilon_{x_i}. \end{aligned}$$

(ii) Clearly degree is C^1-continuous and locally constant.

(iii) Put a Riemannian metric on M. Given $x \in N$, we may find a geodesic curve connecting $f(x)$ to $g(x)$. We now define $\psi(t, x) = \gamma(t; f(x), g(x))$ where $\gamma(t; a, b)$ is defined to be a point on the geodesic connecting a to b with $\gamma(0; a, b) = a$, $\gamma(1; a, b) = b$. This can be done smoothly for a, b sufficiently close.

(iv) Let Ω be a volume form on M with $\int_M \Omega = 1$, $\Omega > 0$. We then have $\deg(f) = \int_N f_*\Omega = \int_{\partial X} f_*\Omega = \int_X d(F_*\Omega) = \int_X F_* d\Omega = 0$. □

Leray–Schander Theory allows us to have a similar notion of degree for functions of the form $f = I + L : \bar{U} \to \mathcal{E}$ with $\mathcal{E}$ a Banach space, U a bounded open subset of $\mathcal{E}$, and L a compact operator. Again we wish to define $\deg(f, U, y)$ provided that $y \notin f(\partial U)$. To do so, first we find a sequence $L_m : \bar{U} \to \mathcal{E}$ such that the range of L_m, denoted by $L_m(\bar{U})$, is a subset of finite dimensional space $\mathcal{E}_m$, and

$$\lim_{m\to\infty} \sup_{x \in \bar{U}} \|L_m(x) - L(x)\| = 0.$$

We then set

$$\deg(f, U, y) = \deg(f_m, \mathcal{E}_m \cap U, y), \tag{A.2}$$

for large m, where $f_m = I + L_m : \mathcal{E}_m \cap \bar{U} \to \mathcal{E}_m$. For this to work, we need to check that $y \notin f_m(\partial U)$ for sufficiently large m. By Exercise A.3(ii), the set $f(\partial U)$ is closed. Since $x \notin f(\partial U)$, we have dist.$(x, f(\partial U)) = \delta > 0$. Then we find m_0 such that if $m > m_0$, then dist.$(x, f_m(\partial U)) \geq \delta/2$. Hence the right-hand side of $(A.2)$ is well-defined by Lemma A.2 below.

Lemma A.2. *Let U be an open bounded subset of $\mathbb{R}^n = \mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}$. Consider $f(x) = x + L(x)$ with $L : \bar{U} \to \mathbb{R}^{n_1}$, $f : \bar{U} \to \mathbb{R}^n$. If $y \notin f(\partial U)$ and $y \in \mathbb{R}^{n_1}$, then*

$$\deg(f, U, y) = \deg(f|_{U_1}, U_1, y)$$

where $U_1 = U \cap \mathbb{R}^{n_1}$.

Proof. We may assume $f \in C^1(U)$ and $y = 0$. Let us take two continuous functions φ_1, φ_2 with $\varphi_i : \mathbb{R}^{n_i} \to \mathbb{R}$, $\int \varphi_i dy_i = 1$, for $i = 1, 2$, and both φ_1, φ_2 have support near 0. Set $\varphi(y_1, y_2) = \varphi_1(y_1)\varphi_2(y_2)$ and $\omega = \varphi(y_1, y_2)dy_1dy_2$ is a volume form of total volume 1. We then have

$$\deg(f, U, y) = \int_U f^*\omega,$$

by Lemma A.1. But $\det(I_n + \nabla L) = \det\left(I_{n_1} + \frac{\partial L}{\partial x_1}\right)$. As a result,

$$\deg(f, U, y) = \int \varphi_1(x_1 + L(x))\varphi_2(x_2)\det\left(I_{n_1} + \frac{\partial L}{\partial x_1}\right)dx_1dx_2.$$

We may send φ_2 to δ_0 to yield

$$\begin{aligned}\deg(f, U, y) &= \int \varphi_1(x_1 + L(x))\det\left(I_{n_1} + \frac{\partial L}{\partial x_1}\right)dx_1 \\ &= \deg(f|_{U_1}, U_1, 0).\end{aligned}$$

□

Exercise A.3. Let $\mathcal{E}$ be a Banach space and $K : \Omega \to \mathcal{E}$ be a compact operator with Ω a bounded closed subset of X.

- (i) Show that K is a uniform limit of finite dimensional transformations. **Hint:** Cover the compact set $\overline{K(\Omega)}$ by finitely many open balls, and use a partition of unity.
- (ii) Show that $I + K$ maps closed sets to closed sets.

□

References

[AM] A. Ambrosetti and A. Malchiodi, "Nonlinear Analysis and Semilinear Elliptic Problem," Cambridge, 2007.

[C] M. Chaperon, "On generating families", in H. Hofer, C.H. Taubes, A. Weinstein, E. Zehnder (Editors), The Floer Memorial Volume, Birkhäuser (1995).

[CHLS] K. Cieliebak, H. Hofer, J. Latschev and F. Schlenk, *Quantitative symplectic geometry*, preprint (2007).

[E] I. Ekeland, "Convexity Methods in Hamiltonian Mechanics," Springer, 1990.

[G] C. Golé, "Symplectic Twist Maps," World Scientific, 2001.

[HZ] Helmut Hofer and Eduard Zehnder, "Symplectic Invariants and Hamiltonian Dynamics," Birkhäuser, 1994.

[MSa] Dusa McDuff and Dietmar Salamon, "Introduction to Symplectic Topology," Oxford, 1998.

[Schl] Felix Schlenk,"Embedding Problems in Symplectic Geometry," de Gryter, 2005.

[Schw] M. Schwarz, "Morse Homology," Birkhäuser, 1993.

[St] M. Struwe, "Variational Methods," Springer, 1990.

[V] C. Villani, "Topics in Optimal Transportation," Graduate Studies in Mathematics series, vol. 58, AMS (2003).