The minimax principle in the calculus of variations

Alberto Abbondandolo

June 30, 2005

1 Introduction

When studying a functional f on an infinite-dimensional function space X, one is often interested in finding critical points which are not local minima. A simple yet powerful method to detect those critical points is the minimax method. The idea consists in detecting some complexity in the topology of X, or in the structure of the sublevels of f, to find a class Γ of subsets of X which somehow reveals such a topological complexity, and to show that the number

$$c := \inf_{\gamma \in \Gamma} \sup_{x \in \gamma} f(x)$$

is finite (even if the functional may be unbounded above and below). If the class Γ is positively invariant under the action of the negative gradient flow of f, and if a suitable compactness assumption known as the Palais-Smale condition holds, c is proved to be a critical value of f. Quite remarkably, the minimax method works also when no topological complexity is present, but the negative gradient flow of f exhibits some kind of rigidity.

In this article we shall describe these ideas, starting from the simplest minimax result, the mountain pass theorem. We will show how to apply the minimax method by discussing the existence question of solutions of a nonlinear elliptic boundary value problem, of closed geodesics on compact manifolds, and of closed characteristics on compact energy hypersurfaces.

2 The mountain pass theorem

Let us start by considering the following familiar fact. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth coercive function (that is, its sublevels have compact closure). If a sublevel $\{f < a\}$ is not connected - say $\{f < a\} = A \cup B$, with A, B disjoint open sets - then f has a critical point x at level

$$f(x) = c := \inf_{\gamma \in \Gamma} \max_{u \in \gamma} f(u) \geq a,$$

where Γ is the class of all continuous curves in $\mathbb{R}^n$ with one end-point in A and the other end-point in B. More figuratively: if there are two valleys, then there must be a mountain pass. Let us examine a possible proof.

First notice that any curve in the class Γ will have to cross the level $\{f = a\}$, so $c \geq a$. If by contradiction c is not a critical value of f, by the compactness of the sublevels there is some $\epsilon > 0$ such that $|\nabla f| \geq \epsilon$ on $\{c - \epsilon \leq f \leq c + \epsilon\}$. Then the negative gradient flow of f, that is the solution of

$$\partial_t \phi(t, u) = -\nabla f(\phi(t, u)), \quad \phi(0, u) = u,$$

pulls the sublevel $\{f \leq c + \epsilon\}$ down into the sublevel $\{f \leq c - \epsilon\}$ in finite time $2/\epsilon$. Indeed, if $\phi([0, t], u) \subset \{c - \epsilon \leq f \leq c + \epsilon\}$, then the inequalities

$$2\epsilon \geq f(u) - f(\phi(t, u)) = -\int_0^t \frac{d}{ds} f(\phi(s, u))\, ds = \int_0^t |\nabla f(\phi(s, u))|^2\, ds \geq \epsilon^2 t$$

imply that $t \leq 2/\epsilon$. By definition of c, we can find a continuous curve $\gamma \in \Gamma$ which is contained in $\{f \leq c + \epsilon\}$. But then the curve $\gamma' := \phi(2/\epsilon, \gamma)$ still has one end-point in A, the other one in B, and lies in $\{f \leq c - \epsilon\}$, contradicting the definition of c.

If we try to generalize this result to functions defined on an infinite dimensional real Hilbert space H, we encounter difficulties due to lack of compactness. Indeed, a continuous function on an infinite dimensional Hilbert space can never have compact sublevels (with respect to the norm topology). If we look back at the proof, we see that we have used coercivity to guarantee that if the level set $\{f = c\}$ contains no critical points, then ∇f is bounded away from zero on the strip $\{c - \epsilon \leq f \leq c + \epsilon\}$, for some small $\epsilon > 0$. A natural idea is then to replace the coercivity assumption by a condition implying the latter fact.

Definition. *Let $f : H \to \mathbb{R}$ be a continuously differentiable function on a real Hilbert space H. A sequence $(u_h) \subset H$ is said a* Palais-Smale sequence *if $f(u_h)$ is bounded and $Df(u_h)$ tends to zero. The function f is said to satisfy the* Palais-Smale condition *if every Palais-Smale sequence has a converging subsequence.*

The Palais-Smale condition readily implies the statement above. Assuming also that f is twice continuously differentiable, the negative gradient flow of f (a well-defined local flow because ∇f is continuously differentiable) pulls the sublevel $\{f \leq c + \epsilon\}$ down into $\{f \leq c - \epsilon\}$ in finite time. These observations lead to the following:

Theorem. (Mountain pass) *Let f be a twice continuously differentiable function on a real Hilbert space H, satisfying the Palais-Smale condition. Assume that a sublevel $\{f < a\}$ is not connected, and let A, B be two disjoint open sets such that $A \cup B = \{f < a\}$. Then f has a critical point x at level*

$$f(x) = c := \inf_{\gamma \in \Gamma} \max_{u \in \gamma} f(u) \geq a,$$

where Γ is the class of all continuous curves in H with one end-point in A and the other one in B.

If we are even more ambitious, and we wish to consider functions defined on a real Banach space E, we also encounter the problem of not having a gradient vector field. Indeed, the differential of f at x, $Df(x)$, is an element of the dual space E^*, but in this case we have no inner product on E by which we can represent $Df(x)$ as the product by some vector of E. This problem can be overcome by the notion of a *pseudo-gradient vector field*. In fact, it can be proved that if f is continuously differentiable on E, then there exists a locally Lipschitz vector field V defined on the complement of the critical points of f, such that

$$\|V(u)\| < \min\{\|Df(u)\|, 1\}, \quad Df(u)[V(u)] > \frac{1}{2} \min\{\|Df(u)\|, 1\} \|Df(u)\|.$$

In other words, even if there is no direction of steepest increase for f, we do have directions along which the increase of f is steep enough, and these directions can be selected in a locally Lipschitz way. Notice that pseudo-gradients are useful also in the case of a continuously differentiable function on a Hilbert space: in this case the gradient of f is just continuous, so it does not generate a flow. The Palais-Smale condition, as stated above, makes perfectly sense on the Banach space E (with the only difference that now $Df(u_h)$ tends to zero in the dual norm of E^*), and the mountain pass theorem holds for functions of class C^1 on a Banach space.

Actually, the fact that the domain of f has a vector structure is not relevant in this statement, and the mountain pass theorem holds also for functions defined on connected infinite dimensional manifolds. Since the essential feature is to dispose of a pseudo-gradient vector field, the right level of generality is to consider Banach manifolds M (that is a manifold modeled on a Banach space) endowed with a complete Finsler structure (that is a Banach norm on each tangent space of M, varying in a suitably regular way, inducing a complete distance on M).

3 A nonlinear elliptic boundary value problem

Let us consider a typical application of the mountain pass theorem to a semilinear elliptic boundary value problem. Let Ω be a smooth bounded domain in $\mathbb{R}^n$, and for $\lambda \in \mathbb{R}$, $p > 2$, consider the problem

$$\begin{cases} -\Delta u &= \lambda u + u|u|^{p-2} \qquad \text{in } \Omega, \\ \quad u &= 0 \qquad\qquad\qquad \text{on } \partial\Omega. \end{cases} \tag{1}$$

Let $0 < \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots$ be the eigenvalues of the Laplace operator $-\Delta$, with domain $H^1_0(\Omega)$, the Sobolev space of L^2 functions on Ω with weak first derivatives in L^2, vanishing on $\partial\Omega$. We claim that if $n = 2$, or if $n \geq 3$ and $2 < p < 2^* := 2n/(n-2)$, then problem (1) with $\lambda < \lambda_1$ has a non-trivial solution.

By elliptic regularity, the solutions of (1) are precisely the critical points of the functional

$$\mathcal{E}(u) = \frac{1}{2}\int_\Omega \left(|\nabla u(x)|^2 - \lambda u(x)^2\right)\, dx - \frac{1}{p}\int_\Omega |u(x)|^p\, dx.$$

We recall that $H^1_0(\Omega)$ continuously embeds into $L^p(\Omega)$, for every $p < +\infty$ if $n = 2$, for every $p \leq 2^*$ if $n \geq 3$. So the functional $\mathcal{E}$ is well defined, and actually continuously differentiable, on $H^1_0(\Omega)$, a Hilbert space with the inner product

$$\langle u, v\rangle_{H^1_0(\Omega)} = \int_\Omega \nabla u(x) \cdot \nabla v(x)\, dx.$$

Since $p > 2$, near zero the quadratic part of the functional $\mathcal{E}$ dominates over the part with the L^p norm. By the Rayleigh characterization of the first eigenvalue of the Laplacian,

$$\lambda_1 = \min_{u \in H_0^1(\Omega)\setminus\{0\}} \frac{\int_\Omega |\nabla u(x)|^2\,dx}{\int_\Omega u(x)^2\,dx},$$

the assumption $\lambda < \lambda_1$ implies that the quadratic part of $\mathcal{E}$ is positive definite. So we can find a small $\rho > 0$ such that

$$a := \inf_{\|u\|_{H_0^1(\Omega)} = \rho} \mathcal{E}(u) > 0.$$

On the other hand, the fact that $p > 2$ implies that

$$\lim_{\mu \to +\infty} \mathcal{E}(\mu u) = -\infty,$$

for every $u \neq 0$. Therefore, the sublevel $\{\mathcal{E} < a\}$ is not connected, and if we can prove the Palais-Smale condition, the mountain pass theorem will imply the existence of a critical point u with $\mathcal{E}(u) \geq a > 0$, that is a non-trivial solution of (1).

In order to prove the Palais-Smale condition, notice that the expression for the differential of $\mathcal{E}$,

$$D\mathcal{E}(u)[v] = \int_\Omega \nabla u(x)\cdot\nabla v(x)\,dx - \int_\Omega \left(\lambda u(x) + |u(x)|^{p-2}u(x)\right) v(x)\,dx,$$

and the compactness of the embedding of $H_0^1(\Omega)$ into $L^p(\Omega)$ for $p < 2^*$ imply that the gradient of $\mathcal{E}$ has the form

$$\nabla\mathcal{E}(u) = u + K(u), \tag{2}$$

where $K : H_0^1(\Omega) \to H_0^1(\Omega)$ is a compact map, i.e. it maps bounded sets into pre-compact ones. It is readily seen that when $\nabla\mathcal{E}$ has such a form, bounded Palais-Smale sequences are compact. Thus, it is enough to show that every Palais-Smale sequence is bounded. But this follows from the identity

$$p\mathcal{E}(u) - D\mathcal{E}(u)[u] = \left(\frac{p}{2} - 1\right) \int_\Omega \left(|\nabla u(x)|^2 - \lambda u(x)^2\right)\,dx,$$

together with the fact that the right-hand side term defines an equivalent norm on $H_0^1(\Omega)$, because $p > 2$ and $\lambda < \lambda_1$. This concludes the proof.

Actually, using the maximum principle one could show that under the same assumptions, problem (1) has a solution which is positive in Ω.

When $n \geq 3$ and $p = 2^* = 2n/(n-2)$, the functional f still exhibits a mountain pass geometry, but the Palais-Smale condition fails. In fact, the embedding of $H_0^1(\Omega)$ into $L^{2^*}(\Omega)$ is not compact, so the map K appearing in (2) is not compact, and bounded Palais-Smale sequences need not have a converging subsequence. We recall that the non-compactness of the embedding of $H_0^1(\Omega)$ into $L^{2^*}(\Omega)$ is due to the fact that the quotient

$$S(u) = \frac{\int_\Omega |\nabla u(x)|^2 \, dx}{\left(\int_\Omega |u(x)|^{2^*} \, dx\right)^{2/2^*}}$$

is invariant under rescaling $u \mapsto u_\mu(x) = u(\mu x)$.

When $\lambda = 0$, the Pohožaev identity - an integral formula obtained by multiplying the equation by $x \cdot \nabla u(x)$ - can be used to prove that problem (1) has no non-trivial solutions, when Ω is a star-shaped domain other than the whole $\mathbb{R}^n$.

When $\lambda \neq 0$, the presence in the functional of an L^2-norm - which rescales differently - breaks the symmetry, and the existence of non-trivial solutions is again possible. Indeed, Brezis and Nirenberg have shown that problem (1) with $p = 2^*$ has a non-trivial solution provided that $n \geq 4$ and $0 < \lambda < \lambda_1$, or $n = 3$ and $\lambda^* < \lambda < \lambda_1$, for some $\lambda^* \in [0, \lambda_1[$ depending on the domain Ω.

The proof is based on the fact that there is a certain threshold $s > 0$, related to the best Sobolev constant obtained by taking the infimum of $S(u)$ over all $u \in H_0^1$ (the domain is irrelevant here), below which the Palais-Smale condition holds. That is, every sequence (u_h) such that $\mathcal{E}(u_h)$ converges to some b less than s, and $D\mathcal{E}(u_h)$ tends to zero, is compact. The proof of the mountain pass theorem shows that the Palais-Smale condition is needed only at the minimax level c. In order to conclude, it is then enough to show that $c < s$. The value of c can be estimated by using the fact that the infimum of the quotient S over functions on the whole $\mathbb{R}^n$ is attained at the family of functions

$$u^*(x) = \left(\frac{\epsilon^2 n(n-2)}{(\epsilon^2 + |x|^2)^2}\right)^{\frac{n-2}{4}},$$

which are then solutions of (1) with $p = 2^*$, $\lambda = 0$, and $\Omega = \mathbb{R}^n$.

Another way to break the symmetry is to keep $\lambda = 0$ but to consider domains with a rich topology. For instance, Bahri and Coron have shown that if Ω is a domain with some non-zero singular homology group $H_k(\Omega; \mathbb{Z}_2)$, $k \geq 1$, then problem (1) with $p = 2^*$ and $\lambda = 0$ has a positive solution.

Elliptic equations having nonlinearities with the critical exponent 2^* arise naturally in some geometric problems. Consider a manifold M of dimension $n \geq 3$, with a metric g having scalar curvature k. The Yamabe problem calls for finding a metric g_0, conformally equivalent to g, having constant scalar curvature. If $g_0 = u^{4/(n-2)}g$, where the positive function u gives the conformal factor, one finds that u must solve the equation

$$-\frac{4(n-1)}{n-2}\Delta_g u = -ku + k_0 u|u|^{2^*-2},$$

where Δ_g is the Laplace-Beltrami operator associated to the metric g, and the constant k_0 is the scalar curvature of g_0. Again, the corresponding functional satisfies the Palais-Smale condition only below a certain threshold (actually, the same number s seen before; this because the lack of compactness is due to local concentration phenomena, and the metric structure of the whole ambient becomes irrelevant). The task is then to show that the minimax level is below that threshold, or equivalently that a certain best Sobolev constant for (M, g) is less than the corresponding constant for $\mathbb{R}^n$ with the flat metric (the latter constant is again the infimum of $S(u)$). This fact was proved by Aubin in the case $n \geq 6$ or (M, g) not locally conformally flat. Schoen has then treated the remaining case, by means of the positive mass theorem, a deep result in differential geometry.

4 A general minimax principle

Let us consider again a twice continuously differentiable function f on a real Hilbert space H. The vector field

$$V(u) = \frac{\nabla f(u)}{\sqrt{1 + \|\nabla f(u)\|^2}},$$

has the same nice properties of the gradient vector field of f, but in addition it is bounded. The advantage is that the flow of $-V$ is globally defined. When talking about the negative gradient flow of f, we will actually refer to such a flow. It will be also useful to dispose of a negative gradient flow *truncated below level b*. This is the flow of the vector field $-V_b$, where

$$V_b(u) = \varphi(f(u))V(u),$$

with φ a smooth function on $\mathbb{R}$ which is identically zero on $]-\infty, b]$, then increases up to reaching the value 1, and afterwards remains constantly equal to 1. This truncated negative gradient flow keeps the points in the sublevel $\{f \leq b\}$ fixed, and behaves as the negative gradient flow above b (except the fact that trajectories slow down as the value of f approaches b).

After these preliminaries, let us consider again the characterization of the critical level c appearing in the mountain pass theorem. This critical level was obtained as the infimum over a certain class Γ of sets γ - the curves with end-points in different components of $\{f < a\}$ - of the maximum of f over γ. But if we look back at the proof, we realize that the fact that these sets were curves was not essential. The important feature was that the negative gradient flow $\phi(t, \cdot)$ mapped a set of the class Γ into a set still belonging to the class Γ, for $t \geq 0$. This observation leads to the following general minimax theorem, due to Palais:

THEOREM. (General minimax) *Let f be a twice continuously differentiable function on a real Hilbert space H, satisfying the Palais-Smale condition. Let Γ be a class of subsets of H which is positively invariant under the action of the negative gradient flow ϕ of f (possibly truncated below level b): that is, if the set γ belongs to Γ, then the set $\phi(t, \gamma)$ belongs to Γ for all $t \geq 0$. Then, if the number*

$$c := \inf_{\gamma \in \Gamma} \sup_{u \in \gamma} f(u)$$

is finite (and larger than b), then c is a critical value of f.

The proof goes along the same lines of the proof of the mountain pass theorem: if c is not a critical value of f, the (possibly truncated) negative gradient flow $\phi(t_0, \cdot)$ pulls a sublevel $\{f \leq c + \epsilon\}$ down into the sublevel $\{f \leq c - \epsilon\}$ (with $c - \epsilon > b$), for some large t_0, by the Palais-Smale condition. Then we achieve a contradiction choosing a set $\gamma \in \Gamma$ on which f does not exceed $c + \epsilon$, and noticing that $\phi(t_0, \gamma)$ is a set which still belongs to the class Γ, by positive invariance, and on which f does not exceed $c - \epsilon$.

As we shall see in the last section, the possibility of working with a truncated negative gradient flow (assuming in this case that $c > b$) makes the application of this theorem easier. Again, an analogous result holds for continuously differentiable functions on Banach spaces, or more generally on Banach manifolds with a complete Finsler structure.

Trivial classes Γ are the class of all points in H, and the class consisting of the single set H, yielding to the infimum and the supremum of f, respectively.

More interesting classes are constructed by fixing a topological space X and considering the images of all continuous maps $h : X \to H$ belonging to a certain relative homotopy class.

5 Closed geodesics on compact manifolds

A typical application of the general minimax theorem is Birkhoff proof of the existence of a closed geodesic on the sphere S^2, endowed with an arbitrary metric g. Closed geodesics are precisely the critical points of the energy functional

$$\mathcal{S}(x) = \frac{1}{2} \int_0^1 g(\dot{x}(t), \dot{x}(t))\, dt$$

on the Hilbert manifold $H^1(\mathbb{T}, S^2)$ consisting of all one-periodic loops on S^2 of Sobolev regularity H^1 (here $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denotes the circle parameterized by $[0,1]$). This functional satisfies the Palais-Smale condition and it is bounded below, but its minima are just the trivial constant loops, on which $\mathcal{S} = 0$.

Let us use angle coordinates (θ, φ) on S^2, $-\pi/2 \leq \theta \leq \pi/2$, $0 \leq \varphi \leq 2\pi$ (θ is th latitude, φ the longitude). A (suitably regular) map $h : S^2 \to S^2$ induces a curve in $H^1(\mathbb{T}, S^2)$ parameterized by θ: the value of this curve at $\theta \in [-\pi/2, \pi/2]$ is the loop $t \mapsto h(\theta, 2\pi t)$. It is a curve that joins two constant loops. Let Γ be the set of curves in $H^1(\mathbb{T}, S^2)$ which are obtained by maps $h : S^2 \to S^2$ of topological degree 1. This class is clearly positively invariant under the action of the negative gradient flow of $\mathcal{S}$ (as of every homotopy fixing the constant loops).

If we can show that the minimax level

$$c := \inf_{\gamma \in \Gamma} \sup_{x \in \gamma} \mathcal{S}(x)$$

is positive, we will get a positive critical value of $\mathcal{S}$ by the general minimax theorem, hence a non-trivial closed geodesic. By considering the fact that loops with small energy also have a small diameter, it is easy to construct a homotopy on $\{\mathcal{S} < a\}$, for some small $a > 0$, which shrinks every loop to a point. If $h : S^2 \to S^2$ determines a curve γ with $\max_{x \in \gamma} \mathcal{S}(x) < a$, composition with this homotopy yields to a homotopy of h to a map whose image is a curve in S^2. A further homotopy then shows that the map h is homotopic to a constant, which is impossible if h has degree 1. This shows that $c \geq a > 0$, concluding the proof.

Actually, Ljusternik and Fet have proved that every compact manifold M has a non-trivial closed geodesic. Indeed, if M has non-zero fundamental group, it is enough to minimize $\mathcal{S}$ on some non-trivial homotopy class of loops. Otherwise, the fact that M is a compact manifold implies that some homotopy group $\pi_{k+1}(M)$, $1 \leq k < \dim M$, does not vanish. A construction similar to the one described above then allows to associate to every non-contractible map $h : S^{k+1} \to M$ a map $u : (B^k, \partial B^k) \to (H^1(\mathbb{T}, M), \{\mathcal{S} = 0\})$ which is not homotopically trivial (here B^k denotes the closed unit ball in $\mathbb{R}^k$, and the notation means that u maps the boundary of the ball B^k into the set of constant loops). Taking a minimax over the set of images of the maps u associated to every non-contractible map $h : S^{k+1} \to M$ yields to the desired critical point of $\mathcal{S}$ with positive energy.

It is conjectured that every compact manifold has infinitely many closed geodesics. Morse theory allows to prove this fact for the vast majority of manifolds, but not for the spheres. Franks has established the existence of infinitely many geodesics on S^2 by proving that every area-preserving homeomorphism of the open disk with two fixed points must have infinitely many periodic points. Proving the existence of infinitely many closed geodesics on higher dimensional spheres is a challenging open problem.

6 A rigidity property of a certain class of maps

It is important that the class Γ in the general minimax theorem is only required to be invariant under the action of the negative gradient flow, and not, say, under the action of any continuous homotopy on which the function f is non-increasing. Indeed, too many bad things can be done on an infinite dimensional Hilbert space by arbitrary continuous maps, whereas the maps arising from our negative gradient flow might show some rigidity, forcing them to behave as maps on finite dimensional spaces.

Let us clarify this point by considering the following example, due to Benci and Rabinowitz. It may sound a bit artificial at this moment (simpler examples could be built), but we will find it useful in the next section. Assume that our Hilbert space is endowed with an orthogonal splitting $H = H^- \oplus H^+$, fix a unit vector u^+ in H^+, and consider the sets

$$S = \left\{ u \in H^+ \mid \|u\| = \rho \right\}, \; Q = \left\{ u + \lambda u^+ \mid u \in H^-, \; \|u\| \leq \sigma, \; 0 \leq \lambda \leq \tau \right\},$$
$$\partial Q = \left\{ u + \lambda u^+ \in Q \mid \lambda \in \{0, \tau\} \text{ or } \|u\| = \sigma \right\},$$

for some positive numbers ρ, σ, τ such that $\tau > \rho$. The latter inequality implies that the intersection $Q \cap S$ is not empty (see figure 1).

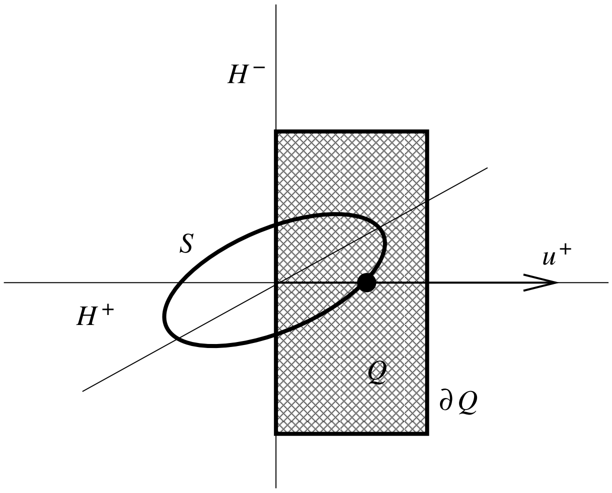

Figure 1: The sets S, Q, ∂Q.

If the linear subspace H^- is finite-dimensional, a simple argument involving the topological degree shows the following fact: the image of any continuous map $h : Q \to H$ which is the identity on ∂Q has non-empty intersection with S.

When H^- is infinite-dimensional, this fact is not true anymore. Indeed, it is not difficult to see that the set Q is homeomorphic to an infinite dimensional closed ball B, by a homeomorphism ψ mapping ∂Q onto the infinite-dimensional sphere ∂B. If B is the closed ball of an infinite dimensional Hilbert space, for instance the space ℓ_2 of all square summable sequences (x_h) endowed with the norm $|x|_2 = \left(\sum_{h=0}^{\infty} |x_h|^2\right)^{1/2}$, the continuous map

$$g(x_0, x_1, x_2, \dots) = (\sqrt{1 - |x|_2^2}, x_0, x_1, x_2, \dots)$$

maps B into ∂B and is a shift operator on ∂B. In particular, it is a continuous map on B without fixed points, and it can be used to define a map $h : B \to$

∂B which is the identity on ∂B, by setting

$$h(x) = \mu(x)x + (1-\mu(x))g(x), \quad \text{with } \mu(x) \geq 1 \text{ such that } |h(x)|_2 = 1.$$

Conjugation by the homeomorphism ψ produces a continuous map from Q to ∂Q which is the identity on ∂Q, providing us with the desired counterexample.

In other terms, when H^- is infinite dimensional the sets ∂Q and S can be unlinked by means of a continuous map. The situation changes if we restrict the class of maps $h : Q \to H$ to those of the form

$$h(u) = u + K(u), \tag{3}$$

where K is a continuous compact map. In this case indeed, the argument for a finite-dimensional H^- can be applied, by replacing the topological degree by the Leray-Schauder degree (which is invariant precisely with respect to homotopies of the form above), and one proves that ∂Q and S cannot be unlinked by means of continuous maps of this form.

7 Closed characteristics on compact energy hypersurfaces

Consider $\mathbb{R}^{2n}$ with coordinates $(p_1, \dots, p_n, q_1, \dots, q_n)$, endowed with the standard symplectic form

$$\omega := dp \wedge dq = \sum_{j=1}^{n} dp_j \wedge dq_j.$$

Let Σ be a compact connected hypersurface in $\mathbb{R}^{2n}$. The restriction of ω to the tangent space $T_x\Sigma$ has a one-dimensional kernel, which varies smoothly with x. In other words, there is a smooth line bundle

$$\mathcal{L}_\Sigma := \{(x, u) \in T\Sigma \mid \omega(u, v) = 0 \ \forall v \in T_x\Sigma\}$$

over Σ. We wish to discuss the classical problem of finding a closed characteristic for $\mathcal{L}_\Sigma$, that is a closed curve everywhere tangent to $\mathcal{L}_\Sigma$.

This geometric problem has a dynamical interpretation. Indeed, let H be a smooth real function on $\mathbb{R}^{2n}$ such that Σ is the inverse image of the regular value 1. The function H - the Hamiltonian - generates a vector field X_H on $\mathbb{R}^{2n}$ by the formula

$$\omega(X_H(x), u) = -DH(x)[u], \quad \forall u \in \mathbb{R}^{2n},$$

or equivalently

$$X_H(x) = J\nabla H(x), \quad \text{with } J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}.$$

The Hamiltonian vector field X_H is tangent to Σ and belongs to $\mathcal{L}_\Sigma$. Therefore, the hypersurface Σ is invariant for the flow of X_H, and the flow orbits are precisely the characteristics. So finding a closed characteristic on Σ is equivalent to finding a periodic orbit of X_H with energy $H = 1$.

Up to changing the Hamiltonian, we may assume that all the values in an interval $]1-\delta_0, 1+\delta_0[$ are regular for H, and that the corresponding level sets $\Sigma_\eta := \{H = \eta\}$ are all connected (hence diffeomorphic to $\Sigma = \Sigma_1$). We would like to sketch Hofer and Zehnder's proof of the fact that there is a dense set of values $\eta \in]1 - \delta_0, 1 + \delta_0[$ for which Σ_η admits a closed characteristic.

This proof is based on the fact that the one-periodic orbits of X_H are critical points of the action functional

$$\mathcal{A}_H(x) = \int_{\mathbb{T}} x^*(p\,dq - H\,dt) = \frac{1}{2}\int_0^1 \dot{x}(t) \cdot Jx(t)\,dt - \int_0^1 H(x(t))\,dt,$$

on the space of loops $x : \mathbb{T} \to \mathbb{R}^{2n}$.

Clearly, it is enough to show that for every $\delta > 0$ there is a closed characteristic on some Σ_η with $|\eta - 1| < \delta$. We can take advantage of the fact that we are free to change the Hamiltonian, as long as it has the level sets Σ_η, $|\eta - 1| < \delta$. Denoting by B the bounded component of the complement of $\{1 - \delta \leq H \leq 1 + \delta\}$, we may assume that B contains the origin. We can modify H in such a way that H vanishes identically on B, then it grows, parameterizing all the hypersurfaces Σ_η, $|\eta - 1| < \delta$, in a strictly increasing way, then it remains constant in a large ball, and finally it smoothly switches to the quadratic form $(3/2)\pi|x|^2$. By choosing H in this way, one can ensure that all the constant orbits and all the one-periodic orbits which do not lie on Σ_η for some $|\eta - 1| < \delta$ have non-positive action. So it is enough to prove that the functional $\mathcal{A}_H$ has a positive critical value.

Using the Fourier series decomposition

$$x(t) = \sum_{k\in\mathbb{Z}} e^{2\pi ktJ}\hat{x}_k, \quad \hat{x}_k \in \mathbb{R}^{2n},$$

one sees that the quadratic part of the action functional has the form

$$\int_0^1 \dot{x}(t) \cdot Jx(t)\, dt = 2\pi \sum_{k \in \mathbb{Z}} k|\hat{x}_k|^2, \tag{4}$$

so it is positive on an infinite dimensional linear space, negative on an infinite dimensional linear space, and null on the $2n$-dimensional space spanned by the constant loops. The specific form of (4) suggests to choose as domain of the action functional the Sobolev space $H^{1/2}(\mathbb{T}, \mathbb{R}^{2n})$, the space of square integrable one-periodic curves x in $\mathbb{R}^{2n}$ with

$$\|x\|^2_{H^{1/2}} := |\hat{x}_0|^2 + 2\pi \sum_{k \in \mathbb{Z}} |k|\, |\hat{x}_k|^2 < +\infty.$$

This is indeed a Hilbert norm on $H^{1/2}(\mathbb{T}, \mathbb{R}^{2n})$. The functional $\mathcal{A}_H$ is smooth on this space, and its gradient takes the form

$$\nabla \mathcal{A}_H(x) = Lx + K(x), \tag{5}$$

where L is the self-adjoint Fredholm operator representing the quadratic form (4) with respect to the $H^{1/2}$-Hilbert product, and K is a compact map. A gradient of the form (5) again implies that bounded Palais-Smale sequences are compact. The Palais-Smale condition then follows from the fact that the Hamiltonian H is quadratic outside a large ball, and has no one-periodic orbits there (the large orbits are all periodic, but their period is 2/3).

Consider the splitting $H^{1/2}(\mathbb{T}, \mathbb{R}^{2n}) = H^- \oplus H^+$, with

$$H^- = \{x \mid \hat{x}_k = 0 \text{ for } k > 0\}, \quad H^+ = \{x \mid \hat{x}_k = 0 \text{ for } k \leq 0\}.$$

Let S, Q, and ∂Q be the sets defined in the previous section, with

$$u^+(t) = \frac{1}{\sqrt{2\pi}} e^{2\pi t J} u_0, \quad u_0 \in \mathbb{R}^{2n},\ |u_0| = 1,$$

and constants ρ, σ, τ to be determined. Since the quadratic form (4) is positive on H^+ and the Hamiltonian H vanishes near the origin, we can find a small $\rho > 0$ such that

$$\inf_{x \in S} \mathcal{A}_H(x) > 0.$$

The fact that the quadratic form (4) is semi-negative on H^- and the behavior of $H(x)$ for large $|x|$ imply that if σ and τ are suitably large (in particular $\tau > \rho$), then

$$\sup_{x\in\partial Q} \mathcal{A}_H(x) \leq 0.$$

Let Γ be the set of all images of maps

$$h : Q \to H^{1/2}(\mathbb{T}, \mathbb{R}^{2n}),$$

which are the identity on ∂Q and are of the form

$$h(x) = e^{\alpha(x)L}(x + K(x)), \tag{6}$$

with α a continuous real valued function, and K a continuous compact map. This class of maps is more general than the one considered in the previous section, but the fact that $e^{\alpha L}$ commutes with the projections onto H^- and H^+ ensures that ∂Q and S cannot be unlinked even inside this class. Therefore, any $\gamma \in \Gamma$ has non-empty intersection with S, so

$$c := \inf_{\gamma\in\Gamma} \sup_{x\in\gamma} \mathcal{A}_H(x) \geq \inf_{x\in S} \mathcal{A}_H(x) > 0.$$

We would like to apply the general minimax theorem, and conclude that c is the desired positive critical value.

The number c being clearly finite, it is enough to show that Γ is positively invariant under the action of the negative gradient flow ϕ of $\mathcal{A}_H$, truncated below level 0. Let $\gamma = h(Q) \in \Gamma$ and $t \geq 0$. Then $\phi(t, \gamma)$ is the image of Q by the map $\phi(t, h(\cdot))$. This map is the identity on ∂Q because ∂Q lies in $\{\mathcal{A}_H \leq 0\}$ and ϕ is truncated below level 0. It is of the form (6) because by (5) the truncated negative gradient flow of $\mathcal{A}_H$ has the form

$$\phi(t, x) = e^{-\theta(t,x)L}(x + K(t, x)),$$

for some continuous function $0 \leq \theta(t, x) \leq t$ and for some continuous compact map K. This concludes the proof.

This result was refined by Struwe, who proved the existence of a closed characteristic on Σ_η for almost every η, in the sense of the Lebesgue measure. We could try to use the abundance of closed characteristics on energy levels near Σ to get the existence of one on Σ by taking a limit. But this process produces a closed characteristic on Σ only if we can bound the periods of the approximating closed orbits, otherwise a more general invariant set comes out. Actually, Ginzburg, Herman, and Gürel have produced examples of compact hypersurfaces without any closed characteristic.

As conjectured by Weinstein and proved by Viterbo, closed characteristics always exist on contact-type compact hypersurfaces (i.e. hypersurfaces Σ on which the restriction of ω is the differential of a 1-form λ such that $\lambda \wedge d\lambda \wedge \cdots \wedge d\lambda$ is a volume form). In this case one should even expect a multiplicity result. For hypersurfaces which bound a strictly convex set in $\mathbb{R}^{2n}$, for instance, the existence of n closed characteristics is conjectured. The best result so far is due to Long, who could prove the existence of $[n/2]+1$ of them. Hofer, Wysocki and Zehnder have proved that when $n = 2$ there are either 2 or infinitely many closed characteristics (for any contact-type hypersurface diffeomorphic to S^3), by using the already mentioned theorem by Franks on periodic points of area-preserving homeomorphisms of the disk. Proving an analogous result for $n \geq 3$ is an intriguing open problem.

References

[1] A. Abbondandolo. *Morse theory for Hamiltonian systems*, volume 425 of *Pitman Research Notes in Mathematics.* Chapman & Hall, London, 2001.

[2] A. Ambrosetti and A. Malchiodi. *Perturbation methods is semilinear elliptic problems in $\mathbb{R}^n$.* Birkhäuser, 2005.

[3] T. Aubin. *Some nonlinear problems in Riemannian geometry.* Springer, New York, 1998.

[4] K. C. Chang. *Infinite-dimensional Morse theory and multiple solution problems.* Birkhäuser, Boston, 1993.

[5] K. C. Chang. *Methods in nonlinear analysis.* Springer, 2005.

[6] N. Ghoussoub. *Duality and perturbation methods in critical point theory.* Cambridge University Press, Cambridge, 1993.

[7] H. Hofer and E. Zehnder. *Symplectic invariants and Hamiltonian dynamics.* Birkhäuser, Basel, 1994.

[8] W. Klingenberg. *Riemannian geometry.* Walter de Gruyter & Co., Berlin, 1982.

[9] J. Mawhin and M. Willem. *Critical point theory and Hamiltonian systems.* Springer, New York, 1989.

[10] P. H. Rabinowitz. *Minimax methods in critical point theory with applications to differential equations*, volume 65 of *CBMS Regional Conference Series in Mathematics.* Amer. Math. Soc., Providence, R.I., 1986.

[11] M. Schechter. *Linking methods in critical point theory.* Birkäuser, Boston, 1999.

[12] M. Struwe. *Variational methods.* Springer-Verlag, Berlin, third edition, 2000.

[13] M. Willem. *Minimax theorems.* Birkhäuser, Boston, 1996.