Morse and Conley theory of flows
A topological introduction

⋆ ⋆ ⋆

Gradient flows and Morse Theory

José Manuel Rodríguez-Sanjurjo, Jesús M. Ruiz
Departamento de Álgebra, Geometría y Topología, UCM

Contents

One of the aims of Morse Theory is to study the global structure of a manifold using a smooth function defined on the manifold. This theory reveals a deep relationship between the topology of the manifold and the properties of the map in the neighborhood of its critical points. So, it establishes a connection between Analysis and Topology and provides an illuminating example of a mathematical model in which some local properties determine the global structure. There is another aspect of this theory which is specially attractive, namely its relation with the theory of flows and dynamical systems. The reason is that the critical points of the map are stationary points of certain flows, the gradient and the Morse flows associated with it. Therefore, the topology of the manifold and the dynamics of the flow are also correlated. Adopting this point of view we obtain a fruitful connection between Topology and Dynamics that allows a double perspective of each situation. In this chapter we undertake the study of this topic. Our treatment of most notions of a dynamical nature is essentially topological.

1.1 Gradient systems

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^∞ map. The *gradient system of* f is the system of differential equations

$$x' = -\operatorname{grad}_x f,$$

where

$$x = (x_1, \dots, x_n) \quad \text{and} \quad \operatorname{grad}_x f = \Big(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)\Big).$$

The function f is sometimes called the *potential function* of the system, a terminology which comes from Physics. In the sequel we shall make use of the fact that, for $y = (y_1, \dots, y_n) \in \mathbb{R}^n$,

$$d_x f(y) = \langle \operatorname{grad}_x f, y \rangle = \sum\nolimits_i \frac{\partial f}{\partial x_i}(x) y_i,$$

where $d_x f : \mathbb{R}^n \to \mathbb{R}$ is the derivative (Jacobian) of f at $x \in \mathbb{R}^n$, and $\langle\, ,\rangle$ denotes the (Euclidean) scalar product.

If $x(t)$ is a solution of the gradient system and we consider the function

$$\gamma(t) = f(x(t)),$$

then, by the chain rule, we have

$$\begin{aligned}\gamma'(t) &= d_{x(t)} f(x'(t)) = \langle \operatorname{grad}_{x(t)} f, x'(t) \rangle = \langle \operatorname{grad}_{x(t)} f, -\operatorname{grad}_{x(t)} f \rangle \\ &= -\|\operatorname{grad}_{x(t)} f\|^2 \leq 0.\end{aligned}$$

In other words, f decreases along the solutions of the system. We remark that $\gamma'(t) = 0$ if and only if $\operatorname{grad}_{x(t)} f = 0$, which happens only when $x(t)$ is a constant solution corresponding to a critical point of f. As a consequence, f is a *strict Lyapunov map,* i.e. it is *strictly decreasing* along nonconstant solutions. It follows from this that a periodic solution of the gradient system must necessarily be constant.

If x_0 is an *isolated critical point and also a local minimum* of f then there is a neighborhood V of x_0 such that: (i) $f(x) > f(x_0)$ for every $x \in V \setminus \{x_0\}$, and (ii) f is strictly decreasing along orbits of points $x \in V \setminus \{x_0\}$. As a consequence, x_0 is *asymptotically stable.*

Indeed, since x_0 is a local minimum and an isolated singularity, it is a strict local minimum. Let us see stability. We pick a neighborhood V of x_0 where $f(x) > f(x_0)$ for $x \neq x_0$ and there is no other critical point. Now fix $\varepsilon > 0$. We can suppose $B[x_0, \varepsilon] \subset V$, so that $\mu = \min\{f(x) : x \in S(x_0, \varepsilon)\} > f(x_0)$ and there is $\delta < \varepsilon$ such that $0 < \|x - x_0\| < \delta$ implies $f(x_0) < f(x) < \mu$. This δ gives stability. Otherwise, $\|xt - x_0\| \geq \varepsilon$ for some $t > 0$, hence $\|xs - x_0\| = \varepsilon$ for some s with $0 < s \leq t$, and then $\mu \leq f(xs)$. Since f decreases along trajectories $f(x) > f(xs) \geq \mu$, a contradiction.

Asymptotic stability now. Suppose $0 < \|x - x_0\| < \delta$. First note that stability and the *escape lemma* imply that the domain of the integral curve of x is not bounded from above. Thus we must see that $\lim_{t\to\infty} xt = x_0$. Suppose, on the contrary, that there are $\eta > 0$ and a sequence $t_k \to \infty$ such that $\|xt_k - x_0\| \geq \eta$. Fix any $s > 0$. By stability $xt_k \in B[x_0, \varepsilon]$, which is compact, hence some subsequence of xt_k has a limit $y \neq x_0$. We can suppose $xt_k \to y$ and $t_k + s < t_{k+1}$. Thus

$$f(xt_k) > f(x(t_k + s)) > f(xt_{k+1}) > f(x(t_{k+1} + s)).$$

Now, since $xt_k \to y$ we have $x(t_k + s) \to ys$, and so

$$f(y) = \lim_k f(xt_k) = \lim_k f(x(t_k + s)) = f(ys).$$

This is a contradiction, since y is not stationary and f must be strictly decreasing on its trajectory. ∎

When c is a regular value of f then $f^{-1}(c)$ is a *level hypersuperface* of $\mathbb{R}^n$ whenever it is nonempty. The tangent hyperplane at a point $x \in f^{-1}(c)$ is given by

$$\{v \in \mathbb{R}^n : v \perp \operatorname{grad}_x f\},$$

hence the solutions of the gradient system are ortogonal to the level hypersurfaces of f. Moreover, f increases most rapidly at x in the direction of $\operatorname{grad}_x f$ (that is, $\|d_x f\| = |d_x f(u)|$ for $u = \operatorname{grad}_x f / \|\operatorname{grad}_x f\|$).

Remark 1.1.1 For $n = 1$ all systems $x' = F(x)$ are gradient as $F(x) = -\operatorname{grad}_x f$ with $f(x) = -\int_0^x F(t)dt$. On the other hand, for $n > 1$, a system $x' = F(x)$ must satisfy the necessary condition

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial F_j}{\partial x_i}$$

in order to be gradient since, in this case,

$$\frac{\partial F_i}{\partial x_j} = -\frac{\partial^2 f}{\partial x_j \partial x_i} = -\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial F_j}{\partial x_i},$$

where f is the potential of the gradient. On the other hand, it is well known that this condition is also sufficient [Fle]. ∎

If A is a real matrix, the system $x' = Ax$ is not, in general a gradient system. The following proposition identifies the only case in which a linear system has this property.

Proposition 1.1.2 *The linear system $x' = Ax$ is gradient if and only if the matrix A is symmetric.*

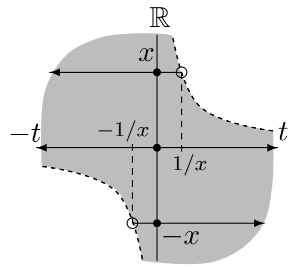
Proof. If $A = [a_{ij}]$ is symmetric then $x_i' = \sum_{j=1}^n a_{ij} x_j$ with $a_{ij} = a_{ji}$. We define the map

$$f : \mathbb{R}^n \to \mathbb{R} : x = (x_1, \ldots, x_n) \mapsto -\sum_{i,j=1}^n \tfrac{1}{2} a_{ij} x_i x_j.$$

Clearly, $-\sum_{j=1}^n a_{ij} x_j$ is the i-th coordinate of $\operatorname{grad}_x f$ and, thus, the system is a gradient system. The converse is a consequence of the second part of the previous remark. ∎

Remark 1.1.3 It must be also observed that the solutions of gradient systems do not define in general *flows* in $\mathbb{R}^n$: the domains of some maximal integral curves can be bounded from below or from above. The escape lemma mentioned before shows that a sufficient (although by no means necessary) condition to have a true flow is that each solution is contained in a compact set.

For instance, the gradient system of the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = -\frac{1}{3}x^3$ is $x' = x^2$. The solution $x = \frac{1}{1-t}$ blows up at the finite time $t = 1$ and cannot be extended to the whole line, hence the solutions of this system do not define a flow in $\mathbb{R}$. The trajectories of this system are $xt = \frac{x}{1-xt}$, with maximal domain $I(x)$ given by $1 - xt > 0$:



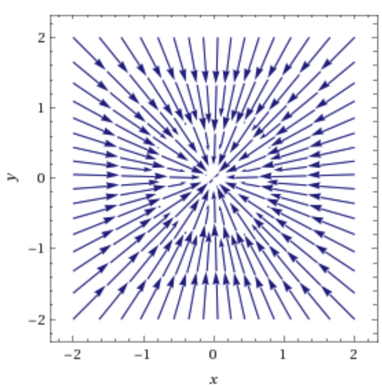
$$I(x) = \begin{cases} (\frac{1}{x},\, \infty) & \text{if } x < 0, \\ (-\infty,\, \infty) = \mathbb{R} & \text{if } x = 0, \\ (-\infty,\, \frac{1}{x}) & \text{if } x > 0. \end{cases}$$

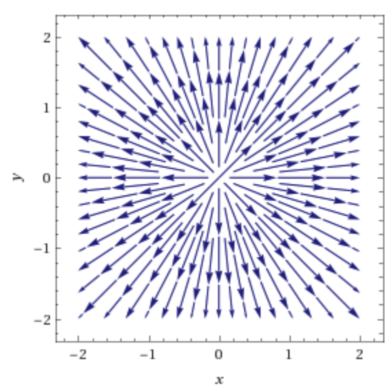
∎

Example 1.1.4 (1) Consider $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$, with $\operatorname{grad}_x f = (x_1, x_2)$. The origin is the only critical point and, also, a minimum of f. Hence as explained before, the origin is asymptotically stable. This is easy to check, as the gradient flow (the flow of $-\operatorname{grad} f$) is given by $\varphi(x, t) = e^{-t}(x_1, x_2)$. Obviously the origin is asymptotically stable. As a matter of fact, any open ball $B(0, \delta)$ is *positively invariant:* if $\|x\| < \delta$ and $t > 0$ then

$$\|\varphi(x, t)\| = e^{-t}\|x\| < \|x\| < \delta.$$

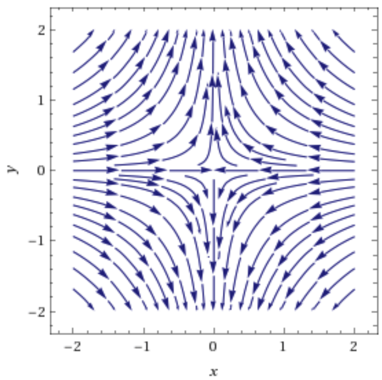
More generally, such an estimation shows that for any $x \neq 0$ and $t > \log(\|x\|/\delta)$, it is $\|\varphi(x, t)\| < \delta$, so that all trajectories enter at a finite time any neigborhood of the origin. We see that the origin is a *global attractor:* it is the ω-limit of all points. The figure below, left, represents this flow.

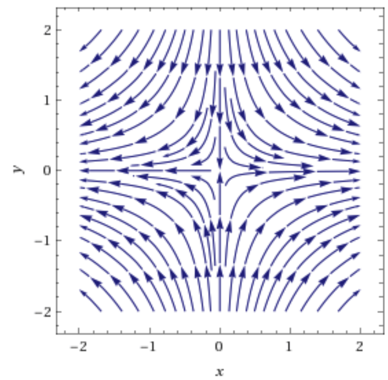




The figure above right represents the opposite situacion: for the function $-f$, the gradient flow is $\varphi(x, -t)$ and the origin is a *global repeller*.

(2) Let now $f : \mathbb{R}^2 \to \mathbb{R}^2$ be $f(x_1, x_2) = \frac{1}{2}(x_1^2 - x_2^2)$, with $\operatorname{grad}_x f = (x_1, -x_2)$. The origin is the only critical point but is not an extreme and not asymptotically stable. The gradient flow is given by $\varphi(x, t) = (e^{-t}x_1, e^{t}x_2)$, depicted below, left. Below right is the opposite gradient flow.





∎

Suppose now that x_0 is a critical point of f. The linearized system of $\operatorname{grad} f$ at x_0 is given by

$$y' = -(d_{x_0} \operatorname{grad} f)(y),$$

where the differential $d_{x_0} \operatorname{grad} f : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map whose jacobian matrix is

$$[a_{ij}] = \Big(\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \Big).$$

Since mixed partial derivatives are equal, we have that $a_{ij} = a_{ji}$ and the matrix $[a_{ij}]$ is symmetric. This matrix is called the *Hessian matrix* of f at x_0. As it is well known, symmetric matrices have only real eigenvalues. The critical point x_0 is said to be *non-degenerate* if its Hessian matrix is non-degenerate. It follows from the inverse function theorem applied to the function $\operatorname{grad} f : \mathbb{R}^n \to \mathbb{R}^n$ that every non-degenerate critical point x_0 has a neighborhood where there are no other critical points (take any neighborhood of 0 diffeomorphically mapped by $\operatorname{grad} f$ onto a neighborhood of 0, so that $\operatorname{grad} f$ is injective in that neighborhood). In other words, *non-degenerate critical points are isolated*. Moreover, non-degenerate critical points are *hyperbolic* points of the gradient system. As a matter of fact, the linearization $-d_{x_0} \operatorname{grad} f$ has only positive and negative eigenvalues and, by the *Hartman-Grobman linearization theorem* (although more directly here we can use the Morse Lemma 1.5.5 to get a local diffeomorphism and not only a homeomorphism), the gradient flow is conjugate to the flow of the linearized system in a neighborhood of x_0.

1.2 Gradients on manifolds

All the previous notions can be defined in the more general framework of C^∞ maps $f : M \to \mathbb{R}$, where $M \subset \mathbb{R}^p$ is a C^∞-manifold of dimension $m \leq p$ (in particular, M can be an open subset of $\mathbb{R}^m$ or a hypersurface of $\mathbb{R}^{m+1}$). For every $x \in M$, the differential $d_x f : T_x M \to \mathbb{R}$ is a linear form on the tangent space at x and, hence, there exists a *unique* vector in $T_x M$, which again we denote by $\operatorname{grad}_x f$, such that $d_x f(y) = \langle \operatorname{grad}_x f, y\rangle$ for every $y \in T_x M$. In this way, we define a function $\operatorname{grad} f : M \to \mathbb{R}^p$, the *gradient field* of f, that maps each point $x \in M$ to a *tangent vector* to M at x, a particular case of *tangent field on* M. Let us specify the local representation.

Let x be a point of M and suppose that $\varphi : U \to M$ is a parametrization of a neighborhood of x (since $M \subset \mathbb{R}^p$ we may also see φ as a function from U to $\mathbb{R}^p$). Consider the basis

$$\Big\{\frac{\partial\varphi}{\partial u_1}(u), \dots, \frac{\partial\varphi}{\partial u_m}(u)\Big\}$$

of $T_x M$, where $u \in U$ and $\varphi(u) = x$. Then $\operatorname{grad}_x f$ can be expressed in terms of this basis

$$\operatorname{grad}_x f = \sum_{i=1}^{m} \rho_i(u)\frac{\partial\varphi}{\partial u_i}(u).$$

We are interested in obtaining the expression of the local coordinates $\rho_i(u)$ of $\operatorname{grad}_x f$.

We have that

$$\begin{aligned}\frac{\partial(f\circ\varphi)}{\partial u_j}(u) &= d_x f\Big(\frac{\partial\varphi}{\partial u_j}(u)\Big) = \Big\langle \operatorname{grad}_x f, \frac{\partial\varphi}{\partial u_j}(u)\Big\rangle \\ &= \Big\langle \sum_{i=1}^{m} \rho_i(u)\frac{\partial\varphi}{\partial u_i}(u), \frac{\partial\varphi}{\partial u_j}(u)\Big\rangle = \sum_{i=1}^{m}\rho_i(u)\Big\langle\frac{\partial\varphi}{\partial u_i}(u), \frac{\partial\varphi}{\partial u_j}(u)\Big\rangle \\ &= \sum_{i=1}^{m}\rho_i(u) g_{ij}(u), \quad g_{ij}(u) = \Big\langle\frac{\partial\varphi}{\partial u_i}(u), \frac{\partial\varphi}{\partial u_j}(u)\Big\rangle.\end{aligned}$$

Thus $G(u) = [g_{ij}(u)]$ is the *Gramm matrix* of the scalar product in $T_x M$ with respect to the basis $\big\{\frac{\partial\varphi}{\partial u_j}(u)\big\}$, and is consequently a nonsingular matrix. The linear system of equations

$$\sum_{i=1}^{m}\rho_i(u) g_{ij}(u) = \frac{\partial(f\circ\varphi)}{\partial u_j}(u),\ j = 1, \dots, m,$$

expressed in matrix form as

$$(\rho_1(u), ..., \rho_m(u))G(u) = \left(\frac{\partial(f\circ\varphi)}{\partial u_1}(u), \dots, \frac{\partial(f\circ\varphi)}{\partial u_m}(u)\right),$$

has a unique solution for the local coordinates $\rho_i(u)$ of $\operatorname{grad}_x f$, namely

$$(\rho_1(u), \ldots, \rho_m(u)) = \left(\frac{\partial (f \circ \varphi)}{\partial u_1}(u), \ldots, \frac{\partial (f \circ \varphi)}{\partial u_m}(u) \right) G(u)^{-1},$$

which are obviously of class C^∞. Hence the gradient field of a C^∞-function on a manifold is C^∞ as well. It is important to note that, in general, it is not true that $(\operatorname{grad} f) \circ \varphi = \operatorname{grad}(f \circ \varphi)$, as we will stress later.

Remark 1.2.1 If $F : W \to \mathbb{R}$ is a C^∞ function on an open set W of $\mathbb{R}^p$ which contains the manifold M, then the gradient (tangent) field of $f = F|M$ can be expressed in the following terms: given $x \in M$ consider $\operatorname{grad}_x F$. Then $\operatorname{grad}_x f = \pi_x(\operatorname{grad}_x F)$, where $\pi_x : \mathbb{R}^p \to T_x M$ denotes the ortogonal projection. Indeed, the component of $\operatorname{grad}_x F$ ortogonal to $T_x M$ does not play any role when multiplied by vectors of $T_x M$. This statement can also be interpreted locally. ∎

As in the Euclidean case, the equation

$$x' = -\operatorname{grad}_x f$$

defines a system of differential equations in the manifold M, also called *gradient system,* such that the function f is a strict Lyapunov map for the system. Again, not every system $x' = \xi(x)$ of differential equations on M defines a flow, that is, there may be maximal integral curves with domain bounded from above or from below. It does define a flow $\varphi : M \times \mathbb{R} \to M$ when the tangent field ξ vanishes off a compact set, which is always the case if M is compact.

Let $x_0 \in M$ be a critical point of f and $\varphi : U \subset \mathbb{R}^m \to M$ a parametrization of a neighborhood of x_0 with $\varphi(u_0) = x_0$, where $u_0 \in U$. The differential $d_{x_0} \operatorname{grad} f$ can be seen as a linear map from $T_{x_0} M$ to $\mathbb{R}^p$. We are interested in its equation and, more specifically, in the local representation (the jacobian) of the differential respect to the basis $\frac{\partial \varphi}{\partial u_j}(u_0)$ of $T_{x_0} M$.

In order to calculate the jacobian, we have

$$\begin{aligned} d_{x_0} \operatorname{grad} f \Big(\frac{\partial \varphi}{\partial u_i}(u_0) \Big) &= \frac{\partial}{\partial u_i} \left((\rho_1, \ldots, \rho_m) \begin{pmatrix} \frac{\partial \varphi}{\partial u_1} \\ \vdots \\ \frac{\partial \varphi}{\partial u_m} \end{pmatrix} \right)(u_0) \\ &= \frac{\partial}{\partial u_i} \left(\Big(\frac{\partial (f \circ \varphi)}{\partial u_1}, \ldots, \frac{\partial (f \circ \varphi)}{\partial u_m} \Big) G(u_0)^{-1} \begin{pmatrix} \frac{\partial \varphi}{\partial u_1} \\ \vdots \\ \frac{\partial \varphi}{\partial u_m} \end{pmatrix} \right)(u_0). \end{aligned}$$

This derivative is simply

$$\Big(\frac{\partial^2(f\circ\varphi)}{\partial u_i\partial u_1}(u_0),\dots,\frac{\partial^2(f\circ\varphi)}{\partial u_i\partial u_m}(u_0)\Big)G(u_0)^{-1}\begin{pmatrix}\frac{\partial\varphi}{\partial u_1}(u_0)\\ \vdots\\ \frac{\partial\varphi}{\partial u_m}(u_0)\end{pmatrix},$$

since the fact that

$$0=(\rho_1(u_0),\dots,\rho_m(u_0))=\Big(\frac{\partial(f\circ\varphi)}{\partial u_1}(u_0),\dots,\frac{\partial(f\circ\varphi)}{\partial u_m}(u_0)\Big)G(u_0)^{-1}$$

(because x_0 is a critical point) implies

$$\frac{\partial(f\circ\varphi)}{\partial u_1}(u_0)=\cdots=\frac{\partial(f\circ\varphi)}{\partial u_m}(u_0)=0.$$

Of course: being a critical point is preserved by diffeomorphisms, in particular by parametrizations. ∎

We get two consequences from the above calculations. First, when x_0 is a critical point of f, the map $d_{x_0}\,\mathrm{grad}\,f$ is, in fact, an *endomorphism* of $T_{x_0}M$. Second, if $x_0=\varphi(u_0)$, the matrix of $d_{x_0}\,\mathrm{grad}\,f$ with respect to the basis $\frac{\partial\varphi}{\partial u_j}(u_0)$ of $T_{x_0}M$ is $H(u_0)G(u_0)^{-1}$, where

$$H(u_0)=[h_{ij}(u_0)],\quad h_{ij}(u_0)=\frac{\partial^2(f\circ\varphi)}{\partial u_i\partial u_j}(u_0)$$

is the Hessian matrix of the map $f\circ\varphi$ at the point u_0 and $G(u_0)$ is the Gramm matrix of the scalar product on the tangent space at the point u_0. Here we use matrices of linear mappings via rows, that is, $u\mapsto uL$, to write formulas better.

Remark 1.2.2 An important consequence of the previous discussion is that when the basis of the partial derivatives $\frac{\partial\varphi}{\partial u_j}(u_0)$ of $T_{x_0}M$ is orthonormal then the matrix of $d_{x_0}\,\mathrm{grad}\,f$ is the Hessian matrix $H(u_0)$. It is easy to see that such a basis always exists. Therefore, since $H(u_0)$ is symmetric, when x_0 is a critical point all the eigenvalues of $d_{x_0}\,\mathrm{grad}\,f$ are real. ∎

Example 1.2.3 An interesting example is that of spherically symmetric fields, for instance the Coulomb field of a negative charge. A field $F:\mathbb{R}^n\setminus\{0\}\to\mathbb{R}^n$ is *spherically symmetric* if there is a C^∞ function $\lambda:\mathbb{R}^+\to\mathbb{R}$ such that $F(x)=$

$\lambda(\|x\|)x$ for all $x \in \mathbb{R}^n \setminus \{0\}$. Consider a primitive h of the function $t\lambda(t)$, i.e. a function $h : \mathbb{R}^+ \to \mathbb{R}$ such that $h'(t) = t\lambda(t)$ for every $t \in \mathbb{R}^+$. Then the system $x' = F(x)$ is gradient with potential $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ defined by $f(x) = -h(\|x\|)$. Indeed,

$$-\frac{\partial f}{\partial x_i} = h'(\|x\|)\frac{\partial \|x\|}{\partial x_i} = \|x\|\lambda(\|x\|)\frac{2x_i}{2\|x\|} = \lambda(\|x\|)x_i,$$

hence $-\operatorname{grad}_x f = \lambda(\|x\|)x = F(x)$. ∎

Back to the previous discussion, we say that the critical point $x_0 \in M$ is *non-degenerate* if the Hessian matrix $H(u_0)$ of the map $f \circ \varphi$ at the point u_0, with $\varphi(u_0) = x_0$, is non-degenerate. Since the matrix of $d_{x_0} \operatorname{grad} f$ is $H(u_0)G(u_0)^{-1}$ and the Gramm matrix $G(u_0)$ is nonsingular, this is equivalent to saying that $d_{x_0} \operatorname{grad} f$ is an *automorphism* of $T_{x_0}M$. In particular, this notion does not depend on the local parametrization φ. Furthermore, since all eigenvalues of $d_{x_0} \operatorname{grad} f$ are real, the non-degenerate critical points are hyperbolic points of the gradient flow.

Obviously, the Hessian matrix of the map $f \circ \varphi$ depends on the parametrizacion φ. We can, however, use it to define an intrinsic notion in $T_{x_0}M$: let $Q_{x_0}(f)$ be the quadratic form whose matrix with respect to the basis of the partial derivatives is the matrix $H(u_0) = [h_{ij}(u_0)]$ above. This quadratic form is called the *Hessian of f at x_0,* and does not depend on φ:

Proposition 1.2.4 *With the notations above, let $v \in T_{x_0}M$. Then for any curve $\gamma(t)$ in M with $\gamma(0) = x_0$ and $v = \gamma'(0)$ (there are many), we have*

$$Q_{x_0}(f)(v) = (f \circ \gamma)''(0).$$

Proof. Set $x = \varphi(u)$. Let $\overline{v} = (v_1, \dots, v_m)$ be the coordinates of v with respect to the basis of the partial derivatives, which means that $v = d_{u_0}\varphi(\overline{v})$. Then

$$Q_x(f)(v) = \sum_{ij} \frac{\partial^2 (f \circ \varphi)}{\partial u_i \partial u_j}(u) v_i v_j.$$

Let $\gamma(t)$ be a curve as in the statement (for instante $\gamma(t) = \varphi(u_0 + t\overline{v})$). We denote $\overline{\gamma}(t) = \varphi^{-1}(\gamma(t)) = (\gamma_1(t), \dots, \gamma_m(t))$, so that $\overline{\gamma}(0) = u_0$, $\gamma_i'(0) = v_i$. Now we have

$$(f \circ \gamma)'(t) = (f \circ \varphi \circ \overline{\gamma})'(t) = d_{\overline{\gamma}(t)}(f \circ \varphi)(\overline{\gamma}\,'(t)) = \sum\nolimits_i \frac{\partial (f \circ \varphi)}{\partial u_i}(\overline{\gamma}(t))\gamma_i'(t),$$

and then

$$(f\circ\gamma)''(t) = \sum_{ij} \frac{\partial^2 (f\circ\varphi)}{\partial u_j \partial u_i}(\bar\gamma(t))\gamma_j{}'(t)\gamma_i{}'(t) + \sum_i \frac{\partial (f\circ\varphi)}{\partial u_i}(\bar\gamma(t))\gamma_i{}''(t).$$

The last summand here vanishes at $t = 0$ because $\overline{\gamma}(0) = u_0$ is a critical point of $f\circ\varphi$, hence we conclude

$$(f\circ\gamma)''(0) = \sum_{ij} \frac{\partial^2 (f\circ\varphi)}{\partial u_j \partial u_i}(x)\overline{\gamma}_j{}'(0)\overline{\gamma}_i{}'(0) = \sum_{ij} \frac{\partial^2 (f\circ\varphi)}{\partial u_j \partial u_i}(x) v_j v_i,$$

as wanted. ∎

(1.2.5) Hamiltonian Systems. A type of systems related to gradient systems of great importance in classical mechanics are the Hamiltonians. Let U be an open set in $\mathbb{R}^n$ and $H : U\times\mathbb{R}^n \to \mathbb{R}$ a C^∞ map. The *Hamiltonian field of* H is the field

$$F : U\times\mathbb{R}^n \to \mathbb{R}^{2n} : (x, y) \mapsto F(x, y) = \left(\tfrac{\partial H(x,y)}{\partial y}, -\tfrac{\partial H(x,y)}{\partial x}\right).$$

The corresponding system of differential equations

$$\begin{cases} x' = \frac{\partial H(x,y)}{\partial y}, \\ y' = -\frac{\partial H(x,y)}{\partial x}, \end{cases}$$

is called *the Hamiltonian system with Hamiltonian function* H. ∎

One of the simplest examples is the system of differential equations in $\mathbb{R}^2$ describing the motion of a harmonic oscillator

$$\begin{cases} x' = y, \\ y' = -x, \end{cases}$$

where x measures the distance from equilibrium and y is the velocity. Here H is the total energy $H(x, y) = \frac{1}{2}(x^2 + y^2)$.

Clearly, the equilibria of a Hamiltonian system are the critical points of the Hamiltonian function H. The main property of Hamiltonian systems, proved in the next proposition, is that H is constant along every solution. In classical terms, H is *a first integral* of the system. Therefore, the solutions lie in the level sets $H(x) =$ constant.

Proposition 1.2.6 *Let $(x', y') = F(x, y)$ be a Hamiltonian system with Hamiltonian function $H : U \times \mathbb{R}^n \to \mathbb{R}$. Then H is constant along every solution.*

Proof. The proof is a simple verification. Supose $(x(t), y(t))$ is a solution ot the system. Then

$$\frac{d}{dt}H(x(t)) = \sum\nolimits_i \frac{\partial H}{\partial x_i}\, x_i'(t) + \frac{\partial H}{\partial y_i}\, y_i'(t) = \sum\nolimits_i \frac{\partial H}{\partial x_i}\frac{\partial H_i}{\partial y_i} - \frac{\partial H}{\partial y_i}\frac{\partial H_i}{\partial x_i} = 0.$$

It follows that $H(x(t))$ is constant. ∎

A significant result, which we do not prove here, is the *Liouville Theorem,* according to which the flow of a Hamiltonian system is volume preserving. In turn, volume preserving systems have important recurrence properties given by the *Poincaré recurrence theorem.*

1.3 Gradientlike flows

The previous notions can be defined in the even more general framework of flows in metric spaces. With this aim we introduce the following definition.

Definition 1.3.1 Suppose that $\varphi : M \times \mathbb{R} \to M : \varphi(x, t) = xt$, is a flow in the metric space M. We say that φ is *gradientlike* if there exists a map $f : M \to \mathbb{R}$ which is strictly decreasing along non-stationary trajectories of φ, i.e. if $f(xs) < f(xt)$ for all non-stationary points $x \in M$ and all $s > t$. We still say that f is a *strict Lyapunov map for φ.*

Note that f is strictly decreasing along non-stationary trajectories if and only if $f(xt) < f(x)$ for every stationary point x and every $t > 0$ (because $f(xs) = f(xt(s - t))$ for $s > t$).

Remark 1.3.2 The examples that interest us here come mainly from gradient fields, whose solution may well not be a flow, that is, the maximal integral curves may be defined in intervals $I \neq \mathbb{R}$. In that case the theory does not apply, or so it seems. However, any system $x' = f(x)$ can be modified to another $x' = g$ which truly defines a flow and its trajectories are positive reparametrizations of those of the initial system. Consequently the dynamics of both systems are the same. For instance, a function strictly decreasing on non-stationary trajectories for $x' = f$ is properly called Lyapunov for the flow of $x' = g$.

The argument is easy. Let $x' = f$ be defined in a manifold M, which we can suppose a closed subset of an affine space $\mathbb{R}^p$. Consider the function $g = \theta f$, where $\theta = 1/(1 + \|f\|^2)$ (the square preserves the class of f). Since $\theta > 0$ the trajectories of $x' = g$ are indeed positive reparametrizations. But furthermore $\|g\| \leq 1$, and this implies those trajectories are defined for all times. Otherwise, let $\gamma(t)$ be a trajectory of $x' = g$ not defined for $t \geq \delta$. But for $t < \delta$ we have

$$d(\gamma(0), \gamma(t)) \leq \int_0^t \|\gamma'(t)\| dt = \int_0^t \frac{\|f(\gamma(t)\|}{1+\|f(\gamma(t))\|^2} dt \leq \int_0^t dt = t \leq \delta,$$

which is not possible by the escape lemma: as M is closed in $\mathbb{R}^p$, the intersection $M \cap B[\gamma(0), \delta]$ is compact. For negative trajectories the argument is analogous. ∎

The next proposition provides an important property of the *omega limit* (as $t \to \infty$) and the *alpha limit* (as $t \to -\infty$) of the trajectories of gradientlike flows.

Proposition 1.3.3 *Suppose that $\varphi : M \times \mathbb{R} \to M$ is a gradientlike flow in the metric space M. Then for every $x \in M$ the omega limit $\omega(x)$ (resp. the alpha limit $\alpha(x)$), if non-empty, is composed of stationary points of φ.*

Proof. If x is stationary the proposition is obvious, thus we assume that x is non-stationary. Let y be an arbitrary point of $\omega(x)$. Then there exists a sequence $t_n \to \infty$, that we can suppose strictly increasing, such that $xt_n \to y$. If f is a strict Lyapunov map for φ, we have that $f(y) < f(xt_{n+1}) < f(xt_n)$ for every n. Suppose, to get a contradiction, that $yt \neq y$ for some $t > 0$. Select for every n a $k_n > n$ such that $t_n + t < t_{k_n}$. Hence $f(xt_{k_n}) < f(x(t_n + t))$. But $f(xt_{k_n}) \to f(y)$ and $f(x(t_n + t)) \to f(yt)$, which implies that $f(y) \leq f(yt)$ and, thus, f is not strictly decreasing along the trajectory of y, in contradiction with the hypothesis. A similar argument can be used when y is a point of $\alpha(x)$. ∎

A consequence is that if a point x is not stationary then $\omega(x) \cap \alpha(x) = \varnothing$ (if $y \in \omega(x) \cap \alpha(x)$, then $f(y) > f(x) > f(y)$!). In particular, gradientlike flows do not have homoclinic trajectories. It is easy to see that they do not have *recurrent* points either, that is, non-stationary points x such that $x \in \omega(x)$.

From 1.3.3 we get:

Corollary 1.3.4 *Suppose that $\varphi : M \times \mathbb{R} \to M$ is a gradientlike flow in the metric space M whose stationary points are isolated. Let x be a point such that its positive semitrajectory $\gamma^+(x)$ is contained in a compact subset of M. Then $\omega(x)$ consists exactly of one point. An analogous statement holds for $\alpha(x)$.*

Proof. Since $\gamma^+(x)$ is contained in a compact subset of M we have that $\omega(x)$ is non-empty, compact and connected. That it is non-empty comes from compactness of $\overline{x[0,\infty)}$. On the other hand, if x is non-stationary, then $\omega(x) \cap x[0,\infty) = \varnothing$ by Lyapunov monotony. Then $\omega(x) = \overline{x[0,\infty)} \setminus x[0,\infty) = \bigcap_n \overline{x[n,\infty)}$ and any intersection of a decreasing family of compact connected sets is connected. Since all points in $\omega(x)$ are stationary, they are isolated and, hence, $\omega(x)$ has the discrete topology. As a consequence $\omega(x)$ must consist exactly of one point. The proof for $\alpha(x)$ is similar. ∎

We see in the following example how the above results can be used to analyse a local difeomorphism of the euclidean space.

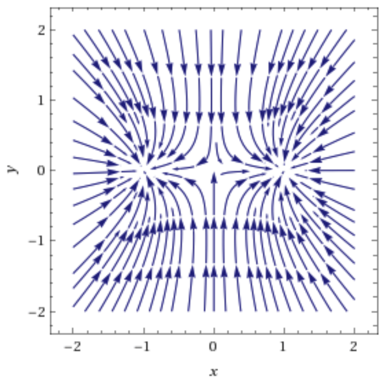
Example 1.3.5 (1) Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a proper local diffeomorphism. Then h is closed an open, hence onto by connectedness. Let us see h is injective, hence a homeomorphism.

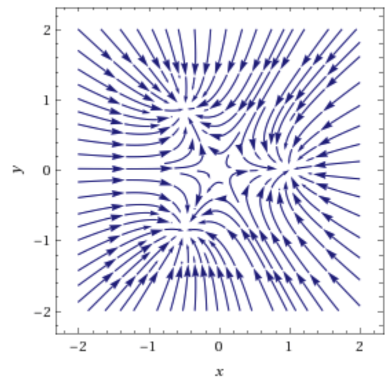
Fix $a \in \mathbb{R}^n$ and consider the function $f = \frac{1}{2}\|h(x) - a\|^2$. Some easy calculations show that the critical (stationary) points of the gradient field $\operatorname{grad} f$ are the local minima of f, that is, the points $x \in \mathbb{R}^n$ with $h(x) = a$, which are isolated. Then consider the flow φ of $-\operatorname{grad} f$ and set

$$W_x = \{y \in \mathbb{R}^n : \omega(y) = x\} \quad \text{for every stationary point } x.$$

Asymptotic stability implies that the W_x's are open. Since h is proper, so is f, and every trajectory $\gamma^+(y) \subset f^{-1}[0, f(y)]$ is bounded. Consequently $\omega(y)$ is a unique stationary point x and $y \in W_x$. Thus we have a covering of $\mathbb{R}^n$ by disjoint open sets, and by connectedness again, there is only one such W_x, or in other words only one stationary point. We are done

(2) Of course, this fact is well known: a proper local homeomorphism is a covering, and $\mathbb{R}^n$ has no proper coverings. Thus, the above argument hides the triviality of the fundamental group. Whence, it is interesting to consider proper local diffeomorphisms of non simply connected spaces. The first example are complex powers in the punctured plane $\mathbb{C} \setminus \{0\}$, say $h(z) = z^k$. One can mimic the construction of f and $-\operatorname{grad} f$ to get a flow and try the argument of (1). Here there are for instance the plots of the flows corresponding to $k = 2$ and 3 for $a = (1, 0)$:





We see the critical points at the roots of the unit: $(\pm 1, 0)$ for $k = 2$ on the left and $(1, 0)$, $\frac{1}{2}(-1, \pm\sqrt{3})$ for $k = 3$ on the right, which are all attractors. But we furthermore have another critical point at the origin, explaining why the W_x's do not cover. Indeed, they miss the points whose ω-limit is the origin: (i) the axis $x = 0$ on the left, and (ii) the three semilines $\{y = 0, x < 0\}$, $\{y = \pm\sqrt{3}, x > 0\}$, on the right. In other words, the missing points escape towards the hole that makes the space not simply connected. ∎

For M noncompact the following notion is useful. We say that a function $f : M \to \mathbb{R}$ is *uniformly unbounded* if for every $L \in \mathbb{R}$ there exists a compact subset $C \subset M$ such that $f(x) > L$ for every $x \in M \setminus C$. In other words, f is *proper from above:* $f^{-1}(-\infty, L]$ is compact for every $L \in \mathbb{R}$.

Proposition 1.3.6 *Suppose that $\varphi : M \times \mathbb{R} \to M$ is a gradientlike flow in the metric space M. Suppose that $f : M \to \mathbb{R}$ is a uniformly unbounded strict Lyapunov map for φ. Then for every $x \in M$ the positive semitrajectory $\gamma^+(x)$ is contained in a compact set and, as a consequence, $\omega(x)$ is non-empty. In particular, if all the stationary points are isolated then $\omega(x)$ consists exactly of one point.*

Proof. Just note that every positive semitrajectory $\gamma^+(x)$ is contained in the compact set $f^{-1}(-\infty, f(x)]$. ∎

The following result states an important attraction property related to the existence of uniformly unbounded Lyapunov maps.

Proposition 1.3.7 *Suppose that $\varphi : M \times \mathbb{R} \to M$ is a gradientlike flow in the locally compact metric space M such that there exists a uniformly unbounded strict*

Lyapunov map for φ. Suppose that the set $\mathcal{K}$ of stationary points is compact. Then there exists a global attractor A of φ. Moreover, this attractor consists of all the points $x \in M$ with non-empty alpha-limit (or, equivalently, A is the unstable manifold $W^u(\mathcal{K})$ of $\mathcal{K}$). In particular, if $\mathcal{K}$ is finite, then A is the union of all trajectories connecting points of $\mathcal{K}$.

Proof. We define the set

$$A = \{y \in M : \text{there are } x_n \to x \in \mathcal{K} \text{ and } t_n \to +\infty \text{ such that } x_n t_n \to y\}$$

We shall prove that A is a global attractor. It is straightforward to see that A is closed and invariant. Since the space is locally compact, balls with small enough radius (depending on the center) are compact, and from this it follows that if $\varepsilon > 0$ is small enough $B[\mathcal{K}, \varepsilon]$ is compact. We show that there exists $T > 0$ such that $A \subset B[\mathcal{K}, \varepsilon][0, T]$.

For every $x \in S(\mathcal{K}, \varepsilon)$ define $\tau_x = \inf\{t > 0 : xt \in B(\mathcal{K}, \varepsilon)\}$. This number is well defined since $\varnothing \neq \omega(x) \subset \mathcal{K}$. Set $T = \sup\{\tau_x : x \in S(\mathcal{K}, \varepsilon)\}$. We claim that $T < +\infty$. Otherwise, there is a sequence $x_n \in S(\mathcal{K}, \varepsilon)$ such that $\tau_{x_n} \to +\infty$. By compactness, we may assume, without loss of generality, that $x_n \to x \in S(\mathcal{K}, \varepsilon)$. Let $\tau > 0$ be such that $x\tau \in B(\mathcal{K}, \varepsilon)$. Then $x_n\tau \in B(\mathcal{K}, \varepsilon)$ and $\tau_{x_n} \leq \tau$ for almost every n, which is a contradiction.

Now suppose that $y \in A \setminus B[\mathcal{K}, \varepsilon]$. Then there are $x_n \in M$ and $t_n \to +\infty$ such that $x_n \to x$ with $x \in \mathcal{K}$ and $x_n t_n \to y$. For almost every n, $x_n \in B(\mathcal{K}, \varepsilon)$ and $x_n t_n \notin B[\mathcal{K}, \varepsilon]$. Now let $s_n = \sup\{s : x_n s \in B(\mathcal{K}, \varepsilon), 0 \leq s \leq t_n\}$, so that $0 < s_n < t_n$ and $x_n s_n \in S(\mathcal{K}, \varepsilon)$. We have $x_n s \notin B(\mathcal{K}, \varepsilon)$ for $s_n \leq s \leq t_n$. Thus for $0 \leq t = s - s_n \leq t_n - s_n$ we have $(x_n s_n)t = x_n s \notin B(\mathcal{K}, \varepsilon)$. This implies $t_n - s_n \leq \tau_{x_n s_n}$, hence $t_n - s_n \leq T$. It follows that $x_n t_n = x_n s_n(t_n - s_n) \in B[\mathcal{K}, \varepsilon][0, T]$, and since this set is closed, $y \in B[\mathcal{K}, \varepsilon][0, T]$. As a consequence, A is a closed subset of $B[\mathcal{K}, \varepsilon][0, T]$, which is a compact set and thus A is compact.

Moreover A is stable. Otherwise, there is an $\varepsilon > 0$, a sequence $x_n \to x \in A$ and times $t_n \geq 0$ such that $x_n t_n \to z \in S(A, \varepsilon)$. If t_n is bounded, then there is a convergent subsequence $t_{n_k} \to t_0$ and, as a consequence, $x_{n_k} t_{n_k} \to xt_0 \in A$. On the other hand, $x_{n_k} t_{n_k} \to z \in S(A, \varepsilon)$, which is a contradiction. If t_n is unbounded we may suppose, without loss of generality, that $t_n \to +\infty$. Since $\omega(x) \subset \mathcal{K}$, then there exists a sequence $t'_n \to +\infty$, such that $t'_n - t_n > n$ and $xt'_n \to w \in \mathcal{K}$. Now, for every k there is n_k such that $d(x_{n_k} t'_k, xt'_k) < 1/k$ and $t_{n_k} - t'_k > k$. By the first condition

$$d(x_{n_k} t'_k, w) \leq d(x_{n_k} t'_k, xt'_k) + d(xt'_k, w) < \tfrac{1}{k} + d(xt'_k, w) \to 0,$$

and $x_{n_k} t'_k \to w$. Since $t_{n_k} - t'_k \to \infty$ and $x_{n_k} t'_k (t_{n_k} - t'_k) = x_{n_k} t_{n_k} \to z$, we conclude that $z \in A$. This is a contradiction with the fact that $z \in S(A, \varepsilon)$. As a consequence,

A is a compact invariant set which is stable and attracts all points in M. Indeed, suppose it does not attract x: $d(xt_n, A) \geq \varepsilon$ for times $t_n \to +\infty$ and $\varepsilon > 0$. Since positive semitrajectories are contained in compact sets, there is a subsequence $xt_{n_k} \to z \in \mathcal{K} \subset A$, hence $d(xt_{n_k}, A) < \varepsilon$ for k large, a contradiction. Hence A is a global attractor.

Now the *moreover*. If there is $y \in \alpha(x) \subset \mathcal{K} \subset A$, we have $y_n = xs_n \to y$ for some $s_n \to -\infty$. Now we use stability. For every $\varepsilon > 0$ there is y_n close enough to y so that $y_n t \in B(A, \varepsilon)$ for all $t > 0$. But for $t = -s_n > 0$, $y_n t = x$ and thus $d(x, A) < \varepsilon$. As A is closed, we conclude $x \in A$. Conversely, consider a point $x \in A$. Since A is invariant, $\gamma^-(x) \subset A$, an since A is compact, $\alpha(x) \neq \varnothing$. ∎

Remark 1.3.8 The last part of the preceding proof could be simplified using the definition of A instead of stability. Indeed, for $y_n \to y \in \mathcal{K}$ and $t_n = -s_n \to \infty$ as there, $y_n t_n = x \to x$. But our claim was that *any* global attractor is that attractor. Furthermore, our little detour gives an argument that works in general to show that *for all gradientlike flows in locally compact metric spaces* (not only for those with unbounded Lyapunov maps) *there is at most one global compact attractor: the unstable manifold of* $\mathcal{K}$. But note that there may not be any attractor at all! *In particular,* if $\mathcal{K}$ is finite, then A is the union of all trajectories connecting points of $\mathcal{K}$. ∎

The next result provides a clear picture of a gradientlike system in the Euclidean space when all its stationary points are isolated.

Proposition 1.3.9 (Structure of isolated stationary points in $\mathbb{R}^n$) *Consider a gradientlike flow $\varphi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ whose stationary points are isolated. Let $x \in \mathbb{R}^n$. If $\gamma^+(x)$ is bounded then $\omega(x)$ consists exactly of one point. Otherwise, $\gamma^+(x) \to \infty$. An analogous statement holds for $\gamma^-(x)$ and $\alpha(x)$.*

Proof. The first statement is a consequence of Corollary 1.3.4. Now, let x be a point such that $\gamma^+(x)$ is not bounded. Assume, to get a contradiction, that $\gamma^+(x)$ does not tend to infinity. Then there exists a closed ball B_{k_0} with center the origin and radius k_0 and a sequence $t_n \to \infty$ such that $xt_n \in B_{k_0}$ for every n. Moreover, since $\gamma^+(x)$ is not bounded, for every $k \in \mathbb{R}$ there exists a sequence $s_n^k \to +\infty$ as $n \to +\infty$ such that $xs_n^k \notin B_k$, the ball with center the origin and radius k. As a consequence, for every $k \geq k_0$ there is a sequence $r_n^k \to +\infty$ for $n \to +\infty$, such that $xr_n^k \in S_k$, the sphere with center the origin and radius k. It follows from this that in every

S_k with $k \geq k_0$ there is a point belonging to $\omega(x)$. An easy argument establishes now that in every S_k with $k \geq k_0$ there is a point which is limit of points of $\omega(x)$ belonging to spheres $S_{k+\varepsilon_n}$ with $\varepsilon_n \to 0$. This point is stationary but nonisolated. This contradiction proves that $\gamma^+(x) \to +\infty$. The proof for $\alpha(x)$ is similar. ∎

The following is one of the classical examples of gradient system. Using the previous results we can get a full picture of the flow.

Example 1.3.10 Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = x^2(x-1)^2 + y^2$$

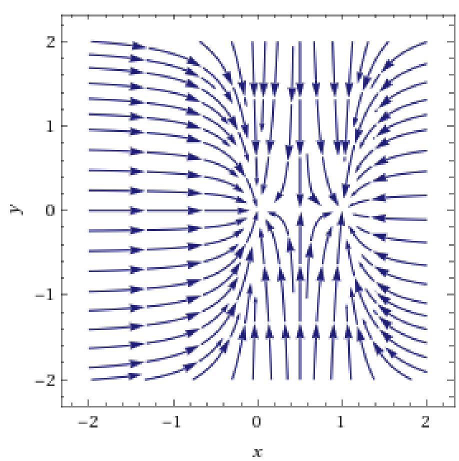
and the corresponding gradient system

$$(x', y') = -\operatorname{grad} f(x, y) = (-2x(x-1)(2x-1), -2y).$$

The function f is uniformly unbounded and, as a consequence, all positive semi-trajectories of the flow are bounded. From the expression above we see that the critical points of f are $(0, 0), (1/2, 0), (1, 0)$ and, thus, $\gamma^+(x, y)$ tends to one of these points for every $(x, y) \in \mathbb{R}^2$. The linearization of $-\operatorname{grad} f(x, y)$ at the critical points provides the following matrices

$$\begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}, \quad \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix},$$

which shows that $(0, 0)$ and $(1, 0)$ are sinks, hence asymptotically stable, and $(1/2, 0)$ is a saddle. On the other hand, it can be readily seen that the x and y axes and the lines $x = 1/2$ and $x = 1$ are invariant. The fact that $y' = -2y$ on the line $x = 1/2$ implies that the stable manifold at $(1/2, 0)$ is the line $x = 1/2$ itself. The unstable manifold at this critical point is the interval $(0, 1)$ on the x-axis.



To complete the description, we have that the negative bounded semitrajectories must also tend to one of the critical points, which means that there are just two, lying in the x-axis with α-limit equal to $(1/2, 0)$. The rest of negative semiorbits tend to ∞ as $t \to -\infty$.

In the preceding discussion we have not cared whether the system defines a flow or not, because our qualitative discussion works the same by Remark 1.3.2. In fact, the solution of the system is

$$\varphi(x,t) = \left(\tfrac{1}{2} + \tfrac{1}{2}\tfrac{2x-1}{\sqrt{(2x-1)^2 - 4x(x-1)e^{2t}}}, -2ye^{-2t}\right),$$

whose domain is the set

$$\Omega \subset \mathbb{R}^2 \times \mathbb{R} : \; (2x-1)^2 - 4x(x-1)e^{2t} > 0.$$

In any case, we do not even need to know this to describe the dynamics! ∎

1.4 Lusternik-Schnirelmann category

An important connection between topology and the theory of gradientlike flows is provided by the *Lusternik-Schnirelmann Theorem,* which gives an estimation from below of the number of stationary points.

First we recall the following notion. Suppose that M is a metric space and X is a closed subset of M. The *Lusternik-Schnirelmann category of X in M*, denoted by $cat_M(X)$, is the smallest k such that X is the union $X = X_1 \cup \cdots \cup X_k$ of k closed subsets $X_1, \ldots, X_k$ which are contractible in M; if no such k exists we write $cat_M(X) = \infty$. If there is no possibility of confusion we write $cat(X) = cat_M(X)$. It is straightforward to see that the Lusternik-Schnirelmann category is *subadditive:* $cat(A \cup B) \leq cat(A) + cat(B)$ for $A, B \subset M$ and that it is *monotonous with respect to deformations:* if $f : A \to M$ is a deformation of a closed set A of M then $cat(A) \leq cat(f(A))$.

For instance, for a n-sphere $cat(S^n) = 2$: it is the union of two contractible hemispheres but not contractible in itself. On the other hand, it is known that the unit sphere S^∞ in an infinite dimensional normed vector space M is a retract of M and from this it follows that S^∞ is contractible in itself and $cat(S^\infty) = 1$. In general, the calculation of the category is not an easy matter, but some important examples are well-know. For instance, for the n-torus $cat(T_n) = n + 1$. Also, if M is a compact manifold of dimension m, $cat(M) \leq m + 1$.

The estimation of stationary points is as follows:

Theorem 1.4.1 (Lusternik-Schnirelmann) *Let $\varphi : M \times \mathbb{R} \to M$ be a gradientlike flow on a locally (or, more generally, semilocally) contractible and compact metric space M. Then the number of stationary points of φ is bounded from below by $cat_M(M)$.*

Proof. Suppose that the number of stationary points is finite (otherwise there is nothing to prove). Label the stationary points $x_1, \ldots, x_r$ in such a way that $f(x_i) \leq f(x_{i+1})$ (where f is the Lyapunov map). Then, by Corollary 1.3.4, for every nonstationary point $x \in M$ the omega-limit $\omega(x)$ and the alpha-limit $\alpha(x)$ are stationary points, x_i and x_j respectively, with $f(x_i) < f(x) < f(x_j)$, hence $i < j$. Note that this implies $\omega(x) \neq \alpha(x)$, that is, $\omega = \alpha$ exactly for stationary points. On the other hand, $f(x) \leq f(x_r)$ for all x, and $f(x) < f(x_r)$ if x is not stationary (because then $f(x) < f(\alpha(x)) = f(x_k) \leq f(x_r)$).

For $0 \leq k \leq r$ we define

$$M_k = \{x \in M : \alpha(x) = x_j, j \leq k\},$$
$$M_k^* = \{x \in M : \omega(x) = x_i, i > k\}.$$

Notice that the opposite flow $\varphi(x, -t)$ has the opposite Lyapunov map $-f$ and the same critical points ordered backwards. Consequently M_k and M_ℓ^* change places with $\ell = r - k$ and everything we prove for the M_k's follows for the M_ℓ^*'s.

Let us list some immediate facts (one sees the duality just mentioned):

1. *$x_\ell \in M_k$ for $\ell \leq k$, $x_\ell \in M_k^*$ for $\ell > k$, $M_k \cap M_k^* = \varnothing$, $M_k \cap M_{k-1}^* = \{x_k\}$.*

Consequently, if $x \notin M_k$ (resp. M_ℓ^*), then $\alpha(x) \in M_k^*$ (resp. $\omega(x) \in M_\ell$).

2. *$M_{k-1} \subset M_k$, $M_k^* \subset M_{k-1}^*$, $M_0 = M_r^* = \varnothing$, $M_1 = \{x_1\}, M_{r-1}^* = \{x_r\}$, $M_r = M_0^* = M$.*
3. *M_k and M_ℓ^* are invariant, and we can restrict the flow to them.*
4. *M_k attracts all points in $M \setminus M_k^*$ and M_ℓ^* repells all points in $M \setminus M_\ell$.*

More delicate is that

5. *M_k and M_ℓ^* are closed, hence compact.*

Indeed, first we consider M_{r-1}. Suppose there is a sequence $x_n \to x$ with $x_n \in M_{r-1}$. Then $\alpha(x_n) = x_j$ with $j < r$ and we can assume the same j for all x_n. Thus we find a decreasing sequence $t_n \to -\infty$ with $x_n t_n \to x_j$. By way of contradiction, let $x \notin M_{r-1}$, that is $\alpha(x) = x_r$. Thus for each n, $x_\nu t_n \to x t_n$, hence for each n there is $\nu_n \geq n$ with $d(x_{\nu_n} t_n, x t_n) < 1/n$ and since $\alpha(x) = x_r$, we get $x_{\nu_n} t_n \to x_r$. Next, choose $\varepsilon > 0$ such that $B[x_j, \varepsilon]$ contains no critical point but x_j. For almost all n we have $x_{\nu_n} t_{\nu_n} \in B(x_j, \varepsilon)$, $x_{\nu_n} t_n \notin B[x_j, \varepsilon]$, and there is $x_{\nu_n} s_n \in S(x_j, \varepsilon)$ with

$t_{\nu_n} < s_n < t_n$; by compactness of $S(x_j, \varepsilon)$ we can suppose $x_{\nu_n} s_n \to y \in S(x_j, \varepsilon)$. We have $f(x_j) > f(x_{\nu_n} s_n) > f(x_{\nu_n} t_n)$, so in the limit $f(x_j) \geq f(y) \geq f(x_r)$. But this implies $f(x_j) = f(y) = f(x_r)$ and y is a critical point. Contradiction, as there is none in $S(x_j, \varepsilon)$.

For $k < r$ the same argument shows that M_k is closed in M_{k+1}, where the flow can be restricted and we are left with the critical points $x_1, \ldots, x_{k+1}$. For M_ℓ^* we use duality.

Finally we prove stability:

6. *M_k (resp. M_ℓ^*) has a neighborhood basis of positively (resp. negatively) invariant neighborhoods.*

Since $M \setminus M_k^*$ is open, it is a neigborhood of M_k, and we have to prove that for any closed neighborhood $U \subset M \setminus M_k^*$ of M_k the set

$$W = \{x \in U : \gamma^+(x) \subset U\}$$

is a neighborhood of every point $x \in M_k$. But otherwise there would be a sequence $x_n \to x$ and $t_n > 0$ such that $x_n t_n \notin U$. Since x is interior to U we get $x_n t_n \in \partial U$ for a new t_n, and ∂U being compact, we can take $t_n = \min\{t \geq 0 : x_n t \in \partial U\}$. If t_n has a convergent subsequence, we can simply suppose $t_n \to t_0$, and then $x_n t_n \to x t_0 \in M_k \cap \partial U$, impossible. Hence let $t_n \to \infty$. As ∂U is compact, we can suppose $x_n t_n \to y \in \partial U$. Then for every $s < 0$ we have $x_n(t_n + s) \to ys$ and $x_n(t_n + s) \in U$ (recall the choice of t_n). Since U is closed, $ys \in U$ for $s < 0$, and $\alpha(y) \in U$. But since $y \notin M_k$, then $x_j = \alpha(y) \in M_k^*$. Since $y \notin M_k$, $j > k$ and thus $x_j \in M_k^* \cap U$, contradiction. The assertiion for M_ℓ^* follows by duality.

Consequently, we can say that M_k is an attractor.and M_k^* is a repeller. In particular each pair (M_{k-1}, x_k) is an attractor-repeller decomposition of the flow restricted to M_k (recall that $M_k \cap M_{k-1}^* = \{x_k\}$).

Now we come to the key construction to estimate $cat(M)$.

There is a sequence of positively invariant closed neighborhoods $U_k \subset M \setminus M_k^*$ of M_k and closed contractible neighborhoods $V_k \subset U_k$ of x_k such that $U_{k-1} \cup V_k$ is a closed neighborhood of M_k:

$$(*) \qquad M_k \subset U_{k-1} \cup V_k \subset U_k \subset M \setminus M_k^*.$$

Here $U_0 = \varnothing$.

We start with $U_r = M$ and by descent suppose U_k given and are to construct U_{k-1} and V_k. Note that U_k is a neighborhhod of M_k, hence of M_{k-1}. Since $x_k \in M_k \subset U_k$ and M is semilocally simply conected, pick any closed neighboorhood $V_k \subset U_k$ of

x_k which is contractible in M, and choose $B = B(x_k, \varepsilon) \subset V_k$. By stability, M_{k-1} has a positively invariant closed neighborhood $U \subset (M \setminus M_k^*) \cap U_k$; let $W \subset W_k$ be the interiors of $U \subset U_k$ respectively. If $M_k \subset W \cup B$, then $U_{k-1} = U$ does the job, hence we suppose $M_k \not\subset W \cup B$. Then the Lyapunov function f has a maximum on the nonempty compact set $M_k \setminus W \cup B$, which is $< f(x_k)$, hence $< a < f(x_k)$ for suitable a. Indeed, if $x \in M_k \setminus W \subset M_k \setminus M_{k-1}$ and $x \neq x_k$, then $\alpha(x) = x_k$ and $f(x) < f(x_k)$. In this situation $U_{k-1} = U \cup \big(f^{-1}(-\infty, a] \cap U_k\big)$ is a positively invariant closed neighborhood of M_{k-1}. Besides,

$$M_k \subset W \cup \big(f^{-1}(-\infty, a) \cap W_k\big) \cup B,$$

hence $U_{k-1} \cup V_k \subset U_k$ a closed neighborhood of M_k. This completes the construction.

As announced, we are now ready to estimate $cat(M)$ using $(*)$ above. Firstly, the flow gives a deformation of U_k into $E = U_{k-1} \cup V_k$, because U_k is positively invariant and there is $a > 0$ such that $\varphi^a(U_k) \subset E$. For, otherwise, we find sequences $x_n \in U_k$, and $t_n \to +\infty$ such that $x_n t_n \notin E$. Then $x_n \notin M_k^*$ and $\omega(x_n) \in M_k \subset E$, hence $x_n t \in W$ for t large, and $x_n t_n \in \partial E$ for a bigger t_n. By compactness, taking subsequences we have $x_n t_n \to y \in \partial E$ and $x_n \to x \in U_k$. Thus $\omega(x) \in M_k$ and a trick used before gives $x_{\nu_n} t_n \to \omega(x)$ with $t_{\nu_n} > t_n$. Choose now, by stability, a positively invariant closed neighborhood F of M_k (hence of $\omega(x)$) contained in the interior of E. We have, for large n,

$$x_{\nu_n} t_{\nu_n} = x_{\nu_n} t_n (t_{\nu_n} - t_n) \in F(t_{\nu_n} - t_n) \subset F.$$

Then $y \in \overline{F} = F$, which is impossible because F does not meet ∂E. Note that this works the same for $k = 1$, so that U_1 can be deformed into $U_0 \cup V_1 = V_1$.

Once we are given this deformation, by the basic properties of the category,

$$cat(U_k) \leq cat(U_{k-1}) + 1,$$

which includes $cat(U_1) = 1$. Hence:

$$\begin{aligned} cat(M) = cat(U_r) &\leq cat(U_{r-1}) + 1 \leq (cat(U_{r-2}) + 1) + 1 \\ &= cat(U_{r-2}) + 2 \leq \cdots \leq cat(U_1) + r - 1 = r. \end{aligned}$$

This completes the proof. ∎

Before the theorem we listed the values of some categories, and from the previous theorem we deduce now that every gradientlike flow in S^n has at least two stationary points, in S^∞ at least one, in T^n at least $n + 1$.

Remark 1.4.2 It is worth noting that in the proof of Theorem 1.4.1 we have constructed a filtration of M by attractors

$$\varnothing = M_0 \subset M_1 \subset \cdots \subset M_k \subset \cdots \subset M_r = M$$

and another filtration by repellers

$$\varnothing = M_r^* \subset M_{r-1}^* \subset \cdots \subset M_{r-k}^* \subset \cdots \subset M_0^* = M$$

such that for every k the set $M \setminus (M_k \cup M_k^*)$ is composed of the *connecting orbits,* i.e. the orbits $\gamma(x)$ such that $\omega(x) \subset M_k$ and $\alpha(x) \subset M_k^*$. In other words, the pair (M_k, M_k^*) defines an *attractor-repeller decomposition of* M. ∎

1.5 Morse functions

In this section, M stands for a compact m-manifold. When needed, we can suppose M embedded in an affine space, say $M \subset \mathbb{R}^p$.

We introduce the following basic definition.

Definition 1.5.1 A smooth function $f : M \to \mathbb{R}$ is called a *Morse function* when all its critical points are non-degenerate.

As we know, non-degenerate critical points are isolated, so Morse functions have only a finite number of critical points. At each one of them the Hessian matrix of a localization $f \circ \varphi$ is symmetric and its eigenvalues are all real $\neq 0$. The number of negative eigenvalues is called the *Morse index* of f at the critical point. This notion does not depend on the particular parametrization φ. By 1.2.2 and the comments that follow, the Morse index of f at a critical point x_0 agrees with the number of negative eigenvalues of $d_{x_0} \operatorname{grad} f$, and x_0 is a hyperbolic point of the gradient flow. This is of course relevant for our later discussion of gradient flows of Morse functions.

It is not difficult to see that Morse functions abound:

Proposition 1.5.2 *Let $f : M \to \mathbb{R}$ be a smooth function. Recall we can assume $M \subset \mathbb{R}^p$, and for every $a \in \mathbb{R}^p$ define $f_a : M \to \mathbb{R}$ by $f_a(x) = f(x) + \langle a, x \rangle$. Let $\mathcal{M} \subset \mathbb{R}^p$ be the set of all $a \in \mathbb{R}^p$ such that f_a is a Morse function. Then $\mathcal{M}$ is a residual, hence dense, subset of $\mathbb{R}^p$.*

Proof. Here residual means that $\mathcal{M}$ contains a countable intersection of dense open sets, which implies dense by the Baire Theorem. Since being Morse is a local matter, and M has a countable basis, we are reduced to consider a parametrization $\varphi : U \to M$, where U is an open set of $\mathbb{R}^m$. Thus let us see that the $a \in \mathbb{R}^p$ such that

$$f_a \circ \varphi(u) = f \circ \varphi(u) + \langle a, \varphi(u) \rangle$$

is a Morse function (on U) form a residual set; we write $g_a = f_a \circ \varphi$ and $g = f \circ \varphi$. We are to define a smooth function $h : N \to \mathbb{R}^p$ such that g_a is Morse if and only if a is a regular value of h, which by the Sard-Brown Theorem implies what we want.

To that end define $G : \mathbb{R}^p \times U \to \mathbb{R}^m : (a, u) \mapsto \operatorname{grad}_u g_a$. Some little calculation gives:

$$G_i = \frac{\partial g}{\partial u_i} + \sum_k a_k \frac{\partial \varphi_k}{\partial u_i}, \quad \frac{\partial G_i}{\partial a_k} = \frac{\partial \varphi_k}{\partial u_i}, \quad \frac{\partial G_i}{\partial u_j} = \frac{\partial^2 g_a}{\partial u_j \partial u_i},$$

and we have the jacobian matrix

$$JG = \big(J\varphi^t \,|\, \operatorname{Hess}(g_a) \big).$$

(where the upperscript t means transpose). Clearly $N = G^{-1}(0)$ collects all pairs (a, u) such that u is a critical point of g_a. This N is a manifold by the implicit function theorem, because $\operatorname{rk}(JG) \geq \operatorname{rk}(J\varphi) = m$ because φ is a parametrization. Finally the most natural h we have at hand is the linear projection $N \to \mathbb{R}^p : (a, u) \mapsto a$.

After this preparation we only have to check what is a regular value a of h. So let a be one, that is all $(a, u) \in N$ are regular points, and recall these u's are exactly the critical points of g_a. Back to regularity, the derivative $d_{(a,u)}h : T_{(a,u)}N \to \mathbb{R}^p$ must be surjective. Now, N has codimension m in $\mathbb{R}^p \times U$, hence dimension p, and that derivative is surjective if and only if it is injective. On the other hand h is the restriction of the linear projection, hence $d_{(a,u)}h$ is the linear projection too. Thus injectivity means that if $(0, v)$ is tangent to N at (a, u) then $v = 0$. But $T_{(a,u)}N$ is the kernel of $d_{(a,u)}G$, hence $(0, v)$ is tangent if and only if

$$0 = d_{(a,u)}G(0, v) = \operatorname{Hess}_u(g_a)(v).$$

Thus (a, u) is a regular point if and only if that Hessian is non-degenerate. We conclude that a is a regular value of h if and only if all critical points u of g_a are non-degenerate. ∎

From this it follows easily a more qualitative density result:

Corollary 1.5.3 *Any continuous function $f : M \to \mathbb{R}$ can be approximated by Morse functions.*

Proof. Again let $M \subset \mathbb{R}^p$. By the previous proposition, there is $a \in \mathbb{R}^p$ with $\|a\|$ arbitrarily small such that f_a is a Morse function. Since M is compact, it is bounded, say $M \subset B[0, R]$, and we have

$$|f(x) - f_a(x)| = |\langle a, x\rangle| \leq \|a\|\|x\| \leq \|a\|R.$$

Hence, for $\|a\|$ small enough f_a is arbitrarily close to f. ∎

By the previous considerations, the critical points of a Morse function $f : M \to \mathbb{R}$ are the stationary points of the gradient flow of $x' = -\operatorname{grad}_x f$, which hence are finitely many. In particular, each *critical level* of f, that is, the set of critical points with a given critical value, is finite. In fact, we can even find Morse functions with a single critical point at each critical level. Namely:

Corollary 1.5.4 *Any continuous function $f : M \to \mathbb{R}$ can be approximated by Morse functions g whose critical levels are singletons.*

Proof. By the previous corollary we can suppose f is a Morse function. Then it has finitely many critical points, say $a_1, \ldots, a_r$. Choose diffeomorphisms $\varphi_i : \mathbb{R}^m \to U_i$ onto disjoint open neighborhoods U_i of the a_i's, with $a_i = \varphi_i(0)$. Fix any bump function $\theta : \mathbb{R}^n \to [0, 1]$ which is $\equiv 1$ on $\|x\| \leq 1$ and $\equiv 0$ on $\|x\| \geq 2$, and define $\theta_i = \theta \circ \varphi_i^{-1} : M \to \mathbb{R}$ ($\theta_i \equiv 0$ off U_i). We claim that for generic $t = (t_1, \ldots, t_r)$ near $0 \in \mathbb{R}^r$ the function

$$g = f + \sum_i t_i\theta_i$$

solves our problem.

Indeed, first note that $\|f - g\| \leq \sum_i |t_i|$ is arbitrarily small for t is close to 0, hence g approximates f. Now let us look at the critical points of g.

Firstly, $g \equiv f$ off the sets $\varphi_i(\|x\| \geq 2)$, hence g has no critical point there. Secondly, $g \equiv f + t_i$ on each $\varphi_i(\|x\| \leq 1)$, hence a_i is the unique critical point of g in that set. Thus we are left with the sets $\varphi_i(1 < \|x\| < 2)$. There we write $g_i = g \circ \varphi_i, f_i = f \circ \varphi_i$, so that $g_i = f_i + t_i\theta$. Now, for t_i small enough

$$\|\operatorname{grad}_x g_i\| = \|\operatorname{grad}_x f_i + t_i \operatorname{grad}_x \theta\| \geq \|\operatorname{grad}_x f_i\| - |t_i|\|\operatorname{grad}_x \theta\| > 0,$$

because on $1 \leq \|x\| \leq 2, \operatorname{grad} f_i$ does not vanish and $\operatorname{grad} \theta$ is bounded. Thus for t small enough, the critical points of g are the a_i's and

$$g(a_i) = f(a_i) + \sum_j t_j\varphi_j(a_i) = f(a_i) + t_i.$$

Consequently, if $g(a_i) = g(a_j)$, we have $t_i - t_j = f_j(a_j) - f_i(a_i)$. If $f_j(a_j) \neq f_i(a_i)$ this fails for t close enough to 0, and if $f_j(a_j) = f_i(a_i)$ we impose $t_i \neq t_j$. All in all we get what we want for t small enough off some hyperplanes $t_i = t_j$. ∎

The local structure of the critical points of a Morse function is perfectly understood by the following famous lemma.

Lemma 1.5.5 (Morse Lemma) *Let $f : M \to \mathbb{R}$ be a Morse function and let $x_0 \in M$ be a critical point of f of index d. There is a parametrization $\varphi : U \to M$ such that*

$$f \circ \varphi(u) = f(x_0) + u_1^2 + \cdots + u_{m-d}^2 - u_{m-d+1}^2 - \cdots - u_m^2.$$

Proof. Again, this is a local question, hence we can suppose M an open ball in $\mathbb{R}^m$ centered at $x_0 = 0$. We claim that $f(x) = f(0) + xQ(x)x^t$ for a smooth symmetric matrix $Q(x)$.

This is just a cheap Taylor expansion. First we have

$$f(x) - f(0) = \int_0^1 \frac{\partial}{\partial s} f(sx)ds = \sum_i x_i \int_0^1 \frac{\partial f}{\partial x_i}(sx)ds,$$

which gives $f(x) = f(0) + \sum_i x_i g_i(x)$ for the obvious g_i's. Next we do the same for each g_i to get

$$g_i(x) = g_i(0) + \sum_j x_j g_{ij}(x),$$

and insert it in the previous formula:

$$f(x) = f(0) + \sum_i x_i g_i(0) + \sum_{ij} x_i x_j g_{ij}(x).$$

Now, since $x_0 = 0$ is a critical point of f the linear part vanishes. As for the quadratic remainder, we symmetrize via the matrix $Q(x) = [\frac{1}{2}(g_{ij}(x) + g_{ji}(x)]$ to obtain the formula claimed.

Next we seek a local diffeomorfism $x = h(y)$ such that $h(y)Q(h(y))h(y)^t = yQ(0)y^t$.

Look at $\mathbb{R}^{m\times m}$ as the space of $m \times m$ matrices. Consider the linear subspace Σ of symmetric matrices and the smooth mapping

$$\zeta : \mathbb{R}^{m\times m} \to \Sigma : A \mapsto AQ(0)A^t.$$

An easy computation gives the derivative at the identity matrix:

$$d_I\zeta(A) = Q(0)A^t + AQ(0),$$

which is surjective: any $B \in \Sigma$ is $d_I\zeta(A)$ for $A = \frac{1}{2}BQ(0)^{-1}$ (recall $Q(0)$ is symmetric as well as B). By the implicit function theorem, ζ has a smooth left inverse $\xi : W \to \mathbb{R}^m \times \mathbb{R}^m$ defined on an open neighborhood $W \subset \Sigma$ of $\zeta(I) = Q(0)$, that is:

$$\begin{cases} T = \zeta \circ \xi(T) = \xi(T)Q(0)\xi(T)^t & \text{for } T \in W, \\ \xi(Q(0)) = I. \end{cases}$$

Back to our situation we have $T = Q(x) \in W$ for x close to 0, hence

$$xQ(x)x^t = x\xi(Q(x))Q(0)\xi^t(Q(x))x^t = yQ(0)y^t,$$

where $y = x\xi(Q(x))$. This is a local diffeomorphism at $y = 0$:

$$d_0y = I\xi(Q(0)) + 0 \cdot d_0(\xi \circ Q) = \xi(Q(0)) = I,$$

and its local inverse is the local diffeomorphism $x = h(y)$ we sought.

Finally, by derivation of $f(x) = f(0) + xQ(x)x^t$ we find $Q(0) = \frac{1}{2}\operatorname{Hess}_0(f)$, hence $Q(0)$ and $\operatorname{Hess}_0(f)$ have the same index d. Consequently, a linear change of coordinates transforms $Q(0)$ to diagonal form $D = \langle 1, \ldots, 1, -1, \ldots, -1 \rangle$ with d signs -1. In the new coordinates $(u_1, \ldots, u_m)$ we have

$$xQ(0)x^t = uDu^t = u_1^2 + \cdots + u_{m-d}^2 - u_{m-d+1}^2 - \cdots - u_m^2,$$

and we are done. ∎

1.6 Morse fields

We keep the setting of the preceding section: M is a compact m-manifold, embedded in some $\mathbb{R}^p$.

Let $f : M \to \mathbb{R}$ be a Morse function, $x_0 \in M$ a critical point of f and φ a parametrization at x_0 as provided by the Morse Lemma:

$$f \circ \varphi(u) = f(x_0) + u_1^2 + \cdots + u_{m-d}^2 - u_{m-d+1}^2 - \cdots - u_m^2.$$

As remarked in Section 1.2, the expression of the field $\operatorname{grad}_{\varphi(u)} f$ in these local coordinates need not be

$$\operatorname{grad}_u(f \circ \varphi) = 2(u_1, \ldots, u_{m-d}, -u_{-m-d+1}, \ldots, -u_m).$$

Thus, in order to make full use of the Morse Lemma for the study of gradients we introduce the following notion.

Definition 1.6.1 Let $f : M \to \mathbb{R}$ be a Morse function and let $F : M \to \mathbb{R}^p$ be a tangent vector field on M. We say that F is a *Morse field* for f if:

1. For every point $x \in M$ we have $d_x f(F(x)) \leq 0$, with equality if and only if x is a critical point of f, and
2. For every critical point $x \in M$ of f there is a parametrization $\varphi : U \to M$ with $\varphi(0) = x$ such that $f \circ \varphi(u) = f(x) + u_1^2 + \cdots + u_{m-d}^2 - u_{m-d+1}^2 - \cdots - u_m^2$ and $F \circ \varphi(u) = -2(u_1, \ldots, u_{m-d}, -u_{-m-d+1}, \ldots, -u_m)$.

It is clear from the definition that the stationary points of the flow are the critical points of f, and the flow of the equation $x' = F(x)$ is gradientlike with strict Lyapunov map f.

Proposition 1.6.2 *Every Morse function $f : M \to \mathbb{R}$ has a Morse field F.*

Proof. We remark compactness is not used in the proof. For each critical point x_0 of f the Morse Lemma gives a parametrization $\varphi : U \to M$ with $\varphi(0) = x_0$ and $f \circ \varphi(u) = f(x_0) + u_1^2 + \cdots + u_{m-d}^2 - u_{m-d+1}^2 - \cdots - u_m^2$. We can reduce the U's for the images $\varphi(U)$ to be disjoint. Next we cover all non-critical points with parametrizations $\varphi : U \to M$ whose images $\varphi(U)$ contain no critical point. These U of two types cover M and we consider a partition of unity $\{\theta_U\}$ subordinated to the cover $\{U\}$.

Next, por each U we transport the field $-\operatorname{grad}(f \circ \varphi)$ from U to $\varphi(U)$ by the formula

$$F_U(x) = d_u\varphi(-\operatorname{grad}_u(f \circ \varphi)) \text{ for } x = \varphi(u) \in \varphi(U).$$

As is customary, we can extend $\theta_U F_U$ by zero off U and then the finite sum $F = \sum_U \theta_U F_U$ defines a tangent field on M.

Let us check that F is a Morse field for f. Condition 2 of the definition is guaranteed by construction. For Condition 1 we compute $d_x f(F(x))$ at a given $x \in M$. The sum $F(x) = \sum \theta_U(x) F_U(x)$ is finite in a neighborhood of $x = \varphi(u)$ and we compute as follows:

$$\begin{aligned} d_x f(F(x)) &= \sum \theta_U(x) d_x f(F_U(x)) = \sum \theta_U(x) d_x f(d_u\varphi(-\operatorname{grad}_u(f \circ \varphi))) \\ &= \sum \theta_U(x) d_u(f \circ \varphi)(-\operatorname{grad}_u(f \circ \varphi))) \\ &= -\sum \theta_U(x) \| \operatorname{grad}_u(f \circ \varphi)\|^2 \leq 0 \end{aligned}$$

since all θ_U are ≥ 0. Now suppose the sum vanishes. Then all summands do, and since not all $\theta_U(x)$ can vanish (their sum is 1!)) we have some $\operatorname{grad}_u(f \circ \varphi) = 0$ and

$x = \varphi(u)$ is a critical point. ∎

Next we recall that, in the usual setting of a flow with Lyapunov function $f : M \to \mathbb{R}$, the *unstable manifold* $W^u(x_0)$ *of a stationary point* x_0 consists of the points x whose α-limit is the point x_0, and the *stable manifold* $W^s(x_0)$ consists of the points x whose ω-limit is the point x_0. Note here that unstable becomes stable and conversely if we substitute $-f$ for f and reparametrize the flow by $-t$.

For Morse fields we have the following result:

Proposition 1.6.3 (Unstable Manifold Theorem for Morse fields) *Consider a Morse function $f : M \to \mathbb{R}$ and a Morse field F for f. Suppose that x_0 is a critical point of f. Then the unstable manifold $W^u(x_0)$ for the flow of $x' = F(x)$ is diffeomorphic to $\mathbb{R}^d$, where d is the index of x_0.*

Proof. By definition, there is a parametrization $\varphi : U \to M$ with $\varphi(0) = x_0$ such that

$$\begin{cases} f \circ \varphi(u) = f(x_0) + u_1^2 + \cdots + u_{m-d}^2 - u_{m-d+1}^2 - \cdots - u_m^2, \\ F \circ \varphi(u) = d_u\varphi(-\operatorname{grad}_u(f \circ \varphi)), \quad u \in U. \end{cases}$$

Write $v = (u_1, \ldots, u_{m-d}), w = (u_{m-d+1}, \ldots, u_m)$. For later discussions of the involved flows we can suppose U is an open ball in $\mathbb{R}^m$ and $f(x_0) = 0$. Consequently, in $\varphi(U)$ the flow of $x' = F(x)$ is the image by φ of the flow of

$$-\operatorname{grad}_u(f \circ \varphi) = 2(-u_1, \ldots, -u_{m-d}, u_{m-d+1}, \ldots, u_m),$$

that is $\varphi(u)t = \varphi(ut)$ for $u, ut \in U$. Now we choose $\varepsilon > 0$ small enough for U to contain the closed m-ball $B[0, \varepsilon]$.

The flow ut is well known, and actually defined on the whole $\mathbb{R}^m$:

$$ut = (ve^{-2t}, we^{2t}) \quad \text{for } u = (v, w) \in \mathbb{R}^{m-d} \times \mathbb{R}^d.$$

Its unstable manifold is $\{v = 0\} = \{0\} \times \mathbb{R}^d \equiv \mathbb{R}^d$ and the restriction

$$\varphi| : \mathbb{R}^d \cap B[0, \varepsilon] \to W^u(x_0)$$

is well defined. It is smooth and injective, and we claim it extends to a diffeomorphism $\mathbb{R}^d \to W^u(x_0)$. For this we find a different description of φ off the origin.

For each $u = (0, w) \neq 0$ there is a unique s such that $\|us\| = \varepsilon$. We compute it explicitely: from

$$\varepsilon = \|us\| = \|(0, we^{2s})\| = \|w\|e^{2s} = \|u\|e^{2s},$$

it follows $s = \frac{1}{2}\log(\varepsilon/\|u\|)$. In case $0 < \|u\| \leq \varepsilon$ we get

$$\varphi(u) = \varphi(us(-s)) = \varphi(us)(-s).$$

Now define

$$\psi : \mathbb{R}^d \setminus \{0\} \to W^u(x_0) \setminus \{x_0\} : w \equiv (0, w) = u \mapsto x = \varphi(us)(-s),$$

using the flow of $x' = F(x)$. Clearly this is well defined and smooth, and coincides with φ on $\mathbb{R}^d \cap B[0,\varepsilon] \setminus \{0\}$. Thus, we set $\psi(0) = \varphi(0) = x_0$.

Now, ψ is bijective. We must see that the trajectory of any given $x \in W^u(x_0)$, $x \neq x_0$, meets $S' = \varphi(\mathbb{R}^d \cap S(0,\varepsilon))$ at a unique point. Note that $f \equiv -\varepsilon^2$ on S', from which uniqueness follows readily because f is a Lyapunov map for the flow, hence injective on the trajectory of x. Let us check that indeed the trajectory xt hits S'.

Since $\lim_{t\to-\infty} xt = x_0$ there is t_0 such that $xt \in \varphi(B(0,\varepsilon))$ for all $t \leq t_0$. In particular $xt_0 = \varphi(u)$ for some $u \in B(0,\varepsilon)$ and since $\varphi^{-1}(xt)$ is defined for $t \leq t_0$, it is the negative trajectory of u: $\varphi^{-1}(xt) = u(t - t_0)$. Hence

$$\lim_{t\to-\infty} u(t-t_0) = \lim_{t\to-\infty} \varphi^{-1}(xt) = \varphi^{-1}\big(\lim_{t\to-\infty} xt\big) = \varphi^{-1}(x_0) = 0$$

and u belongs to the unstable manifold $\mathbb{R}^d$. Thus $xt_0 = \varphi(u) \in \varphi(\mathbb{R}^d \cap B(0,\varepsilon))$. Then, for the s found above such that $\varepsilon = \|us\|$, we have

$$x(t_0 + s) = xt_0 s = \varphi(u)s = \varphi(us) \in \varphi(\mathbb{R}^d \cap S[0,\varepsilon]) = S'.$$

It remains to show that $\psi^{-1} : W^u(x_0) \to \mathbb{R}^d$ is differentiable. For $x \in W^u(x_0)$, $x \neq x_0$, consider the implicit equation $f(xt) = -\varepsilon^2$. Since F is a Morse field for f we have

$$\tfrac{\partial}{\partial t} f(xt) = d_{xt}f(F(xt)) < 0,$$

and by the implicit functions theorem, there is a smooth solution t_x such that $f(xt_x) = -\varepsilon^2$. Then

$$\psi^{-1}(x) = \varphi^{-1}(xt_x)(-t_x),$$

is differentiable for $x \neq x_0$.

Finally, we show ψ^{-1} is differentiable at the critical point x_0. Since ψ and φ coincide on $\mathbb{R}^d \cap U$, ψ^{-1} and φ^{-1} coincide on $V = \psi(\mathbb{R}^d \cap B(0,\varepsilon))$. As the parametrization φ is a diffeomorphism, it is enough to see that V is a neighborhood of x_0 in $W^u(x_0)$, or equivalently that x_0 is not adherent to $W^u(x_0)\setminus V = \psi(\mathbb{R}^d \setminus B(0,\varepsilon))$.

For suppose it is: there is a sequence $u_n \in \mathbb{R}^d \setminus B(0,\varepsilon)$ with $x_n = \psi(u_n) \to x_0$. If there is a limit $u_{n_k} \to u_0$, then $\psi(u_0) = x_0$ and $u_0 = 0$, so that $u_{n_k} \in B(0,\varepsilon)$ for

k large, which is not the case. Consequently, we can suppose $\|u_n\| \to \infty$, or merely $\|u_n\| > \varepsilon$. By definition $x_n = \varphi(u_n s_n)(-s_n)$ with $\|u_n s_n\| = \varepsilon$. We know also that $s_n = \frac{1}{2}\log(\varepsilon/\|u_n\|)$ and, since $\|u_n\| > \varepsilon$, it is $s_n < 0$. Thus

$$-\varepsilon^2 = f(\varphi(u_n s_n)) > f(\varphi(u_n s_n)(-s_n)) = f(x_n) \to f(x_0) = 0.$$

This contradiction ends the proof. ∎

A similar result can be deduced for the stable manifold, replacing f by $-f$ and F by $-F$. In this case, the manifold $W^s(x_0)$ is diffeomorphic to $\mathbb{R}^{m-d}$, where m is the dimension of M.

The unstable (stable) manifold theorem also holds for gradient fields of Morse functions, although the proof is considerably more difficult.

Corollary 1.6.4 *Consider a Morse function $f : M \to \mathbb{R}$ and a Morse field F for f. Then M is the disjoint union of the unstable manifolds $W^u(x_i)$ of the critical points x_i of f, hence M is a disjoint union of open cells of dimension d_i, where d_i is the index of x_i. The same statement is valid for the stable manifolds.*

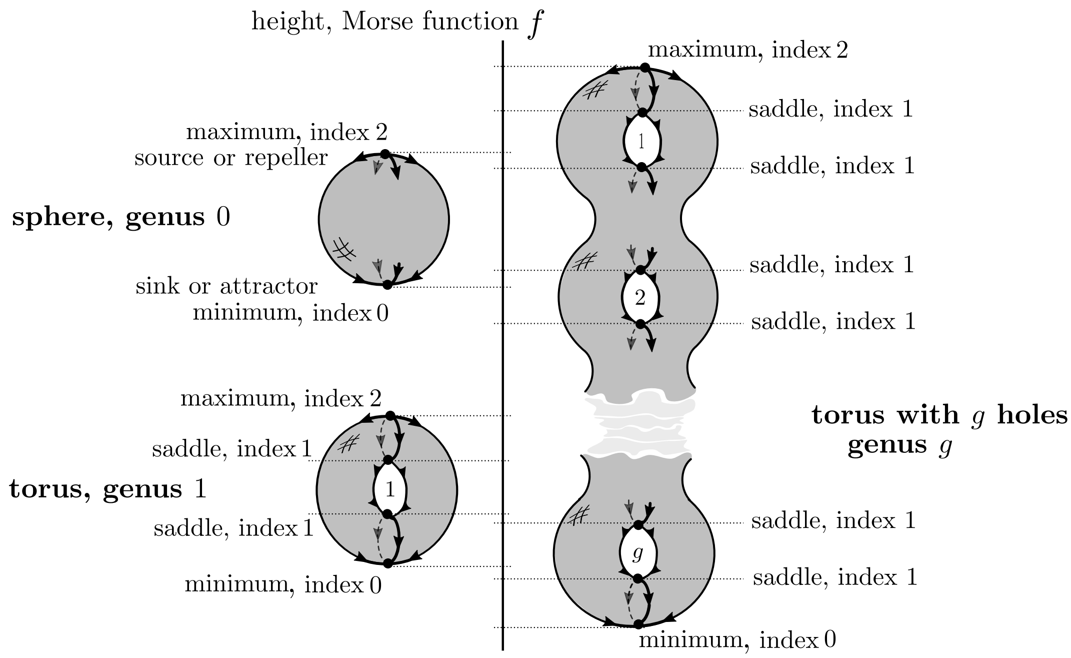
Proof. As we know, for every point $x \in M$ the $\alpha-$limit of the trajectory $\gamma(x)$ consists of a critical point x_i. Therefore $x \in W^u(x_i)$. On the other hand, the unstable manifolds of different critical points are disjoint. The result is then a consequence of the preceding Unstable Manifold Theorem. The argument is the same for the stable manifolds. ∎

To illustrate the simplest case, consider a compact and connected manifold M of dimension 1. The previous corollary provides (from any chosen Morse function on M) a decomposition of M into a finite number of points and Jordan arcs. Using the fact that M is locally homeomorphic to $\mathbb{R}$, we obtain that every point must be the endpoint of at most two arcs (not belonging to them) and using connectednes we easily obtain that M is topologically the circle S^1. This is a well known fact. For a direct and elementary proof see [ORRz] or [SjRz].

1.7 Examples

We have proved that Morse functions abound, but it is interesting to exhibit explicit examples. Here we present some, of the type described in Corollary 1.5.4. These examples should give a good idea of the general purpose behind the scenes.

The first three of them formalise the intuition in the figure below.



(1.7.1) Height functions on spheres. Let $S^n \subset \mathbb{R}^{n+1}$ be the unit sphere $x_0^2 + \cdots + x_n^2 = 1$. The simplest Morse function $f : S^n \to \mathbb{R}$ is the restriction of the 0-th projection: $x \mapsto x_0$. For $n = 2$ we can think of x_0 as the vertical axis; in general, we take $p_+ = (1, 0, \ldots, 0)$ for north pole and $p_- = (-1, 0, \ldots, 0)$ for south pole. Thus the name *height* for f is only natural. We analise the critical points of f as follows.

The parametrizations defined by the stereographic projection from p_- and p_+ are

$$\begin{cases} \varphi_-(x_1, \ldots, x_n) = \left(\frac{1-\|x\|^2}{1+\|x\|^2}, \frac{2x_1}{1+\|x\|^2}, \ldots, \frac{2x_n}{1+\|x\|^2}\right), \\ \varphi_+(x_1, \ldots, x_n) = \left(\frac{-1+\|x\|^2}{1+\|x\|^2}, \frac{2x_1}{1+\|x\|^2}, \ldots, \frac{2x_n}{1+\|x\|^2}\right), \end{cases}$$

so that

$$\begin{cases} f \circ \varphi_-(x_1, \ldots, x_n) = \frac{1-\|x\|^2}{1+\|x\|^2} = -1 + \frac{2}{1+\|x\|^2}, \\ f \circ \varphi_+(x_1, \ldots, x_n) = \frac{-1+\|x\|^2}{1+\|x\|^2} = 1 - \frac{2}{1+\|x\|^2}. \end{cases}$$

We get

$$\frac{\partial f \circ \varphi_-}{\partial x_i}(x_1, \ldots, x_n) = \frac{-4x_i}{(1 + \|x\|^2)^2}, \quad \frac{\partial f \circ \varphi_+}{\partial x_i}(x_1, \ldots, x_n) = \frac{4x_i}{(1 + \|x\|^2)^2},$$

and the only critical point in both cases is the origin, so that f has two critical points: the poles. We already knew they are the only local extremal points of f, and both global, p_N maximum and p_S minimum ($f(p_N) = 1, f(p_S) = -1$). On the other hand, a little more calculation and we see

$$\mathrm{Hess}_0(f \circ \varphi_-) = -4I, \quad \mathrm{Hess}_0(f \circ \varphi_+) = 4I,$$

and conclude p_N has index n and p_S has index 0.

This is a quite particular situation, and in fact characterizes *topological* spheres. Indeed, we will prove that *a compact connected smooth manifold which has a Morse function with exactly two critical points is homeomorphic to a sphere*. Note that by compactness f has two global extrema that must be the two critical points.

This is the Reeb lemma, with gives a homeomorphism, not a diffeomorphism. Actually, a diffeomorphism need not exist, and here enter Milnor's *exotic spheres*. These are spheres with non-standard differential structures. And there are a lot, although their distribution by dimension is quite chaotic. It is known that there are exotic differentiable structures in spheres of any odd dimension $n \neq 1, 3, 5, 13, 29, 61$, and in spheres of any even dimension $n < 124, n \neq 2, 4, 6, 56$. The case $n = 4$ remains misterious, although the specialist consider there must be exotic 4-spheres. However many proposed examples have been excluded (one by J.M. Montesinos in 1983). On the other hand, the distribution of exotic structures by dimension is rather amazing too. For instance, for dimension $n = 7$ their number is 8, for $n = 11$, 992, for $n = 15$, 16256, for $n = 16$, 2, ... ∎

(1.7.2) Height functions on the torus. We consider the parametrization of the torus $M \subset \mathbb{R}^3$ given by

$$\begin{cases} x = (2 + \cos u)\cos v \\ y = -\sin u \\ z = (2 + \cos u)\sin v + 3, \end{cases}$$

which describes the torus in "vertical" position, resting over the plane $z = 0$.

The tangent plane at a point $p \in M$ of local coordinates (u, v) is generated by the ortogonal vectors

$$\begin{cases} \varphi_u = (-\sin u \cos v, -\cos u, -\sin u \sin v), \\ \varphi_v = (-(2 + \cos u)\sin v, 0, (2 + \cos u)\cos v). \end{cases}$$

The map $H : \mathbb{R}^3 \to \mathbb{R} : a = (x, y, z) \mapsto z$ measures the height of a point over the plane $z = 0$. We consider $h = H|M : M \to \mathbb{R}$. Obviously $\mathrm{grad}_a H = (0, 0, 1)$ for every point $a \in \mathbb{R}^3$. We calculate $\mathrm{grad}_a h = \pi_a(\mathrm{grad}_a H)$, where π_a is the ortogonal projection of $\mathbb{R}^3$ over the tangent plane of M at a. The vector $\operatorname{grad} h(a)$ can be expressed as a linear combination $\alpha\varphi_u + \beta\varphi_v$, where

$$\begin{cases} \alpha = (\operatorname{grad}_a h)\varphi_u/\|\varphi_u\|^2 = (\operatorname{grad}_a H)\varphi_u/\|\varphi_u\|^2 = -\sin u \sin v/\|\varphi_u\|^2, \\ \beta = (2+\cos u)\cos v/\|\varphi_v|^2. \end{cases}$$

There are four critical points, with local coordinates $(0, 3\pi/2)$, $(\pi, 3\pi/2)$, $(\pi, \pi/2)$ and $(0, \pi/2)$, and different critical values $0, 2, 4$ and 6. Now, it is a routine matter to compute the Hessian matrix:

$$\begin{pmatrix} h_{uu} & h_{uv} \\ h_{vu} & h_{vv} \end{pmatrix} = \begin{pmatrix} -\cos u \sin v & -\sin u \cos v \\ -\sin u \cos v & -(2+\cos u)\sin v \end{pmatrix}.$$

For the critical points $(0, 3\pi/2), (\pi, 3\pi/2), (\pi, \pi/2), (0, \pi/2)$ we obtain:

$$\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -3 \end{pmatrix},$$

of course nonsingular matrices. We find, respectively, an attractor, a saddle, another saddle and a repeller for the gradient sistem $-\operatorname{grad} h$ in the torus. ∎

(1.7.3) Height functions on tori. Let us think of tori T_g of any genus $g \geq 1$. We can use equations in $\mathbb{R}^3$ of the form

$$T_g : x^2 + h(y,z)^2 = a^2,$$

where $h = 0$ defines a compact curve with $g-1$ simple crossings, like g circles centered at the z axis, each tangent to the next. Then for $\varepsilon > 0$ small, the nonsingular perturbations $h = \varepsilon$ and $h = -\varepsilon$ are, respectively, 1 and g Jordan curves, the latters inside the former. This explains why the equation above (for suitable $a > 0$) represents a torus with g holes. We claim that the height function $f : T_g \to \mathbb{R} : (x,y,z) \mapsto z$ is a Morse function with the predictable critical points.

First note that the critical points are the points of T_g whose tangent planes are orthogonal to the z axis, i.e. the points $(0,y,z)$ at which $h = \pm a$ and $h_y = 0$ (and then $h_z \neq 0$). These are the points in the plane $x = 0$ where the curves $h = \pm a$ are normal to the z axis: 2 points for $+a$ and $2g$ for $-a$. Some reflection shows that at those points $h_{yy} > 0$. On the other hand, implicit derivation of $z = z(x,y)$ in the equation $x^2 + h(x,y)^2 = a^2$ gives the Hessians at those critical points:

$$\begin{pmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{pmatrix} = \frac{-1}{h_z} \begin{pmatrix} \frac{1}{h} & 0 \\ 0 & h_{yy} \end{pmatrix},$$

which are non-degenerate.

In conclusion, f is a Morse function with $2+2g$ critical points at different heights with indices $2, 0$ (attractor and repeller, extrema of f) for $h = +a$, and 1 (saddles) for $h = -a$. ∎

(1.7.4) Morse functions on real projective spaces. The real projective space $P_n(\mathbb{R})$ is the quotient of $\mathbb{R}^{n+1} \setminus \{0\}$ by the relation $x \sim x'$ if and only if $x' = \lambda x$ for some $\lambda \in \mathbb{R}$ (hence $\lambda \neq 0$). We denote $x = (x_0, \dots, x_n)$ points in R^{n+1} and $z = (z_0 : \dots : z_n)$ points in $P_n(\mathbb{R})$. The projective space is also the quotient of the unit sphere $S^n \subset \mathbb{R}^{n+1}$ by the same relation, which in the sphere reduces to antipodal identification. We get a covering of two sheets $S^n \to P_n(\mathbb{R})$. The differential structure on $P_n(\mathbb{R})$ makes this covering a local diffeomorphism. It can also be described by the affine parametrizations of the open sets $U_k = \{x \in P_n(\mathbb{R}) : x_k \neq 0\}$:

$$\varphi_k : \mathbb{R}^n \to U_k : x \mapsto (x_0 : \dots : 1 : \dots : x_n),$$

where 1 is inserted as k-th component that is missing in x.

We want to define a Morse function in $P_n(\mathbb{R})$ by

$$f : P_n(\mathbb{R}) \to \mathbb{R} : x \mapsto \frac{1}{\|x\|^2} \sum_k a_k x_k^2.$$

Since the antipodal identification is a local diffeomorphism, we have to analise the critical points of

$$g : S^n \to \mathbb{R} : x \mapsto \sum_k a_k x_k^2.$$

We can parametrize each semisphere $x_k > 0$ (note that $x_k < 0$ is the same in $P_n(\mathbb{R})$) by

$$\psi_k : B^n \to S^n : x \mapsto (x_0, \dots, \sqrt{1 - \|x\|^2}, \dots, x_n),$$

where B^n is the unit open ball and x does not have the k-th component. Then:

$$g \circ \psi_k(x) = a_k + \sum_{\ell \neq k} (a_\ell - a_k) x_\ell^2.$$

This is a quadratic function, hence it is Morse if it is non degenerated. This requires $a_\ell \neq a_k$. Then the only critical point is the origin, and its index is the number of negative differences $a_\ell - a_k$. To simplify suppose $a_0 < \dots < a_n$ and the index is k. Since $\psi_k(0) = (0, \dots, 1, \dots, 0) \in S^n$, down to $P_n(\mathbb{R})$, we find the critical point $(0 : \dots : 1 : \dots : 0)$ with critical value a_k, non degenerated of index k.

Hence, with the assumption $a_0 < \cdots < a_n$, f is a Morse function on $P_n(\mathbb{R})$ with $n+1$ critical points of indices from 0 to n. ∎

(1.7.5) Morse functions on complex projective spaces. As usual, we will use the identification $\mathbb{C} \equiv \mathbb{R}^2$ by real and imaginary parts: $z = x + yi \equiv (x, y)$, so that $|z| = \sqrt{x^2 + y^2} = \|(x, y)\|$. We have the conjugate $\overline{z} = x - yi$ and $z\overline{z} = |z|^2$. This goes to $\mathbb{C}^{n+1} \equiv \mathbb{R}^{2n+2}$:

$$z = (z_0, \dots, z_n), \quad z_k \equiv (x_k, y_k), \quad \|z\| = \sqrt{\sum_{k=0}^{n} |z_k|^2} = \sqrt{\sum_{k=0}^{n} (x_k^2 + y_k^2)}.$$

We denote $z^{(k)} \in \mathbb{C}^n$ the vector z above without the k-th component.

As is well known, $P_n(\mathbb{C})$ is the quotient of $\mathbb{C}^{n+1} \setminus \{0\}$ by the relation $z \sim z'$ if and only if $z' = \lambda z$ for some $\lambda \in \mathbb{C}$ (hence $\lambda \neq 0$). We denote $z = (z_0, \dots, z_n)$ a point in $\mathbb{C}^{n+1}$ and $z = (z_0 : \dots : z_n)$ the corresponding point in $P_n(\mathbb{C})$. The smooth (in fact holomorphic) structure of $P_n(\mathbb{C})$ is given by the affine open sets $U_k : z_k \neq 0$ and the parametrizations

$$\varphi_k : \mathbb{C}^n \to U_k \subset P_n(\mathbb{C}) : z \mapsto \varphi_k(z) = (z_0 : \dots : 1 : \dots : z_n),$$

the k-th component missing in z, equal to 1 in $\varphi_k(z)$.

All of this is standard. Now we claim that

$$f : P_n(\mathbb{C}) \to \mathbb{R} : z = (z_0 : \dots : z_n) \mapsto \frac{1}{\|z\|^2} \sum_{\ell=0}^{n} \ell |z_\ell|^2,$$

is a well defined smooth function and that it is a Morse function. To prove it one can use the affine parametrizations and much patience, or do something different. Indeed, we can consider $P_n(\mathbb{C})$ the quotient of S^{2n+1} by the same relation $z' = \lambda z$, which restricted to that sphere imposes $|\lambda| = 1$. In particular, the inverse image of $z \in P_n(\mathbb{C})$ is a circle cut in S^{2n+2} by the complex line z, which is a real plane. This is the so-called Hopf fibration $S^{2n+1} \to P_n(\mathbb{C})$ of projective space by circles, a very important map. But let us come back to our Morse function.

As said we will use a different parametrization of the U_k. Note that any $z \in S^{2n+1}$ with $z_k \neq 0$ is proportional to another with k-component real and positive. In fact, multiply by $\lambda = \overline{z_k}/|z_k|$ and for the new z it holds

$$z_k = \sqrt{1 - \sum_{\ell \neq k}^{n} |z_\ell|^2} = \sqrt{1 - \|z^{(k)}\|^2}.$$

Furthermore, note that $0 < z_k \leq 1$, hence $0 \leq \|z^{(k)}\|^2 < 1$ and we are in the unit open ball $B^{2n} \subset \mathbb{C}^n \equiv \mathbb{R}^{2n}$. Thus we obtain the parametrization

$$\psi_k : B^{2n} \to U_k : z^{(k)} \mapsto (z_0 : \ldots : \sqrt{1 - \|z^{(k)}\|^2} : \ldots : z_n).$$

With this we compute the localization of f:

$$f \circ \psi_k(z^{(k)}) = k + \sum_{\ell \neq k} (\ell - k)|z_\ell|^2 = k + \sum_{\ell \neq k} (\ell - k)(x_\ell^2 + y_\ell^2).$$

This is a non degenerated quadratic function, hence a Morse function whose only critical point is the origin and has index $2k$. Thus the critical points of f are $\psi_k(0) = (0 : \ldots : 1 : \ldots : 0)$, $0 \leq k \leq n$, with $f(\psi_k(0)) = k$, non degenerated of indices $2k$.

Thus indeed, f is a Morse function. ∎

1.8 Morse and Poincaré polynomials

One of the aims of Morse theory is to study the relation between the dynamics of the gradient system of a Morse function defined on a manifold and the topology of the manifold. In order to develop this theory we introduce the following notions. Let M be a compact m-manifold.

Definition 1.8.1 The *Morse polynomial* of the Morse function $f : M \to \mathbb{R}$ is

$$\mathtt{M}_f(t) = \sum_{x \in \mathcal{K}} t^{\mu(x)},$$

where $\mathcal{K}$ denotes the set of critical points of f and $\mu(x)$ is the Morse index of x.

By $\mathtt{c}_j$ we denote the number of critical points whose index is equal to $j \geq 0$. Then

$$\mathtt{M}_f(t) = \sum_{j=0}^{m} \mathtt{c}_j t^j.$$

Observe that $\mathtt{M}_f(1) = \sum_{j=0}^{m} \mathtt{c}_j$ is the total number of critical points.

Example 1.8.2 The polynomials of the Morse function given in the previous section are the following:

(1) For the height $f : S^n \to \mathbb{R}$ of the sphere, $\mathtt{M}_f(t) = 1 + t^n$.

(2) For the height $f : T_g \to \mathbb{R}$ on a torus of genus g, $\mathtt{M}_f(t) = 1 + 2gt + t^2$.

(3) For the Morse function $f : P_n(\mathbb{R}) \to \mathbb{R}$ in a real projective space, $\mathtt{M}_f(t) = 1 + t + \cdots + t^n$.

(4) For the Morse function $f : P_n(\mathbb{C}) \to \mathbb{R}$ in a complex projective space, $\mathtt{M}_f(t) = 1 + t^2 + \cdots + t^{2n}$. ∎

This Morse polynomial involves the dynamic of a gradient field on M. Next we will define a polynomial of purely topological nature. We denote by β_j the j-th Betti number of M.

Definition 1.8.3 The *Poincaré polynomial* of M is the polynomial

$$\mathtt{P}_M(t) = \sum_{j=0}^{m} \beta_j t^t.$$

Observe that $\mathtt{P}_M(-1) = \sum_{j=0}^{m}(-1)^j \beta_j$ is the Euler characteristic $\chi(M)$ of M.

We are now in a position to state the following theorem, which is one of the central results of the Morse theory.

Theorem 1.8.4 *Let $f : M \to \mathbb{R}$ be a Morse function. Then*

$$\mathtt{M}_f(t) = \mathtt{P}_M(t) + (1 + t)Q(t),$$

where $Q(t)$ is a polynomial with non-negative integer coefficients.

We remark that the left-hand side contains information on the dynamical invariants of the gradient flow near the critical points and the right-hand side contains information on the topology of the manifold M. As a consequence, this result establishes a deep relation between dynamics and topology. For instance, we will compute the Euler characteristic $\chi(M)$:

$$\mathtt{M}_f(-1) = \mathtt{P}_M(-1) = \chi(M).$$

Example 1.8.5 In view of the Morse polynomials given above we get:

$$\chi(S^n) = 1 + (-1)^n = \begin{cases} 0 & \text{for } n \text{ odd,} \\ 2 & \text{for } n \text{ even,} \end{cases}$$
$$\chi(T_g) = 1 - 2g + (-1)^2 = 2 - 2g,$$
$$\chi(P_n(\mathbb{R})) = 1 + (-1) + \cdots + (-1)^n = \begin{cases} 0 & \text{for } n \text{ odd,} \\ 1 & \text{for } n \text{ even,} \end{cases}$$
$$\chi(P_n(\mathbb{C})) = 1 + (-1)^2 + \cdots + (-1)^{2n} = n + 1.$$

∎

The proof of the theorem runs very geometrically studying the decomposition of M defined by the Morse function f. For every regular value a of f, the level set $f^{-1}(a)$ is a hypersurface of M and these level sets divide M into pieces whose topology evolves when a increases from the minimum to the maximum of f. Let us illustrate this.

Let $c_0 < \cdots < c_k < \cdots < c_r$ be the critical values of f, and let us suppose f has one critical point z_k at each critical level $f^{-1}(c_k)$ (1.5.4). Now choose regular values a_{k-1} in between: $c_1 < a_1 < c_2 < \cdots < a_{k-1} < c_k < a_k < \cdots < a_{r-1} < c_r$. Note that M being compact c_1 must be its minimum and c_r its maximum. Then we have the decomposition

$$M = M_{-\infty}^{a_1} \cup M_{a_1}^{a_2} \cup \cdots \cup_{r-1} M_{a_{r-1}}^{a_r} \cup M_{a_r}^{+\infty},$$

where $M_{a_{k-1}}^{a_k} = f^{-1}[a_{k-1}, a_k]$ is a compact manifold whose boundary consists of the hypersurfaces $f^{-1}(a_{k-1})$ and $f^{-1}(a_k)$, and the unions are made glueing through the corresponding common critical level $f^{-1}(a_k)$ (with a small abuse of notation for $a_0 = -\infty$ and $a_{r+1} = +\infty$).

The idea is to reconstruct the topology of M from the understanding of the topology of the pieces.

With this in mind, we start at the minimum $c_1 = 0$ at the unique point z_1. By the Morse lemma,

$$f(x) = c_1 + x_1^2 + \cdots + x_m^2$$

in some local coordinates x in an open neighborhood U of z_1. Since the minimum is unique, $f > 0$ off U, and we can choose $a_1 > 0$ close enough to 0 so that $f > a_1$ on the compact set $M \setminus U$. Thus $M_{-\infty}^{a_1} = f^{-1}(-\infty, a_1] \subset U$ and

$$M_{-\infty}^{a_1} : x_1^2 + \cdots + x_m^2 \leq a_1 - c_1, \quad f^{-1}(a_1) : x_1^2 + \cdots + x_m^2 = a_1 - c_1.$$

Hence $M_{-\infty}^{a_1}$ is a closed ball with boundary the sphere $f^{-1}(a_1)$.

Thus we have described the first piece, but we need $a_1 \to c_1$. What if $a_1 \to c_2$? Let $a_1 < b_1 < c_2$ with b_1 close enough to a_1 to have $M_{a_1}^{b_1} = f^{-1}[b_1, a_1] \subset U$. Then

$$M_{a_1}^{b_1} : a_1 - c_1 \leq x_1^2 + \cdots + x_m^2 \leq b_1 - c_1$$

is a cylinder with boundary the two spheres $f^{-1}(a_1)$ and $f^{-1}(b_1)$. It is not difficult to deduce that

$$M_{-\infty}^{a_1} \approx M_{-\infty}^{a_1} \cup M_{a_1}^{b_1} = M_{-\infty}^{b_1}.$$

Now accept the topology does not change while $b_1 \to c_2$, so that $M_{a_1}^{b_1}$ is still a cylinder and the above homeomorphism remains. In other words, we can suppose a_1 arbitrarily close to c_2 and go for the piece $M_{a_1}^{a_2}$. Accept that for a_2 close enough to c_1 we can describe the topology of $M_{a_1}^{a_2}$. Next turn to a b_2 between a_2 and c_3 to find a cylinder joining $f^{-1}(a_2)$ and $f^{-1}(b_2)$ which remains so when $b_2 \to c_3$ ... And go on.

In this way we reach the last piece with a_{r-1} close to c_r, the maximum of f at z_r. We mimic the argument for the minimum. In some local coordinates $f(x) = c_r - x_1^2 - \cdots - x_m^2$, and

$$M_{a_{r-1}}^{+\infty} : x_1^2 + \cdots + x_m^2 \leq c_r - a_{r-1}, \quad f^{-1}(a_{r-1}) : x_1^2 + \cdots + x_m^2 = c_r - a_{r-1}.$$

Again we find a ball with boundary a sphere.

This procedure leads to a decomposition of M into pieces of known topology glued through their boundaries. In the way we asked to accept that: (i) the topology does not change between consecutive critical points, and (ii) the change through a critical value can be described well. This is the concern of the following sections.

1.9 Level sets without stationary points in between

Here we start the formalization of the ideas scketched at the end of the previous section. We introduce some general notations for a given function $f : M \to \mathbb{R}$:

$$\begin{aligned} M_a &= f^{-1}[a, \infty) = \{x \in M : a \leq f(x)\}, \\ M^b &= f^{-1}(-\infty, b] = \{x \in M : f(x) \leq b\}, \\ M_a^b &= f^{-1}[a, b] = \{x \in M | a \leq f(x) \leq b\}, \end{aligned}$$

where $a < b$.

We will consider strict Lyapunov maps f of gradientlike flows, which is the situation for a Morse function f.

As we asked to accept, nothing happens when stationary points do not occur:

Proposition 1.9.1 *Let $\varphi : M \times \mathbb{R} \to M$ be a gradientlike flow in the metric space M with strict Lyapunov map $f : M \to \mathbb{R}$. Suppose that M_a^b is compact and that there is no stationary point in M_a^b. Then M^a is a strong deformation retract of M^b relative to M^a. Furthermore, there is a homeomorphism $h : f^{-1}(b) \times [0,1] \to M_a^b$ mapping $f^{-1}(b) \times \{0\}$ onto $f^{-1}(b)$ and $f^{-1}(b) \times \{1\}$ onto $f^{-1}(a)$.*

Proof. The deformation is performed by means of the flow. For every point $x \in M_a^b$ we have an *exit time* $t_x \geq 0$ defined in the following way: t_x is the unique nonnegative number which satisfies $f(xt_x) = a$. We prove that the exit map $M_a^b \to \mathbb{R} : x \mapsto t_x$ is well defined and continuous. On the one hand, there is at least a $t \geq 0$ such that $f(xt) = a$. Otherwise $x[0, \infty)$ would be contained in M_a^b and, since M_a^b is compact, $\omega(x)$ would be nonempty and consist of stationary points contained in M_a^b, in contradiction with the hypothesis. Moreover, the fact that f is strictly decreasing along non stationary orbits implies that such a t is unique.

On the other hand, the exit map is continuous. Let us see that $x_n \to x$ implies $t_{x_n} \to t_x$ for $x_n, x \in M_a^b$. If not, taking a subsequence, there is $\varepsilon > 0$ such that $d(t_{x_n}, t_x) \geq \varepsilon$. We now show that the sequence t_{x_n} is bounded. Otherwise we can suppose $t_{x_n} \to \infty$. Now $x_n t_{x_n} \in f^{-1}(a)$ for every n and, by compactness, we can also assume $x_n t_{x_n} \to y_0 \in f^{-1}(a)$. We say that $y_0(-\infty, 0]$ is contained in M_a^b. Indeed, for any fixed $s < 0$ we have $x_n t_{x_n} s \to y_0 s$, hence there is a subsequence n_k such that $d(x_{n_k} t_{x_{n_k}} s, y_0 s) < 1/k$. As $t_{n_k} + s > 0$ for large k and $s < 0$, we have

$$b \geq f(x_{n_k}) > f(x_{n_k} t_{x_{n_k}} s) > f(x_{n_k} t_{x_{n_k}}) = a$$

and $x_{n_k} t_{x_{n_k}} s \in M_a^b$. Since M_a^b is closed, $y_0 s \in M_a^b$. Thus $y_0(-\infty, 0] \subset M_a^b$ as said. But then $\alpha(y_0) \subset M_a^b$ and consists of stationary points, in contradiction with the hypothesis. We have showed that t_{x_n} is bounded and can assume $t_{x_n} \to t^*$. Hence $x_n t_{x_n} \to xt^*$ and $a = f(x_n t_{x_n}) \to f(xt^*)$. This implies $a = f(xt^*)$ and $t_x = t^*$, that is, $t_{x_n} \to t_x$ against the condition $d(t_{x_n}, t_x) \geq \varepsilon$.

Once we know t_x is continuous, we define $r(x) = xt_x$ for $x \in M_a^b$ and $r(x) = x$ for $x \in M^a$. This map is continuous because $t_x = 0$ for $x \in f^{-1}(a) = M_a^b \cap M^a$, and is obviously a retraction. The deformation $H : M^b \times I \to M^b$ is defined by the formula $H(x, s) = x(st_x)$ for $x \in M_a^b$ and $H(x, s) = x$ for $x \in M^a$ (again this is continuous because $t_x = 0$ for $x \in f^{-1}(a)$).

Finally, $h = H|f^{-1}(b) \times I$ is a homeomorphism onto M_a^b. Indeed, for $y \in M_a^b$ we can define the *entry time:* the unique $s_y \leq 0$ with $f(ys_y) = b$. Using $\alpha(y)$ for this entry time as $\omega(x)$ was used for the exit time, one sees s_y is continuous. Then we have $y = x(st_x)$ if and only if $x = ys_y$, $s = -s_y/t_x$ ($t_x \neq 0$ because $f(x) = b$). These formulae define the inverse of $y = h(x, s)$, and h is a homeomorphism with the required properties. ∎

As deformation retracts are homotopy equivalences, we get:

Corollary 1.9.2 *In the preceding situation, the inclusion $i : M^a \to M^b$ induces isomorphisms in cohomology relative to any closed set $K_0 \subset M^a$. In particular we have $H^j(M^b, M^a) = \{0\}$ for every j.*

Remark 1.9.3 Suppose our flow is the gradient flow of a smooth Morse function $f : M \to \mathbb{R}$ on a compact manifold M. Then the flow is smooth, and the deformation retract H is smooth too. Also, the restriction $h = H|f^{-1}(b) \times [0,1]$ is a diffeomorphism.

For the exit map t_x is smooth. Indeed, t_x is the solution of the smooth implicit function problem $F(x,t) = f \circ \varphi(x,t)) = a$ and, as computed in the first section to show f is a strict Lyapunov map, $\partial F/\partial t < 0$, hence $\neq 0$. Similarly, the entry map s_y is smooth too. ∎

Remark 1.9.4 The homeomorphism $h_1 : f^{-1}(b) \to f^{-1}(a)$ given by the furthermore of the proposition is $h_1(x) = xt_x$. Now suppose that there are no stationary points in $M^b_{a'}$ for some $a' < a$ (for instance, this happens for a' close enough to a when the set of stationary points is compact or when f is proper). Then h_1 extends to a homeomorphism $M^b \to M^a$.

Indeed, in that case the proposition can be applied to the couple b, a' and the couple a, a' to get homeomorphisms

$$M^b_{a'} \approx f^{-1}(b) \times I \approx f^{-1}(a) \times I \approx M^a_{a'},$$

where the one among cylinders comes from h_1 above, hence extends h_1 to $M^b_{a'}$. But one immediately checks h_1 is the identity on $f^{-1}(a')$, hence it glues well with the identity on $M^{a'}$ to obtain the homeomorphism we sought.

Whether there is a diffeomorphism $M^b \to M^a$ for the gradient flow of a Morse function is a crucial matter. ∎

We conclude the section with this very important fact:

Proposition 1.9.5 (Reeb's Lemma) *Let M be a compact manifold and suppose there is a Morse function $f : M \to \mathbb{R}$ with exactly two critical points. Then M is homeomorphic to a sphere.*

Proof. Since M is compact, f has a maximum b_0 and a minimum $a_0 < b_0$. They are critical points of f, hence there are no more. Consequently, there is a unique point x_0 with value a_0 and a unique x_1 with value b_0. As explained at the end of the previous section, for $a > a_0$ close enough to a_0, the set M^a is a closed ball with boundary $f^{-1}(a)$, and arguing similarly with the maximum b_0, for $b < b_0$ close enough to b_0, the set M_b is a closed ball with boundary $f^{-1}(b)$. We have

$$M = M^a \cup_a M_a^b \cup_b M_b.$$

Here the first piece is a ball, the second a cylinder and the third another ball, and the unions go through the spheres $f^{-1}(a)$ and $f^{-1}(b)$ pointed by the subindices.

Consider the unit sphere $S = S^m : y_1^2 + \cdots + y_{m+1}^2 = 1$ in $\mathbb{R}^{m+1}$ and mimic the decomposition of M using the last coordinate as Morse function:

$$S^a : y_{m+1} \leq -\tfrac{1}{2}, \quad S_a^b : -\tfrac{1}{2} \leq y_{m+1} \leq \tfrac{1}{2}, \quad S_b : y_{m+1} \geq \tfrac{1}{2}.$$

This three pieces are of course a ball, a cylinder and a ball and we have homeomorphisms (in fact diffeomorphisms)

$$h_a : M^a \to S^a, \quad h : M_a^b \to S_a^b \quad h_b : M_b \to S_b,$$

but of course they need not glue to a global homeomorphism $M \to S^m$. So let us modify h_a and h_b to amend this. Since the middle pieces are cylinders, we can suppose that h maps the upper boundary $f^{-1}(b)$ onto $\partial S_b : y_{m+1} = \frac{1}{2}$ and the lower one $f^{-1}(a)$ onto $\partial S^a : y_{m+1} = -\frac{1}{2}$ and so h restricts to homeomorphisms

$$g_a : f^{-1}(a) \to \partial S^a, \quad g_b : f^{-1}(b) \to \partial S_b.$$

Then g_a (resp. g_b) extend to homeomorphisms $h'_a : M^a \to S^a, h'_b : M_b \to S_b$. Let us do it for g_a. Since M^a and S^a are balls, be can suppose $g_a : S^{m-1} \to S^{m-1}$, and define the extension to the ball B^m by

$$h'_a : B^m \to B^m : y \mapsto \begin{cases} \|y\| g_a(y/\|y\|) & \text{for } x \neq 0, \\ 0 & \text{for } x \neq 0. \end{cases}$$

Finally, h'_a, h, h'_b glue well to give a homeomorphism from M onto S^m. ∎

The key is that the above proof gives a homeomorphism, not a diffeomorphism: h'_a and h'_b need not be differentiable. And this is definite: Milnor found his exotic spheres by constructing Morse functions with exactly two critical points on compact manifolds not diffeomorphic to spheres.

1.10 Level sets with one stationary point in between

Once settled the case when there is no stationary point between two levels, we turn to the more delicate situation when there is one.

In Theorem 1.6.3 we discussed the unstable and stable manifolds of a stationary point x_0 of a flow with strict Lyapunov function $f : M \to \mathbb{R}$. Here we consider for $a < b$ the *truncated* unstable and stable manifolds:

$$\begin{cases} W^u(x_0) \cap M_a^b = \{x \in W^u(x_0) : a \leq f(x) \leq b\}, \\ W^s(x_0) \cap M_a^b = \{x \in W^s(x_0) : a \leq f(x) \leq b\}. \end{cases}$$

We now introduce a notion that will play an important role in the next results. The terminology is motivated by Shape Theory, a homotopy theory introduced and developed by K. Borsuk which has proved to be very useful in Dynamics. For more information see [B2], [DySe] and [MaSe].

Let X be a metric space and $K_0 \subset K$ closed subsets of X. We say that the inclusion $i : K \to X$ is a *strict shape equivalence relative to* K_0 if there exists a sequence of maps $h_k : X \to X$ such that:

(sh1) $h_k \simeq \mathrm{Id}_X$,

(sh2) $h_k(K) \subset K$ and $h_k|K \simeq \mathrm{Id}_K$ in K, and

(sh3) for every neighborhood U of K in X, $h_k(X) \subset U$ and $h_k \simeq h_{k+1}$ in U for k large enough,

and all homotopies involved (hence all h_k's) fix every point in K_0.

Now we can state and prove the main result in this section.

Proposition 1.10.1 *Let $\varphi : M \times \mathbb{R} \to M$ be a gradientlike flow in the metric space M with strict Lyapunov map $f : M \to \mathbb{R}$. Suppose that M_a^b is compact, that there is a unique stationary point x_0 in M_a^b, and that $f(x_0) = c$ with $a < c < b$; set $W_a^u(x_0) = W^u(x_0) \cap M_a^b$. Then the inclusion $i : M^a \cup W_a^u(x_0) \to M^b$ is a strict shape equivalence relative to M^a.*

Proof. To prove that i is a strict shape equivalence we must define maps and homotopies on M^b and various subsets, all fixing M^a. Clearly, it is enough to define them all on M_a^b and check they fix $f^{-1}(a)$. We will do this sistematically.

Concerning $W_a^u(x_0)$ we remark that

$$W^u(x_0) \cap M_a^b = W^u(x_0) \cap M_a, \quad M^a \cap W_a^u(x_0) = f^{-1}(a).$$

The first equality explains why b is missing in the notation $W_a^u(x_0)$: if $x \in W^u(x_0)$ then always $b > f(x_0) \geq f(x)$. The second one explains why we write $M^a \cup W_a^u(x_0)$ instead $M^a \cup W^u(x_0)$: both sets are equal but we prefer the two pieces to meet exactly at the a-level $f^{-1}(a)$

This said, set $X = M^b$, $K = M^a \cup W_a^u(x_0)$, $K_0 = M^a$ and suitable maps h_k. Clearly K_0 is closed, and K too, because $W_a^u(x_0)$ is closed in M_a^b (so compact). Indeed, let $x_n \in W_a^u(x_0)$ be a sequence with $x_n \to x \in M_a^b$. Since $x_n(-\infty, 0] \subset M_a^b$ for every n, fixed $t \leq 0$, the sequence $x_n t$ is in M_a^b. As $x_n t \to xt$, we get $xt \in M_a^b$. Thus $x(-\infty, 0] \subset M_a^b$, hence $\alpha(x) \subset M_a^b$ and $\alpha(x) = x_0$, That is, $x \in W_a^u(x_0)$ and $W_a^u(x_0)$ is closed in M_a^b.

We need maps $h_k : X = M^b \to X$ and homotopies as in the conditions (sh1,2,3) of shape equivalence, All these maps and homotopies must be the identity on $K_0 = M^a$.

(1) The maps h_k.

We define, for $b \geq f(x) \geq a$:

$$h_k(x) = \begin{cases} xk & \text{when } f(xk) \geq a, \\ xt_x & \text{when } f(xk) \leq a, \end{cases}$$

where the exit time $t_x \geq 0$ is the (unique) time with $f(xt_x) = a$, which does exists when $f(xk) \leq a$ (recall the proof of Proposition 1.9.1). Clearly $h_k(x) \in M^b$. Moreover, if $f(xk) = a$ then $t_x = k$, and both definitions coincide. Finally, h_k fixes all points in $f^{-1}(a)$: if $f(x) = a$ then $f(xk) \leq f(x) = a$ and $t_x = 0$. This guarantees that h_k extends by the identity to M^a.

We see that h_k is continuous. Working on closed pieces, we are reduced to prove continuity on the piece $f(xk) \leq a$, and there to prove continuity of the exit time. To do that we consider a sequence $x_n \to x$ with $f(x_n k) \leq a$, $f(xk) \leq a$ and $t_{x_n} \to t_x$. Indeed, otherwise, up to a subsequence, we have $|t_{x_n} - t_x| \geq \varepsilon$ for some $\varepsilon > 0$. Now, k bounds this sequence t_{x_n}: if $t_{x_n} > k$ for some n, then $f(x_n k) > f(x_n t_{x_n}) = a$, which is not the case. Hence, we can suppose $t_{x_n} \to t^*$, so that $x_n t_{x_n} \to xt^*$. As $f(x_n t_{x_n}) = a$ for all n, $f(xt^*) = a$ and $t^* = t_x$. So $t_{x_n} \to t^* = t_x$, against the ε inequality above.

(2) The homotopies $h_k \simeq \mathrm{Id}$.

These appear in (sh1,2). with various domains. The common definition for $b \geq f(x) \geq a$ and $0 \leq s \leq 1$ is

$$H_k(x, s) = \begin{cases} x(sk) & \text{when } f(xk) \geq a, \\ x(st_x) & \text{when } f(xk) \leq a. \end{cases}$$

We have $b \geq f(x) \geq f(H_k(x,s))$, and $t_x = k$ if $f(xk) = a$. On the other hand, if $f(x) = a$, $f(xk) \leq a, t_x = 0$ and $H_k(x,s) = x$, which again enables extension by the identity to M^a. Obviously $H_k : \mathrm{Id}_X \simeq h_k$ and we have (sh1). But also (sh2) because $H_k(K \times I) \subset K$. It suffices to see it for images of points $x \in W^u_a(x_0)$. For $x = x_0$ is trivial, hence suppose $x \neq x_0$. Then $\alpha(H_k(x,s)) = \alpha(x)$ is x_0 and $H_k(x,s) \in W^u(x_0) \subset K$ (we recall our initial remarks). Thus we have (sh2).

But we can prove an additional condition. Set $X_k = h_k(X)$, so that $h_k(X_k) \subset X_k$. Note that $X_k = h_k(M^b_a) \cup M^a$, hence it is closed. We have:

(sh4) $K \subset X_k$ and $h_k|X_k \simeq \mathrm{Id}_{X_k}$.

For the inclusion, we only must consider points $x \in K$ with $f(x) > a$. Then $x \in W^u_a(x_0)$, hence $x(-k) \in W^u_a(x_0)$ and $x = h_k(x(-k)) \in X_k$. For the homotopy we have to check that

$$H_k(X_k \times I) \subset X_k.$$

Let $z = H_k(x_k, s)$, $x_k \in X_k$. If $f(z) \leq a$, $z = h_k(z) \in X_k$. If $f(z) > a$, then $f(x_k) > a$ too, hence $x_k = h_k(x) = xk$, $x \in X$. Now, $z = H_k(x_k, s) = x_k t$ for some $t \geq 0$, so that $z = (xk)t = x(k+t)$ and $z(-k) = xt$. Consequently

$$b \geq f(x) \geq f(z(-k)) \geq f(z) > a$$

and $z = h_k(z(-k)) \in X_k$.

(3) The homotopies $h_k \simeq h_{k+1}$.

These are in (sh3), and they are defined by

$$H_k(x,s) = \begin{cases} x(k+s) & \text{when } f(x(k+s)) \geq a, \\ xt_x & \text{when } f(x(k+s)) \leq a, \end{cases}$$

for $b \geq f(x) \geq a$, $0 \leq s \leq 1$. If $f(x) = a$, then $f(x(k+s)) \leq a$ and $t_x = 0$, hence $H_k(x,s) = x$ and H_k extends to M^a by the identity. Clearly, $H_k : h_k \simeq h_{k+1}$.

Next, fix any open neighborhood U of $K = M^a \cup W^u_a(x_0)$ in M^b. We must see that $H_k(X \times I) \subset U$ for k large. If $f(H_k(x,s)) \leq a$ any k works, hence we have to see that for k large and $f(x(k+s)) > a$ it is $x(k+s) \in U$. If that is not the case, there exist sequences $x_n \in X$, $t_n = n + s_n \geq n$ with $x_n t_n \notin U$. In particular $t_n \to +\infty$ and $f(x_n t_n) > a$, so that $x_n[0,t_n] \subset M^b_a$. By compactness we can suppose $x_n t_n \to x \in M^b_a$, and then $x(-\infty, 0] \subset M^b_a$. Indeed, we apply the usual trick: for every $s < 0$, $x_n t_n s \to xs$ and for n large $0 \leq t_n + s \leq t_n$ so that $x_n t_n s \in M^b_a$. As M^b_a is closed, $xs \in M^b_a$. But $x(-\infty,0] \subset M^b_a$ implies $\alpha(x) \subset M^b_a$, hence it reduces to x_0, and $x \in W^u_a(x_0) \subset U$. However x is adherent to the closed set $M^b \setminus U$, a contradiction.

All in all, we have (sh3), but also

(sh5) $X_{k+1} \subset X_k$ and $h_k \simeq h_{k+1}$ in X_k.

This is because $H_k(X \times I) \subset X_k = h_k(X)$. Indeed: (i) if $f(x(k+s)) \geq a$, then $f(xs) \geq a$ and $H_k(x,s) = h_k(xs)$, (ii) if $f(x(k+s)) \leq a$, then $H_k(x,s) = xt_x = h_k(xt_x)$, and (iii) if $f(x) \leq a$, then $H_k(x) = x = h_k(x)$.

Thus we have proved that $M^a \cup W^u_a(x_0) \to M^b$ is a strict shape equivalence, with some refinements needed in the next proposition. ∎

A shape equivalence is not in general a homotopy equivalence, however it works fine on cohomology.

Proposition 1.10.2 *In the preceding situation, the inclusion $i : M^a \cup W^u_a(x_0) \to M^b$ induces isomorphisms in cohomology relative to any closed set $K_0 \subset M^a$ (in particular $K_0 = \varnothing$).*

Proof. As in the preceding proof, set $X = M^b$, $K = M^a \cup W^u_a(x_0)$, but here $K_0 \subset M^a$ (not equal as there). We have shown that $i : K \to M^b_a$ is a shape equivalence relative to K_0 via maps $h_k : X \to X$ that verify (sh1,2,3) plus some additional properties (sh4,5) involving the closed sets $X_k = h_k(X)$.

Consider the sequence of homomorphisms induced in cohomology by the h_k's:

$$h_k^* : H^*(X_k, K_0) \to H^*(X, K_0).$$

Let $j_{k+1} : X_{k+1} \to X_k$ be the inclusion. Observe that $h_k = j_{k+1} \circ h_{k+1}$ only up to homotopy (sh5) (for k large), but this guarantees that the corresponding equality $h_k^* = h_{k+1}^* \circ j_{k+1}^*$ holds in cohomology (for k large). Consequently we have a limit homomorphism

$$h^* = \lim_k h_k^* : \lim_k H^*(X_k, K_0) \to H^*(X, K_0).$$

These X_k form a *nested system,* because by (sh3) the neighborhoods V of all X_k's are the neighborhoods V of K. Then, since in metric spaces the cohomology of a closed subset is the limit of the cohomology of its neighborhoods [Br, G], we have

$$\lim_k H^*(X_k, K_0) = \lim_k \lim_{V \supset X_k} H^*(V, K_0) = \lim_{V \supset K} H^*(V, K_0) = H^*(K, K_0).$$

Hence we have a homomorphism $h^* : H^*(K, K_0) \to H^*(X, K_0)$. Consider now the inclusions $i_k : X_k \to X$ and the following diagrams

$$\begin{array}{ccc} H^*(X_k,K_0) & \xrightarrow{\mathrm{Id}} & H^*(X_k,K_0) \\ {\scriptstyle i_k^*}\uparrow & \swarrow {\scriptstyle h_k^*} & \uparrow{\scriptstyle i_k^*} \\ H^*(X,K_0) & \xrightarrow{\mathrm{Id}} & H^*(X,K_0) \end{array} \qquad\qquad \begin{array}{ccc} H^*(K,K_0) & \xrightarrow{\mathrm{Id}} & H^*(K,K_0) \\ {\scriptstyle i^*}\uparrow & \swarrow {\scriptstyle h^*} & \uparrow{\scriptstyle i^*} \\ H^*(X,K_0) & \xrightarrow{\mathrm{Id}} & H^*(X,K_0) \end{array}$$

The first one is commutative because $h_k \circ i_k \simeq \mathrm{Id}_{X_k}$ (sh4) and $h_k \simeq \mathrm{Id}_X$ (sh1), the second one by the previous remark on $\lim_k H^*(X_k,K_0)$. But the second one says that the diagonal arrow is an isomorphism with inverse the vertical one. ∎

In the situation of Proposition 1.10.1 consider k with $a \le k < c$. The trajectory of every $x \in W^u(x_0)\setminus\{x_0\}$ hits $f^{-1}(k)$. We repeat the argument used before. As $\alpha(x) = x_0$, $f(xs) > a$ for some s (maybe $s < 0$), and $x[s,+\infty)$ escapes de compact M_k^b. For otherwise $\omega(x) \subset M_k^b$ and $\omega(x) = x_0$, impossible. After this, there is a time t_x with $f(xt_x) = k$. This time is unique because f is injective on trajectories, and the map $x \mapsto t_x$ is continuous.

The proof is very alike others done before, so we explain only the variation. Let $x_n \to x$. The variation is on why t_{x_n} is bounded. Consider $t_0 > t_x$ and $r = f(xt_0) < k$. By continuity $f(x_n t_0) \to f(xt_0) = r$ and, thus, $f(x_n t_0) < k$ for almost every n. Hence $t_{x_n} < t_0$ for almost every n and t_{x_n} is bounded.

The set $S = W^u(x_0) \cap f^{-1}(k)$ is called a *section* of the unstable manifold. We have seen that $W_a^u(x_0)$ is compact, hence so is S. The set $S \times [0,\infty)/S \times \{0\}$ is called the *cone over the section* S. All sections are homeomorphic. If S' is another section, say for k', then $x \mapsto xt_x$ is a homeomorphism from S' to S.

Sections provide some insight on the structure of the unstable manifold of a stationary point for flows more general than those associated to Morse fields. The remark that all are topologically the same fits in the following result:

Proposition 1.10.3 *Let $\varphi : M \times \mathbb{R} \to M$ be a gradientlike flow in the metric space M with strict Lyapunov map $f : M \to \mathbb{R}$. Suppose that M_a^b is compact, that there is a unique stationary point x_0 in M_a^b and that $f(x_0) = c$ with $a < c < b$. Consider a section $S = W^u(x_0) \cap f^{-1}(k)$ of the unstable manifold. Then $W^u(x_0)$ is homeomorphic to the cone over S.*

Proof. We define the map $h : S \times \mathbb{R} \to W^u(x_0)\setminus\{x_0\}$ by $h(x,t) = xt$. It is easily seen that h is a homeomorphism with inverse $x \mapsto xt_x$ given by the t_x just described. Denote by $\theta : (0,\infty) \to \mathbb{R}$ the map $t \mapsto \log t$. Now define $\widehat{h} : S \times [0,\infty) \to W^u(x_0)$

by $\widehat{h}|S \times \{0\} \equiv x_0$ and $\widehat{h}(x,t) = h(x, \theta(t))$. This $\widehat{h}$ induces in the cone the homeomorphism we are looking for. ∎

Sections of $W^u(x_0)$ can be topologically identified in some important cases, in particular in the case of flows associated to Morse fields. We are to see that in this case, the sections are topologically spheres of dimensión $d-1$, where d is the index of x_0. Thus the cones are, by the last proposition, open d cells, what corroborates the Unstable Manifold Theorem.

Corollary 1.10.4 *Let $f : M \to \mathbb{R}$ be a Morse function on the compact m-manifold M. Suppose that f has a unique critical point x_0 in M_a^b with $a < f(x_0) < b$; let d be the index of x_0. Then:*

1. *If $F : M \to \mathbb{R}^p$ is a Morse field for f, the truncated manifold $W_a^u(x_0)$ of the flow of $x' = F(x)$ is a d-cell attached to M^a.*
2. *$H^j(M^b, M^a)$ is trivial for $j \neq d$, and $\mathbb{Z}$ for $j = d$.*

Proof. As we know, f is a strict Lyapunov map for the flow of $x' = F(x)$, whose stationary points are the critical points of f. Suppose, without loss of generality, that $f(x_0) = 0$. Since F is a Morse field for f, there are local coordinates x in an open neighborhood $U \subset M_a^b$ of x_0 such that

$$f(x) = x_1^2 + \cdots + x_{m-d}^2 - x_{m-d+1}^2 - \cdots - x_m^2.$$

In these coordinates x_0 is the origin, and we can reduce U so that $x(U) = B$ is an open ball $B = B(0, \delta)$. Let $0 < \varepsilon < \delta$. We claim that

$$W_{-\varepsilon}^u(x_0) = W_{-\varepsilon}^u(x_0) \cap U = \{x \in U : x_1 = \cdots = x_{m-d} = 0, -x_{m-d+1}^2 - \cdots - x_m^2 \geq -\varepsilon\},$$

which is a d-cell with boundary

$$W_{-\varepsilon}^u(x_0) \cap f^{-1}(-\varepsilon) = \{x \in U : x_{m-d+1}^2 + \cdots + x_m^2 = \varepsilon\}.$$

Indeed, the expression of f in our coordinates gives everything but the inclusion $W_{-\varepsilon}^u(x_0) \subset U$. Let $x \in W_{-\varepsilon}^u(x_0)$. Then $x_0 \in \alpha(x)$ and there is some point $y \in x(-\infty, 0] \cap U$. Then $x(-\infty, 0]$ contains the whole trajectory from y to x. If that trajectory reaches x in U we are done. If not, it reaches a point y^* with $f(y^*) = -\varepsilon$, which implies $f(x) < f(y^*) = -\varepsilon$, a contradiction. This proved, $M^{-\varepsilon} \cap W_{-\varepsilon}^u(x_0)$ is the boundary of that cell and $W_{-\varepsilon}^u(x_0)$ is a d-cell attached to $M^{-\varepsilon}$.

In particular we see that de section $W^u(x_0) \cap f^{-1}(-\varepsilon)$ is a $(d-1)$-sphere. Hence all sections are $(d-1)$-spheres, as was announced.

Now, the flow defines a homeomorphism $h : W^u_{-\varepsilon}(x_0) \to W^u_a(x_0)$ as follows.

As explained before, every $x \in W^u_{-\varepsilon}(x_0)$, $x \neq x_0$, has an exit time t_x with $f(xt_x) = -\varepsilon$ and another t'_x with $f(xt'_x) = a$, and these exit times are continuous functions. We use them to define

$$h : W^u_{-\varepsilon}(x_0) \to W^u_a(x_0) : x \mapsto h(x) = x(t'_x - t_x).$$

This brings back the fact that the unstable manifold is the cone over every section and the homeomorphisms between sections. In each trajectory h is the translation that takes the point in $f^{-1}(-\varepsilon)$ to the point in $f^{-1}(a)$, hence h is a bijection. We have seen this is continuous for $x \neq x_0$, and now we see it extends continuously by $h(x_0) = x_0$.

Indeed, note first that $t'_x - t_x$ is the exit time t'_y of the point $y = xt_x \in f^{-1}(-\varepsilon)$:

$$f(y(t'_x - t_x)) = f(xt_x(t'_x - t_x)) = f(xt'_x) = a,$$

and since the exit time t' is continuous on the compact level set $f^{-1}(-\varepsilon)$, it is bounded there. Now consider $x_n \to x_0$. By the above remark we may assume $t'_{x_n} - t_{x_n} \to t^*$, and then

$$h(x_n) = x_n(t'_{x_n} - t_{x_n}) \to x_0 t^* = x_0,$$

because x_0 is stationary. In the end, we have a continuous bijection $h : W^u_{-\varepsilon}(x_0) \to W^u_a(x_0)$ between compact sets, hence a homeomorphism.

Now, since there are no stationary points in $M^{-\varepsilon}_a$, the flow defines a homeomorphism $M^{-\varepsilon} \to M^a$ which coincides with h where needed to glue well (1.9.4). Thus we have a homeomorphism $\overline{h} : M^{-\varepsilon} \cup W^u_{-\varepsilon}(x_0) \to M^a \cup W^u_a(x_0)$ that shows $W^u_a(x_0)$ is a d-cell attached to M^a as says the first part of the statement.

On the other hand, since $i : M^a \cup W^u_a(x_0) \to M^b$ is a strict shape equivalence relative to M^a we have that

$$H^j(M^b, M^a) \cong H^j(M^a \cup W^u_a(x_0), M^a) \cong H^j(S^d, *),$$

where S^d is the d-sphere obtained by collapsing the boundary of the d-cell into a point $*$. Hence the second part of the statement follows. ∎

Remark 1.10.5 The preceding results hold true for finitely many stationary/critical points x_i at the same level $f^{-1}(c)$. The proof works the same dealing simultaneously with all the unstable manifolds $W^u_a(x_i)$, which are disjoint. Indeed, if

$x \in W_a^u(x_i) \cap W_a^u(x_j)$, then $x(-\infty, 0] \subset M_a^b$, which is a compact set where stationary points are isolated, hence $\alpha(x)$ is a singleton and $x_i = x_j$. Let us give a few hints more and the generalized statement.

(1) For Propositions 1.10.1 and 1.10.2 the formulation is that the inclusion $i : M^a \cup \bigcup_i W_a^u(x_i) \to M^b$ is a strict shape equivalence relative to M^a and induces isomorphisms in cohomology relative to any $K_0 \subset M^a$.

The only additional care here comes at the discussion of the homotopies $h_k \simeq \mathrm{Id}$, to see that the exit time exists if $x \in W_a^u(x_i)$ is not x_i. But if t_x does not exist, then $\omega(x) \subset M_a^b$, hence $\omega(x) = x_j$. We get $c = f(x_i) > f(x) > f(x_j) = c$, contradiction.

(2) For Corollary 1.10.4 let d_i be the index of the critical point x_i. Then each $W_a^u(x_i)$ is a d_i-cell attached to M^a, and $H^j(M^b, M^a) = \mathbb{Z}^{\mathsf{c}_j}$, where c_j is the number of critical points x_i with index $d_i = j$.

Here we choose disjoint neighborhhoods U_i of the critical points x_i with the same δ and ε for all of them. The computation of the cohomology groups is

$$H^j(M^b, M^a) \cong H^j\big(M^a \cup \bigcup_i W_a^u(x_i), M^a\big) \cong H^j\big(\bigvee_i S^{d_i}, *\big) \cong \mathbb{Z}^{\mathsf{c}_j},$$

where the d_i-sphere S^{d_i} comes from collapsing the boundary of the d_i-cell $W_a^u(x_i)$ to the wedge point $*$.

1.11 Morse polynomial and Morse inequalities

To start with we prove the formula relating Morse and Poincaré polynomials:

Proof of Theorem 1.8.4. Let $f : M \to \mathbb{R}$ be a Morse function on a compact manifold M, with critical values $c_1 < \cdots < c_k < \cdots < c_r$. Choose regular values d_i in between: $d_0 < c_1 < d_1 < \cdots < d_{k-1} < c_k < d_k < \cdots < d_{r-1} < c_r < d_r$. Note that M being compact c_1 must be its minimum and c_r its maximum. Then we have the filtration

$$\varnothing = M^{d_0} \subset M^{d_1} \subset \cdots \subset M^{d_r} = M,$$

where each M^{d_k} is a compact manifold with boundary $f^{-1}(d_k)$. The theorem is a standard consequence of the properties of cohomology applied to this filtration. We provide the argument now.

Consider for each $k = 1, \dots, r$ the long exact sequence of the pair $(M^{d_k}, M^{d_{k-1}})$:

$$\cdots \to H^{j-1}(M^{d_{k-1}}) \xrightarrow{\delta_{j-1}^k} H^j(M^{d_k}, M^{d_{k-1}}) \xrightarrow{\lambda_j^k} H^j(M^{d_k}) \xrightarrow{\varrho_j^k} H^j(M^{d_{k-1}}) \xrightarrow{\delta_j^k} \cdots$$

where the ends are $H^{-1}(M^{d_{k-1}}) = 0$ and $H^{m+1}(M^{d_k}, M^{d_{k-1}}) = 0$, hence $\delta^k_{-1} = 0$ and $\delta^k_m = 0$. We can compute the ranks of these groups from exactness as follows:

$$\begin{aligned} H^j(M^{d_k}, M^{d_{k-1}})\big/\operatorname{im}\delta^k_{j-1} &\cong \operatorname{im}\lambda^k_j \quad \text{hence} \quad \operatorname{rk} H^j(M^{d_k}, M^{d_{k-1}}) = \operatorname{rk}\operatorname{im}\delta^k_{j-1} + \operatorname{rk}\operatorname{im}\lambda^k_j, \\ H^j(M^{d_k})\big/\operatorname{im}\lambda^k_j &\cong \operatorname{im}\varrho^k_j \quad \text{hence} \quad \operatorname{rk} H^j(M^{d_k}) = \operatorname{rk}\operatorname{im}\lambda^k_j + \operatorname{rk}\operatorname{im}\varrho^k_j, \\ H^j(M^{d_{k-1}})\big/\operatorname{im}\varrho^k_j &\cong \operatorname{im}\delta^k_j \quad \text{hence} \quad \operatorname{rk} H^j(M^{d_{k-1}}) = \operatorname{rk}\operatorname{im}\varrho^k_j + \operatorname{rk}\operatorname{im}\delta^k_j. \end{aligned}$$

Now from the three equalities we obtain

$$\operatorname{rk} H^j(M^{d_k}, M^{d_{k-1}}) = \operatorname{rk} H^j(M^{d_k}) - \operatorname{rk} H^j(M^{d_{k-1}}) + (q^k_j + q^k_{j-1}),$$

where $q^k_\ell = \operatorname{rk}\operatorname{im}\delta^k_\ell \geq 0$, and in particular $q^k_{-1} = q^k_m = 0$.

Now we add everything:

$$\begin{aligned} \sum\nolimits_k \operatorname{rk} H^j(M^{d_k}, M^{d_{k-1}}) &= \sum\nolimits_k (\operatorname{rk} H^j(M^{d_k}) - \operatorname{rk} H^j(M^{d_{k-1}})) + \sum\nolimits_k (q^k_j + q^k_{j-1}) \\ &= \operatorname{rk} H^j(M^{d_r}) - \operatorname{rk} H^j(M^{d_0}) + \sum\nolimits_k (q^k_j + q^k_{j-1}) \\ &= \operatorname{rk} H^j(M) + \sum\nolimits_k (q^k_j + q^k_{j-1}). \end{aligned}$$

By 1.10.5(2) in the previous section, the first sum is the number $\mathtt{c}_j$ of critical points of f of index j, and $\beta_j = \operatorname{rk} H^j(M)$ is the j-th Betti number of M. Consequently

$$\begin{aligned} \mathtt{M}_f(t) - \mathtt{P}_M(t) &= \sum\nolimits_j (\mathtt{c}_j - \beta_j) t^j = \sum\nolimits_j \sum\nolimits_k (q^k_j + q^k_{j-1}) t^j \\ &= \sum\nolimits_k \Big(\sum\nolimits_j q^k_j t^j + \sum\nolimits_j q^k_{j-1} t^j \Big) = (1+t) Q(t), \end{aligned}$$

where $Q(t) = \sum_k \sum_j q^k_j t^j$ has coefficients ≥ 0. ∎

The following are some consequences of the theorem just proved. This is the main result in Morse's original presentation to the theory.

Corollary 1.11.1 (Morse inequalities) *Let $f : M \to \mathbb{R}$ be a Morse function on a compact m-manifold M. As usual, let $\mathtt{c}_j$ the number of critical points of index j of f (hence $\mathtt{c}(f) = \sum_j \mathtt{c}_j$ is the number of critical points), and β_j the j-th Betti number of M. Then*

1. *$\beta_j \leq \mathtt{c}_j$, hence $\sum_j \beta_j \leq \mathtt{c}(f)$.*
2. *The Euler characteristic of M is $\chi(M) = \sum_j (-1)^j \mathtt{c}_j$.*
3. *$\sum_{j=0}^r (-1)^{r-j} \beta_j \leq \sum_{j=0}^r (-1)^{r-j} \mathtt{c}_j$ for every $r \leq m$.*

Proof. The first assertion follows from the fact that the coefficients of $Q(t)$ are all ≥ 0. The second assertion was already mentioned: $\chi(M) = \mathtt{P}_M(-1) = \mathtt{M}_f(-1)$. For the third formula set $Q(t) = \sum q_j t^j$. Then

$$\sum_{j=0}^{r} \mathtt{c}_j t^j = \sum_{j=0}^{r} \beta_j t^j + q_r t^r + (1+t) \sum_{j=0}^{r-1} q_j t^j.$$

Multiplying by t^{-r} both members of the equality we get

$$\sum_{j=0}^{r} \mathtt{c}_j t^{j-r} = \sum_{j=0}^{j} \beta_j t^{j-r} + q_r + (1+t) \sum_{j=0}^{r-1} q_j t^{j-r}.$$

Evaluating at $t = -1$, since $q_k \geq 0$, we obtain the required inequality. ∎

A first immediate corollary:

Corollary 1.11.2 *The Euler characteristic of a compact manifold of odd dimension is zero.*

Proof. Let M be a compact manifold of odd dimension m. Consider a Morse function $f : M \to \mathbb{R}$. Then $-f$ is also a Morse function and $\mathtt{c}_j(-f) = \mathtt{c}_{m-j}(f)$. From this we get $\mathtt{M}_f(t) = t^m \mathtt{M}_{-f}(1/t)$, hence

$$\chi(M) = \mathtt{M}_f(-1) = (-1)^m \mathtt{M}_{-f}(-1) = -\chi(M).$$

Thus $\chi(M) = 0$. ∎

A second:

Corollary 1.11.3 *Any Morse function $f : S^n \to \mathbb{R}$ always has an even number of critical points.*

The notion of index of a critical point can be seen from different points of view. For instance, if x_0 is a critical point of the Morse function f then it is a singular point of the gradient field and also a singular point of every Morse field for f. If d is the index of this critical point and F is one of those vector fields then, traditionally, the number $(-1)^d$ is called the index of F at x_0 and denoted $\operatorname{ind}_{x_0}(F)$. Furthermore, the sum $\sum \operatorname{ind}_{x_i}(F)$, extended to all singularities x_i of F, is called the *total index of* F and denoted $\operatorname{Ind}(F)$. Another interpretation of the index d, as we have pointed out before, is dynamic in terms of the dimension of the unstable manifold of the flow of each Morse field for f or the flow of the field $-\operatorname{grad} f$. To illustrate the power of the Morse inequalities, we note that the equality 2 of Corollary 9.1 is nothing other than the Poincaré-Hopf theorem for the field F.

Theorem 1.11.4 (Poincaré-Hopf for gradient fields) *Let F be the gradient field of the Morse function $f : M \to \mathbb{R}$. Then* $\mathrm{Ind}(F) = \chi(M)$.

Another direct consequence of the Morse inequalities is the following:

Corollary 1.11.5 *Every Morse function on an orientable compact surface of genus g has at least $2g+2$ critical points.*

There are smooth functions on surfaces with less than $2g+2$ critical points. The following is an example. If we represent the torus as $T^2 = \mathbb{R}^2/\mathbb{Z}^2$, it is easy to see that the function $f(x,y) = \sin(\pi x)\sin(\pi y)\sin\pi(x+y)$ has exactly three critical points, the maximum, the minimum and a third degenerate critical point. Note that this is the minimum possible. Since the category of the torus T^2 is 3, according to the Lusternik-Schnirelmann Theorem every smooth function on T^2 must have at least three critical points.

Also from the inequalities:

Corollary 1.11.6 (Lacunary principle) *If the Morse polynomial $\mathtt{M}_f(t)$ has no consecutive powers of t, then $\mathtt{c}_j = \beta_j$ for every j.*

Proof. Suppose not, and let r be the first index with $\mathtt{c}_r \neq \beta_r$, that is $\mathtt{c}_r > \beta_r$ (by the third Morse inequality). Since the alternate sum of the $\mathtt{c}_j$'s is the Euler characteristic (the second Morse inequality, which is an equality), necessarily $r < m$. Then $\mathtt{c}_{r+1} = 0$ and applying the third Morse inequality we get:

$$-\mathtt{c}_r + \beta_r - \beta_{r+1} = \sum_{j=0}^{r+1} (-1)^{r+1-j}(\mathtt{c}_j - \beta_j) \geq 0,$$

so that $\mathtt{c}_r \leq \beta_r - \beta_{r+1} \leq \beta_r$, a contradiction. ∎

For instance, the Morse polynomial of the complex projective space given in Example 1.8.2 has no consecutive powers, hence the non-zero Betti numbers of $P_n(\mathbb{C})$ are $\beta_{2i} = 1$ for $i = 1, \ldots, n$.

Corollary 1.11.6 could be seen as a consequence of the following related result whose proof we leave as an exercice.

Corollary 1.11.7 *If $c_{j-1} = c_{j+1} = 0$ then $c_j = \beta_j$.*

1.12 Homotopy type associated to Morse functions

Now we would like to know a little more about the relation between the topological structure of a manifold M and the Morse function $f : M \to \mathbb{R}$. We present an important result stating that M has, in fact, the homotopy type of a CW-complex with a particular structure determined by the Morse function. To prove it, we need three lemmas. The first of them improves Proposition 1.10.1 and Remark 1.10.5 in the case that M is a compact manifold.

Lemma 1.12.1 *Let f be a Morse function on the manifold M and consider the flow of a Morse field of f. Suppose that $a < b$ and that there is a unique critical value c with $a < c < b$. Let $x_1, \ldots, x_l$ be all the critical points contained in $f^{-1}(c)$. Then the inclusion $i : M^a \cup W_a^u(x_1) \cup \cdots \cup W_a^u(x_l) \to M^b$ is a homotopy equivalence.*

Proof. Set $W_a^u = W_a^u(x_1) \cup \cdots \cup W_a^u(x_l)$. There exists a retraction $r : U \to M^a \cup W_a^u$, where U is a neighborhood of $M^a \cup W_a^u$ in M^b, such that $i \circ r$ is homotopic to the inclusion $j : U \to M^b$. This is a consequence of the fact that $M^a \cup W_a^u$ and M^b are Euclidean Neighborhood Retracts.

Since $i : M^a \cup W_a^u \to M^b$ is a strict shape equivalence relative to M^a by Proposition 1.10.1 and Remark 1.10.5(1), there is a map $h_k : M^b \to U$ such that: (i) $h_k | M^a$ is the identity Id_{M^a}, (ii) h_k restricts to a map $M^a \cup W_a^u \to M^a \cup W_a^u$ homotopic to the identity $\mathrm{Id}_{M^a \cup W_a^u}$ relative to M^a, and (iii) $j \circ h_k$ is homotopic to the identity Id_{M^b}.

Then the composition $r \circ h_k$ is a homotopy equivalence with inverse i. ∎

The following two technical but classical lemmas refer to the attachment of cells to a topological space.

Lemma 1.12.2 (Whitehead) *Let X be a topological space and suppose that h_0 and h_1 are homotopic maps from the sphere $\dot{e}^k$ to X. Then the identity map of X extends to a homotopy equivalence $f : X \cup_{h_0} e^k \to X \cup_{h_1} e^k$.*

Proof. We denote by h_t a homotopy between h_0 and h_1 and by $\hat{h}_0 : e^k \to X \cup_{h_0} e^k$ and $\hat{h}_1 : e^k \to X \cup_{h_1} e^k$ the characteristic maps corresponding to h_0 and h_1 respectively. Now define a mapping $f : X \cup_{h_0} e^k \to X \cup_{h_1} e^k$ by

$$\begin{cases} f(x) = x & \text{for } x \in X, \\ f(\hat{h}_0(su)) = \hat{h}_1(2su) & \text{for } 0 \leq s \leq \frac{1}{2},\ u \in \dot{e}^k, \\ f(\hat{h}_0(su)) = h_{2-2s}(u) & \text{for } \frac{1}{2} \leq s \leq 1,\ u \in \dot{e}^k, \end{cases}$$

and another $g : X \cup_{h_1} e^k \to X \cup_{h_0} e^k$ by

$$\begin{cases} g(x) = x & \text{if } x \in X, \\ g(\hat{h}_1(su)) = \hat{h}_0(2su) & \text{for } 0 \le s \le \frac{1}{2},\ u \in \dot{e}^k, \\ g(\hat{h}_1(su)) = h_{2s-1}(u) & \text{for } \frac{1}{2} \le s \le 1,\ u \in \dot{e}^k. \end{cases}$$

It is easy to see that f and g are well defined and continuous, and that

$$(g \circ f)(\hat{h}_0(su)) = \begin{cases} \hat{h}_0(4su) & \text{for } 0 \le s \le \frac{1}{4},\ u \in \dot{e}^k, \\ g(\hat{h}_1(2su) = h_{4s-1}(u) & \text{for } \frac{1}{4} \le s \le \frac{1}{2},\ u \in \dot{e}^k, \\ g(h_{2-2s}(u)) = h_{2-2s}(u) & \text{for } \frac{1}{2} \le s \le 1,\ u \in \dot{e}^k. \end{cases}$$

Now we define a homotopy $\theta_t : X \cup_{h_0} e^k \to X \cup_{h_0} e^k$ by

$$\begin{cases} \theta_t(x) = x & \text{for } x \in X, \\ \theta_t(\hat{h}_0(su)) = \hat{h}_0((4-3t)su) & \text{for } 0 \le s \le \frac{1}{4-3t},\ u \in \dot{e}^k, \\ \theta_t(\hat{h}_0(su)) = h_{(4-3t)s-1}(u) & \text{for } \frac{1}{4-3t} \le s \le \frac{2-t}{4-3t},\ u \in \dot{e}^k, \\ \theta_t(\hat{h}_0(su)) = h_{\frac{1}{2}(4-3t)(1-s)}(u) & \text{for } \frac{2-t}{4-3t} \le s \le 1,\ u \in \dot{e}^k. \end{cases}$$

It is easy to verify that θ_t is a well defined homotopy connecting $g \circ f$ and the identity on $X \cup_{h_0} e^k$. A similar homopy can be defined connecting $f \circ g$ and the identity on $X \cup_{h_1} e^k$. Hence both f and g are mutually inverse homotopy equivalences. ∎

Lemma 1.12.3 (Hilton) *Let X and Y be topological spaces and let $h : \dot{e}^k \to X$ be an attaching map. Then any homotopy equivalence $f : X \to Y$ extends to a homotopy equivalence $F : X \cup_h e^k \to Y \cup_{fh} e^k$.*

Proof. The map F is defined in a natural way: $F|X = f$ and $F|e^k$ is the identity (we use a less detailed notation here than in the previous lemma). We consider a homotopy inverse $g : Y \to X$ of the map f and define, in a similar way, $G : Y \cup_{fh} e^k \to X \cup_{gfh} e^k$ by $G|Y = g$ and $G|e^k =$ identity.

Consider a homotopy ϕ_t between gf and the identity, and then $\phi_t \circ h$ is a homotopy between gfh and h. By using the same specific construction as in the Whitehead Lemma we obtain a homotopy equivalence $H : X \cup_{gfh} e^k \to X \cup_h e^k$ which extends the identity of X. We show that $HGF : X \cup_h e^k \to X \cup_h e^k$ is homotopic to the identity. We remark first that

$$\begin{cases} HGF(x) = gf(x) & \text{for } x \in X, \\ HGF(su) = 2su & \text{for } 0 \le s \le \frac{1}{2},\ u \in \dot{e}^k, \\ HGF(su) = (\phi_{2-2s} \circ h)(u) & \text{for } \frac{1}{2} \le s \le 1,\ u \in \dot{e}^k. \end{cases}$$

Then the sought homotopy $\psi_t : X \cup_h e^k \to X \cup_h e^k$ is

$$\begin{cases} \psi_t(x) = \phi_t(x) & \text{for } x \in X, \\ \psi_t(su) = \frac{2}{1+t} su & \text{for } 0 \leq s \leq \frac{1+t}{2}, u \in \dot{e}^k, \\ \psi_t(su) = (\phi_{2-2s+t} \circ h)(u) & \text{for } \frac{1+t}{2} \leq s \leq 1, u \in \dot{e}^k. \end{cases}$$

As a consequence, F has a left homotopy inverse F^{-1} and a similar proof shows that G has a left homotopy inverse G^{-1}. Also, we know that H has a homotopy inverse H^{-1}. Now we use carefully these inverses to have the following sequence of homotopies starting at ψ_t:

$$HGF \simeq 1 \xrightarrow{H^{-1}\circ} GF \simeq H^{-1} \xrightarrow{\circ H} GFH \simeq 1 \xrightarrow{G^{-1}\circ} FH \simeq G^{-1} \xrightarrow{\circ G} FHG \simeq 1.$$

Thus F has also a right homotopy inverse. It is a standard fact in homotopy theory that a mapping with homotopy inverses on both sides is a homotopy equivalence. This completes the proof. ∎

Proposition 1.12.4 *Let $f : M \to \mathbb{R}$ be a Morse function on a compact manifold M. Then M has the homotopy type of a CW-complex with one cell of dimension d for each critical point of index d.*

Proof. Let $c_1 < \cdots < c_r$ be the critical values of f. Then c_1 is the minimum of f, and $M^{c_1} = f^{-1}(c_1)$ are finitely many critical points with index 0. Let $c_1 < b \leq c_2$. By Remark 1.10.5 and Lemma 1.12.1, M^b has the homotopy type of finitely points (the critical points in $f^{-1}(c_1)$). Suppose now, for the purpose of induction, that $c_{i-1} < b < c_i$ and that M^b has the homotopy type of a CW-complex K. Then, by the same Remark and Lemma, for every b' with $c_i < b' < c_{i+1}$ the set $M^{b'}$ has the homotopy type of $M^b \cup_{\phi_1} e^{d_1} \cup \cdots \cup_{\phi_j} e^{d_j}$, where the e^{d_k}'s are cells of dimension d_k corresponding to the indexes of the critical points in $f^{-1}(c_i)$ and the ϕ_k's are attaching maps. Consider now a homotopy equivalence $\psi : M^b \to K$. By the cellular approximation theorem, the map $\psi \circ \phi_k : \dot{e}^{d_k} \to K$ is homotopic to a map $h_k : \dot{e}^{d_k} \to (d_k - 1)$-skeleton of K. Thus $K \cup_{h_1} e^{d_1} \cup \cdots \cup_{h_j} e^{d_j}$ is a CW-complex and has the same homotopy type as $M^{b'}$ by the Hilton Lemma. By induction it follows that M has the homotopy type of a CW-complex with the required properties. ∎

Notes

Morse published in 1925 his first paper [M1] on the subject of critical points of a function. The paper contains, among other things, the Morse inequalities and was inspired by the minimax principle of Birkhoff. According to Raoul Bott [Bt1], *"through his teacher, G.D. Birkhoff, Morse had inherited the dynamical tradition —in no uncertain terms— and it was in this framework that he understood Analysis Situs."* In fact, Morse considered himself a mathematical descendant of Poincaré. He published more than 50 papers on this subject and found fundamental applications to the calculus of variations [M1, M2, M3]. For example, he showed that for any Riemann structure on S^n there must be an infinite number of geodesics joining any two fixed points. Smale [Sm1, Sm2, Sm3] used Morse theory in a substantial way in his proofs of the Poincaré conjecture in dimensions greater than 4 and of the h-cobordim theorem and he wrote on the occasion of Morse's death in 1977, *"I believe that Morse theory is the single greatest contribution of American mathematics."* Other important contributions of Smale regarding the applications of Morse theory to Dynamics can be found in [Sm5]. See also [Flo], [IzSt], [KaRo], [Sa], [Sj2], [SaZe] for related subjects.

The Lusternik-Schnirelmann theorem was proved with a view to its applications in the calculus of variations in the large. The original results are contained in [Lu], [LuSch1, LuSch2] and [Sch], see also [B1], [F] and [J] for useful information. As mentioned in [CLOT], *"the basic idea in the Lusternik-Schnirelmann approach to critical point theory is that critical points are obstructions to collapsing a manifold down to a point via the flow associated to the the gradient."* According to these authors, *"critical points are the places in the manifold where topological complexity arises and this particular complexity is well measured by category."* Obviously, this is also the general philosophy of Morse Theory. For some other relationships between the Lusternik-Schnirelmann category and Morse theory and other subjects see [CoMa], [GGM], [LiWa], [Sj1] and [T].

Gradient (or more generally gradientlike) flows have been an important source of inspiration in the theory of dynamical systems (see, for instance, [Sm4, Sm5]). Perhaps the definitive result in this regard has been found by C. Conley. He proved that (we use his words) *"each flow dominates a unique gradient flow which in turn dominates any other gradient flow dominated by the original flow."* This result is a consequence of the fact that every flow φ on a compact metric space has a *gradient part* and a *recurrent part*. If every component of the recurrent part is identified to a point then we obtain the maximal gradient flow dominated by φ. This result, often referred to as the *Fundamental Theorem of the theory of Dynamical Systems,* is proved in Conley's paper [C], *The gradient structure of a flow.*

Finally, we recommend the books and papers [AuDa], [BaHu], [B2], [Bt2], [Jo], [Lau], [Mat], [Maz], [Mi], [N], [Schw] as excellent sources of information for the material presented in this chapter.

Bibliography

[AuDa] M. Audin, M. Damian: *Morse Theory and Floer Homology.* Springer, London-Heidelberg-New York-Dordrecht 2014. 58

[BaHu] A. Banyaga, D. Hurtubise: *Lectures on Morse homology.* Kluwer Academic Publishers, Dordrech-Boston-London 2004 58

[B1] K. Borsuk: *Über den Lusternik-Schnirelmannschen Begriff der Kategorie.* Math. 26 (1936) 123–136. 57

[B2] K. Borsuk: *Theory of Shape.* Monografie Matematyczne 59, Polish Scientific Publishers, Warsaw 1975. 43, 58

[Bt1] R. Bott: *Marston Morse and his mathematical works.* Bull. AMS vol 3, no 3 (1980) 907–950. 57

[Bt2] R. Bott: *Lectures on Morse theory, old and new.* Bull. AMS vol 7, no 2 (1982), 331–358. 58

[Br] G.E. Bredon: *Sheaf theory.* GTM 170, Springer, New York-Berlin-Heidelberg 1997. 46

[CoMa] H. Colman, E. Macias-Virgós: *Transverse Lusternik-Schnirelmann category of foliated manifolds.* Topology 40 (2001) 419–430. 57

[C] C. Conley: *The gradient structure of a flow.* Ergodic Theory and Dynamical Systems, vol 8, Charles Conley Memorial Issue (1988) 11–26. 58

[CLOT] O. Cornea, G. Lupton, J. Oprea, D. Tanré: *Lusternik-Schnirelmann Category.* Mathematical Surveys and Monographs, vol 103, AMS, Providence, RI, 2003. 57

[DySe] J. Dydak, J. Segal: *Shape Theory. An Introduction.* Lecture Notes in Mathematics 688, Springer, Berlin 1978. 43

[Fle] W. Fleming: *Functions of several variables*. Springer-Verlag, New York-Heidelberg-Berlin 1977. 3

[Flo] A. Floer: *Witten's complex and infinite dimensional Morse theory*. J. Differ. Geom. 30 (1989) 207–221. 57

[F] R.H. Fox: *On the Lusternik-Schnirelmann Category*. Ann. of Math. Second Series, vol 42, no 2 (1941) 333–370. 57

[GGM] J.M. García-Calcines, P.R. García-Díaz, A. Murillo Mas: *A Whitehead–Ganea approach for proper Lusternik–Schnirelmann category*. Math. Proc. Cambridge Philos. Soc. 142 (2007) no. 3, 439–457. 57

[G] R. Godement: *Théorie des faisceaux*. Act. Scient. Industr. 1252, Hermann, Paris 1964. 46

[IzSt] M. Izydorek, M. Styborski: *Morse inequalities via Conley index theory*. Topological methods in nonlinear analysis, 37–60, Lect. Notes Nonlinear Anal. 12, Juliusz Schauder Cent. Nonlinear Stud. 2011. 57

[J] I.M. James: *On category in the sense of Lusternik-Schnirelmann*. Topology 17 (1978) 331–348. 57

[Jo] J. Jost: *Riemannian Geometry and Geometric Analysis*. Springer-Verlag, Berlin-Heidelberg 1995. 58

[KaRo] L. Kapitanski, I. Rodnianski: *Shape and Morse theory of attractors*. Comm. Pure Appl. Math. 53 (2000) no. 2, 218–242. 57

[Lau] F. Laudenbach: *Transversalité, Courants et Théorie de Morse*. Éditions de l'École Polytechnique, 2011. 58

[Lu] L. Lusternik: *Topologische Grundlagen der allgemeinen Eigenwerttheorie*. Monatshefte für Math. u. Physik 37 (1930, 125–130. 57

[LuSch1] L. Lusternik, L. Schnirelmann: *Sur un principe topologique en analyse*. Comptes Rend. Acad. Sc. Paris 188 (1929), 295–297. 57

[LuSch2] L. Lusternik, L. Schnirelmann: *Méthodes topologiques dans les problèmes variationnels*. Inst. for Math. and Mechanics (1930) Moscow (in Russian), Actualités Scientifiques et Industrielles (1934) Paris. 57

[LiWa] D. Li, J. Wang: *On relative category and Morse decompositions for infinite-dimensional dynamical systems*. Topology Appl 291 (2021) Paper No. 107624, 16 pp. 57

[Mat] Y. Matsumoto: *An introduction to Morse theory*. Translated from the 1997 Japanese original by Kiki Hudson and Masahico Saito. Translations of Mathematical Monographs, 208. Iwanami Series in Modern Mathematics. AMS, Providence, RI, 2002. xiv+219 pp. 58

[Maz] B. Mazur: *Morse theory*. Differential and combinatorial topology, A symposium in honor of Marston Morse (ed. by S. S. Cairns), Princeton Univ. Press, 1965, 145–165. 58

[Mi] J. Milnor: *Morse theory*. Princeton Univ. Press, Princeton, NJ, 1963. 58

[M1] M. Morse: *Relations between the critical points of a real function of n independent variables*. Trans. AMS 27 (1925) 345–396. 57

[M2] M. Morse: *Calculus of variations in the large*. AMS Colloquium Publ., no. 18, Providence, RI, 1934. 57

[M3] M. Morse: *Variational analysis: Critical extremals and Sturmian extensions*. Wiley, New York 1973. 57

[MaSe] S. Mardešić, J. Segal: *Shape theory. The Inverse System Approach*. North-Holland Mathematical Library, 26. North-Holland Publishing Co, Amsterdam-New York 1982. 43

[N] L. Nicolaescu: *An invitation to Morse theory*. Universitext, Springer, New York 2007, xiv+241 pp. 58

[ORRz] E. Outerelo, J.A. Rojo, J.M. Ruiz: *Topología diferencial*. Sanz y Torres, Madrid 2014. 30

[Sa] D. Salamon: *Morse theory, the Conley index and Floer homology*. Bull. London Math. Soc. 22 (1990) 113–140. 57

[SaZe] D. Salamon, E. Zehnder: *Morse theory for periodic solutions of Hamiltonian systems and the Maslov index*. Comm. Pure Appl. Math. 45 (1992) 1303–1360. 57

[Sj1] J.M.R. Sanjurjo: *Lusternik-Schnirelmann category and Morse decompositions*. Mathematika 47 (1-2) (2000), 299–305. 57

[Sj2] J.M.R. Sanjurjo: *Morse equations and unstable manifolds of isolated invariant sets*. Nonlinearity 16 (2003) no. 4, 1435–1448. 57

[SjRz] J.M.R. Sanjurjo, J.M. Ruiz: *Introducción a la Geometría Diferencial I: Curvas*. Sanz y Torres, Madrid 2012. 30

[Sch] L. Schnirelmann: *Über eine neue kombinatorische Invariante*. Monatshefte fur Math. u. Physik 37 (1930) 131–134. 57

[Schw] M. Schwarz: *Morse homology*. Birkhauser Verlag, Basel 1993. 58

[Sm1] S. Smale: *Generalized Poincaré conjecture in higher dimensions*. Bull. AMS 66 (1960) 373–375. 57

[Sm2] S. Smale: *Generalized Poincare's conjecture in dimensions greater than four*. Ann. of Math. 74 (1961) 391–406. 57

[Sm3] S. Smale: *Differentiable dynamical systems*. Bull. AMS 69 (1963) 131–145. 57

[Sm4] S. Smale: *On gradient dynamical systems*. Ann. of Math. (2) 74 (1961) 199–206. 57

[Sm5] S. Smale: *Morse inequalities for a dynamical system*. Bull. AMS 66 (1960) 43–49. 57

[T] F. Takens: *The minimal number of critical points of a function on a compact manifold and the Lusternik-Schnirelman category*. Invent. Math. 6 (1968) 197–244. 57