Chapter 2
The Topology of Morse Functions

The present chapter is the heart of Morse *theory*, which is based on two fundamental principles. The "weak" Morse principle states that as long as the real parameter t varies in an interval containing only regular values of a smooth function $f : M \to \mathbb{R}$, the topology of the sublevel set $\{f \leq t\}$ is independent of t. We can turn this on its head and state that a change in the topology of $\{f \leq t\}$ is an indicator of the presence of a critical point.

The "strong" Morse principle describes precisely the changes in the topology of $\{f \leq t\}$ as t crosses a critical value of f. These changes are known in geometric topology as *surgery operations* or *handle attachments*.

The surgery operations are more subtle than they first appear, and we thought it wise to devote an entire section to this topic. It will give the reader a glimpse at the potential "zoo" of smooth manifolds that can be obtained by an iterated application of these operations.

2.1 Surgery, Handle Attachment, and Cobordisms

To formulate the central results of the Morse theory, we need to introduce some topological terminology. Denote by $\mathbb{D}^k$ the k-dimensional, *closed* unit disk and by $\dot{\mathbb{D}}^k$ its interior. We will refer to $\mathbb{D}^k$ as the *standard k-cell*. The cell attachment technique is one of the most versatile methods of producing new topological spaces out of existing ones.

Given a topological space X and a continuous map $\varphi : \partial\mathbb{D}^k \to X$, we can attach a k-cell to X to form the topological space $X \cup_\varphi \mathbb{D}^k$. The compact spaces obtained by attaching finitely many cells to a point are homotopy equivalent to finite CW-complexes. We would like to describe a related operation in the more restricted category of *smooth* manifolds.

L. Nicolaescu, *An Invitation to Morse Theory*, Universitext,
DOI 10.1007/978-1-4614-1105-5_2,

We begin with the operation of *surgery*. Suppose that M is a smooth m-dimensional manifold. The operation of surgery requires several additional data:

- An embedding $S \hookrightarrow M$ of the standard k-dimensional sphere S^k, $k < m$, with trivializable normal bundle $T_S M$
- A *framing* of the normal bundle $T_S M$, i.e., a bundle isomorphism

$$\varphi : T_S M \to \underline{\mathbb{R}}_S^{m-k} = \mathbb{R}^{m-k} \times S.$$

Equivalently, a framing of S defines an isotopy class of embeddings

$$\varphi : \mathbb{D}^{m-k} \times S^k \to M \;\text{ such that }\; \varphi(\{0\} \times S^k) = S.$$

Set $U := \varphi(\dot{\mathbb{D}}^{m-k} \times S^k)$. Then, U is a tubular neighborhood of S in M. We can now define a new *topological* manifold $M(S, \varphi)$ by removing U and then gluing instead $\hat{U} = S^{m-k-1} \times \mathbb{D}^{k+1}$ along $\partial U = \partial(M \setminus U)$ via the identifications

$$\partial \hat{U} \xrightarrow{\varphi} \partial U = \partial(M \setminus U).$$

For every $e_0 \in \partial\mathbb{D}^{m-k} = S^{m-k-1}$, the sphere $\varphi(e_0 \times S^k) \subset M$ will bound the disk $e_0 \times \mathbb{D}^{k+1}$ in $M(S, \varphi)$. Note that $e_0 \times S^k$ can be regarded as the graph of a section of the trivial bundle $\mathbb{D}^{m-k} \times S^k \to S^k$.

To see that $M(S, \varphi)$ is indeed a smooth manifold, we observe that

$$U \setminus S \cong (\dot{\mathbb{D}}^{m-k} \setminus 0) \times S^k.$$

Using spherical coordinates, we obtain diffeomorphisms

$$(\dot{\mathbb{D}}^{m-k} \setminus 0) \times S^k \cong (0,1) \times S^{m-k-1} \times S^k,$$
$$S^{m-k-1} \times (0,1) \times S^k \cong S^{m-k-1} \times (\dot{\mathbb{D}}^{k+1} \setminus 0).$$

Now, attach $(S^{m-k-1} \times \mathbb{D}^{k+1})$ to U along $U \setminus S$ using the obvious diffeomorphism

$$(0,1) \times S^{m-k-1} \times S^k \to S^{m-k-1} \times (0,1) \times S^k.$$

The diffeomorphism type of $M(S, \varphi)$ depends on the isotopy class of the embedding $S \hookrightarrow M$ and on the regular homotopy class of the framing φ. We say that $M(S, \varphi)$ is obtained from M by a surgery of type (S, φ).

Example 2.1 (Zero-dimensional surgery). Suppose M is a smooth m-dimensional manifold consisting of two connected components $M_\pm$. A zero-dimensional sphere S^0 consists of two points $p_\pm$. Fix an embedding $S^0 \hookrightarrow M$ such that $p_\pm \in M_\pm$. Fix open neighborhoods $U_\pm$ of $p_\pm \in M_\pm$ diffeomorphic to $\dot{\mathbb{D}}^m$ and set $U = U_- \cup U_+$. Then,

$$\partial(M \setminus U) \cong \partial U_- \cup \partial U_+ \cong S^0 \times S^{m-1}.$$

If we now glue $\mathbb{D}^1 \times S^{m-1} = [-1, 1] \times S^{m-1}$ such that $\{\pm 1\} \times S^{m-1}$ is identified with $\partial U_{\pm}$, we deduce that the surgery of $M_- \cup M_+$ along the zero sphere $\{p_{\pm}\}$ is diffeomorphic to the connected sum $M_- \# M_+$. Equivalently, we identify $(-1, 0) \times S^{m-1} \subset \mathbb{D}^1 \times S^{m-1}$ with the punctured neighborhood $U_- \setminus \{p_-\}$ (so that for $s \in (-1, 0)$ the parameter $-s$ is the radial distance in U_-) and then identify $(0, 1) \times S^{m-1}$ with the punctured neighborhood $U_+ \setminus \{p_+\}$ (so that $s \in (0, 1)$ represents the radial distance). □

Example 2.2 (Codimension two surgery). Suppose M^m is a compact, *oriented* smooth manifold $m \geq 3$ and $i : S^{m-2} \hookrightarrow M$ is an embedding of a $(m-2)$-sphere with trivializable normal bundle. Set $S = i(S^{m-2})$. The natural orientation on S^{m-2} (as boundary of the unit disk in $\mathbb{R}^{m-1}$) induces an orientation on S. We have a short exact sequence

$$0 \to TS \to TM|_S \to T_S M \to 0$$

of vector bundles over S.

The orientation on S together with the orientation on M induce via the above sequence an orientation on the normal bundle $T_S M$. Fix a metric on this bundle and denote by $\mathbb{D}_S M$ the associated unit disk bundle. Since the normal bundle has rank two, the orientation on $T_S M$ makes it possible to speak of *counterclockwise* rotations in each fiber. A trivialization is then uniquely determined by a choice of section

$$\mathbf{e} : S \to \partial \mathbb{D}_S M.$$

Given such a section $\mathbf{e}$, we obtain a positively oriented orthonormal frame $(\mathbf{e}, \mathbf{f})$ of $T_S M$, where $\mathbf{f}$ is obtained from $\mathbf{e}$ by a $\pi/2$ counterclockwise rotation. In particular, we obtain an embedding

$$\varphi_{\mathbf{e}} : \mathbb{D}^2 \times S^{m-2} \cong \mathbb{D}_S M \hookrightarrow M.$$

Once we fix such a section $\mathbf{e}_0 : S \to \partial \mathbb{D}_S M$, we obtain a trivialization

$$\partial \mathbb{D}_S M \cong S^1 \times S,$$

and then any other framing is described by a smooth map $S^{m-2} \to S^1$. We see that the homotopy classes of framings are classified by $\pi_{m-2}(S^1)$. In particular, this shows that the choice of framing becomes relevant only when $m = 3$.

The surgery on the framed sphere $(S, \mathbf{e}_0)$ has the effect of removing a tubular neighborhood $U \cong \varphi_{\mathbf{e}_0}(\mathbb{D}^2 \times S^{m-2})$ and replacing it with the manifold $\hat{U} = S^1 \times \mathbb{D}^{m-1}$, which has identical boundary.

The section $\mathbf{e}_0$ of $\partial \mathbb{D}_S \to S$ traces a submanifold $L_0 \subset \partial \mathbb{D}_S M$ diffeomorphic to S^{m-2}. Via the trivialization $\varphi_{\mathbf{e}_0}$, it traces a sphere $\varphi_{\mathbf{e}_0}(L_0) \subset \partial U$ called the *attaching sphere* of the surgery. After the surgery, this attaching sphere will bound the disk $\{1\} \times \mathbb{D}^{m-1} \subset \hat{U}$. □

Example 2.3 (Surgery on knots in S^3). Suppose that $M = S^3$ and that K is a smooth embedding of a circle S^1 in S^3. Such embeddings are commonly referred to as *knots*.

A classical result of Seifert (see [Rolf, 5.A]) states that any such knot bounds an orientable Riemann surface X smoothly embedded in S^3. The interior-pointing unit normal along $\partial X = K$ defines a nowhere vanishing section of the normal bundle $T_K S^3$ and thus defines a framing of this bundle. This is known as the *canonical framing*[1] of the knot. It defines a diffeomorphism between a tubular neighborhood U of the knot and the solid torus $\mathbb{D}^2 \times S^1$.

The canonical framing traces the curve

$$\ell = \ell_K = \{1\} \times S^1 \subset \partial\mathbb{D}^2 \times S^1.$$

The curve ℓ is called the *longitude* of the knot, while the boundary $\partial\mathbb{D}^2 \times \{1\}$ of a fiber of the normal disk bundle defines a curve called the *meridian* of the knot and is denoted by $\mu = \mu_K$.

Any other framing of the normal bundle will trace a curve φ on $\partial U \cong \partial\mathbb{D}^2 \times S^1$ isotopic inside U to the axis $K = \{0\} \times S^1$ of the solid torus U. Thus in $H_1(\partial\mathbb{D}_2 \times S^1, \mathbb{Z})$, it has the form

$$[\varphi] = p[\mu] + [\ell],$$

where the integer p is the winding number of φ in the meridional plane $\mathbb{D}^2$. The curve φ is called the *attaching curve* of the surgery.

The integer p completely determines the isotopy class of φ. Thus, every surgery on a knot in S^3 is uniquely determined by an integer p called the *coefficient of the surgery*, and the surgery with this framing coefficient will be called *p-surgery*. We denote by $S^3(K, p)$ the result of a p-surgery on the knot K.

The attaching curve of the surgery φ is a *parallel* of the knot K. By definition, a parallel of K is a knot K' located in a thin tubular neighborhood of K with the property that the radial projection onto K defines a homeomorphism $K' \to K$. Conversely, every parallel K' of the knot K can be viewed as the attaching curve of a surgery. The surgery coefficient is then the *linking number* of K and K', denoted by $\mathbf{lk}(K, K')$.

When we perform a p-surgery on K, we remove the solid torus $U = \mathbb{D}^2 \times S^1$ and we replace it with a new solid torus $\hat{U} = S^1 \times \mathbb{D}^2$, so that in the new manifold the attaching curve $K_p = \ell + p\mu$ will bound the disk $\{1\} \times \mathbb{D}^2 \subset \hat{U}$.

Let us look at a very simple yet fundamental example. Think of S^3 as the round sphere

$$\big\{(z_0, z_1) \in \mathbb{C}^2;\ \ |z_0|^2 + |z_1|^2 = 2\big\}.$$

[1] Its homotopy class is indeed independent of the choice of the Seifert surface X.

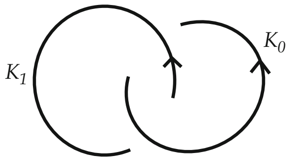

Fig. 2.1 The Hopf link

Consider the closed subsets $U_i = \{(z_0, z_1) \in S^3;\ \ |z_i| \leq 1\}, i = 0, 1$. Observe that U_0 is a solid torus via the diffeomorphism

$$U_0 \ni (z_0, z_1) \mapsto \left(z_0, \frac{z_1}{|z_1|}\right) \in \mathbb{D}^2 \times S^1.$$

Denote by K_i the knot in S^3 defined by $z_i = 0$. For example, K_0 admits the parametrization

$$[0, 1] \ni t \mapsto \left(0, \sqrt{2}e^{2\pi \mathbf{i} t}\right) \in S^3.$$

The knots K_0, K_1 are disjoint and form the *Hopf link*. Both are unknotted (see Fig. 2.1).

For example, K_0 bounds the embedded 2-disk

$$X_0 := \left\{\zeta \in \mathbb{C};\ \ |\zeta|^2 \leq 2\right\} \hookrightarrow \left\{(z_0, z_1) = \left(\sqrt{2 - |\zeta|^2}, \zeta\right), \in S^3\right\}.$$

Observe that U_0 is a tubular neighborhood of K_0, and the above isomorphism identifies it with the trivial 2-disk bundle, thus defining a framing of K_0. This framing is the canonical framing of U_0. The longitude of this framing is the curve

$$\ell_0 = \partial U_0 \cap X_0 = \left\{\left(1, e^{2\pi \mathbf{i} t}\right);\ \ t \in [0, 1]\right\}.$$

The meridian of K_0 is the curve $z_0 = e^{2\pi \mathbf{i} t}, z_1 = 1, t \in [0, 1]$. Via the diffeomorphism

$$U_1 \to \mathbb{D}^2 \times S^1,\ \ U_1 \ni (z_0, z_1) \mapsto \left(z_1, \frac{1}{|z_0|} z_0\right) \in \mathbb{D}^2 \times S^1,$$

this curve can be identified with the *meridian* μ_1 of K_1.

Set $M_p := S^3(K, p)$. The manifold M_p is obtained by removing U_0 from S^3 and gluing back a solid torus $\hat{U}_0 = S^1 \times \mathbb{D}^2$ to the complement of U_0, which is the solid torus U_1, so that,

$$\partial \hat{U}_0 \supset \hat{\mu}_0 = \{1\} \times \partial \mathbb{D}^2 \longmapsto p[\mu_0] + [\ell_0] = p[\mu_0] + [\mu_1].$$

For $p = 0$, we see that the disk $\{1\} \times \partial \mathbb{D}^2 \in S^1 \times \mathbb{D}^2 = \hat{U}_0$ bounds a disk in $\hat{U}_0$ and a meridional disk in U_1. The result of zero surgery on the unknot will then be $S^1 \times S^2$.

If $p \neq 0$, we can compute the fundamental group of M_p using the van Kampen theorem. Denote by T the solid torus $\partial \hat{U}_0$, by j_0 the inclusion induced morphism $\pi_1(T) \to \pi_1(\hat{U}_0)$, and by j_1 the inclusion induced morphism $\pi_1(T) \to \pi_1(U_1)$. As generators of $\pi_1(T)$, we can choose μ_0 and the attaching curve of the surgery $\varphi = \mu_0^p \ell_0$ because the intersection number of these two curves is ± 1. As generator of $\pi_1(U_1)$, we can choose $\ell_1 = \mu_0$ because the longitude of K_1 is the meridian of K_0. As generator of $\pi_1(\hat{U}_0)$ we can choose $j_0(\mu_0)$ because j_0 is surjective and $\varphi \in \ker j_0$. Thus, $\pi_1(M_p)$ is generated by μ_0, φ with the relation

$$1 = j_0(\hat{\mu}_0) = j_p(\hat{\mu}_0) = \mu_0^p \ell_0, \;\; \ell_0 = j_0(\ell_0) = j_p(\ell_0), \;\; j_p(\mu_0) = j_0(\mu_0).$$

Hence, $\pi_1(M_p) \cong \mathbb{Z}/p$. In fact, M_p is a lens space. More precisely, we have an *orientation preserving* diffeomorphism

$$S^3(K_0, \pm|p|) \cong L(|p|, |p| \pm 1). \qquad \square$$

Example 2.4 (Surgery on the trefoil knot). Suppose that K is a knot in S^3. Choose a closed tubular neighborhood U of K. The canonical framing of K defines a diffeomorphism $U = \mathbb{D}^2 \times S^1$. Denote by E_K the exterior

$$E_K = S^3 \setminus \operatorname{int}(U).$$

Let $T = \partial E_K = \partial U$, and denote by $i_* : \pi_1(T) \to \pi_1(E_K)$ the inclusion induced morphism. Let $K' \subset T$ be a *parallel* of K, i.e., a simple closed curve that intersects each meridian $\mu = \partial \mathbb{D}^2 \times \{pt\}$ of the knot exactly once.

The parallel K' determines a surgery on the knot K with surgery coefficient $p = \mathbf{lk}(K, K')$. To compute the fundamental group of $S^3(K, p)$, we use as before the van Kampen theorem.

Suppose $\pi_1(E_K)$ has a presentation with the set of generators $\mathcal{G}_K$ and relations $\mathcal{R}_K$. Let $\hat{U} = S^1 \times \mathbb{D}^2$ and denote by j the natural map

$$\partial U = \partial \mathbb{D}^1 \times S^1 \to S^1 \times \mathbb{D}^2 = \hat{U}.$$

Then, $\pi_1(\hat{U})$ is generated by $\hat{\ell} = j_*(\mu)$ and we deduce that $S^3(K, p)$ has a presentation with generators $\mathcal{G} \cup \{\hat{\ell}\}$ and relation

$$i_*(K') = 1, \;\; \hat{\ell} = j_*(\mu).$$

Equivalently, a presentation of $S^3(K, p)$ is obtained from a presentation of $\pi_1(E_K)$ by adding a single relation

$$i_*(K') = 1.$$

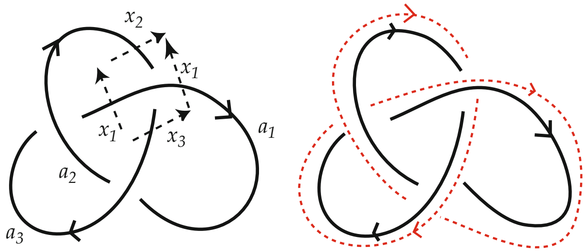

Fig. 2.2 The (left-handed) trefoil knot and its blackboard parallel

The fundamental group of the complement of the knot is called the *group of the knot*, and we will denote it by G_K. Let us explain how to compute a presentation of G_K and the morphism i_*.

Observe first that $\pi_1(T)$ is a free Abelian group of rank 2. As basis of $\pi_1(T)$, we can choose any pair (μ, γ), where γ is a parallel of K situated on T. Then, we can write

$$K' = a\mu + b\gamma.$$

If w denotes the linking number of γ and K , and ℓ denotes the longitude of K, then we can write $\gamma = w\mu + \ell$,

$$K' = p\mu + \ell = a\mu + b(w\mu + \ell) \Longrightarrow b = 1, \ \ a = p - w, \ \ K' = (p - w)\mu + \gamma.$$

Thus, i_* is completely understood if we know $i_*(\mu)$ and $i_*(\gamma)$ for some parallel γ of K.

The group of the knot K can be given an explicit presentation in terms of the *knot diagram*. This algorithmic presentation is known as the *Wirtinger presentation*. We describe the special case of the (left-handed) trefoil knot depicted in Fig. 2.2 and we refer to [Rolf, III.A] for proofs.

The Wirtinger algorithm goes as follows.

- Choose an orientation of the knot and a basepoint $*$ situated off the plane of the diagram. Think of the basepoint as the location of the eyes of the reader.
- The diagram of the knot consists of several disjoint arcs. Label them by

$$a_1, a_2, \ldots, a_\nu,$$

  in increasing cyclic order given by the above chosen orientation of the knot. In the case of the trefoil knot, we have three arcs, a_1, a_2, a_3.
- To each arc a_k, there corresponds a generator x_k represented by a loop starting at $*$ and winding around a_k once in the positive direction, where the positive direction is determined by the right-hand rule: if you point your right-hand thumb in the direction of a_k, then the rest of your palm should be wrapping around a_k in the direction of x_k (see Fig. 2.3).

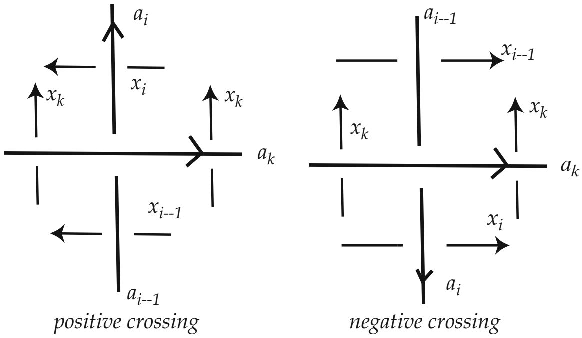

Fig. 2.3 The Wirtinger relations

- For each crossing of the knot diagram we have a relation. The crossings are of two types, positive $(+)$ (or right-handed) and negative $(-)$ (or left-handed) (see Fig. 2.3). Label by i the crossing, where the arc a_i begins and the arc a_{i-1} ends. Denote by $a_{k(i)}$ the arc going over the ith crossing and set

$$\epsilon(i) = \pm 1 \ \text{ if } i \text{ is a } \pm\text{-crossing.}$$

  Then, the relation introduced by the ith crossing is

$$x_i = x_{k(i)}^{-\epsilon(i)} x_{i-1} x_{k(i)}^{\epsilon(i)}.$$

The knot diagram defines a parallel of K called the *blackboard parallel* and denoted by K_{bb}. It is obtained by tracing a contour parallel and very close to the diagram of K and situated to the left of K with respect to the chosen orientation. In Fig. 2.2 the blackboard parallel of the trefoil knot is depicted with a thin line.

The linking number of K and K_{bb} is called the *writhe* of the knot diagram and it is denoted by $w(K)$. *It is not an invariant of the knot.* It is equal to the signed number of crossings of the diagram, i.e., the difference between the number of positive crossings and the number of negative crossings. One can show that

$$i_*(K_{\text{bb}}) = \prod_{i=1}^{\nu} x_{k(i)}^{\epsilon(i)}, \quad i_*(\mu) = x_\nu. \tag{2.1}$$

Set $G = G_K$, where K is the (left-handed) trefoil knot. In this case all the crossings in the diagram depicted in Fig. 2.2 are negative and we have $w(K) = -3$. The group G has three generators x_1, x_2, x_3, and since all the crossings are negative, we conclude that $\epsilon(i) = -1$, $\forall i = 1, 2, 3$, so that we have three relations

$$x_1 = x_2 x_3 x_2^{-1} \qquad x_2 = x_3 x_1 x_3^{-1}, \qquad x_3 = x_1 x_2 x_1^{-1}, \tag{2.2}$$

$$k(1) = 2, \qquad k(2) = 3, \qquad k(3) = 1. \tag{2.3}$$

From the equalities (2.3) we deduce

$$c = i_*(K_{\text{bb}}) = x_2^{-1}x_3^{-1}x_1^{-1}, \quad i_*(\mu) = x_3. \tag{2.4}$$

For $x \in G$ we denote by $T_x : G \to G$ the conjugation $g \mapsto xgx^{-1}$. We deduce

$$x_i = T_{x_{k(i)}}x_{i-1}, \quad \forall i = 1,2,3 \Longrightarrow x_3 = T_{x_{k(1)}^{-1}x_{k(2)}^{-1}x_{k(3)}^{-1}}x_3 = T_c x_3,$$

i.e., x_3 commutes with $c = x_2^{-1}x_3^{-1}x_1^{-1}$. Set for simplicity

$$a = x_1, \quad b = x_2, \quad x_3 = T_a b = aba^{-1}.$$

We deduce from (2.2) that G has the presentation

$$G = \langle a, b \mid \; aba = bab \rangle.$$

Consider the group

$$H = \langle \, x, y \mid \; x^3 = y^2 \, \rangle.$$

We have a map

$$H \to G, \quad x \mapsto ab, \quad y \mapsto aba.$$

It is easily seen to be a morphism with inverse $a = x^{-1}y$, $b = a^{-1}x = y^{-1}x^2$, so that $G \cong H$.

If we perform -1 surgery on the (left-handed) trefoil knot, then the attaching curve of the surgery is isotopic to

$$K' = -1 - w\mu + K_{\text{bb}}, \quad w = \operatorname{lk}(K_{\text{bb}}, \ell) = -3,$$

and we conclude

$$i_*(K_{\text{bb}}) = c = x_2^{-1}x_3^{-1}x_1^{-1} = b^{-1}ab^{-1}a^{-1}a = b^{-1}ab^{-1}, \quad i_*(\mu) = aba^{-1}.$$

The fundamental group $\pi_1\big(S^3(K,-1)\big)$ is obtained from G by introducing a new relation

$$i_*(\mu)^{-1-w} = c^{-1} \overset{w=-3}{\Longleftrightarrow} ab^2a^{-1} = ba^{-1}b.$$

Hence, the fundamental group of $S^3(K,-1)$ has the presentation

$$\langle a, b \mid \; aba{=}bab, ab^2a^{-1}{=}ba^{-1}b \rangle \Longleftrightarrow \langle a, b \mid \; aba = bab, \; a^2b^2 = aba^{-1}ba \rangle.$$

Observe that its abelianization is trivial. However, this group is nontrivial. It has order 120 and it can be given the equivalent presentation

$$\langle x, y \mid x^3 = y^5 = (xy)^2 \rangle.$$

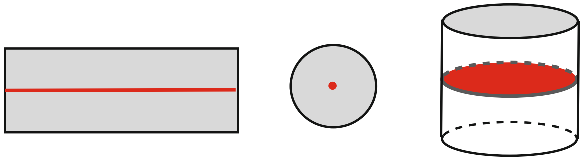

Fig. 2.4 A 1-handle of dimension two, a 0-handle of dimension two and a 2-handle of dimension three. The midsection disks are the cores of these handles

It is isomorphic to the binary icosahedral group I^*. This is the finite subgroup of $SU(2)$ that projects onto the subgroup $I \subset SO(3)$ of isometries of a regular icosahedron via the $2:1$ map $SU(2) \to SO(3)$.

The manifold $S^3(K, -1)$ is called the *Poincaré sphere*, and it is traditionally denoted by $\Sigma(2,3,5)$ because it is diffeomorphic to

$$\big\{ z = (z_0, z_1, z_2) \in \mathbb{C}^3;\ \ z_0^2 + z_1^3 + z_2^5 = 0,\ \ |z| = \varepsilon \big\}.$$

It is a $\mathbb{Z}$-*homology sphere*, meaning that its homology is isomorphic to the $\mathbb{Z}$-homology of S^3. □

Suppose that X is an m-dimensional smooth manifold with boundary. We want to describe what it means to attach a k-handle to X. This operation will produce a new *smooth* manifold with boundary.

A k*-handle* of dimension m (or a *handle of index* k) is the manifold with corners

$$\mathbf{H}_{k,m} := \mathbb{D}^k \times \mathbb{D}^{m-k}.$$

The disk $\mathbb{D}^k \times \{0\} \subset \mathbf{H}_{k,m}$ is called the *core*, while the disk $\{0\} \times \mathbb{D}^{m-k} \subset \mathbf{H}_{k,m}$ is called the *cocore*. The boundary of the handle decomposes as

$$\partial \mathbf{H}_{k,m} = \partial_- \mathbf{H}_{k,m} \cup \partial_+ \mathbf{H}_{k,m},$$

where

$$\partial_- \mathbf{H}_{k,m} := \partial \mathbb{D}^k \times \mathbb{D}^{m-k},\ \ \partial_+ \mathbf{H}_{k,m} := \mathbb{D}^k \times \partial \mathbb{D}^{k-m}.$$

The operation of attaching a k-handle (of dimension m) requires several additional data (Figs. 2.4 and 2.5).

- A $(k-1)$-dimensional sphere $\Sigma \hookrightarrow \partial X$ embedded in ∂X with *trivializable* normal bundle $T_\Sigma \partial X$. This normal bundle has rank $m - k = \dim \partial X - \dim \Sigma$.
- A framing φ of the normal bundle $T_\Sigma \partial X$.

The framing defines a diffeomorphism from $\mathbb{D}^{m-k} \times S^{k-1}$ to a tubular neighborhood N of Σ in ∂X. Using this identification, we detect inside N a copy of $\partial_- \mathbf{H}_{k,m} = \Sigma \times \mathbb{D}^{m-k}$. Now attach $\mathbf{H}_{k,m}$ to ∂X by identifying $\partial_- \mathbf{H}_{k,m}$ with its copy inside N and denote the resulting manifold by $X^+ = X(\Sigma, \varphi)$.

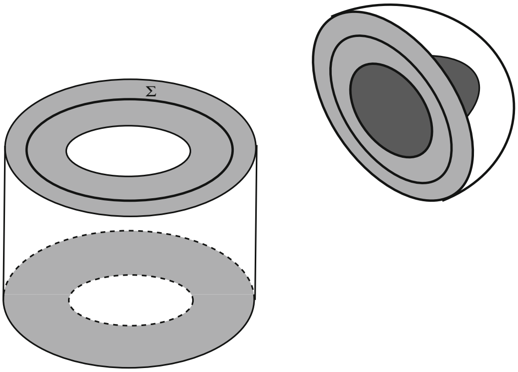

Fig. 2.5 Attaching a 2-handle of dimension three

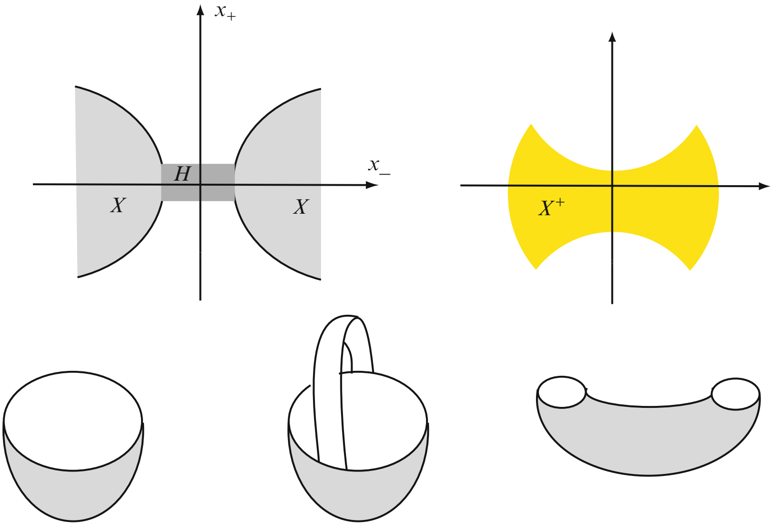

Fig. 2.6 Attaching a 1-handle of dimension two and smoothing the corners

The manifold X^+ has corners, but they can be smoothed out (see Fig. 2.6). The smoothing procedure is local, so it suffices to understand it in the special case

$$X \cong (-\infty, 0] \times \partial\mathbb{D}^k \times \mathbb{R}^{m-k}, \quad \partial X = \{0\} \times \partial\mathbb{D}^k \times \mathbb{R}^{m-k} (\cong N).$$

Consider the decomposition

$$\mathbb{R}^m = \mathbb{R}^k \times \mathbb{R}^{m-k}, \quad \mathbb{R}^m \ni x = (x_-, x_+) \in \mathbb{R}^k \times \mathbb{R}^{m-k}.$$

We have a homeomorphism

$$(-\infty, 0] \times \partial\mathbb{D}^k \times \mathbb{R}^{m-k} \longrightarrow \big\{ x \in \mathbb{R}^m; \;\; |x_+|^2 - |x_-|^2 \leq -1 \big\},$$

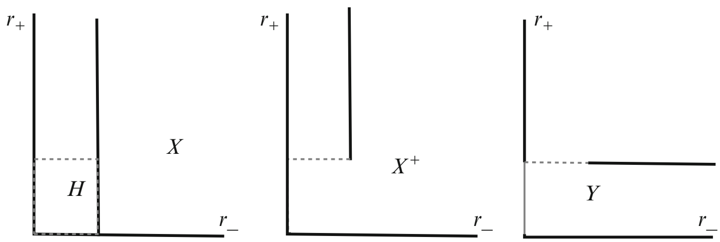

Fig. 2.7 Smoothing corners

defined by

$$(-\infty, 0] \times \partial\mathbb{D}^k \times \mathbb{R}^{m-k} \ni (t, \theta, x_+) \mapsto \big((e^{-2t} + |x_+|^2)^{1/2} \cdot \theta,\ x_+ \big) \in \mathbb{R}^k \times \mathbb{R}^{m-k}.$$

The manifold X^+ obtained after the surgery is homeomorphic to

$$\{ x \in \mathbb{R}^m;\ \ |x_+|^2 - |x_-|^2 \leq 1 \},$$

which is a *smooth* manifold with boundary.

This homeomorphism is visible in Fig. 2.6, but a formal proof can be read from Fig. 2.7.

Let us explain Fig. 2.7. We set $r_\pm = |x_\pm|$ and observe that

$$X \cong \{ r_- \geq 1 \},\quad \mathbf{H}_{k,m} = \{ r_-, r_+ \leq 1 \}.$$

After we attach the handle, we obtain

$$X_+ = \{ r_- \geq 1 \} \cup \{ r_- \leq 1,\ \ r_+ \leq 1 \}.$$

Now, fix a homeomorphism

$$X_+ \to Y = \{ r_+ \leq 1 \},$$

which is the identity in a neighborhood of the region $\{ r_- \cdot r_+ = 0 \}$. Clearly Y is homeomorphic to the region $r_+^2 - r_-^2 \leq 1$ via the homeomorphism

$$Y \ni (x_-, x_+) \mapsto \Big(x_-, \big(1 + r_-^2\big)^{1/2} x_+ \Big).$$

Let us analyze the difference between the topologies of ∂X^+ and ∂X.

Observe that we have a decomposition

$$\partial X^+ = (\partial X \setminus \partial_- \mathbf{H}_{k,m}) \cup_\varphi \partial_+ \mathbf{H}_{k,m}.$$

Above, $(\partial X \setminus \partial_{-}\mathbf{H}_{k,m})$ is a manifold with boundary diffeomorphic to $\partial\mathbb{D}^{m-k} \times S^{k-1}$, which is identified with the boundary of $\partial_{+}\mathbf{H}_{k,m} = \mathbb{D}^{k} \times \partial\mathbb{D}^{m-k}$ via the chosen framing φ. In other words, ∂X^{+} is obtained from ∂X via the surgery given by the data (S, φ).

In general, if M_1 is obtained from M_0 by a surgery of type (S, φ), then M_1 is cobordant to M_0. Indeed, consider the manifold

$$X = [0,1] \times M_0.$$

We obtain an embedding $S \hookrightarrow \{1\} \times M_0 \hookrightarrow \partial X$ and a framing φ of its normal bundle. Then,

$$\partial X(S, \varphi) = M_0(S, \varphi) \sqcup M_0.$$

The above cobordism $X(S, \varphi)$ is called the *trace* of the surgery.

2.2 The Topology of Sublevel Sets

Suppose M is a smooth connected m-dimensional manifold and $f : M \to \mathbb{R}$ is an *exhaustive* Morse function, i.e., the sublevel set

$$M^{c} = \{x \in M;\ \ f(x) \leq c\}$$

is compact for every $c \in \mathbb{R}$. We fix a smooth vector field X on M that is *gradient-like* with respect to f. This means that

$$X \cdot f > 0 \ \text{ on } \ M \setminus \mathbf{Cr}_{f},$$

and for every critical point p of f there exist coordinates adapted to p and X, i.e., coordinates (x^{i}) such that

$$X = -2\sum_{i=1}^{\lambda} x^{i}\partial_{x^{i}} + 2\sum_{j>\lambda} x^{j}\partial_{x^{j}}, \ \ \lambda = \lambda(f, p).$$

In these coordinates near p the flow Γ_t generated by $-X$ is described by

$$\Gamma_t(x) = e^{2t}x_{-} + e^{-2t}x_{+},$$

where $x = x_{-} + x_{+}$,

$$x_{-} := \big(x^{1}, \dots, x^{\lambda}, 0, \dots, 0\big), \ \ x_{+} := \big(0, \dots, 0, x^{\lambda+1}, \dots, x^{m}\big).$$

To see that there exist such vector fields, choose a Riemannian metric g *adapted* to f, i.e., a metric with the property that for every critical point p of f there exist coordinates (x^i) adapted to p such that near p we have

$$g = \sum_{i=1}^{m} (\mathrm{d}x^i)^2, \quad f = f(p) + \sum_{j=1}^{\lambda} (x^j)^2 - \sum_{k>\lambda} (x^k)^2.$$

We denote by $\nabla f = \nabla^g f \in \mathrm{Vect}(M)$ the *gradient* of f with respect to the metric g, i.e., the vector field g-dual to the differential $\mathrm{d}f$. More precisely, ∇f is defined by the equality

$$g(\nabla f, X) = \mathrm{d}f(X) = X \cdot f, \quad \forall X \in \mathrm{Vect}(M).$$

In local coordinates (x^i), if

$$\mathrm{d}f = \sum_i \frac{\partial f}{\partial x^i} \mathrm{d}x^i, \quad g = \sum_{i,j} g_{ij} \, \mathrm{d}x^i \mathrm{d}x^j,$$

then

$$\nabla f = \sum_j g^{ij} \partial_{x^j} f,$$

where $(g^{ij})_{1 \le i,j \le m}$ denotes the matrix inverse to $(g_{ij})_{1 \le i,j \le m}$. In particular, near a critical point p of index λ the gradient of f in the above coordinates is given by

$$\nabla f = -2 \sum_{i=1}^{\lambda} x^i \partial_{x^i} + 2 \sum_{j>\lambda} x^j \partial_{x^j}.$$

This shows that $X = \nabla f$ is a gradient-like vector field.

Remark 2.5. As explained in [Sm, Theorem B], any gradient-like vector field can be obtained by the method described above. □

Notation. In the sequel, when referring to $f^{-1}\big((a,b)\big)$, we will use the more suggestive notation $\{a < f < b\}$. The same goes for $\{a \le f < b\}$, etc. □

Theorem 2.6. *Suppose that the interval* $[a,b] \in \mathbb{R}$ *contains no critical values of* f. *Then the sublevel sets* M^a *and* M^b *are diffeomorphic. Furthermore,* M^a *is a deformation retract of* M^b, *so that the inclusion* $M^a \hookrightarrow M^b$ *is a homotopy equivalence.*

Proof. Since there are no critical values of f in $[a,b]$ and the sublevel sets M^c are compact, we deduce that there exists $\varepsilon > 0$ such that

$$\{a - \varepsilon < f < b + \varepsilon\} \subset M \setminus \mathbf{Cr}_f.$$

Fix a gradient-like vector field Y and construct a smooth function

$$\rho : M \to [0, \infty)$$

such that

$$\rho(x) = \begin{cases} |Yf|^{-1}, & a \le f(x) \le b, \\ 0, & f(x) \notin (a - \varepsilon, b + \varepsilon). \end{cases}$$

We can now construct the vector field $X := -\rho Y$ on M, and we denote by

$$\Phi : \mathbb{R} \times M \to M, \quad (t, x) \longmapsto \Phi_t(x)$$

the flow generated by X. If $u(t)$ is an integral curve of X, i.e., $u(t)$ satisfies the differential equation

$$\dot{u} = X(u),$$

then differentiating f along $u(t)$, we deduce that in the region $\{a \le f \le b\}$ we have the equality

$$\frac{\mathrm{d}f}{\mathrm{d}t} = Xf = -\frac{1}{Yf} Yf = -1.$$

In other words, in the region $\{a \le f \le b\}$, the function f decreases at a rate of one unit per second. This implies

$$\Phi_{b-a}(M^b) = M^a, \quad \Phi_{a-b}(M^a) = M^b,$$

so that Φ_{b-a} establishes a diffeomorphism between M^b and M^a.

To show that M^a is a deformation retract of M^b, we consider

$$H : [0, 1] \times M^b \to M^b, \quad H(t, x) = \Phi_{t \cdot (f(x)-a)^+}(x),$$

where for every real number r we set $r^+ := \max(r, 0)$. Observe that if $f(x) \le a$, then

$$H(t, x) = x, \quad \forall t \in [0, 1],$$

while for every $x \in M^b$ we have

$$H(1, x) = \Phi_{(f(x)-a)^+}(x) \in M^a.$$

This proves that M^a is a deformation retract of M^b. □

Theorem 2.7 (Fundamental structural theorem). *Suppose c is a critical value of f containing a single critical point p of Morse index λ. Then for every $\varepsilon > 0$, sufficiently small the sublevel set $\{f \le c + \varepsilon\}$ is diffeomorphic to $\{f \le c - \varepsilon\}$ with* *a λ-handle of dimension m attached. If $x = (x_-, x_+)$ are coordinates adapted to the critical point, then the core of the handle is given by*

$$e_\lambda(p) := \{ x_+ = 0,\ |x_-|^2 \leq \varepsilon \}.$$

In particular, $\{f \leq c + \varepsilon\}$ is homotopic to $\{f \leq c - \varepsilon\}$ with the λ-cell e_λ attached.

Proof. We follow the elegant approach in [M3, Sect. I.3]. There exist $\varepsilon > 0$ and local coordinates (x^i) in an open neighborhood U of p with the following properties.

- The region $\{ |f - c| \leq \varepsilon \}$ is compact and contains no critical point of f other than p.
- $x^i(p) = 0$, $\forall i$, and the image of U under the diffeomorphism

  $$(x^1, \dots, x^m) : U \to \mathbb{R}^m$$

  contains the closed disk

  $$D = \Big\{ \sum (x^i)^2 \leq 2\varepsilon \Big\}.$$

- $$f|_D = c - \sum_{i \leq \lambda} (x^i)^2 + \sum_{j > \lambda} (x^j)^2.$$

  We set

  $$x_- := (x^1, \dots, x^\lambda, 0, \dots, 0),\quad u_- := \sum_{i \leq \lambda} (x^i)^2,$$

  $$x_+ := (0, \dots, 0, x^{\lambda+1}, \dots, x^m),\quad u_+ := \sum_{j > \lambda} (x^j)^2.$$

We have

$$f|_D = c - u_- + u_+.$$

We fix a smooth function $\mu : [0, \infty) \to \mathbb{R}$ with the following properties (see Fig. 2.8).

$$\mu(0) > \varepsilon,\quad \mu(t) = 0,\quad \forall t \geq 2\varepsilon, \tag{2.5}$$

$$-1 < \mu'(t) \leq 0,\quad \forall t \geq 0. \tag{2.6}$$

Now let (see Fig. 2.8)

$$h := \mu(0) > \varepsilon,\quad r := \min\{t;\ \mu(t) = 0\} \leq 2\varepsilon.$$

Define

$$F : M \to \mathbb{R},\quad F = f - \mu(u_- + 2u_+),$$

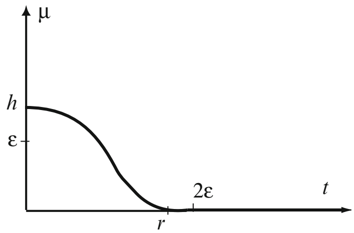

Fig. 2.8 The cutoff function μ

so that along D we have

$$F\,|_U = c - u_- + u_+ - \mu(u_- + 2u_+),$$

while on $M \setminus D$ we have $F = f$.

Lemma 2.8. *The function F satisfies the following properties.*

(a) F is a Morse function,

$$\mathbf{Cr}_F = \mathbf{Cr}_f, \quad F(p) < c - \varepsilon, \quad \text{and} \quad F(q) = f(q), \quad \forall q \in \mathbf{Cr}_f \setminus \{p\}.$$

(b) $\{f \leq a\} \subset \{F \leq a\}$, $\forall a \in \mathbb{R}$, $\{F \leq c + \delta\} = \{f \leq c + \delta\}$, $\forall \delta \geq \varepsilon$.

Proof. (a) Clearly $\mathbf{Cr}_F \cap (M \setminus D) = \mathbf{Cr}_f \cap (M \setminus U)$. To show that $\mathbf{Cr}_F \cap D = \mathbf{Cr}_f \cap D$, we use the fact that along D we have

$$F = f - \mu(u_- + 2u_+), \quad \mathrm{d}F = -(1 + \mu')\mathrm{d}u_- + (1 - 2\mu')\mathrm{d}u_+.$$

The condition (2.6) implies that $\mathrm{d}u_- = 0 = \mathrm{d}u_+$ at every critical point q of F in U, so that $x_-(q) = 0$, $x_+(q) = 0$, i.e., $q = p$. Clearly $F(p) = f(p) - \mu(0) < c - \varepsilon$. Clearly p is a nondegenerate critical point of F.

(b) Note first that

$$F \leq f \Longrightarrow \{f \leq a\} \subset \{F \leq a\}, \quad \forall a \in \mathbb{R}.$$

Again we have

$$\{F \leq c + \delta\} \cap (M \setminus D) = \{f \leq c + \delta\} \cap (M \setminus D),$$

so we have to prove

$$\{F \leq c + \delta\} \cap D \subset \{f \leq c + \delta\} \cap D.$$

Suppose $q \in \{F \leq c + \delta\} \cap D$ and set $u_\pm = u_\pm(q)$. This means that

$$u_- + u_+ \leq 2\varepsilon, \quad u_+ \leq u_- + \delta + \mu(u_- + 2u_+).$$

Using the condition $-1 < \mu'$, we deduce

$$\mu(t) = \mu(t) - \mu(2\varepsilon) \leq 2\varepsilon - t \leq 2\delta - t, \quad \forall t \leq 2\varepsilon.$$

If $u_- + 2u_+ \leq 2\varepsilon$, we have

$$\begin{aligned} u_- + \delta + \mu(u_- + 2u_+) \leq 3\delta - 2u_+ &\Rightarrow u_+ \leq \delta \\ &\Rightarrow u_+ - u_- \leq \delta \Rightarrow f(q) \leq c + \delta. \end{aligned}$$

If $u_- + 2u_+ \geq 2\varepsilon$, then $f(q) = F(q) \leq c + \varepsilon$. □

The above lemma implies that F is an exhaustive Morse function such that the interval $[c-\varepsilon, c+\varepsilon]$ consists only of regular values. We deduce from Theorem 2.6 that $\{F \leq c+\varepsilon\}$ is diffeomorphic to $\{F \leq c-\varepsilon\}$. Since

$$\big\{ F \leq c+\varepsilon \big\} = \big\{ f \leq c+\varepsilon \big\},$$

it suffices to show that $\big\{ F \leq c-\varepsilon \big\}$ is diffeomorphic to $\big\{ f \leq c-\varepsilon \big\}$ with a λ-handle attached.

Denote by H the closure of

$$\big\{ F \leq c-\varepsilon \big\} \setminus \{ f \leq c-\varepsilon \} = \big\{ F \leq c-\varepsilon \big\} \cap \{ f > c-\varepsilon \}.$$

Observe that

$$H = \big\{ F \leq c-\varepsilon \big\} \cap \big\{ f \geq c-\varepsilon \big\} \subset D.$$

The region H is described by the system of inequalities

$$\begin{cases} u_- + u_+ \leq 2\varepsilon, \\ f = -u_- + u_+ \geq -\varepsilon, \\ F = -u_- + u_+ - \mu \leq -\varepsilon, \end{cases} \qquad \mu = \mu(u_- + 2u_+).$$

Its boundary decomposes as $\partial H = \partial_- H \cup \partial_+ H$, where

$$\partial_- H = \begin{cases} u_- + u_+ \leq 2\varepsilon \;\; f = -u_- + u_+ = -\varepsilon, \\ F = -u_- + u_+ - \mu \leq -\varepsilon, \end{cases}$$

and

$$\partial_+ H = \begin{cases} u_- + u_+ \leq 2\varepsilon, \\ f = -u_- + u_+ \geq -\varepsilon, \\ F = -u_- + u_+ - \mu = -\varepsilon. \end{cases}$$

Let us analyze the region R in the Cartesian plane described by the system of inequalities

$$x, y \geq 0, \quad x + y \leq 2\varepsilon, \quad -x + y - \mu(x + 2y) \leq -\varepsilon, \quad -x + y \geq -\varepsilon.$$

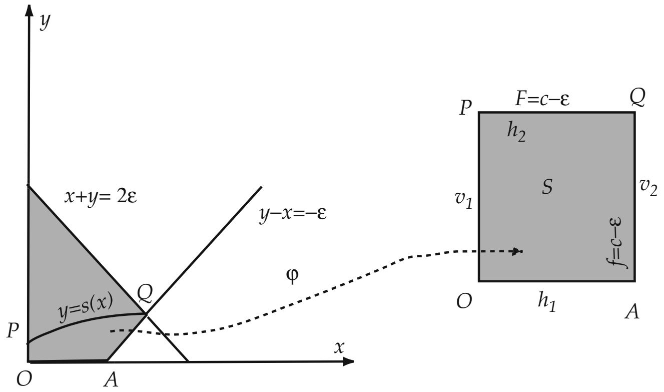

Fig. 2.9 A planar convex region

The region

$$\{y - x \geq -\varepsilon, \quad x + y \leq 2\varepsilon, \quad x, y \geq 0\}$$

is the shaded polygonal area depicted in Fig. 2.9. The two lines $y - x = -\varepsilon$ and $x + y = 2\varepsilon$ intersect at the point $Q = (\frac{3\varepsilon}{2}, \frac{\varepsilon}{2})$. We want to investigate the equation

$$-x + y - \mu(x + 2y) + \varepsilon = 0.$$

Set

$$\eta_x(y) := -x + y - \mu(x + 2y) + \varepsilon.$$

Observe that since $\mu(x) > \mu(0) - x$, we have

$$\eta_x(0) = -x - \mu(x) + \varepsilon < -\mu(0) + \varepsilon < 0,$$

while

$$\lim_{y \to \infty} \eta_x(y) = \infty.$$

Since $y \mapsto \eta_x(y)$ is strictly increasing there exists a unique solution $y = s(x)$ of the equation $\eta_x(y) = 0$. Using the implicit function theorem, we deduce that $s(x)$ depends smoothly on x and

$$\frac{\mathrm{d}s}{\mathrm{d}x} = \frac{1 + \mu'}{1 - 2\mu'} \in [0, 1].$$

The point Q lies on the graph of the function $y = s(x)$, $s(0) > 0$, and since $s'(x) \in [0, 1]$, we deduce that the slope-1 segment AQ lies below the graph of $s(x)$. We now see that the region R is described by the system of inequalities

$$x, y \geq 0, \quad y \leq s(x), \quad y - x \geq -\varepsilon.$$

Fix a homeomorphism φ from R to the standard square

$$S = \left\{ (t_-, t_+) \in \mathbb{R}^2; \;\; 0 \le t_\pm \le 1 \right\}$$

such that the vertices O, A, P, Q are mapped to the vertices

$$(0,0), \;\; (1,0), \;\; (1,1), \;\; (0,1)$$

(see Fig. 2.9). Denote by h_i and v_j the horizontal and vertical edges of S (see Fig. 2.9). Observe that we have a natural projection

$$u : H \to \mathbb{R}^2, \;\; H \ni q \mapsto (x, y) = (u_-(q), u_+(q)).$$

Its image is precisely the region R, and we denote by $t = (t_-, t_+)$ the composition $\varphi \circ u$. We now have a homeomorphism

$$H \mapsto \mathbf{H}_\lambda = \mathbb{D}^\lambda \times \mathbb{D}^{m-\lambda},$$

$$H \ni q \longmapsto \big(t_-(q)\theta_-(q), t_+(q)\theta_+(q) \big) \in \mathbb{D}^\lambda \times \mathbb{D}^{m-\lambda},$$

where

$$\theta_\pm(q) = u_\pm^{-1/2}(q) x_\pm(q)$$

denote the angular coordinates in

$$\Sigma_- = \left\{ u_- = 1, \;\; x_+ = 0 \right\} \cong S^{\lambda-1}$$

and

$$\Sigma_+ = \left\{ u_+ = 1, \;\; x_- = 0 \right\} \cong S^{m-\lambda-1}.$$

Then $\partial_+ H$ corresponds to the part of H mapped by u onto h_2, and $\partial_- H$ corresponds to the part of H mapped by u onto v_2. The core is the part mapped onto the horizontal segment h_1, while the cocore is the part of H mapped onto v_1. This discussion shows that indeed $\left\{ F \le c - \varepsilon \right\}$ is obtained from $\left\{ f \le c - \varepsilon \right\}$ by attaching the λ-handle H. $\square$

Remark 2.9. Suppose that c is a critical value of the exhaustive Morse function $f : M \to \mathbb{R}$ and the level set $f^{-1}(c)$ contains critical points $p_1, \dots, p_k$ with Morse indices $\lambda_1, \dots, \lambda_k$. Then the above argument shows that for $\varepsilon > 0$ sufficiently small the sublevel set $\{f \le c + \varepsilon\}$ is obtained from $\{f \le c - \varepsilon\}$ by attaching handles $H_1, \dots, H_k$ of indices $\lambda_1, \dots, \lambda_k$. $\square$

Corollary 2.10. *Suppose M is a smooth manifold and $f : M \to \mathbb{R}$ is an exhaustive Morse function on M. Then M is homotopy equivalent to a CW-complex that has exactly one λ-cell for every critical point of f of index λ.* $\square$

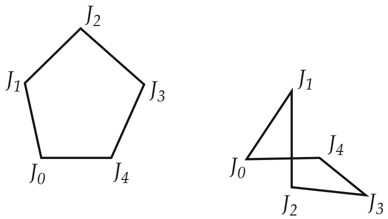

Fig. 2.10 Planar pentagons

Example 2.11 (Planar pentagons). Let us show how to use the fundamental structural theorem in a simple yet very illuminating example. We define a *regular planar pentagon* to be a closed polygonal line in the plane consisting of five straight-line segments of equal length 1. We would like to understand the topology of the space of all possible regular planar pentagons.

Consider one such pentagon with vertices J_0, J_1, J_2, J_3, J_4 such that

$$\operatorname{dist}(J_i, J_{i+1}) = 1.$$

There are a few trivial ways of generating new pentagons out of a given one. We can translate it, or we can rotate it about a fixed point in the plane. The new pentagons are not that interesting, and we will declare all pentagons obtained in this fashion from a given one to be equivalent. In other words, we are really interested in orbits of pentagons with respect to the obvious action of the affine isometry group of the plane.

There is a natural way of choosing a representative in such an orbit. We fix a cartesian coordinate system and we assume that the vertex J_0 is placed at the origin, while the vertex J_4 lies on the positive x-semiaxis, i.e., J_4 has coordinates $(1, 0)$ (Fig. 2.10).

Note that we can regard such a pentagon as a robot arm with four segments such that the last vertex J_4 is fixed at the point $(0, 1)$. Now recall some of the notation in Example 1.5.

A possible position of such a robot arm is described by four complex numbers,

$$z_1, \dots, z_4, \quad |z_i| = 1, \quad \forall i = 1, 2, 3, 4.$$

Since all the segments of such a robot arm have length 1, the position of the vertex J_k is given by the complex number $z_1 + \cdots + z_k$.

The space C of configurations of the robot arm constrained by the condition that J_4 can only slide along the positive x-semiaxis is a three-dimensional manifold. On C we have a Morse function

$$h : C \to \mathbb{R}, \quad h(\mathbf{z}) = \operatorname{Re}(z_1 + z_2 + z_3 + z_4),$$

which measures the distance of the last joint to the origin. The space of pentagons can be identified with the level set $\{ h = 1 \}$.

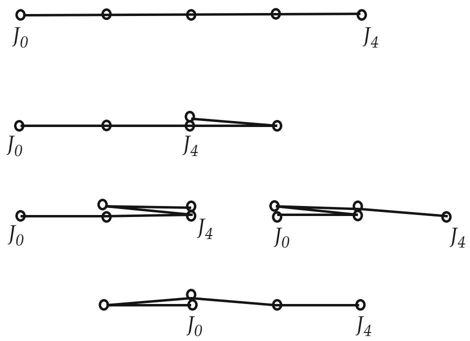

Fig. 2.11 Critical positions

Consider the function $f = -h : C \to \mathbb{R}$. The sublevel sets of f are compact. Moreover, the computations in Example 1.10 show that f has exactly five critical points, a local minimum

$$(1, 1, 1, 1),$$

and four critical configurations of index 1

$$(1, 1, 1, -1), \quad (1, 1, -1, 1), \quad (1, -1, 1, 1), \quad (-1, 1, 1, 1),$$

all situated on the level set $\{h = 2\} = \{f = -2\}$. The corresponding positions of the robot arm are depicted in Fig. 2.11.

The level set $\{f = -1\}$ is not critical, and it is obtained from the sublevel set $\{f \le -3\}$ by attaching four 1-handles.

The sublevel set $\{f \le -3\}$ is a closed three-dimensional ball, and thus the sublevel set $\{f \le -1\}$ is a 3-ball with four 1-handles attached. Its boundary, $\{f = -1\}$, is, therefore, a Riemann surface of genus 4. We conclude that the space of orbits of regular planar pentagons is a Riemann surface of genus 4. For more general results on the topology of the space of planar polygons, we refer to the very nice papers [FaSch, KM]. We will have more to say about this problem in Sect. 3.1. □

Remark 2.12. We can use the fundamental structural theorem to produce a new description of the trace of a surgery. We follow the presentation in [M4, Sect. 3].

Consider an orthogonal direct sum decomposition $\mathbb{R}^m = \mathbb{R}^\lambda \oplus \mathbb{R}^{m-\lambda}$. We denote by x the coordinates in $\mathbb{R}^\lambda$ and by y the coordinates in $\mathbb{R}^{m-\lambda}$. Then identify

$$\mathbb{D}^\lambda = \{x \in \mathbb{R}^\lambda;\ |x| \le 1\},\ \ \mathbb{D}^{m-\lambda} = \{y \in \mathbb{R}^{m-\lambda};\ |y| \le 1\},$$

$$\mathbf{H}_{\lambda,m} = \{(x, y) \in \mathbb{R}^m;\ |x|, |y| \le 1\}.$$

Consider the regions (see Fig. 2.12)

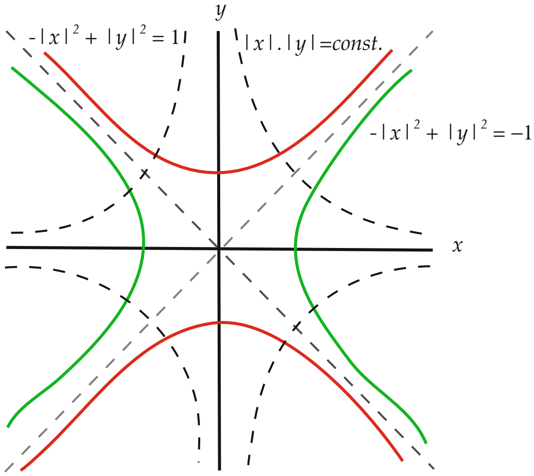

Fig. 2.12 A Morse theoretic picture of the trace of a surgery

$$\begin{aligned}\hat{B}_\lambda &:= \big\{(x,y) \in \mathbb{R}^m;\ \ -1 \le -|x|^2 + |y|^2 \le 1,\ \ 0 \le |x| \cdot |y| < r\big\},\\ B_\lambda &= \big\{(x,y) \in \hat{B}_\lambda;\ \ |x| \cdot |y| > 0\big\}.\end{aligned}$$

The region B_λ has two boundary components (see Fig. 2.12)

$$\partial_\pm B_\lambda = \big\{(x,y) \in B_\lambda;\ \ -|x|^2 + |y|^2 = \pm 1\big\}.$$

Consider the functions

$$f, h : \mathbb{R}^m \to \mathbb{R},\quad f(x,y) = -|x|^2 + |y|^2,\quad h(x,y) = |x| \cdot |y|,$$

so that

$$B_\lambda = \big\{-1 \le f \le 1,\ \ 0 < h < r\big\},\quad \partial_\pm B_\lambda = \big\{f = \pm 1,\ \ 0 < h < r\big\}.$$

Denote by U the gradient vector field of f. We have

$$U = -U_x + U_y,\quad U_x = 2\sum_i x^i \partial_{x_i},\quad U_y = 2\sum_j y^j \partial_{y^j}.$$

The function h is differentiable in the region $h > 0$, and

$$\nabla h = \frac{|y|}{|x|}x + \frac{|x|}{|y|}y.$$

We deduce

$$U \cdot h = (\nabla h, U) = 0.$$

Define $V = \frac{1}{U \cdot f} U$. We have

$$V \cdot f = 1, \quad V \cdot h = 0.$$

Denote by Γ_t the flow generated by V. We have

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\Gamma_t z) = 1, \quad \forall z \in \mathbb{R}^m, \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t} h(\Gamma_t z) = 0, \quad \forall z \in \mathbb{R}^m, \quad h(z) > 0.$$

Thus, h is constant along the trajectories of V, and along such a trajectory f increases at a rate of one unit per second. We deduce that for any $z \in \partial_- B_\lambda$ we have

$$f(\Gamma_t z) = -1 + t, \quad h(\Gamma_t z) = h(z) \in (0, 1).$$

We obtain a diffeomorphism

$$\Psi : [-1, 1] \times \partial_- B_\lambda \to B_\lambda, \quad (t, z) \longmapsto \Gamma_{t+1}(z).$$

Its inverse is

$$B_\lambda \ni w \longmapsto \big(f(w), \Gamma_{-1-f(z)} w \big).$$

This shows that the pullback of $f : B_\lambda \to \mathbb{R}$ to $[-1, 1] \times \partial_- B_\lambda$ via Ψ coincides with the natural projection

$$[-1, 1] \times \partial_- B_\lambda \to [-1, 1].$$

Moreover, we have a diffeomorphism

$$\{1\} \times \partial_- B_\lambda \xrightarrow{\Psi} \partial_+ B_\lambda.$$

Now observe that we have a diffeomorphism

$$\Phi : \big(\dot{\mathbb{D}}^{m-\lambda} \setminus \{0\} \big) \times S^{\lambda-1} \to \partial_- B_\lambda,$$

$$\big(\dot{\mathbb{D}}^{m-\lambda} \setminus \{0\} \big) \times S^{\lambda-1} \ni (u, v) \mapsto \big(\cosh(|u|) v, \, \sinh(|u|) \theta_u \big) \in \mathbb{R}^\lambda \times \mathbb{R}^{m-\lambda},$$

$$\theta_u := \frac{u}{|u|}.$$

Suppose M is a smooth manifold of dimension $m - 1$ and we have an embedding

$$\varphi : \mathbb{D}^{m-\lambda} \times S^{\lambda-1} \hookrightarrow M.$$

Consider the manifold $X = [-1, 1] \times M$ and set

$$X' = X \setminus \varphi\Big([-1, 1] \times \{0\} \times S^{\lambda-1} \Big).$$

Denote by W the manifold obtained from the disjoint union $X' \sqcup \hat{B}_\lambda$ by identifying $B_\lambda \subset \hat{B}_\lambda$ with an open subset of $[-1,1] \times M$ via the gluing map $\gamma = \varphi \circ \Phi^{-1} \circ \Psi^{-1}$,

$$B_\lambda \xrightarrow{\Psi^{-1}} [-1,1] \times \partial_- B_\lambda \xrightarrow{\Phi^{-1}} [-1,1] \times \big(\dot{\mathbb{D}}^{m-\lambda} \setminus \{0\}\big) \times S^{\lambda-1} \xrightarrow{\varphi} [-1,1] \times M.$$

Via the above gluing, the restriction of f to B_λ is identified with the natural projection $\pi : X' \to [-1,1]$, i.e.,

$$\gamma^*(f|_{B_\lambda}) = \pi|_{\gamma(B_\lambda)}.$$

Gluing π and $\gamma^* f$ we obtain a smooth function

$$F : W \to [-1,1]$$

that has a unique critical point p with critical value $F(p) = 0$ and Morse index λ. Set

$$W^a = \big\{ w \in W;\ \ F(w) \leq a \big\}.$$

We deduce from the fundamental structural theorem that $W^{1/2}$ is obtained from $W^{-1/2} \cong M$ by attaching a λ-handle with framing given by φ. The region $\big\{ -\frac{1}{2} \leq F \leq \frac{1}{2} \big\}$ is, therefore, diffeomorphic to the trace of the surgery $M \longrightarrow M(S^{\lambda-1}, \varphi)$. □

2.3 Morse Inequalities

To formulate these important algebraic consequences of the topological facts established so, far, we need to introduce some terminology.

Denote by $\mathbb{Z}[[t, t^{-1}]$ the ring of formal Laurent series with integral coefficients. More precisely,

$$\sum_{n \in \mathbb{Z}} a_n t^n \in \mathbb{Z}[[t, t^{-1}] \Longleftrightarrow a_n = 0 \ \ \forall n \ll 0, \ \ a_m \in \mathbb{Z}, \ \ \forall m.$$

Suppose $\mathbb{F}$ is a field. A graded $\mathbb{F}$-vector space

$$A_\bullet = \bigoplus_{n \in \mathbb{Z}} A_n$$

is said to be *admissible* if $\dim A^n < \infty$, $\forall n$, and $A_n = 0$, $\forall n \ll 0$. To an admissible graded vector space $A_\bullet$ we associate its *Poincaré series*

$$P_{A_\bullet}(t) := \sum_n (\dim_{\mathbb{F}} A_n) t^n \in \mathbb{Z}[[t, t^{-1}].$$

We define an order relation $\succ$ on the ring $\mathbb{Z}[[t, t^{-1}]$ by declaring that

$$X(t) \succ Y(t) \iff \text{there exists } Q \in \mathbb{Z}[[t, t^{-1}] \text{ with } \textit{nonnegative coefficients}$$

such that

$$X(t) = Y(t) + (1 + t)Q(t). \tag{2.7}$$

Remark 2.13. (a) Assume that

$$X(t) = \sum_n x_n t^n \in \mathbb{Z}[[t, t^{-1}], \;\; Y(t) = \sum_n y_n t^n \in \mathbb{Z}[[t, t^{-1}]$$

are such that $X \succ Y$. Then there exists $Q \in \mathbb{Z}[[t, t^{-1}]$ such that

$$X(t) = Y(t) + (1 + t)Q(t), \quad Q(t) = \sum_n q_n t^n, \;\; q_n \geq 0.$$

Then we can rewrite the above equality as

$$(1 + t)^{-1} X(t) = (1 + t)^{-1} Y(t) + Q(t).$$

Using the identity

$$(1 + t)^{-1} = \sum_{n \geq 0} (-1)^n t^n$$

we deduce

$$\sum_{k \geq 0} (-1)^k x_{n-k} - \sum_{k \geq 0} (-1)^k y_{n-k} = q_n \geq 0.$$

Thus the order relation $\succ$ is equivalent to the *abstract Morse inequalities*

$$X \succ Y \iff \sum_{k \geq 0} (-1)^k x_{n-k} \geq \sum_{k \geq 0}^{n} (-1)^k y_{n-k}, \;\; \forall n \geq 0. \tag{2.8}$$

Note that (2.7) implies immediately the *weak Morse inequalities*

$$x_n \geq y_n, \;\; \forall n \geq 0. \tag{2.9}$$

(b) Observe that $\succ$ is an order relation satisfying

$$X \succ Y \iff X + R \succ Y + R, \;\; \forall R \in \mathbb{Z}\big[[t, t^{-1}\big],$$

$$X \succ Y, \;\; Z \succ 0 \Longrightarrow X \cdot Z \succ Y \cdot Z.$$

□

Lemma 2.14 (Subadditivity). *Suppose we have a long exact sequence of admissible graded vector spaces $A_\bullet$, $B_\bullet$, $C_\bullet$*

$$\cdots \to A_k \xrightarrow{i_k} B_k \xrightarrow{j_k} C_k \xrightarrow{\partial_k} A_{k-1} \to \cdots .$$

Then,

$$P_{A_\bullet} + P_{C_\bullet} \succ P_{B_\bullet}. \tag{2.10}$$

Proof. Set

$$a_k = \dim A_k,\ \ b_k = \dim B_k,\ \ c_k = \dim C_k,$$
$$\alpha_k = \dim\ker i_k,\ \ \beta_k = \dim\ker j_k,\ \ \gamma_k = \dim\ker\partial_k.$$

Then,

$$\begin{cases} a_k = \alpha_k + \beta_k \\ b_k = \beta_k + \gamma_k \\ c_k = \gamma_k + \alpha_{k-1} \end{cases} \Longrightarrow a_k - b_k + c_k = \alpha_k + \alpha_{k-1}$$

$$\Longrightarrow \sum_k (a_k - b_k + c_k)t^k = \sum_k t^k(\alpha_k + \alpha_{k-1})$$

$$\Longrightarrow P_{A_\bullet}(t) - P_{B_\bullet}(t) + P_{C_\bullet}(t) = (1+t)Q(t),\ \ Q(t) = \sum_k \alpha_k t^k.$$

□

For every compact topological space X, we denote by $b_k(X) = b_k(X,\mathbb{F})$ the kth *Betti number* (with coefficients in $\mathbb{F}$)

$$b_k(X) := \dim H_k(X,\mathbb{F}),$$

and by $P_X(t) = P_{X,\mathbb{F}}(t)$ the Poincaré polynomial

$$P_{X,\mathbb{F}}(t) = \sum_k b_k(X,\mathbb{F})t^k.$$

If Y is a subspace of X then the relative Poincaré polynomial $P_{X,Y}(t)$ is defined in a similar fashion. The *Euler characteristic* of X is

$$\chi(X) = \sum_{k\geq 0}(-1)^k b_k(X),$$

and we have the equality

$$\chi(X) = P_X(-1).$$

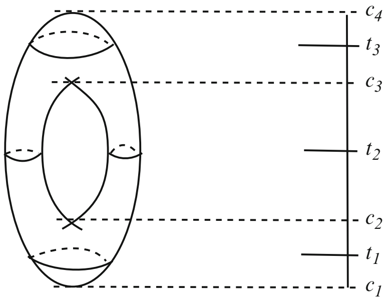

Fig. 2.13 Slicing a manifold by a Morse function

Corollary 2.15 (Topological Morse inequalities). *Suppose $f : M \to \mathbb{R}$ is a Morse function on a smooth compact manifold of dimension M with Morse polynomial*

$$P_f(t) = \sum_\lambda \mu_f(\lambda) t^\lambda.$$

Then, for every field of coefficients $\mathbb{F}$ we have

$$P_f(t) \succ P_{M,\mathbb{F}}(t).$$

In particular,

$$\sum_{\lambda \geq 0} (-1)^\lambda \mu_f(\lambda) = P_f(-1) = P_{M,\mathbb{F}}(-1) = \chi(M).$$

Proof. Let $c_1 < c_1 < \cdots < c_\nu$ be the critical values of f. Set (see Fig. 2.13)

$$t_0 = c_1 - 1, t_\nu = c_\nu + 1, \quad t_k = \frac{c_k + c_{k+1}}{2}, \quad k = 1, \dots, \nu - 1,$$
$$M_i = \{f \leq t_i\}, \quad 0 \leq i \leq \nu.$$

For simplicity, we drop the field of coefficients from our notations.

From the long exact homological sequence of the pair (M_i, M_{i-1}) and the subadditivity lemma, we deduce

$$P_{M_{i-1}} + P_{M_i, M_{i-1}} \succ P_{M_i}.$$

Summing over $i = 1, \dots, \nu$, we deduce

$$\sum_{i=1}^\nu P_{M_{i-1}} + \sum_{i=1}^\nu P_{M_i, M_{i-1}} \succ \sum_{i=1}^\nu P_{M_i} \Longrightarrow \sum_{k=1}^\nu P_{M^k, M^{k-1}} \succ P_{M^\nu}.$$

Using the equality $M_\nu = M$, we deduce

$$\sum_{i=1}^{\nu} P_{M_i,M_{i-1}} \succ P_M.$$

Denote by $\mathbf{Cr}_i \subset \mathbf{Cr}_f$ the critical points on the level set $\{f = c_i\}$. From the fundamental structural theorem and the excision property of the singular homology, we deduce

$$H_\bullet(M_i, M_{i-1}; \mathbb{F}) \cong \bigoplus_{p\in\mathbf{Cr}_i} H_\bullet(\mathbf{H}_{\lambda(p)}, \partial_-\mathbf{H}_{\lambda(p)}; \mathbb{F}) \cong \bigoplus_{p\in\mathbf{Cr}_i} H_\bullet(e_{\lambda(p)}, \partial e_{\lambda(p)}; \mathbb{F}).$$

Now observe that $H_k(e_\lambda, \partial e_\lambda; \mathbb{F}) = 0$, $\forall k \neq \lambda$, while $H_\lambda(e_\lambda, \partial e_\lambda; \mathbb{F}) \cong \mathbb{F}$. Hence

$$P_{M_i,M_{i-1}}(t) = \sum_{p\in\mathbf{Cr}_i} t^{\lambda(p)}.$$

Hence,

$$P_f(t) = \sum_{i=1}^{\nu} P_{M_i,M_{i-1}}(t) \succ P_M.$$

□

Remark 2.16. The above proof yields the following more general result. If

$$X_1 \subset \cdots \subset X_\nu = X$$

is an increasing filtration by closed subsets of the compact space X, then

$$\sum_{i=1}^{\nu} P_{X_i,X_{i-1}}(t) \succ P_X(t).$$

□

Suppose $\mathbb{F}$ is a field and f is a Morse function on a compact manifold. We say that a critical point $p \in \mathbf{Cr}_f$ of index λ is $\mathbb{F}$-*completable* if the boundary of the core $e_\lambda(p)$ defines a trivial homology class in $H_{\lambda-1}(M^{c-\varepsilon}, \mathbb{F})$, $c = f(p)$, and $0 < \varepsilon \ll 1$. We say that f is $\mathbb{F}$-*completable* if all its critical points are $\mathbb{F}$-completable.

We say that f is an $\mathbb{F}$-*perfect* Morse function if its Morse polynomial is equal to the Poincaré polynomial of M with coefficients in $\mathbb{F}$, i.e., all the Morse inequalities become equalities.

Proposition 2.17. *Any $\mathbb{F}$-completable Morse function on a smooth, closed, compact manifold is $\mathbb{F}$-perfect.*

Proof. Suppose $f : M \to \mathbb{R}$ is a Morse function on the compact, smooth m-dimensional manifold. Denote by $c_1 < \cdots < c_\nu$ the critical values of M and set (see Fig. 2.13)

$$t_0 = c_1 - 1, t_\nu = c_\nu + 1, \;\; t_i := \frac{c_i + c_{i+1}}{2}, \;\; i = 1, \ldots, \nu - 1.$$

Denote by $\mathbf{Cr}_i \subset \mathbf{Cr}_f$ the set of critical points on the level set $\{f = c_i\}$. Set $M_i := \{f \leq t_i\}$. From the fundamental structural theorem and the excision property of the singular homology, we deduce

$$H_\bullet(M_i, M_{i-1}; \mathbb{F}) \cong \bigoplus_{p \in \mathbf{Cr}_i} H_\bullet(\mathbf{H}_{\lambda(p)}, \partial_- \mathbf{H}_{\lambda(p)}; \mathbb{F}) \cong \bigoplus_{p \in \mathbf{Cr}_i} H_\bullet(e_{\lambda(p)}, \partial e_{\lambda(p)}; \mathbb{F}).$$

Now, observe that $H_k(e_\lambda, \partial e_\lambda; \mathbb{F}) = 0$, $\forall k \neq \lambda$, while $H_\lambda(e_\lambda, \partial e_\lambda; \mathbb{F}) \cong \mathbb{F}$. This last isomorphism is specified by fixing an orientation on $e_\lambda(p)$, which then produces a basis of $H_\lambda(H_\lambda, \partial_- H_\lambda; \mathbb{F})$ described by the relative homology class $[e_\lambda, \partial e_\lambda]$.

The connecting morphism

$$H_\bullet(M_i, M_{i-1}; \mathbb{F}) \xrightarrow{\partial} H_{\bullet-1}(M_{i-1}, \mathbb{F})$$

maps $[e_\lambda, \partial e_{\lambda(p)}]$ to the image of $[\partial e_\lambda]$ in $H_{\lambda(p)-1}(M_{i-1}, \mathbb{F})$. Since f is $\mathbb{F}$-completable, we deduce that these connecting morphisms are trivial. Hence for every $1 \leq i \leq \nu$ we have a short exact sequence

$$0 \to H_\bullet(M_{i-1}, \mathbb{F}) \to H_\bullet(M_i, \mathbb{F}) \to \bigoplus_{p \in \mathbf{Cr}_i} H_\bullet(e_{\lambda(p)}, \partial e_{\lambda(p)}; \mathbb{F}) \to 0.$$

Hence,

$$P_{M_i, \mathbb{F}}(t) = P_{M_{i-1}, \mathbb{F}}(t) + \sum_{p \in \mathbf{Cr}_i} t^{\lambda(p)}.$$

Summing over $i = 1, \dots, \nu$ and observing that $M_0 = \emptyset$ and $M_\nu = M$, we deduce

$$P_{M, \mathbb{F}}(t) = \sum_{i=1}^{\nu} \sum_{p \in \mathbf{Cr}_i} t^{\lambda(p)} = P_f(t). \qquad \square$$

Let us describe a simple method of recognizing completable functions.

Proposition 2.18. *Suppose $f : M \to \mathbb{R}$ is a Morse function on a compact manifold satisfying the* gap condition*:*

$$|\lambda(p) - \lambda(q)| \neq 1, \quad \forall p, q \in \mathbf{Cr}_f .$$

Then, f is $\mathbb{F}$-completable for any field $\mathbb{F}$.

Proof. We continue to use the notation in the proof of Proposition 2.17. Set

$$\Lambda := \{\lambda(p); \ \ p \in \mathbf{Cr}_f\}, \quad \Lambda_i = \{\lambda(p); \ \ p \in \mathbf{Cr}_i\} \subset \mathbb{Z}.$$

The gap condition shows that

$$\lambda \in \Lambda \Longrightarrow \lambda \pm \in \mathbb{Z} \setminus \Lambda. \tag{2.11}$$

Note that the fundamental structural theorem implies

$$H_k(M_i, M_{i-1}; \mathbb{F}) = 0 \Longleftrightarrow k \in \mathbb{Z} \setminus \Lambda, \tag{2.12}$$

since M_i/M_{i-1} is homotopic to a wedge of spheres of dimensions belonging to Λ.

We will prove by induction over $i \geq 0$ that

$$k \in \mathbb{Z} \setminus \Lambda \Longrightarrow H_k(M_i, \mathbb{F}) = 0, \tag{A_i}$$

and that the connecting morphism

$$\partial : H_k(M_i, M_{i-1}; \mathbb{F}) \longrightarrow H_{k-1}(M_{i-1}, \mathbb{F}) \tag{B_i}$$

is trivial for every $k \geq 0$.

The above assertions are trivially true for $i = 0$. Assume $i > 0$. We begin by proving (B_i).

This statement is obviously true if $H_k(M_i, M_{i-1}; \mathbb{F}) = 0$, so we may assume $H_k(M_i, M_{i-1}; \mathbb{F}) \neq 0$. Note that (2.12) implies $k \in \Lambda$, and thus the gap condition (2.11) implies that $k - 1 \in \mathbb{Z} \setminus \Lambda$.

The inductive assumption (A_{i-1}) implies that $H_{k-1}(M_{i-1}, \mathbb{F}) = 0$, so that the connecting morphism

$$\partial : H_k(M_i, M_{i-1}; \mathbb{F}) \to H_{k-1}(M_{i-1}, \mathbb{F})$$

is zero. This proves (B_i). In particular, for every $k \geq 0$ we have an exact sequence

$$0 \to H_k(M_{i-1}, \mathbb{F}) \to H_k(M_i, \mathbb{F}) \to H_k(M_i, M_{i-1}; \mathbb{F}).$$

Suppose $k \in \mathbb{Z} \setminus \Lambda$. Then $H_k(M_i, M_{i-1}; \mathbb{F}) = 0$, so that $H_k(M_i, \mathbb{F}) \cong H_k(M_{i-1}, \mathbb{F})$. From (A_{i-1}), we now deduce $H_k(M_{i-1}, \mathbb{F}) = 0$. This proves (A_i) as well.

To conclude the proof of the proposition observe that (B_i) implies that f is $\mathbb{F}$-completable. □

Corollary 2.19. *Suppose $f : M \to \mathbb{R}$ is a Morse function on a compact manifold whose critical points have only even indices. Then, f is a perfect Morse function.* □

Example 2.20. Consider the round sphere

$$S^n = \left\{ (x^0, \ldots, x^n) \in \mathbb{R}^{n+1};\ \sum_i |x^i|^2 = 1 \right\}.$$

The height function

$$h_n : S^n \to \mathbb{R}, \quad (x^0, \ldots, x^n) \mapsto x^0$$

is a Morse function with two critical points: a global maximum at the north pole $x^0 = 1$ and a global minimum at the south pole $x^0 = -1$.

For $n > 1$, this is a perfect Morse function, and we deduce

$$P_{S^n}(t) = P_{h_n}(t) = 1 + t^n.$$

Consider the manifold $M = S^m \times S^n$. For $|n - m| \geq 2$ the function

$$h_{m,n} : S^m \times S^n \to \mathbb{R}, \quad S^m \times S^n \ni (x, y) \mapsto h_m(x) + h_n(y),$$

is a Morse function with Morse polynomial

$$P_{h_{m,n}}(t) = P_{h_m}(t)P_{h_n}(t) = 1 + t^m + t^n + t^{m+n},$$

and since $|n - m| \geq 2$, we deduce that it is a perfect Morse function. □

Example 2.21. Consider the complex projective space $\mathbb{CP}^n$ with projective coordinates $[z_0, \ldots, z_n]$ and define

$$f : \mathbb{CP}^n \to \mathbb{R}, \quad f([z_0, z_1, \ldots, z_n]) = \frac{\sum_{j=1}^n j|z_j|^2}{|z_0|^2 + \cdots + |z_n|^2}.$$

We want to prove that f is a perfect Morse function.

The projective space $\mathbb{CP}^n$ is covered by the coordinate charts

$$V_k = \{z_k \neq 0,\}, \quad k = 0, 1, \ldots, n,$$

with affine complex coordinates

$$v^i = v^i(k) = \frac{z_i}{z_k}, \quad i \in \{0, 1, \ldots, n\} \setminus \{k\}.$$

Fix k and set

$$|v|^2 := |v(k)|^2 = \sum_{i \neq k} |v^i|^2.$$

Then,

$$f|_{V_k} = \underbrace{\Big(k + \sum_{j \neq k} j|v^j|^2\Big)}_{=:k+a(v)} \underbrace{(1 + |v|^2)^{-1}}_{=:b(v)}.$$

Observe that $\mathrm{d}b = -b^2\mathrm{d}|v|^2$ and

$$\mathrm{d}f|_{V_k} = bda - (k+a)b^2\mathrm{d}|v|^2 = b^2\sum_{j\neq k}\big(j(1+|v|^2)-(k+a)\big)\mathrm{d}|v^j|^2$$
$$= \sum_{j\neq k}\big((j-k)+(|v|^2-a)\big)\mathrm{d}|v^j|^2.$$

Since

$$\mathrm{d}|v^j|^2 = \bar{v}^j\mathrm{d}v^j + \bar{v}^j\mathrm{d}v^j,$$

and the collection $\{\mathrm{d}v^j, \mathrm{d}\bar{v}^j;\quad j \neq k\}$ defines a trivialization of $T^*V_k \otimes \mathbb{C}$ we deduce that v is a critical point of $f|_{V_k}$ if and only if

$$\big(j(1+|v|^2)-(k+a)\big)v^j = 0,\ \ \forall j \neq k.$$

Hence, $f|_{V_k}$ has only one critical point p_k with coordinates $v(k) = 0$. Near this point we have the Taylor expansions

$$(1+|v|^2)^{-1} = 1-|v|^2+\cdots,$$
$$f|_{V_k} = (k+a(v))(1-|v|^2+\cdots) = k+\sum_{j\neq j}(j-k)|v^j|^2+\cdots.$$

This shows that Hessian of f at p_k is

$$H_{f,p_k} = 2\sum_{j\neq k}(j-k)\Big(x_j^2+y_j^2\Big),\ \ v^j = x_j + y_j\mathbf{i}.$$

Hence, p_k is nondegenerate and has index $\lambda(p_k) := 2k$. This shows that f is a $\mathbb{Q}$-perfect Morse function with Morse polynomial:

$$P_{\mathbb{CP}^n}(t) = P_f(t) = \sum_{j=0}^{n} t^{2j} = \frac{1-t^{2(n+1)}}{1-t^2}.$$

Let us point out an interesting fact which suggests some of the limitations of the homological techniques we have described in this section.

Consider the perfect Morse function $h_{2,4} : S^2 \times S^4 \to \mathbb{R}$ described in Example 2.20. Its Morse polynomial is

$$P_{2,4} = 1 + t^2 + t^4 + t^6$$

and thus coincides with the Morse polynomial of the perfect Morse function $f:\mathbb{CP}^3 \to \mathbb{R}$ investigated in this example. However, $S^2 \times S^4$ is not even homotopic to $\mathbb{CP}^3$, because the cohomology ring of $S^2 \times S^4$ is not isomorphic to the cohomology ring of $\mathbb{CP}^3$. □

Remark 2.22. The above example may give the reader the impression that on any smooth compact manifold, there should exist perfect Morse functions. This is not the case. In Exercise 6.20, we describe a class of manifolds which do not admit perfect Morse functions. The Poincaré sphere is one such example. □

2.4 Morse–Smale Dynamics

Suppose $f : M \to \mathbb{R}$ is a Morse function on the compact manifold M and ξ is a gradient-like vector field relative to f. We denote by Φ_t the flow on M determined by $-\xi$. We will refer to it as the *descending flow* determined by the gradient, like vector field ξ.

Lemma 2.23. *For every $p_0 \in M$ the limits*

$$\Phi_{\pm\infty}(p_0) := \lim_{t\to\pm\infty} \Phi_t(p_0)$$

exist and are critical points of f. □

Proof. Set $\gamma(t) := \Phi_t(p_0)$. If $\gamma(t)$ is the constant path, then the statement is obvious. Assume that $\gamma(t)$ is not constant.

Since $\xi \cdot f \geq 0$ and $\dot{\gamma}(t) = -\xi(\gamma(t))$, we deduce that

$$\dot{f} := \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t)) = \mathrm{d}f(\dot{\gamma}) = -\xi \cdot f \leq 0.$$

From the condition $\xi \cdot f > 0$ on $M \setminus \mathbf{Cr}_f$ and the assumption that $\gamma(t)$ is not constant, we deduce

$$\dot{f}(t) < 0, \quad \forall t.$$

Define $\Omega_{\pm\infty}$ to be the set of points $q \in M$ such that there exists a sequence $t_n \to \pm\infty$ with the property that

$$\lim_{n\to\infty} \gamma(t_n) = q.$$

Since M is compact, we deduce $\Omega_{\pm\infty} \neq \emptyset$. We want to prove that $\Omega_{\pm\infty}$ consist of a single point which is a critical point of f. We discuss only Ω_∞, since the other case is completely similar.

Observe first that

$$\Psi_t(\Omega_\infty) \subset \Omega_\infty, \quad \forall t \geq 0.$$

Indeed, if $q \in \Omega_\infty$ and $\gamma(t_n) \to q$, then

$$\gamma(t_n + t) = \Psi_t(\gamma(t_n)) \to \Psi_t(q) \in \Omega_\infty.$$

Suppose q_0, q_1 are two points in Ω_∞. Then there exists an increasing sequence $t_n \to \infty$ such that

$$\gamma(t_{2n+i}) \to q_i, \quad i = 0, 1, \quad t_{2n+1} \in (t_{2n}, t_{2n+2}).$$

We deduce

$$f(\gamma(t_{2n})) > f(\gamma(t_{2n+1})) > f(\gamma(t_{2n+2})).$$

Letting $n \to \infty$, we deduce $f(q_0) = f(q_1)$, $\forall q_0, q_1 \in \Omega_\infty$, so that there exists $c \in \mathbb{R}$ such that

$$\Omega_\infty \subset f^{-1}(c).$$

If $q \in \Omega_\infty \setminus \mathbf{Cr}_f$, then $t \mapsto \Psi_t(q) \in \Omega_\infty$ is a nonconstant trajectory of $-\xi$ contained in a level set $f^{-1}(c)$. This is impossible since f decreases strictly on such nonconstant trajectories. Hence,

$$\Omega_\infty \subset \mathbf{Cr}_f\,.$$

To conclude it suffices to show that Ω_∞ is connected. Denote by $\mathcal{C}$ the set of connected components of Ω_∞. Assume that $\#\mathcal{C} > 1$. Fix a metric d on M and set

$$\delta := \min\big\{ \operatorname{dist}(C, C');\;\; C, C' \in \mathcal{C},\;\; C \neq C' \big\} > 0.$$

Let $C_0 \neq C_1 \in \mathcal{C}$ and $q_i \in C_i$, $i = 0, 1$. Then, there exists an increasing sequence $t_n \to \infty$ such that

$$\gamma(t_{2n+i}) \to q_i, \quad i = 0, 1, \quad t_{2n+1} \in (t_{2n}, t_{2n+2}).$$

Observe that

$$\lim \operatorname{dist}\big(\gamma(t_{2n}), C_0\big) = \operatorname{dist}(q_0, C_0) = 0,$$
$$\lim \operatorname{dist}\big(\gamma(t_{2n+1}), C_0\big) = \operatorname{dist}(q_1, C_0) \geq \delta.$$

From the intermediate value theorem we deduce that for all $n \gg 0$ there exists $s_n \in (t_{2n}, t_{2n+1})$ such that

$$\operatorname{dist}\big(\gamma(s_n), C_0\big) = \frac{\delta}{2}.$$

A subsequence of $\gamma(s_n)$ converges to a point $q \in \Omega_\infty$ such that $\operatorname{dist}(q, C_0) = \frac{\delta}{2}$. This is impossible since $q \in \Omega_\infty \subset \mathbf{Cr}_f \setminus C_0$. This concludes the proof of Lemma 2.23. □

Suppose $f : M \to \mathbb{R}$ is a Morse function and $p_0 \in \mathbf{Cr}_f$, $c_0 = f(p_0)$. Fix a gradient-like vector field ξ on M and denote by Φ_t the flow on M generated by $-\xi$. We set

$$W_{p_0}^{\pm} = W_{p_0}^{\pm}(\xi) := \Phi_{\pm\infty}^{-1}(p_0) = \left\{ x \in M;\ \lim_{t \to \pm\infty} \Phi_t(x) = p_0 \right\}.$$

$W_{p_0}^{\pm}(\xi)$ is called the *stable/unstable manifold* of p_0 (relative to the gradient-like vector field ξ). We set

$$S_{p_0}^{\pm}(\varepsilon) = W_{p_0}^{\pm} \cap \{ f = c_0 \pm \varepsilon \}.$$

Proposition 2.24. *Let $m = \dim M$, $\lambda = \lambda(f, p_0)$. Then $W_{p_0}^{-}$ is a smooth manifold homeomorphic to $\mathbb{R}^{\lambda}$, while $W_{p_0}^{+}$ is a smooth manifold homeomorphic to $\mathbb{R}^{m-\lambda}$.*

Proof. We will only prove the statement for the unstable manifold, since $-\xi$ is a gradient-like vector field for $-f$ and $W_{p_0}^{+}(\xi) = W_{p_0}^{-}(-\xi)$. We will need the following auxiliary result.

Lemma 2.25. *For any sufficiently small $\varepsilon > 0$, the set $S_{p_0}^{-}(\varepsilon)$ is a sphere of dimension $\lambda - 1$ smoothly embedded in the level set $\{ f = c_0 - \varepsilon \}$ with trivializable normal bundle.*

Proof. Pick local coordinates $x = (x_-, x_+)$ adapted to p_0. Fix $\varepsilon > 0$ sufficiently small so that in the neighborhood

$$U = \left\{ |x_-|^2 + |x_-|^2 < r \right\}$$

the vector field ξ has the form

$$-2x_- \partial_{x_-} + x_+ \partial_{x_+} = -2 \sum_{i \le \lambda} x^i \partial_{x^i} + 2 \sum_{j > \lambda} x^j \partial_{x^j}.$$

A trajectory $\Phi_t(q)$ of $-\xi$, which converges to p_0 as $t \to -\infty$ must stay inside U for all $t \ll 0$. Inside U, the only such trajectories have the form $e^{2t} x_-$, and they are all included in the disk

$$\mathbb{D}_{p_0}^{-}(r) = \left\{ x_+ = 0,\ \ |x_-|^2 \le r \right\}.$$

Moreover, since f decreases strictly on nonconstant trajectories, we deduce that if $\varepsilon < r$, then

$$S_{p_0}^{-}(\varepsilon) = \partial \mathbb{D}_{p_0}^{-}(\varepsilon). \qquad \square$$

Fix now a diffeomorphism $u : S^{\lambda-1} \to S^-_{p_0}(\varepsilon)$. If (r, θ), $\theta \in S^{\lambda-1}$, denote the polar coordinates on $\mathbb{R}^\lambda$, we can define

$$F : \mathbb{R}^\lambda \to W^-_{p_0}, \quad F(r, \theta) = \Phi_{\frac{1}{2}\log r}(u(\theta)).$$

The arguments in the proof of Lemma 2.25 show that F is a diffeomorphism. □

Remark 2.26. The stable and unstable manifolds of a critical point *are not closed subsets of M*. In fact, their closures tend to be quite singular, and one can say that the topological complexity of M is hidden in the structure of these singularities. □

We have the following fundamental result of Smale [Sm].

Theorem 2.27. *Suppose $f : M \to \mathbb{R}$ is a Morse function on a compact manifold. Then there exists a gradient-like vector field ξ such that for any $p_0, p_1 \in \mathbf{Cr}_f$ the unstable manifold $W^-_{p_0}(\xi)$ intersects the stable manifold $W^+_{p_1}(\xi)$ transversally.*

Proof. For the sake of clarity we prove the theorem only in the special case when f is nonresonant, i.e., every level set of f contains at most one critical point. The general case is only notationally more complicated. Let

$$\Delta_f = \{ c_1 < \cdots < c_\nu \}$$

be the set of critical values of f. Denote by p_i the critical point of f on the level set $\{f = c_i\}$. Clearly W^-_p intersects W^+_p transversally at p, $\forall p \in \mathbf{Cr}_f$.

In general, $W^+_{p_i} \cap W^-_{p_j}$ is a union of trajectories of $-\xi$ and

$$W^+_{p_i} \cap W^-_{p_j} \neq \emptyset \Longrightarrow f(p_i) \leq f(p_j) \Longleftrightarrow i \leq j.$$

Note that if r is a regular value of f, then the manifolds $W^\pm_p(\xi)$ intersect the level set $\{f = r\}$ transversally, since ξ is transversal to the level set and tangent to $W^\pm$. For every regular value r we set

$$W^\pm_{p_i}(\xi)_r := W^\pm_{p_i}(\xi) \cap \{f = r\}.$$

Observe that

$$W^-_{p_j}(\xi) \pitchfork W^+_{p_i}(\xi) \Longleftrightarrow W^-_{p_j}(\xi)_r \pitchfork W^+_{p_i}(\xi)_r,$$

for some regular value $f(p_i) < r < f(p_j)$.

For any real numbers $a < b$ such that the interval $[a, b]$ contains only regular values and any gradient-like vector field ξ, we have a diffeomorphism

$$\Phi^\xi_{b,a} : \{f = a\} \longrightarrow \{f = b\}$$

obtained by following the trajectories of the flow of the vector field

$$\langle \xi \rangle := \frac{1}{\xi \cdot f}\xi \tag{2.13}$$

along which f increases at a rate of one unit per second. We denote by $\Phi^{\xi}_{a,b}$ its inverse. Note that

$$W^{\pm}_{p_i}(\xi)_a = \Phi^{\xi}_{a,b}\big(W^{\pm}_{p_i}\xi)_b \big), \quad W^{\pm}_{p_i}(\xi)_b = \Phi^{\xi}_{b,a}\big(W^{\pm}_{p_i}(\xi)_a \big).$$

For every $r \in \mathbb{R}$, we set $M_r := \{f = r\}$.

Lemma 2.28 (The main deformation lemma). *Suppose $a < b$ are such that $[a, b]$ consists only of regular values of f. Suppose $h : M_b \to M_b$ is a diffeomorphism of M_b* isotopic *to the identity. This means that there exists a smooth map*

$$H : [0,1] \times M_b \to M_b, \quad (t, x) \mapsto h_t(x),$$

such that $x \mapsto h_t(x)$ is a diffeomorphism of M_b, $\forall t \in [0, 1]$, $h_0 = \mathbb{1}_{M_b}$, and $h_1 = h$. Then, there exists a gradient-like vector field η for f which coincides with ξ outside $\{a < f < b\}$ and such that the diagram below is commutative:

$$\begin{tikzcd} M_b \arrow[rr, "h"] & & M_b \\ & M_a \arrow[ul, "\Phi^{\xi}_{b,a}"] \arrow[ur, "\Phi^{\eta}_{b,a}"'] & \end{tikzcd}$$

Proof. For the simplicity of exposition we assume that $a = 0$, $b = 1$, and that the correspondence $t \mapsto h_t$ is independent of t for t close to 0 and 1. Note that we have a diffeomorphism

$$\Psi : [0,1] \times M_1 \to \{0 \leq f \leq 1\}, \quad (t, x) \mapsto \Phi^{\xi}_{t,1}(x) \in \{f = t\}.$$

Its inverse is

$$y \mapsto \big(f(y), \Phi^{\xi}_{1,f(y)}(y) \big).$$

Using the isotopy H we obtain a diffeomorphism

$$\hat{H} := [0,1] \times M_1 \to [0,1] \times M_1, \quad \hat{H}(t, x) = (t, h_t(x)).$$

It is now clear that the pushforward of the vector field $\langle \xi \rangle$ in (2.13) via the diffeomorphism

$$F = \Psi \circ \hat{H} \circ \Psi^{-1} : \{0 \leq f \leq 1\} \to \{0 \leq f \leq 1\}$$

is a vector field $\hat{\eta}$, which coincides with $\langle\xi\rangle$ near M_0, M_1, and satisfies $\hat{\eta}\cdot f = 1$. The vector field

$$\eta = (\xi\cdot f)\hat{\eta}$$

extends to a vector field that coincides with ξ outside $\{0 < f < 1\}$ and satisfies $\langle\eta\rangle = \hat{\eta}$. Moreover, the flow of $\langle\eta\rangle$ fits in the commutative diagram

$$\begin{tikzcd} M_1 \arrow[r, "F"] & M_1 \\ M_0 \arrow[u, "\Phi^{\xi}_{1,0}"] \arrow[r, "F"'] & M_0 \arrow[u, "\Phi^{\eta}_{1,0}"'] \end{tikzcd}$$

Now, observe that $F|_{M_0} = \mathbb{1}_{M_0}$ and

$$F_{M_1} = \Phi^{\xi}_{1,1} h_1 \Phi^{\xi}_{1,1} = h_1 = h. \qquad \square$$

Lemma 2.29 (The moving lemma). *Suppose X and Y are smooth submanifolds of the compact smooth manifold V, and X is compact. Then, there exists a diffeomorphism of $h : V \to V$ isotopic to the identity*[2] *such that $h(X)$ intersects Y transversally.* $\square$

We omit the proof which follows from the transversality results in [Hir, Chap. 3] and the isotopy extension theorem [Hir, Chap. 8].

We can now complete the proof of Theorem 2.27. Let $1 \le k \le \nu$. Suppose we have constructed a gradient-like vector field ξ such that

$$W^+_{p_i}(\xi) \pitchfork W^-_{p_j}(\xi), \quad \forall i < j \le k.$$

We will show that for $\varepsilon > 0$ sufficiently small there exists a gradient-like vector field η which coincides with ξ outside the region $\{c_{k+1} - 2\varepsilon < f < c_{k+1} - \varepsilon\}$ and such that

$$W^-_{p_{k+1}}(\eta) \pitchfork W^+_{p_j}(\eta), \quad \forall j \le k.$$

For $\varepsilon > 0$ sufficiently small, the manifold $W^-_{p_{k+1}}(\xi)_{c_{k+1}-\varepsilon}$ is a sphere of dimension $\lambda(p_{k+1}) - 1$ embedded in $\{f = c_{k+1} - \varepsilon\}$. We set

$$a := c_{k+1} - 2\varepsilon, \quad b := c_{k+1} - \varepsilon,$$

and

$$X_b = \bigcup_{j\le k} W^+_{p_j}(\xi)_b.$$

[2]The diffeomorphism h can be chosen to be arbitrarily C^0-close to the identity.

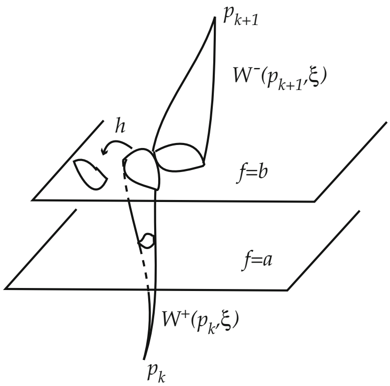

Fig. 2.14 Deforming a gradient-like flow

Using the moving lemma, we can find a diffeomorphism $h : M_b \to M_b$ isotopic to the identity such that (see Fig. 2.14)

$$h(X_b) \pitchfork W^-_{p_{k+1}}(\xi)_b. \tag{2.14}$$

Using the main deformation lemma we can find a gradient-like vector field η, which coincides with ξ outside $\{a < f < b\}$ such that

$$\Phi^{\eta}_{b,a} = h \circ \Phi^{\xi}_{b,a}.$$

Since η coincides with ξ outside $\{a < f < b\}$, we deduce

$$W^+_{p_j}(\eta)_a = W^+_{p_j}(\xi)_a, \quad \forall j \leq k, \quad W^-_{p_{k+1}}(\xi)_b = W^-_{p_{k+1}}(\eta)_b.$$

Now observe that

$$W^+_{p_j}(\eta)_b = \Phi^{\eta}_{b,a} W^+_{p_j}(\eta)_a = h\Phi^{\xi}_{b,a} W^+_{p_j}(\xi)_a = hW^+_{p_j}(\xi)_b,$$

and we deduce from (2.14) that

$$W^+_{p_j}(\eta)_b \pitchfork W^-_{p_{k+1}}(\eta)_b, \quad \forall j \leq k.$$

Performing this procedure gradually, from $k = 1$ to $k = \nu$, we obtain a gradient-like vector field with the properties stipulated in Theorem 2.27. $\square$

Definition 2.30. (a) If $f : M \to \mathbb{R}$ is a Morse function and ξ is a gradient like vector field such that

$$W^-_p(\xi) \pitchfork W^+_q(\xi), \quad \forall p, q \in \mathbf{Cr}_f,$$

then we say that (f,ξ) is a *Morse–Smale pair* on M and that ξ is a *Morse–Smale vector field* adapted to f.

Remark 2.31. Observe that if (f,ξ) is a Morse–Smale pair on M and $p,q \in \mathbf{Cr}_f$ are two *distinct* critical points such that $\lambda_f(p) \leq \lambda_f(q)$, then

$$W_p^-(\xi) \cap W_q^+(\xi) = \emptyset.$$

Indeed, suppose this is not the case. Then

$$\dim W_p^-(\xi) + \dim W_q^+(\xi) = \dim M + \big(\lambda(p) - \lambda(q)\big) \leq \dim M,$$

and because $W_p^-(\xi)$ intersects $W_q^+(\xi)$ transversally, we deduce that

$$\dim\big(W_p^-(\xi) \cap W_q^+(\xi)\big) = 0.$$

Since the intersection $W_p^-(\xi) \cap W_q^+(\xi)$ is flow invariant and $p \neq q$, this zero dimensional intersection must contain at least one nontrivial flow line. □

Definition 2.32. A Morse function $f : M \to \mathbb{R}$ is called *self-indexing* if

$$f(p) = \lambda_f(p), \quad \forall p \in \mathbf{Cr}_f.$$

Theorem 2.33 (Smale). *Suppose M is a compact smooth manifold of dimension m. Then there exist Morse–Smale pairs (f,ξ) on M such that f is self-indexing.*

Proof. We follow closely the strategy in [M4, Sect. 4]. We begin by describing the main technique that allows us to gradually modify f to a self-indexing Morse function.

Lemma 2.34 (Rearrangement lemma). *Suppose $f : M \to \mathbb{R}$ is a Morse function such that $0, 1$ are regular values of f and the region $\{0 < f < 1\}$ contains precisely two critical points p_0 and p_1. Furthermore, assume that ξ is a gradient-like vector field on M such that*

$$W(p_0,\xi) \cap W(p_1,\xi) \cap \{0 \leq f \leq 1\} = \emptyset,$$

where we have used the notation $W(p_i) = W_{p_i}^+ \cup W_{p_i}^-$.

Then for any real numbers $a_0, a_1 \in [0,1]$ there exists a Morse function $g : M \to \mathbb{R}$ with the following properties:

(a) g coincides with f outside the region $\{0 < f < 1\}$.
(b) $g(p_i) = a_i$, $\forall i = 0, 1$.
(c) $f - g$ is constant in a neighborhood of $\{p_0, p_1\}$.
(d) ξ is a gradient-like vector field for g.

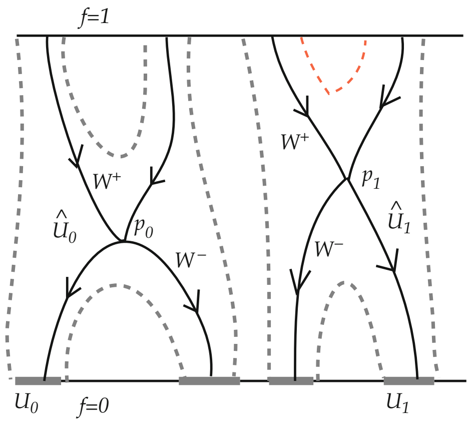

Fig. 2.15 Decomposing a Morse flow

Proof. Let

$$W := \Big(W^+_{p_0}(\xi) \cup W^-_{p_0}(\xi) \cup W^+_{p_1}(\xi) \cup W^-_{p_1}(\xi)\Big) \cap \big\{0 \leq f \leq 1\big\},$$

$$M_0 := \{f = 0\},\ M_0' = M_0 \setminus \big(W^-_{p_0}(\xi) \cup W^-_{p_1}(\xi)\big),$$

$$W^-_{p_i}(\xi)_0 := W^-_{p_i}(\xi) \cap M_0.$$

Denote by $\langle\xi\rangle$ the vector field $\frac{1}{\xi \cdot f}\xi$ on $\{0 \leq f \leq 1\} \setminus W$ and by Φ^ξ_t its flow. Then, Φ^ξ_t defines a diffeomorphism

$$\Psi : [0,1] \times M_0' \to \{0 \leq f \leq 1\} \setminus W, \quad (t,x) \mapsto \Phi^\xi_t(x).$$

Its inverse is

$$y \mapsto \Psi^{-1}(y) = \Big(f(y), \Phi^\xi_{-f(y)}(y)\Big).$$

Choose open neighborhoods U_i of $W^-_{p_i}(\xi)_0$ in M_0 such that $U \cap U' = \emptyset$. This is possible since $W(p_0) \cap W(p_1) \cap M_0 = \emptyset$.

Now, fix a smooth function $\mu : M_0 \to [0,1]$ such that $\mu = i$ on U_i. Denote by $\hat{U}_i$ the set of points y in $\{0 \leq f \leq 1\}$ such that either $y \in W^+_{p_i}(\xi)$, or the trajectory of $-\xi$ through y intersects M_0 in U_i, $i = 0, 1$ (see Fig. 2.15). We can extend μ to a smooth function $\hat{\mu}$ on $\{0 \leq f \leq 1\}$ as follows.

If $y \notin (\hat{U}_0 \cup \hat{U}_1)$, then $\Psi^{-1}(y) = (t,x)$, $x \in M_0 \setminus (U_0 \cup U_1)$, and we set

$$\hat{\mu}(y) := \mu(x).$$

Then, we set $\hat{\mu}(y) = i$, $\forall y \in \hat{U}_i$.

Now, fix a smooth function $G : [0, 1] \times [0, 1] \to [0, 1]$ satisfying the following conditions:

- $\frac{\partial G}{\partial t}(s, t) > 0, \forall 0 \leq s, t \leq 1$.
- $G(s, 0) = 0, G(s, 1) = 1$.
- $G(s, t) = t$ near the segments $t = 0, 1$.
- $G(i, t) - t = (a_i - f(p_i))$ for t near $f(p_i)$.

We can think of G as a one-parameter family of increasing diffeomorphisms

$$G_s : [0, 1] \to [0, 1], \quad s \mapsto G_s(t) = G(s, t)$$

such that $G_0(f(p_0)) = a_0$ and $G_1(f(p_1)) = a_1$.

Now define

$$h : \{0 \leq f \leq 1\} \to [0, 1], \quad h(y) = G(\hat{\mu}(y), f(y)).$$

It is now easy to check that g has all the desired properties. □

Remark 2.35. (a) To understand the above construction it helps us to think of the Morse function f as a clock, i.e., a way of indicating the time when a flow line reaches a point. For example, the time at the point y is $f(y)$.

We can think of the family $s \to G_s$ as one-parameter family of "clock modifiers." If a clock indicates time $t \in [0, 1]$, then by modifying the clock with G_s it will indicate the time $G_s(t)$.

The function h can be perceived as a different way of measuring time, obtained by modifying the "old clock" f using the modifier G_s. More precisely, the new time at y will be $G_{\hat{\mu}(y)}(f(y))$.

(b) The rearrangement lemma works in the more general context, when instead of only two critical points, we have a partition $C_0 \sqcup C_1$ of the set of critical points in the region $\{0 < f < 1\}$ such that f is constant on C_0 and on C_1, and $W(p_0, \xi) \cap W(p_1, \xi) = \emptyset, \forall p_0 \in C_0, \forall p_1 \in C_1$. □

We can now complete the proof of Theorem 2.33. Suppose that (f, ξ) is a Morse–Smale pair on M such that f is nonresonant. Remark 2.31 shows that

$$p \neq q \text{ and } \lambda(p) \leq \lambda(q) \Longrightarrow W_p^-(\xi) \cap W_q^+(\xi) = \emptyset.$$

We say that a pair of critical points $p, q \in \mathbf{Cr}_f$ is an *inversion* if

$$\lambda(p) < \lambda(q) \text{ and } f(p) > f(q).$$

We see that if (p, q) is an inversion, then

$$W_p^-(\xi) \cap W_q^+(\xi) = \emptyset.$$

Using the rearrangement lemma and Theorem 2.27 we can produce inductively a new Morse–Smale pair (g, η) such that $\mathbf{Cr}_g = \mathbf{Cr}_f$, and g is nonresonant and has no inversions.

To see how this is done, define the *level function*

$$\ell_f : \mathbf{Cr}_f \to \mathbb{Z}_{\geq 0}, \quad \ell(p) := \#\big\{q \in \mathbf{Cr}_f;\ \ f(q) < f(p)\big\},$$

denote by $\nu(f)$ the number of inversions of f, and then set

$$\mu(f) = \max\big\{\,\ell_f(q);\ \ (p,q)\ \text{ inversion of } f\,\big\}.$$

If $\nu(f) > 0$, then there exists an inversion (p, q) such that $\ell_f(p) = \mu(f) + 1$ and $\ell_f(q) = \mu(f)$. We can then use the rearrangement lemma to replace f with f' such that $\nu(f') < \nu(f)$.

This implies that there exist regular values $r_0 < r_1 < \cdots < r_m$ such that all the critical points in the region $\{r_\lambda < g < r_{\lambda+1}\}$ have the same index λ.

Using the rearrangement lemma again (see Remark 2.35(b)) we produce a new Morse–Smale pair (h, τ) with critical values $c_0 < \cdots < c_m$, and all the critical points on $\{h = c_\lambda\}$ have the same index λ.

Finally, via an increasing diffeomorphism of $\mathbb{R}$ we can arrange that $c_\lambda = \lambda$. □

Observe that the above arguments prove the following slightly stronger result.

Corollary 2.36. *Suppose (f, ξ) is a Morse–Smale pair on the compact manifold M. Then, we can modify f to a smooth Morse function $g : M \to \mathbb{R}$ with the following properties:*

(a) $\mathbf{Cr}_g = \mathbf{Cr}_f$ *and* $\lambda(f, p) = \lambda(g, p) = g(p)$, $\forall p \in \mathbf{Cr}_f = \mathbf{Cr}_g$.
(b) ξ *is a gradient-like vector field for* g.

In particular, (g, ξ) *is a self-indexing Morse–Smale pair.* □

Here is a simple application of this corollary. We define a *handlebody* to be a three-dimensional manifold with boundary obtained by attaching one-handles to a three-dimensional ball. A *Heegard decomposition* of a smooth, compact, connected 3-manifold M is a quadruple (H_-, H_+, f, Φ) satisfying the following conditions.

- $H_\pm$ are handlebodies.
- f is an orientation reversing diffeomorphism $f : \partial H_- \to \partial H_+$.
- Φ is a homeomorphism from M to the space $H_- \cup_f H_+$ obtained by gluing H_- to H_+ along their boundaries using the identification prescribed by f.

Theorem 2.37. *Any smooth compact connected 3-manifold admits a Heegard decomposition.*

Proof. Fix a self-indexing Morse–Smale pair (f, ξ) on M. The critical values of f are contained in $\{0, 1, 2, 3\}$. To prove the claim in the theorem it suffices to show that the manifolds with boundary

$$H_-(f) := \left\{ f \leq \frac{3}{2} \right\} \text{ and } H_+(f) := \left\{ f \geq \frac{3}{2} \right\}$$

are handlebodies. We do this only for H_-. The case $H_+(f)$ is completely similar since $H_+(f) = H_-(3 - f)$.

Observe first that H_- is connected. Indeed, the connected manifold M is obtained from H_- by attaching 2 and 3-handles and these operations do not change the number of connected components.

The sublevel set $\{f \leq \varepsilon\}$, $\varepsilon \in (0, 1)$, is the disjoint union of a collection of three-dimensional balls, one ball for every minimum point of f. The manifold H_- is obtained from this disjoint union of balls by attaching 1-handles, one for each critical point of index 1.

We can encode this description as a graph Γ. The vertices of Γ correspond to the connected components of $\{f \leq \varepsilon\}$, while the edges correspond to the attached one-handles. The endpoint(s) of an edge indicate how the attaching is performed. The graph Γ may have loops, i.e., edges that start and end at the same vertex. To such a loop it corresponds a 1-handle attached to a single component of $\{f \leq \varepsilon\}$.

Since H_- is connected, so is Γ. Let T be a spanning tree of Γ, i.e., a simply connected subgraph of Γ with the same vertex set as Γ. By attaching first the 1-handles corresponding to the edges of T we obtain a manifold $H(T)$ diffeomorphic to a three-dimensional ball. This shows that H_- is obtained by attaching 1-handles to the three-dimensional ball $H(T)$, so that H_- is a handlebody. □

2.5 Morse–Floer Homology

Suppose that (f, ξ) is a Morse–Smale pair on the compact m-dimensional manifold M such that f is self-indexing. In particular, the real numbers $k + \frac{1}{2}$ are regular values of f. We set

$$M_k = \left\{ f \leq k + \frac{1}{2} \right\}, \quad Y_k = \left\{ k - \frac{1}{2} \leq f \leq k + \frac{1}{2} \right\}.$$

Then, Y_k is a smooth manifold with boundary (see Fig. 2.16)

$$\partial Y_k = \partial_- Y_k \cup \partial_+ Y_k, \quad \partial_\pm Y_k = \left\{ f = k \pm \frac{1}{2} \right\}.$$

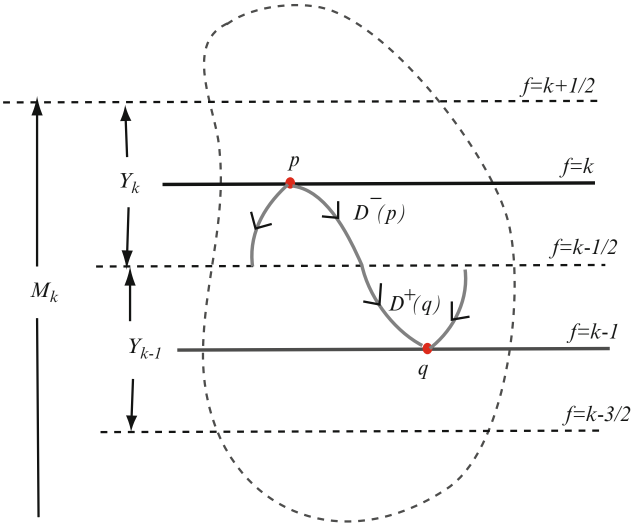

Fig. 2.16 Constructing the Thom–Smale complex

Set

$$C_k(f) := H_k(M_k, M_{k-1}; \mathbb{Z}), \quad \mathbf{Cr}_{f,k} := \big\{ p \in \mathbf{Cr}_f;\ \lambda(p) = k \big\} \subset \{ f = k \}.$$

Finally, for $p \in \mathbf{Cr}_{f,k}$ denote by $D_p^{\pm}$ the unstable disk

$$D_p^{\pm} := W_p^{\pm}(\xi) \cap Y_k.$$

Using the excision theorem and the fundamental structural theorem of Morse theory, we obtain an isomorphism

$$C_k(f) \cong \bigoplus_{p \in \mathbf{Cr}_k} H_k\big(D_p^-, \partial D_p^-; \mathbb{Z} \big).$$

By fixing an orientation $\mathbf{or}^-(p)$ on each unstable manifold W_p^- we obtain isomorphisms

$$H_k(D_p^-, \partial D_p^-; \mathbb{Z}) \to \mathbb{Z}, \quad p \in \mathbf{Cr}_{f,k}\,.$$

We denote by $\langle p|$ the generator of $H_k(D_p^-, \partial D_p^-; \mathbb{Z})$ determined by the choice of orientation $\mathbf{or}^-(p)$.

Observe that we have a natural morphism $\partial : C_k \to C_{k-1}$ defined as the composition

$$H_k(M_k, M_{k-1}; \mathbb{Z}) \to H_{k-1}(M_{k-1}, \mathbb{Z}) \to H_{k-1}(M_{k-1}, M_{k-2}; \mathbb{Z}). \qquad (2.15)$$

Arguing exactly as in the proof of [Ha, Theorem 2.35] (on the equivalence of cellular homology with the singular homology)[3] we deduce that

$$\cdots \to C_k(f) \xrightarrow{\partial} C_{k-1}(f) \to \cdots \tag{2.16}$$

is a chain complex whose homology is isomorphic to the homology of M. This is called the *Thom–Smale complex* associated to the self-indexing Morse function f.

We would like to give a more geometric description of the Thom–Smale complex. More precisely, we will show that it is isomorphic to a chain complex which can be described entirely in terms of Morse data.

Observe first that the connecting morphism

$$\partial_k : H_k(M_k, M_{k-1}) \to H_{k-1}(M_{k-1})$$

can be geometrically described as follows. The relative class $\langle p| \in C_k$ is represented by the fundamental class of the *oriented* manifold with boundary $(D_p^-, \partial D_p^-)$. The orientation $\mathbf{or}_p^-$ induces an orientation on ∂D_p^-, and thus the oriented closed manifold ∂D_p^- defines a homology class in $H_{k-1}(M_{k-1}, \mathbb{Z})$ which represents $\partial\langle p|$.

Assume for simplicity that the ambient manifold M is *oriented*. (As explained in Remark 2.40 (a), this assumption is not needed.) The orientation $\mathbf{or}_M$ on M and the orientation $\mathbf{or}_p^-$ on D_p^- determine an orientation $\mathbf{or}_p^+$ on D_p^+ via the equalities

$$T_pM = T_pD_p^- \oplus T_pD_p^+, \quad \mathbf{or}_p^- \wedge \mathbf{or}_p^+ = or_M.$$

Since ξ is a Morse–Smale gradient-like vector field, we deduce that ∂D_p^- and D_q^+ intersect transversally. In particular, if $p \in \mathbf{Cr}_{f,k}$ and $q \in \mathbf{Cr}_{f,k-1}$, then

$$\dim \partial D_p^- + \dim D_q^+ = (k-1) + \dim M - (k-1) = m,$$

so that ∂D_p^- intersects D_q^+ transversally in finitely many points. We denote by $\langle p|q\rangle$ the *signed* intersection number

$$\langle p|q\rangle := \#\left(\partial D_p^- \cap D_q^+\right), \quad p \in \mathbf{Cr}_{f,k}, \quad q \in \mathbf{Cr}_{f,k-1}.$$

Observe that each point s in $\partial D_p^- \cap D_q^+$ corresponds to a unique trajectory $\gamma(t)$ of the flow generated by $-\xi$ such that $\gamma(-\infty) = p$ and $\gamma(\infty) = q$. We will refer

---

[3] *For the cognoscienti.* The increasing filtration $\cdots \subset M_{k-1} \subset M_k \subset \cdots$ defines an increasing filtration on the singular chain complex $C_\bullet(M, \mathbb{Z})$. The associated (homological) spectral sequence has the property that $E^2_{p,q} = 0$ for all $q > 0$ so that the spectral sequence degenerates at E^2 and the edge morphism induces an isomorphism $H_p(M) \to E^2_{p,0}$. The E^1 term is precisely the chain complex (2.16).

to such a trajectory as a *tunneling* from p to q. Thus $\langle p|q\rangle$ is a signed count of tunnelings from p to q.

Proposition 2.38 (Thom–Smale). *There exist* $\epsilon_k \in \{\pm 1\}$ *such that*

$$\partial\langle p| = \epsilon_k \sum_{q\in \mathbf{Cr}_{f,k-1}} \langle p|q\rangle \cdot \langle q|, \quad \forall p \in \mathbf{Cr}_{f,k}\,. \tag{2.17}$$

Proof. We have

$$\partial\langle p| \in H_{k-1}(M_{k-1}, M_{k-2};\mathbb{Z}) \cong H_{k-1}(Y_{k-1}, \partial_- Y_{k-1};\mathbb{Z}).$$

From the Poincaré–Lefschetz duality theorem, we deduce

$$H_{k-1}(Y_{k-1}, \partial_- Y_{k-1};\mathbb{Z}) \cong H^{m-(k-1)}(Y_{k-1}, \partial_+ Y_{k-1};\mathbb{Z}).$$

Since $H_j(Y_{k-1}, \partial_+ Y_{k-1};\mathbb{Z})$ is a free Abelian group nontrivial only for $j = m - (k-1)$, we deduce that the canonical map

$$H^{m-(k-1)}(Y_{k-1}, \partial_+ Y_{k-1};\mathbb{Z}) \longrightarrow \operatorname{Hom}\big(H_{m-(k-1)}(Y_{k-1}, \partial_+ Y_{k-1};\mathbb{Z}),\ \mathbb{Z}\big)$$

given by the Kronecker pairing is an isomorphism.

The group $H_{m-(k-1)}(Y_{k-1}, \partial_+ Y_{k-1};\mathbb{Z})$ is freely generated by[4]

$$|q\rangle := [D_q^+, \partial D_q^+, \mathbf{or}_q^+], \quad q \in \mathbf{Cr}_{f,k-1}\,.$$

If we view $\partial\langle p|$ as a morphism $H_{m-(k-1)}(Y_{k-1}, \partial_+ Y_{k-1};\mathbb{Z}) \longrightarrow \mathbb{Z}$, then its value on $|q\rangle$ is given (up to a sign ϵ_k which depends only on k) by the above intersection number $\langle p|q\rangle$. □

Given a Morse–Smale pair (f, ξ) on an oriented manifold M and orientations of the unstable manifolds, we can form the *Morse–Floer complex*

$$(C_\bullet(f), \partial), \quad C_k(f) = \bigoplus_{p\in \mathbf{Cr}_k(f)} \mathbb{Z}\cdot\langle p|,$$

where the boundary operator has the tunneling description (2.17). Note that the definitions of $C_k(f)$ and ∂ depend *on* ξ *but not on* f.

In view of Corollary 2.36 we may as well assume that f is self-indexing. Indeed, if this is not the case, we can replace f by a different Morse function g with the same critical points and indices such that g is self-indexing and ξ is a gradient-like vector field for both f and g.

[4]There is no typo! $|q\rangle$ is a *ket* vector and *not* a *bra* vector $\langle q|$.

We conclude that ∂ is indeed a boundary operator, i.e., $\partial^2 = 0$, because it can alternatively be defined as the composition (2.15). We have thus proved the following result.

Corollary 2.39. *For any Morse–Smale pair (f, ξ) on the compact oriented manifold M there exists an isomorphism from the homology of the Morse–Floer complex to the singular homology of M.* □

Remark 2.40. (a) The orientability assumption imposed on M is not necessary. We used it only for the ease of presentation. Here is how one can bypass it.

Choose for every $p \in \mathbf{Cr}_f$ orientations of the vector subspaces $T_p^- M \subset T_p M$ of spanned by the eigenvectors of the Hessian of f corresponding to negative eigenvalues. The unstable manifold W_p^- is homeomorphic to a vector space and its tangent space at p is precisely $T_p^- M$. Thus, the chosen orientation on $T_p^- M$ induces an orientation on W_p^-. Similarly, the chosen orientation on $T_p^- M$ defines an orientation on the normal bundle $T_{W_p^+} M$ of the embedding $W_p^+ \hookrightarrow M$.

Now observe that if X and Y are submanifolds in M intersecting transversally, such that TX is oriented and the normal bundle $T_Y M$ of $Y \hookrightarrow M$ is oriented, then there is a canonical orientation of $X \cap Y$. Indeed, the normal bundle of $X \cap Y \hookrightarrow X$ is naturally isomorphic to the restriction to $X \cap Y$ of the normal bundle of Y in M, i.e., we have a natural short exact sequence of bundles

$$0 \to T(X \cap Y) \hookrightarrow (TX)|_{X \cap Y} \to (T_Y M)|_{X \cap Y} \to 0.$$

Hence, if $\lambda(p) - \lambda(q) = 1$, then $W_p^- \cap W_q^+$ is an oriented one-dimensional manifold.

On the other hand, each component of $W_p^- \cap W_q^+$ is a trajectory of the gradient flow and thus comes with another orientation given by the direction of the flow.

We conclude that on each component of $W_p^- \cap W_q^+$, we have a pair of orientations which differ by a sign ϵ. We can now define $n(p, q)$ to be the sum of all these ϵ's. We then get an operator

$$\hat{\partial} : C_k(f) \to C_{k-1}(f), \quad \hat{\partial}\langle p| = \sum_q n(p, q)\langle q|.$$

One can prove that it coincides, up to an overall sign, with the previous boundary operator.

(b) For different proofs of the above corollary we refer to [BaHu, Sal, Sch].

(c) Corollary 2.39 has one unsatisfactory feature. The isomorphism is not induced by a morphism between the Morse–Floer complex and the singular chain complexes and thus does not highlight the geometric nature of this construction.

For any homology class in a smooth manifold M, the Morse–Smale flow Φ_t on M selects a very special singular chain representing this class. For example, if a

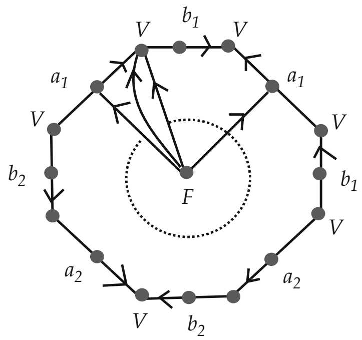

Fig. 2.17 The polyhedral structure determined by a Morse function on a Riemann surface of genus 2

homology class is represented by the singular cycle c, then it is also represented by the cycle $\Phi_t(c)$ and, stretching our imagination, by the cycle $\Phi_\infty(c) = \lim_{t\to\infty} \Phi_t(c)$.

The Morse–Floer complex is, loosely speaking, the subcomplex of the singular complex generated by the family of singular simplices of the form $\Phi_\infty(\sigma)$, where σ is a singular simplex. The supports of such asymptotic simplices are invariant subsets of the Morse–Smale flow and thus must be unions of orbits of the flow.

The isomorphism between the Morse–Floer homology and the singular homology suggests that the subcomplex of the singular chain complex generated by asymptotic simplices might be homotopy equivalent to the singular chain complex. For a rigorous treatment of this idea we refer to [BFK], [Lau], or [HL].

There is another equivalent way of visualizing the Morse flow complex, which goes back to Thom [Th]. Think of a Morse–Smale pair (f, ξ) on M as defining a "polyhedral structure," and then the Morse–Floer complex is the complex naturally associated to this structure. The faces of this "polyhedral structure" are labeled by the critical points of f, and their interiors coincide with the unstable manifolds of the corresponding critical point.

The boundary of a face is a union (with integral multiplicities) of faces of one dimension lower. To better understand this point of view, it helps to look at the simple situation depicted in Fig. 2.17. Let us explain this figure.

First, we have the standard description of a Riemann surface of genus 2 obtained by identifying the edges of an octagon with the gluing rule

$$a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1}.$$

This poyhedral structure corresponds to a Morse function on the Riemann surface, which has the following structure.

- There is a single critical point of index 2, denoted by F, and located in the center of the two-dimensional face. The relative interior of the top face is the unstable manifold of F, and all the trajectories contained in this face will leave F and end up either at a vertex or in the center of some edge.

- There are four critical points of index one, a_1, b_1, b_2, located at the center of the edges labeled by the corresponding letter. The interiors of the edges are the corresponding one-dimensional unstable manifolds. The arrows along the edges describe orientations on these unstable manifolds. The gradient flow trajectories along an edge point away from the center.
- There is a unique critical point of index 0 denoted by V.

In the picture there are two tunnelings connecting F with a_1, but they are counted with opposite signs. In general, we deduce

$$\langle F|a_i\rangle = \langle F|b_j\rangle = 0, \quad \forall i, j.$$

Similarly,

$$\langle a_i|V\rangle = \langle b_j|V\rangle = 0, \quad \forall i, j.$$

The existence of a similar polyhedral structure in the general case was recently established in [Qin]. We refer to Chap. 4 for more details.

(d) The dynamical description of the boundary map of the Morse–Floer complex in terms of tunnelings is due to Witten [Wit] (see the nice story in [B3]), and it has become popular through the groundbreaking work of Floer [Fl]. In Sect. 4.5 we will take a closer look at this dynamical interpretation.

The tunneling approach has been used quite successfully in infinite dimensional situations leading to various flavors of the so-called *Floer homologies*.

These are situations when the stable and unstable manifolds are infinite dimensional yet they intersect along finite dimensional submanifolds. One can still form the operator ∂ using the description in Proposition 2.38, but the equality $\partial^2 = 0$ is no longer obvious, because in this case an alternative description of ∂ of the type (2.15) is lacking. For more information on this aspect, we refer to [ABr, Sch]. □

2.6 Morse–Bott Functions

Suppose $f : M \to \mathbb{R}$ is a smooth function on the m-dimensional manifold M.

Definition 2.41. A smooth submanifold $S \hookrightarrow M$ is said to be a *nondegenerate critical submanifold* of f if the following hold.

- S is compact and *connected*.
- $S \subset \mathbf{Cr}_f$.
- $\forall s \in S$ we have $T_sS = \ker H_{f,s}$, i.e.,

$$H_{f,s}(X, Y) = 0, \quad \forall Y \in T_sM \iff X \in T_sS (\subset T_sM).$$

The function f is called a *Morse–Bott* function if its critical set consists of nondegenerate critical submanifolds. □

Suppose $S \hookrightarrow M$ is a nondegenerate critical submanifold of f. Assume for simplicity that $f|_S = 0$. Denote by $T_S M$ the normal bundle of $S \hookrightarrow M$, $T_S M := (TM)|_S / TS$. For every $s \in S$ and every $X, Y \in T_s S$ we have

$$H_{f,s}(X, Y) = 0,$$

so that the Hessian of f at s induces a quadratic form $Q_{f,s}$ on $T_s M / T_s S = (T_S M)_s$. We thus obtain a quadratic form Q_f on $T_S M$, which we regard as a function on the total space of $T_S M$, quadratic along the fibers.

The same arguments in the proof of Theorem 1.12 imply the following *Morse lemma with parameters.*

Proposition 2.42. *There exist an open neighborhood U of $S \hookrightarrow E = T_S M$ and a smooth open embedding $\Phi : U \to M$ such that $\Phi|_S = \mathbb{1}_S$ and*

$$\Phi^* f = \frac{1}{2} Q_f.$$

If we choose a metric g on E, then we can identify the Hessians $Q_{f,s}$ with a symmetric automorphism $Q : E \to E$. This produces an orthogonal decomposition

$$E = E^+ \oplus E^-,$$

where $E_\pm$ is spanned by the eigenvectors of H corresponding to positive/negative eigenvalues. If we denote by $r_\pm$ the restriction to $E_\pm$ of the function

$$u(v, s) = g_s(v, v),$$

then we can choose the above Φ so that

$$\Phi^* f = -u_- + u_+.$$

The topological type of $E^\pm$ is independent of the various choices, and thus it is an invariant of (S, f) denoted by $E^\pm(S)$ or $E^\pm(S, f)$. We will refer to $E^-(S)$ as the negative normal bundle *of S. In particular, the rank of E^- is an invariant of S called the* Morse index *of the critical submanifold S, and it is denoted by $\lambda(f, S)$. The rank of E^+ is called the* Morse coindex *of S, and it is denoted by $\hat{\lambda}(f, S)$.* □

Definition 2.43. Let $\mathbb{F}$ be a field. The $\mathbb{F}$-*Morse–Bott polynomial* of a Morse–Bott function $f : M \to \mathbb{R}$ defined on the compact manifold M is the polynomial

$$P_f(t) = P_f(t; \mathbb{F}) = \sum_S t^{\lambda(f,S)} P_{S,\mathbb{F}}(t),$$

where the summation is over all the critical submanifolds of f. Note that the Morse–Bott polynomial of a Morse function coincides with the Morse polynomial defined earlier. □

Arguing exactly as in the proof of the fundamental structural theorem, we obtain the following result.

Theorem 2.44 (Bott). *Suppose $f : M \to \mathbb{R}$ is an exhaustive smooth function and $c \in \mathbb{R}$ is a critical value such that $\mathbf{Cr}_f \cap f^{-1}(c)$ consists of finitely many critical submanifolds $S_1, \dots, S_k$. For $i = 1, \dots, k$ denote by $D^-_{S_i}$ the (closed) unit disk bundle of $E^-(S_i)$ (with respect to some metric on $E^-(S_i)$). Then for $\varepsilon > 0$ the sublevel set $M^{c+\varepsilon} = \{f \leq c+\varepsilon\}$ is homotopic to the space obtained from $M^{c-\varepsilon} = \{f \leq c-\varepsilon\}$ by attaching the disk bundles $D^-_{S_i}$ to $M^{c-\varepsilon}$ along the boundaries $\partial D^-_{S_i}$. In particular, for every field $\mathbb{F}$ we have an isomorphism*

$$H_\bullet(M^{c+\varepsilon}, M^{c-\varepsilon}; \mathbb{F}) = \bigoplus_{i=1}^{k} H_\bullet(D^-(S_i), \partial D^-(S_i); \mathbb{F}). \tag{2.18}$$

□

Let $\mathbb{F}$ be a field and X a compact CW-complex. For a real vector bundle $\pi : E \to X$ of rank r over X, we denote by $D(E)$ the unit disk bundle of E with respect to some metric. We say that E is $\mathbb{F}$-orientable if there exists a cohomology class

$$\tau \in H^r(D(E), \partial D(E); \mathbb{F})$$

such that its restriction to each fiber $(D(E)_x, \partial D(E)_x)$, $x \in X$ defines a generator of the relative cohomology group $H^r(D(E)_x, \partial D(E)_x; \mathbb{F})$. The class τ is called the *Thom class* of E associated to a given orientation.

For example, every vector bundle is $\mathbb{Z}/2$-orientable, and every complex vector bundle is $\mathbb{Q}$-orientable. Every real vector bundle over a simply connected space is $\mathbb{Q}$-orientable.

The *Thom isomorphism theorem* states that if the vector bundle $\pi : E \to X$ is $\mathbb{F}$-orientable, then for every $k \geq 0$ the morphism

$$H^k(X, \mathbb{F}) \ni \alpha \longmapsto \tau_E \cup \pi^*\alpha \in H^{k+r}(D(E), \partial D(E); \mathbb{F})$$

is an isomorphism for any $k \in \mathbb{Z}$. Equivalently, the transpose map

$$H_{k+r}(D(E), \partial D(E); \mathbb{F}) \to H_k(X, \mathbb{F}), \quad c \mapsto \pi_*(c \cap \tau_E)$$

is an isomorphism. This implies

$$P_{D(E), \partial D(E)}(t) = t^r P_X(t). \tag{2.19}$$

Definition 2.45. Suppose $\mathbb{F}$ is a field, and $f : M \to \mathbb{R}$ is a Morse–Bott function. We say that f is $\mathbb{F}$-*orientable* if for every critical submanifold S the bundle $E^-(S)$ is $\mathbb{F}$-orientable. □

Corollary 2.46. *Suppose $f : M \to \mathbb{R}$ is an $\mathbb{F}$-orientable Morse–Bott function on the compact manifold. Then, we have the Morse–Bott inequalities*

$$P_f(t) \succ P_{M,\mathbb{F}}(t).$$

In particular,

$$\sum_S (-1)^{\lambda(f,S)} \chi(S) = P_f(-1) = P_M(-1) = \chi(M).$$

Proof. Denote by $c_1 < \cdots < c_\nu$ the critical values of f and set

$$t_k = \frac{c_k + c_{k+1}}{2}, \;\; k = 1, \nu - 1, \;\; t_0 = c_1 - 1, \;\; t_\nu = c_\nu + 1, \;\; M_k = \{f \le t_k\}.$$

As explained in Remark 2.16, we have an inequality

$$\sum_k P_{M_k, M_{k-1}} \succ P_M.$$

Using the equality (2.18), we deduce

$$\sum_k P_{M_k, M_{k-1}} = \sum_S P_{D_S^-, \partial D_S^-},$$

where the summation is over all the critical submanifolds of f. Since $E^-(S)$ is orientable for every S, we deduce from (2.19) that

$$P_{D_S^-, \partial D_S^-} = t^{\lambda(f,S)} P_S.$$ □

Definition 2.47. Suppose $f : M \to \mathbb{R}$ is a Morse–Bott function on a compact manifold M. Then f is called $\mathbb{F}$-*completable* if for every critical value c and every critical submanifold $S \subset f^{-1}(c)$ the inclusion

$$\partial D_S^- \to \{f \le c - \varepsilon\}$$

induces the trivial morphism in homology. □

Arguing exactly as in the proof of Proposition 2.17 we obtain the following result.

Theorem 2.48. *Suppose $f : M \to \mathbb{R}$ is a $\mathbb{F}$-completable, $\mathbb{F}$-orientable, Morse–Bott function on a compact manifold. Then, f is $\mathbb{F}$-perfect, i.e., $P_f(t) = P_M(t)$.* □

Corollary 2.49. *Suppose $f : M \to \mathbb{R}$ is an orientable Morse–Bott function such that for every critical submanifold M we have $\lambda(f, S) \in 2\mathbb{Z}$ and $P_S(t)$ is even, i.e.,*

$$b_k(S) \neq 0 \Longrightarrow k \in 2\mathbb{Z}.$$

Then, f is $\mathbb{Q}$-perfect and thus $P_f(t) = P_M(t)$.

Proof. Using the same notation as in the proof of Corollary 2.46, we deduce by induction over k from the long exact sequences of the pairs (M_k, M_{k-1}) that $b_j(M_k) = 0$ if j is odd, and we have short exact sequence

$$0 \to H_j(M_{k-1}) \to H_j(M_k) \to H_j(M_k, M_{k-1}) \to 0$$

if j is even. □

2.7 Min–Max Theory

So far we have investigated how to use information about the critical points of a smooth function on a smooth manifold to extract information about the manifold itself. In this section we will turn the situation on its head. We will use topological methods to extract information about the critical points of a smooth function.

To keep the technical details to a minimum so that the geometric ideas are as transparent as possible, we will restrict ourselves to the case of a smooth function f on a *compact, connected* smooth manifold M without boundary equipped with a Riemannian metric g.

We can substantially relax the compactness assumption, and the same geometrical principles we will outline below will still apply, but that will require additional technical work.

Morse theory shows that if we have some information about the critical points of f we can obtain lower estimates for their number. For example, if all the critical points are nondegenerate, then their number is bounded from below by the sum of Betti numbers of M. What happens if we drop the nondegeneracy assumption? Can we still produce interesting lower bounds for the number of critical points?

We already have a very simple lower bound. Since a function on a compact manifold must have a minimum and a maximum, it must have at least two critical points. This lower bound is in some sense optimal because the height function on the round sphere has precisely two critical points. This optimality is very unsatisfactory since, as pointed out by Reeb in [Re], if the only critical points of f are (nondegenerate) minima and maxima, then M must be homeomorphic to a sphere.

Min–max theory is quite a powerful technique for producing critical points that often are saddle type points. We start with the basic structure of this theory. For simplicity we denote by M^c the sublevel set $\{f \leq c\}$.

The min–max technology requires a special input.

Definition 2.50. A collection of *min–max data* for the smooth function

$$f : M \to \mathbb{R}$$

is a pair $(\mathcal{H}, \mathcal{S})$ satisfying the following conditions.

- $\mathcal{H}$ is a collection of homeomorphisms of M such that for every regular value a of M there exist $\varepsilon > 0$ and $h \in \mathcal{H}$, such that

$$h(M^{a+\varepsilon}) \subset M^{a-\varepsilon}.$$

- $\mathcal{S}$ is a collection of subsets of M, such that

$$h(S) \in \mathcal{S}, \quad \forall h \in \mathcal{H}, \quad \forall S \in \mathcal{S}. \qquad \square$$

The key existence result of min–max theory is the following.

Theorem 2.51 (Min–max principle). *If $(\mathcal{H}, \mathcal{S})$ is a collection of min–max data for the smooth function $f : M \to \mathbb{R}$, then the real number*

$$c = c(\mathcal{H}, \mathcal{S}) := \inf_{S \in \mathcal{S}} \sup_{x \in S} f(x)$$

is a critical value of f.

Proof. We argue by contradiction. Assume that c is a regular value. Then, there exist $\varepsilon > 0$ and $h \in \mathcal{H}$, such that

$$h(M^{c+\varepsilon}) \subset M^{c-\varepsilon}.$$

From the definition of c, we deduce that there exists $S \in \mathcal{S}$, such that $\sup_{x \in S} f(x) < c + \varepsilon$, that is,

$$S \subset M^{c+\varepsilon}.$$

Then, $S' = h(S) \in \mathcal{S}$ and $h(S) \subset M^{c-\varepsilon}$. It follows that $\sup_{x \in S'} f(x) \leq c - \varepsilon$, so that

$$\inf_{S' \in \mathcal{S}} \sup_{x \in S'} f(x) \leq c - \varepsilon.$$

This contradicts the choice of c as a min–max value. $\square$

The usefulness of the min–max principle depends on our ability to produce interesting min–max data. We will spend the remainder of this section describing a few classical constructions of min–max data.

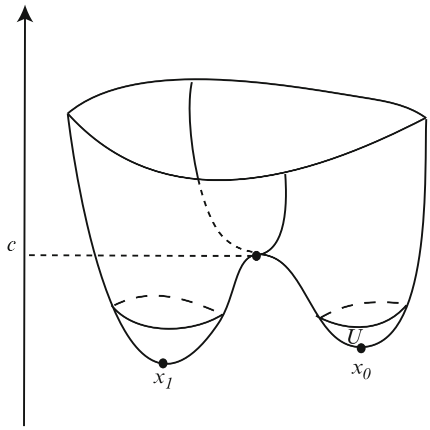

Fig. 2.18 A mountain pass from x_0 to x_1

In all these constructions, the family of homeomorphisms $\mathcal{H}$ will be the same. More precisely, we fix gradient-like vector field ξ and we denote by Φ_t the flow generated by $-\xi$. The condition (a) in the definition of min–max data is clearly satisfied for the family

$$\mathcal{H}_f := \{\Phi_t;\ \ t \geq 0\}.$$

Constructing the family $\mathcal{S}$ requires much more geometric ingenuity.

Example 2.52. Suppose $\mathcal{S}$ is the collection

$$\mathcal{S} = \Big\{ \{x\};\ \ x \in M \Big\}.$$

The condition (b) is clearly satisfied, and in this case we have

$$c(\mathcal{H}_f, \mathcal{S}) = \min_{x \in M} f(x).$$

This is obviously a critical value of f. □

Example 2.53 (Mountain-Pass points). Suppose x_0 is a strict local minimum of f, i.e., there exists a small, closed geodesic ball U centered at $x_0 \in M$, such that

$$c_0 = f(x_0) < f(x),\ \ \forall x \in U \setminus \{x_0\}.$$

Note that

$$c_0' := \min_{x \in \partial U} f(x) > c_0.$$

Assume that there exists another point $x_1 \in M \setminus U$ such that (see Fig. 2.18)

$$c_1 = f(x_1) \leq f(x_0).$$

Now denote by $\mathcal{P}_{x_0}$ the collection of smooth paths $\gamma : [0,1] \to M$, such that

$$\gamma(0) = x_0, \quad \gamma(1) \in M^{c_0} \setminus U.$$

The collection $\mathcal{P}_{x_0}$ is nonempty, since M is connected and $x_1 \in M^{c_0} \setminus U$. Observe that for any $\gamma \in \mathcal{P}_{x_0}$ and any $t \geq 0$, we have

$$\Phi_t \circ \gamma \in \mathcal{P}_{x_0}.$$

Now define

$$\mathcal{S} = \Big\{ \gamma([0,1]); \quad \gamma \in \mathcal{P}_{x_0} \Big\}.$$

Clearly the pair $(\mathcal{H}_f, \mathcal{S})$ satisfies all the conditions in Definition 2.50, and we deduce that

$$c = \inf_{\gamma \in \mathcal{P}_{x_0}} \max_{s \in [0,1]} f(\gamma(s))$$

is a critical value of f such that $c \geq c_0' > c_0$ (see Fig. 2.18).

This statement is often referred to as the *mountain-pass lemma* and critical points on the level set $\{f = c\}$ are often referred to as *mountain-pass points*. Observe that the Mountain Pass Lemma implies that if a smooth function has two strict local minima, then it must admit a third critical point.

The search strategy described in the mountain-pass lemma is very intuitive if we think of f as a height function. The point x_0 can be thought of as a depression and the boundary ∂U as a mountain range surrounding x_0. We look at all paths γ from x_0 to points of lower altitude, and on each of them we pick a point x_γ of greatest height. Then, we select the path γ such that the point x_γ has the smallest possible altitude.

It is perhaps instructive to give another explanation of why there should exist a critical value greater than c_0. Observe that the sublevel set M^{c_0} is disconnected while the manifold M is connected. The change in the topological type in going from M^{c_0} to M can be explained only by the presence of a critical value greater than c_0. □

To produce more sophisticated examples of min–max data we will use a technique pioneered by Lusternik and Schnirelmann. Denote by $\mathcal{C}_M$ the collection of closed subsets of M. For a closed subset $C \subset M$ and $\varepsilon > 0$ we denote by $N_\varepsilon(C)$ the open tube of radius ε around C, i.e., the set of points in M at distance $< \varepsilon$ from C, with respect to a fixed Riemann metric on M.

Definition 2.54. An *index theory* on M is a map

$$\gamma : \mathcal{C}_M \to \bar{\mathbb{Z}}_{\geq 0} := \{0, 1, \ldots\} \cup \{\infty\}$$

satisfying the following conditions.

- **Normalization.** For every $x \in M$, there exists $r = r(x) > 0$ such that

$$\gamma(\{x\}) = 1 = \gamma(\overline{N_\varepsilon(x)}), \quad \forall x \in M, \quad \forall \varepsilon \in (0, r).$$

- **Topological invariance.** If $f : M \to M$ is a homeomorphism, then

$$\gamma(C) = \gamma(f(C)), \quad \forall C \in \mathcal{C}_M.$$

- **Monotonicity.** If $C_0, C_1 \in \mathcal{C}_M$ and $C_0 \subset C_1$, then $\gamma(C_0) \leq \gamma(C_1)$.
- **Subadditivity.** $\gamma(C_0 \cup C_1) \leq \gamma(C_0) + \gamma(C_1)$.

□

Suppose we are given an index theory $\gamma : \mathcal{C}_M \to \bar{\mathbb{Z}}_{\geq 0}$. For every positive integer k we define

$$\Gamma_k := \Big\{ C \in \mathcal{C}_M; \;\; \gamma(C) \geq k \Big\}.$$

The axioms of an index theory imply that for each k the pair $(\mathcal{H}_f, \Gamma_k)$ is a collection of min–max data. Hence, for every k the min–max value

$$c_k = \inf_{C \in \Gamma_k} \max_{x \in C} f(x)$$

is a critical value. Since

$$\Gamma_1 \supset \Gamma_2 \supset \cdots,$$

we deduce that

$$c_1 \leq c_2 \leq \cdots.$$

Observe that the decreasing family $\Gamma_1 \supset \Gamma_2 \supset \cdots$ stabilizes at Γ_m, where $m = \gamma(M)$. If by accident it happens that

$$c_1 > c_2 > \cdots > c_{\gamma(M)},$$

then we could conclude that f has at least $\gamma(M)$ critical points. We want to prove that this conclusion holds even if some of these critical values are equal.

Theorem 2.55. *Suppose that for some $k, p > 0$ we have*

$$c_k = c_{k+1} = \cdots = c_{k+p} = c,$$

and denote by K_c the set of critical points on the level set c. Then either c is an isolated critical value of f and K_c contains at least $p+1$ critical points or c is an accumulation point of $\mathbf{Cr}_f$, i.e., there exists a sequence of critical values $d_n \neq c$ converging to c.

Proof. Assume that c is an isolated critical value. We argue by contradiction. Suppose K_c contains at most p points. Then, $\gamma(K_c) \le p$. At this point we need a deformation result whose proof is postponed. Set

$$T_r(K_c) := \overline{N_r(K_c)}.$$

Lemma 2.56 (Deformation lemma). *Suppose c is an isolated critical value of f and $K_c = \mathbf{Cr}_f \cap \{f = c\}$ is finite. Then for every $\delta > 0$, there exist $0 < \varepsilon, r < \delta$ and a homeomorphism $h = h_{\delta,\varepsilon,r}$ of M such that*

$$h\big(\overline{M^{c+\varepsilon} \setminus T_r(K_c)}\big) \subset M^{c-\varepsilon}.$$

Consider ε, r sufficiently small as in the deformation lemma. Then the normalization and subadditivity axioms imply

$$\gamma(T_r(K_c)) \le \gamma(K_c) = p.$$

We choose $C \in \Gamma_{k+p}$ such that

$$\max_{x \in C} f(x) \le c_{k+p} + \varepsilon = c + \varepsilon.$$

Note that

$$C \subset T_r(K_c) \cup \overline{C \setminus T_r(K_c)},$$

and from the subadditivity of the index we deduce

$$\gamma(\overline{C \setminus T_r(K_c)}) \ge \gamma(C) - \gamma(T_r(K_c)) \ge k.$$

Hence

$$\gamma\big(h(\overline{C \setminus T_r(K_c)})\big) = \gamma(\overline{C \setminus T_r(K_c)}) \ge k,$$

so that

$$C' := h(\overline{C \setminus T_r(K_c)}) \in \Gamma_k.$$

Since

$$\overline{C \setminus T_r(K_c)} \subset \overline{M^{c+\varepsilon} \setminus T_r(K_v)},$$

we deduce from the deformation lemma that

$$C' \subset M^{c-\varepsilon}.$$

Now observe that the condition $C' \in \Gamma_k$ implies

$$c = c_k \le \max_{x \in C'} f(x),$$

which is impossible since $C' \subset M^{c-\varepsilon}$. $\square$

Proof of the deformation lemma. The strategy is a refinement of the proof of Theorem 2.6. The homeomorphism will be obtained via the flow determined by a carefully chosen gradient-like vector field.

Fix a Riemannian metric g on M. For r sufficiently small, $N_r(K_c)$ is a finite disjoint union of open geodesic balls centered at the points of K_c. Let $r_0 > 0$ such that $N_{r_0}(K_c)$ is such a disjoint union and the only critical points of f in $N_{r_0}(K_c)$ are the points in K_c. Fix ε_0 such that c is the only critical value in the interval $[c-\varepsilon_0, c+\varepsilon_0]$. For $r \in (0, r_0)$ define

$$b = b(r) := \inf\bigl\{ |\nabla f(x)|, \quad x \in M^{c+\varepsilon_0} \setminus (M^{c-\varepsilon_0} \cup N_{r/8}(K_c)) \bigr\} > 0.$$

Choose $\varepsilon = \varepsilon(r) \in (0, \varepsilon_0)$ satisfying.

$$2\varepsilon < \min\Bigl(\frac{b(r)r}{8}, b(r)^2, 1 \Bigr) \Longrightarrow \frac{2\varepsilon}{b(r)} < \frac{r}{8}, \quad \frac{2\varepsilon}{\min(1, b(r)^2)} \leq 1. \tag{2.20}$$

Define smooth cutoff functions

$$\alpha : M \to [0,1], \quad \beta : M \to [0,1]$$

such that

- $\alpha(x) = 0$ if $|f(x) - c| \geq \varepsilon_0$ and $\alpha(x) = 1$ if $|f(x) - c| \leq \varepsilon$;
- $\beta(x) = 1$ if $\mathrm{dist}\,(x, K_c) \geq r/4$ and $\beta(x) = 0$ if $\mathrm{dist}\,(x, K_c) < r/8$.

Finally, define a rescaling function

$$\varphi : [0,\infty) \to [0,\infty), \quad \varphi(s) := \begin{cases} 1 & s \in [0,1], \\ s^{-1} & s \geq 1. \end{cases}$$

We can now construct the vector field ξ on M by setting

$$\xi(x) := -\alpha \cdot \beta \cdot \varphi\bigl(|\nabla^g f|^2 \bigr) \nabla^g f.$$

Observe that ξ vanishes outside the region $\{c - \varepsilon_0 < f < c - \varepsilon_0\}$ and also vanishes in $\frac{r}{8}$-neighborhood of K_c. This vector field is not smooth, but it still is Lipschitz continuous. Note also that (Fig. 2.19)

$$|\xi(x)| \leq 1, \quad \forall x \in M.$$

The existence theorem for ODEs shows that for every $x \in M$ there exist $T_\pm(x) \in (0,\infty]$ and a C^1-integral curve $\gamma_x : (-T_-(x), T_+(x)) \to M$ of ξ through x,

$$\gamma_x(0) = x, \quad \dot{\gamma}_x(t) = \xi(\gamma_x(t)), \quad \forall t \in (-T_-(x), T_+(x)).$$

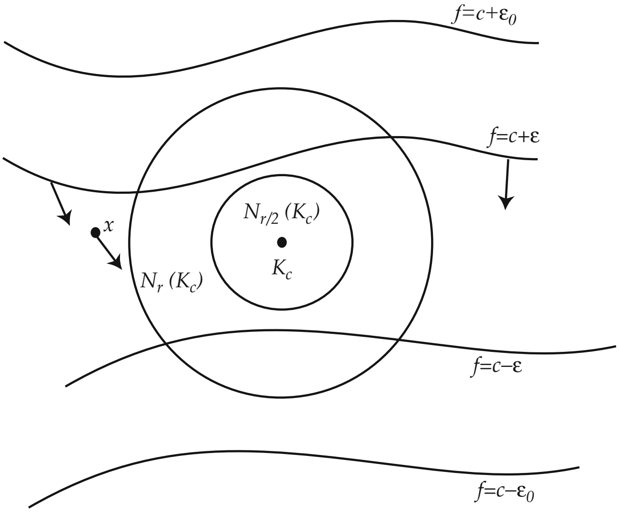

Fig. 2.19 A gradient-like flow

The compactness of M implies that the integral curves of ξ are defined for all $t \in \mathbb{R}$, i.e., $T_{\pm}(x) = \infty$. In particular, we obtain a (topological) flow Φ_t on M. To prove the deformation lemma it suffices to show that

$$\Phi_1\big(M^{c+\varepsilon} \setminus N_r(K_c)\big) \subset M^{c-\varepsilon}.$$

Note that by construction we have

$$\frac{d}{dt} f(\Phi_t(x)) \leq 0, \quad \forall x \in M,$$

so that

$$\Phi_1(M^{c-\varepsilon}) \subset M^{c-\varepsilon}.$$

Let $x \in M^{c+\varepsilon} \setminus \big(N_r(K_c) \cup M^{c-\varepsilon}\big)$. We need to show that $\Phi_1(x) \in M^{c-\varepsilon}$. We will achieve this in several steps.

For simplicity we set $x_t := \Phi_t(x)$. Consider the region

$$Z = \big\{c - \varepsilon \leq f \leq c + \varepsilon\big\} \setminus N_{r/2}(K_c),$$

and define

$$\mathcal{T}_x := \big\{t \geq 0; \quad x_s \in Z, \quad \forall s \in [0, t]\big\}.$$

Clearly $\mathcal{T}_x \neq \emptyset$.

Step 1. We will prove that if $t \in \mathcal{T}_x$, then

$$\operatorname{dist}(x, x_s) < \frac{r}{8}, \quad \forall s \in [0, t].$$

In other words, during the time interval $\mathcal{T}_x$, the flow line $t \mapsto x_t$ cannot stray too far from its initial point.

Observe that α and β are equal to 1 in the region Z and thus for every $t \in \mathcal{T}_x$ we have

$$\begin{aligned} 2\varepsilon \geq f(x) - f(x_t) &= -\int_0^t g\big(\nabla f(x_s), \xi(x_s)\big)\,\mathrm{d}s \\ &= \int_0^t |\nabla f(x_s)|^2 \varphi\big(|\nabla f(x_s)|^2\big)\mathrm{d}s \\ &\geq b(r)\int_0^t |\nabla f(x_s)|\varphi\big((|\nabla f(x_s)|^2\big)\mathrm{d}s = b(r)\int_0^t \left|\frac{\mathrm{d}x_s}{\mathrm{d}s}\right|\mathrm{d}s \\ &\geq b(r)\cdot \operatorname{dist}(x, x_t). \end{aligned}$$

From (2.20) we deduce

$$\operatorname{dist}(x, x_t) \leq \frac{2\varepsilon}{b(r)} < \frac{r}{8}.$$

Step 2. We will prove that there exists $t > 0$ such that $\Phi_t(x) \in M^{c-\varepsilon}$. Loosely speaking, we want to show that there exists a moment of time t when the energy $f(x_t)$ drops below $c - \varepsilon$. Below this level the rate of decrease in the energy f will pickup.

We argue by contradiction, and thus we assume $f(x_t) > c - \varepsilon$, $\forall t > 0$. Thus

$$0 \leq f(x) - f(x_t) \leq 2\varepsilon, \quad \forall t > 0.$$

Since $x_s \in \{c - \varepsilon \leq f \leq c + \varepsilon\}$, $\forall s \geq 0$, we deduce

$$\mathcal{T}_x = \Big\{ t \geq 0;\ \ \operatorname{dist}(x_s, K_c) \geq \frac{r}{2},\ \ \forall s \in [0, t] \Big\}.$$

Hence

$$\operatorname{dist}(x_t, K_c) \geq \operatorname{dist}(x, K_c) - d(x, x_t) > r - \frac{r}{8}, \quad \forall t \in \mathcal{T}_x$$

This implies that $T = \sup \mathcal{T}_x = \infty$. Indeed, if $T < \infty$, then

$$\begin{aligned} &\operatorname{dist}(x_T, K_c) \geq r - \frac{r}{8} > \frac{r}{2} \\ \Longrightarrow\ &\operatorname{dist}(x_t, K_c) > \frac{r}{2}, \quad \forall t \ \text{ sufficiently close to } T. \end{aligned}$$

This contradicts the maximality of T. We deduce

$$x_t \in Z \Longleftrightarrow c - \varepsilon < f(x_t) \leq c + \varepsilon,\ \ \operatorname{dist}(x_t, K_c) > \frac{r}{2}, \quad \forall t \geq 0.$$

This is impossible, since there exists a positive constant ν such that

$$|\xi(x)| > \nu, \quad \forall x \in Z,$$

which implies that

$$\frac{\mathrm{d} f(x_t)}{\mathrm{d}t} \leq -b(r)\nu \Longrightarrow \lim_{t\to\infty} f(x_t) = -\infty,$$

which is incompatible with the condition $0 \leq f(x) - f(x_t) \leq 2\varepsilon$ for every $t \geq 0$.

Step 3. We will prove that $\Phi_1(x) \in M^{c-\varepsilon}$ by showing that there exists $t \in (0,1]$ such that $x_t \in M^{c-\varepsilon}$. Let

$$t_0 := \inf\{t \geq 0;\ \ x_t \in M^{c-\varepsilon}\}.$$

From Step 2 we see that t_0 is well defined and $f(x_{t_0}) = c - \varepsilon$. We claim that the path

$$[0, t_0] \ni s \mapsto x_s$$

does not intersect the neighborhood $N_{r/2}(K_c)$, i.e.,

$$\operatorname{dist}(x_s, K_c) \geq \frac{r}{2}, \quad \forall s \in [0, t_0].$$

Indeed, from Step 1 we deduce

$$\operatorname{dist}(x_s, K_c) > r - \frac{r}{8}, \quad \forall s \in [0, t_0).$$

Now observe that

$$\frac{\mathrm{d} f(x_s)}{\mathrm{d}s} = -|\nabla f|^2 \varphi(|\nabla f|^2) \geq -\max(1, b(r)^2).$$

Thus, for every $s \in [0, t_0]$ we have

$$f(x) - f(x_s) \geq s\max(1, b(r)^2) \Longrightarrow f(x_s) \leq c + \varepsilon - s\max(1, b(r)^2).$$

If we let $s = t_0$ in the above inequality and use the equality $f(x_{t_0}) = c - \varepsilon$, we deduce

$$c - \varepsilon \leq c + \varepsilon - t_0\max(1, b(r)^2) \Longrightarrow t_0 \leq \frac{2\varepsilon}{\max(1, b(r)^2)} \overset{(2.20)}{\leq} 1.$$

This completes the proof of the deformation lemma. □

We now have the following consequence of Theorem 2.55.

Corollary 2.57. *Suppose $\gamma : \mathcal{C}_M \to \bar{\mathbb{Z}}_{\geq 0}$ is an index theory on M. Then any smooth function on M has at least $\gamma(M)$ critical points.* $\square$

To complete the story we need to produce interesting index theories on M. It turns out that the Lusternik–Schnirelmann category of a space is such a theory.

Definition 2.58. (a) A subset $S \subset M$ is said to be *contractible* in M if the inclusion map $S \hookrightarrow M$ is homotopic to the constant map.
(b) For every closed subset $C \subset M$ we define its *Lusternik–Schnirelmann category of C in M* and denote it by $\mathrm{cat}_M(C)$, to be the smallest positive integer k such that there exists a cover of C by closed subsets

$$S_1, \dots, S_k \subset M$$

that are contractible in M. If such a cover does not exist, we set

$$\mathrm{cat}_M(C) := \infty. \qquad \square$$

Theorem 2.59 (Lusternik–Schnirelmann). *If M is a compact smooth manifold, then the correspondence*

$$\mathcal{C}_M \ni C \mapsto \mathrm{cat}_M(C)$$

defines an index theory on M. Moreover, if R denotes one of the rings $\mathbb{Z}/2, \mathbb{Z}, \mathbb{Q}$ then

$$\mathrm{cat}(M) := \mathrm{cat}_M(M) \geq \mathrm{CL}\,(M, R) + 1,$$

where $\mathrm{CL}\,(M, R)$ *denotes the* cuplength *of M with coefficients in R, i.e., the largest integer k such that there exist*

$$\alpha_1, \dots, \alpha_k \in H^\bullet(M, R)$$

with the property that

$$\prod_{j=1}^{k} \deg \alpha_j \neq 0, \quad \alpha_1 \cup \cdots \cup \alpha_k \neq 0.$$

Proof. It is very easy to check that cat_M satisfies all the axioms of an index theory: normalization, topological invariance, monotonicity, and subadditivity, and we leave this task to the reader. The lower estimate of $\mathrm{cat}(M)$ requires a bit more work. We argue by contradiction. Let

$$\ell := \mathrm{CL}\,(M, R)$$

and assume that $\mathrm{cat}(M) \leq \ell$. Then, there exist $\alpha_1, \dots, \alpha_\ell \in H^\bullet(M, R)$ and closed sets $S_1, \dots, S_\ell \subset M$, contractible in M, such that

$$M = \bigcup_{k=1}^{\ell} S_k, \quad \alpha_1 \cup \cdots \cup \alpha_\ell \neq 0, \quad \prod_{j=1}^{k} \deg \alpha_j \neq 0.$$

Denote by j_k the inclusion $S_k \hookrightarrow M$.

Since S_k is contractible in M, we deduce that the induced map

$$j_k^* : H^\bullet(M, R) \to H^\bullet(S_k, R)$$

is trivial. In particular, the long exact sequence of the pair (M, S_k) shows that the natural map

$$i_k : H^\bullet(M, S_k; R) \to H^\bullet(M)$$

is onto. Hence there exist $\beta_k \in H^\bullet(M, S_k)$ such that

$$i_k(\beta_k) = \alpha_k.$$

Now we would like to take the cup products of the classes β_k, but we hit a technical snag. The cup product in *singular* cohomology,

$$H^\bullet(M, S_i; R) \times H^\bullet(M, S_j; R) \to H^\bullet(M, S_i \cup S_j; R),$$

is defined only if the sets S_i, S_j are "reasonably well behaved" ("excisive" in the terminology of [Spa, Sect. 5.6]). Unfortunately, we cannot assume this. There are two ways out of this technical conundrum. Either we modify the definition of cat_M to allow only covers by closed, contractible, and *excisive* sets, or we work with a more supple concept of cohomology. We adopt this second option and we choose to work with Alexander cohomology $\bar{H}^\bullet(-, R)$, [Spa, Sect. 6.4].

This cohomology theory agrees with the singular cohomology for spaces which are not too "wild." In particular, we have an isomorphism $\bar{H}^\bullet(M, R) \cong H^\bullet(M, R)$, and thus we can think of the α_k's as Alexander cohomology classes.

Arguing exactly as above, we can find classes $\beta_k \in \bar{H}^\bullet(M, S_k; R)$, such that

$$i_k(\beta_k) = \alpha_k.$$

In Alexander cohomology there is a cup product

$$\cup : \bar{H}^\bullet(M, A; R) \times \bar{H}^\bullet(M, B; R) \to \bar{H}^\bullet(M, A \cup B; R),$$

well defined for *any* closed subsets of M. In particular, we obtain a class

$$\beta_1 \cup \cdots \cup \beta_l \in \bar{H}^\bullet(M, S_1 \cup \cdots \cup S_\ell; R)$$

that maps to $\alpha_1 \cup \cdots \cup \alpha_\ell$ via the natural morphism

$$\bar{H}^\bullet(M, S_1 \cup \cdots \cup S_\ell; R) \to \bar{H}^\bullet(M, R).$$

Now observe that $\hat{H}^\bullet(M, S_1 \cup \cdots, \cup S_\ell; R) = 0$, since $S_1 \cup \cdots \cup S_\ell = M$. We reached a contradiction since $\alpha_1 \cup \cdots \cup \alpha_\ell \neq 0$. □

Example 2.60. Since $\mathrm{CL}\,(\mathbb{RP}^n, \mathbb{Z}/2) = \mathrm{CL}((S^1)^n, \mathbb{Z}) = \mathrm{CL}\,(\mathbb{CP}^n, \mathbb{Z}) = n$, we deduce

$$\operatorname{cat}(\mathbb{RP}^n) \geq n+1, \quad \operatorname{cat}(\,(S^1)^n\,) \geq n+1, \quad \operatorname{cat}(\mathbb{CP}^n) \geq n+1. \qquad \square$$

Corollary 2.61. *Every even smooth function $f : S^n \to \mathbb{R}$ has at least $2(n+1)$ critical points.*

Proof. Observe that f descends to a smooth function $\bar{f}$ on $\mathbb{RP}^n$, which has at least $\operatorname{cat}(\mathbb{RP}^n) \geq n+1$ critical points. Every critical point of $\bar{f}$ is covered by precisely two critical points of f. $\square$

http://www.springer.com/978-1-4614-1104-8

An Invitation to Morse Theory
Nicolaescu, L.
2011, XVI, 353 p. 47 illus., Softcover
ISBN: 978-1-4614-1104-8