Morse Theory and the h-Cobordism Theorem

An honors thesis on classical topics of differential topology

Nikhil Sahoo, UC Berkeley
Advised by David Nadler

May 16, 2020

Acknowledgements

First and foremost, I would like to extend the utmost thanks to David Nadler for advising me and for being an invaluable mentor for the past two years. Throughout Math 250A, independent study, SURF, and this honors thesis, he has taught me such a variety of interesting things (two of my favorites are moment maps and Morse theory). I often joke that his catch-phrase is

"everything is a tautology once you understand it,"

a remark that would sound pretty daunting coming from almost anyone else. But what makes it endearing is that he always makes sure to transmit such a level of understanding, which is a pedagogical rarity and gift. Math aside, his patience and kindness have made me feel better about making mistakes and helped me to see myself as a valued member of our academic community.

I would be remiss if I did not also thank several others who have aided me far beyond their prescribed responsibilities as instructors: Zvezda Stankova, for creating and running a program that fostered my interest in mathematics, for hiring me, so that I could share my interests in turn, and for constantly encouraging and challenging me; Brent Nelson, for teaching me three times before I was even accepted to UC Berkeley, for readily talking to me about all sorts of problems, and for being a great mentor and friend; Julian Chaidez, for introducing me to symplectic geometry and for challenging me to think about hard topics outside of my coursework; Larsen Linov and Kevin O'Neill, for teaching me through DRP and for cementing my interest in topology; Mark Rowles, for introducing me to Banach spaces and for many interesting and helpful conversations over coffee; and Mike Valdez, for navigating me through my first venture outside of high-school math. Of course, there are many other great instructors who taught me at UC Berkeley and the Peralta Community Colleges, in mathematics and other subjects. Thank you all.

I would also like to thank my loved ones who have helped me through this project and all my college years. To my parents: it is impossible to list all that you have done for me, so thank you for the continuum of support. To Clara: thank you for persisting with rollerblade-enhanced superpowers to offer endless morale boosts, and for being the best sister I could ever have hoped for. To Berkeley: thank you for demonstrating the powerful impact of caring in both pedagogy and friendship. To the Kaminskas: thank you for welcoming me into your family with such warmth and candor. To Kamila: thank you for frequently arising from your diegetic worlds to offer much needed support and inspiration; without your encouragement and urging, this project would never have bloomed. Also, thank you for the carrot cake.

Contents

1 Introduction

This document seeks to give an expository introduction to the ideas needed to understand the h-cobordism theorem. Ultimately, the goal is to present the essential ideas at a level accessible to anyone with a course in smooth manifolds and a course in algebraic topology. There are certainly some details that go beyond this and must be omitted or stated without proof. In particular, I have omitted all but the most basic Riemannian geometry.

In what follows, the term "manifold" will refer to a smooth manifold with boundary, usually compact. A "closed manifold" is a compact manifold without boundary. We write $\bar{D}^n$ for the closed unit disk with its standard smooth structure.[1] This manifold has interior D^n and boundary S^{n-1}.

[1] It is not known if any exotic (i.e. different than the standard) structures on $\bar{D}^n$ exist, but they may exist, so it is still necessary to specify this. Moreover, there are known exotic structures on D^4 and S^7, so it is essential that we specify the structure in question.

1.1 The cobordism category

The histories of algebraic topology and differential topology are inextricably tied to cobordism. Poincaré's initial idea for homology involved embedded k-manifolds, which were identified if they "co-bounded" a $(k+1)$-manifold. For a time, this was imprecise and difficult to describe, so the formulations of simplicial, singular and cellular cohomology took the place of cobordism. But in 1954, Thom defined a new extraordinary (co)homology theory that realized Poincaré's ambition (and completely described its coefficient ring), based on Pontryagin's existing work on (co)bordism. This (and other ideas, many also originating with Thom) led to a veritable torrent of fascinating developments in the decade that followed. We will review this history a little more below, but we first acquaint ourselves with the definition of cobordism.

Definition 1.1.1. A triad is a triple $(W; V_0, V_1)$, where W is a compact manifold and $\partial W = V_0 \sqcup V_1$. Two triads $(W; V_0, V_1)$ and $(W'; V_0', V_1')$ are said to be equivalent if there is a diffeomorphism $f : W \to W'$ with $f(V_0) = V_0'$ and $f(V_1) = V_1'$. A cobordism is an equivalence class of triads. Figure 1(a) shows an example of a triad, which defines the "pair of pants" cobordism.

The definition of cobordisms as equivalence classes of triads at first seems unnecessarily pedantic. The distinction is necessary because lots of typical operations, such as gluing, are only well-defined up to a diffeomorphism.[2] But this will generally cause us no trouble whatsoever, and we view triads and cobordisms interchangeably, unless otherwise noted. For example:

Definition 1.1.2. An h-cobordism is a cobordism $(W; V_0, V_1)$, such that the inclusion maps $V_0 \subset W$ and $V_1 \subset W$ are both homotopy equivalences. Two closed manifolds V_0 and V_1 are said to be (h-)cobordant if there exists a(n) (h-)cobordism $(W; V_0, V_1)$.

The main goal of this exposition is to prove the h-cobordism theorem, which states that any two closed manifolds of dimension ≥ 5 are h-cobordant if and only if they are diffeomorphic. The journey will be replete with cutting and pasting operations, which respect smooth structure. They are as follows:

- If we have two triads $c = (W; V_0, V_1)$ and $c' = (W'; V_1, V_2)$, we can glue them along collars of V_1 to form a cobordism $cc' = (W \sqcup_{V_1} W'; V_0, V_2)$,

[2] On the other hand, sometimes we want things to only be defined up to diffeomorphism. In the subject of "topological quantum field theories," we get numerical invariants of closed manifolds, which we view as triads $(W; \emptyset, \emptyset)$. For these to be diffeomorphism-invariants, we must only consider these triads up to equivalence, i.e. as cobordisms.

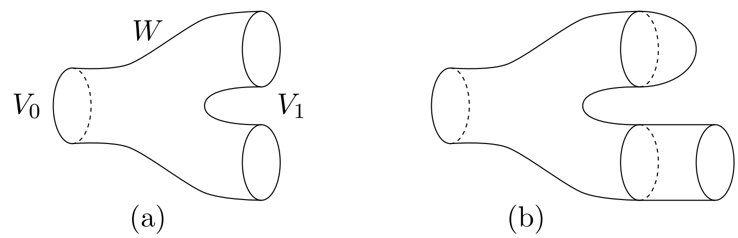


Figure 1: A cobordism and a composition

called their "composition" (this gluing is unique up to diffeomorphism, so the composed cobordism is well-defined). In Figure 1(b), we show the composition of the pair of pants with $\big(D^2 \sqcup (S^1 \times I); S^1 \sqcup S^1, S^1\big)$. Notice that the resulting cobordism is just $(S^1 \times I; S^1, S^1)$.

- Suppose $(W; V_0, V_1)$ is a triad and $f : W \to \mathbb{R}$ is a smooth function, with $f^{-1}(a) = V_0$ and $f^{-1}(b) = V_1$ for some $a < b$. If $c, d \in \mathbb{R}$ are regular values of f and $a \le c < d \le b$, then we can form a new triad
$$\big(f^{-1}[c, d]; f^{-1}(c), f^{-1}(d)\big).$$
In particular, if $a < c < b$, we can decompose $(W; V_0, V_1)$ into triads
$$\big(f^{-1}[a, c]; f^{-1}(a), f^{-1}(c)\big) \quad \text{and} \quad \big(f^{-1}[c, b]; f^{-1}(c), f^{-1}(b)\big).$$
Their composition returns $\big(f^{-1}[a, b]; f^{-1}(a), f^{-1}(b)\big) = (W; V_0, V_1)$.

Given any triad $(W; V_0, V_1)$, note that composing on the left (resp. right) with the cylinder $(V_0 \times I; V_0 \times 0, V_0 \times 1)$ (resp. $(V_1 \times I; V_1 \times 0, V_1 \times 1)$) returns the cobordism $(W; V_0, V_1)$, since the cylinder can be absorbed into a collar neighborhood. Thus we have a "cobordism category," defined as follows:

- objects are diffeomorphism classes of closed n-manifolds;
- morphisms $V_0 \to V_1$ are cobordisms $(W; V_0, V_1)$.

The identity morphism of any V is just the cylinder $(V \times I; V \times 0, V \times 1)$.

Exercise 1.1.3. Show that composition of cobordisms is associative.

Any cobordism $c = (W; V_0, V_1)$ can be "reversed" to form a morphism $\bar{c} = (W; V_1, V_0)$ in the opposite direction. But this operation is *not* inversion. In fact, any invertible cobordism is an h-cobordism [HJ18], and it follows from Theorem 5.0.1 that for any $n \ge 5$, the identity is the only invertible morphism whose domain or codomain is a simply-connected n-manifold. Rather, what this "reversal" operation actually shows is that the cobordism category is isomorphic to its opposite.

1.2 Some history

Now we return to the history of cobordism, after a few more definitions:

Definition 1.2.1. Consider a manifold M and some submanifold $N \subset M$. We call N a neat submanifold if $\partial N = N \pitchfork \partial M$ (transverse intersection). A framing of N is a trivialization of the normal bundle TM/TN.

The tubular neighborhood theorem states that any neat submanifold $N \subset M$ admits a tubular neighborhood, i.e. a neatly embedded open $U \subset M$ containing N, such that the inclusion $N \subset U$ extends to a diffeomorphism of the normal bundle: $TM/TN \cong U$ [Kos93]. In particular, if N is framed, then this tubular neighborhood is just diffeomorphic to $N \times \mathbb{R}^k \cong N \times D^k$, where $k = \dim M - \dim N$. If $\partial N = \emptyset$, then this theorem is easier to prove (see [Lee12]) and we can also find closed tubular neighborhoods (in general, these would only be manifolds with corners). In particular, a neat, closed, framed submanifold N admits a closed neighborhood $N \times \bar{D}^k$, by restricting to vectors of norm ≤ 1 in $N \times \mathbb{R}^k$. We will often use such neighborhoods.

Definition 1.2.2. Let M be a closed n-manifold and $N_0, N_1 \subset M$ be closed, framed submanifolds. We say that N_0 and N_1 are framed cobordant if there is a triad $(W; V_0, V_1)$ and a neat embedding $\varphi : W \to M \times I$, such that $\varphi(V_i) = N_i \times i$ for $i = 0, 1$ and there exists a framing on $\varphi(W) \subset M \times I$, which extends the framings on $N_i \times i \subset M \times I$ for $i = 0, 1$.

Pontryagin first studied framed cobordism, showing that: for any closed manifold M, framed cobordism classes of (framed, closed) k-dimensional submanifolds possess a bijection with homotopy classes of maps $M \to S^k$. This gave a description of $\pi_n(S^k)$ for any $k, n \in \mathbb{N}$, which he used to calculate the stable homotopy groups $\pi_k(S)$ (i.e. $\pi_k(S^{n+k})$, n large) for $k = 1$ and 2. Rokhlin extended these results to describe $\pi_k(S)$ for $k = 3$ and 4 [VK91]. But in this time of extreme flux (the axiomatic characterization of homology was less than ten years old, as were spectral sequences and category theory), the tides of inference soon switched directions entirely. As codimension grew, Pontryagin's approach to stable homotopy became increasingly impractical, especially in comparison to the new algebraic tools that were cropping up. But the bijection could work in the opposite direction, so Thom developed the necessary generalization to form an extraordinary (co)homology theory, whose coefficient ring consisted of the cobordism groups that Pontryagin and Rokhlin had studied. These results are described in [Sto68]; a wonderfully accessible introduction to Pontryagin's construction is given in [Mil65b].

Meanwhile, Milnor was thinking about another question originating with Poincaré: the classification of closed manifolds homotopy equivalent to S^n. In 1956, he gave the first example of a smooth structure on the 7-sphere that is not diffeomorphic to the standard S^7. This was the first real example of a distinction between topological and smooth manifolds. Remarkably, by 1963, Milnor and Kervaire[3] had conceived of a classification of smooth structures on S^n for all $n \geq 5$. Or rather, they classified h-cobordism classes of smooth structures on S^n, which are equivalent to diffeomorphism classes by the h-cobordism theorem, which Smale had proven two years previously. Smale's result provided an answer to the topological classification question of homotopy spheres as well (essentially as a corollary to his main theorem). These amazing results illuminated whole new avenues of research, including the entire field of surgery theory. The objective of this thesis is to present a sketched proof of Smale's result; the exposition closely follows [Mil65a], with the main differences arising from notational preferences and the context provided by results and terminology arising since the publication of [Mil65a].

2 Morse Theory

2.1 Surgery and handles

Suppose $S^{k-1} \subset M$ is a framed sphere inside an n-dimensional manifold M. The framing gives a closed tubular neighborhood $S^{k-1} \times \bar{D}^{n-k+1} \subset M$ and

$$\partial\big(M \setminus (S^{k-1} \times D^{n-k+1})\big) = \partial M \sqcup (S^{k-1} \times S^{n-k}).$$

But we also have $\partial(\bar{D}^k \times S^{n-k}) = S^{k-1} \times S^{n-k}$. Therefore, we can glue $M \setminus (S^{k-1} \times D^{n-k+1})$ and $\bar{D}^k \times S^{n-k}$ along $S^{k-1} \times S^{n-k}$, giving the manifold

$$\big(M \setminus (S^{k-1} \times D^{n-k+1})\big) \sqcup_{S^{k-1} \times S^{n-k}} (\bar{D}^k \times S^{n-k}). \tag{1}$$

Definition 2.1.1. The above operation is called a k-surgery (for reasons that will soon become clear). The resulting manifold is denoted $\chi(M, S^{k-1})$.

By itself, the complex expression (1) is not very enlightening. However, we immediately notice that the resulting manifold just has boundary ∂M; in particular, this construction preserves closed manifolds. Furthermore,

[3]Kervaire proved another important distinction between the topological and smooth categories in 1960, when he constructed a topological 10-manifold that does not admit any smooth structure. Thus Kervaire proved that the functor which forgets smooth structure is not essentially surjective, whereas Milnor proved that it is not "essentially injective."

we can view k-surgery as replacing a framed $(k-1)$-sphere with a framed $(n-k)$-sphere. With this in mind, it seems reasonable that surgery could be used to kill certain homotopy groups, which was its initial use in [Mil61]. Milnor first introduced surgery in this paper, crediting Thom with the idea.

Example 2.1.2. In the following examples, let M denote an n-manifold.

(a) 0-surgery: Any manifold M obviously contains a framed $S^{-1} = \emptyset$. Since $S^{-1} \times S^n = \emptyset$ and $\bar{D}^0 \times S^n = S^n$, we have $\chi(M, S^{-1}) = M \sqcup S^n$.

(b) 1-surgery: Let M and N be connected n-manifolds and let $S^0 \subset M \sqcup N$ consist of two points $p \in M$ and $q \in N$. Then $\chi(M \sqcup N, S^0) = M \# N$. (The framing is simply a choice of ordered bases at T_pM and T_qN. When M and N are both oriented, T_qN should be negatively oriented and T_pM should be positively oriented, with respect to these bases.[4])

(c) For any framed $S^{k-1} \subset M$, we have a framed $S^{n-k} \subset \chi(M, S^{k-1})$ replacing S^{k-1}. Note that $\chi\big(\chi(M, S^{k-1}), S^{n-k}\big)$ is formed as follows:

   i. remove $S^{k-1} \times D^{n-k+1}$;
   ii. glue in $\bar{D}^k \times S^{n-k}$;
   iii. remove $D^k \times S^{n-k}$;
   iv. glue in $S^{k-1} \times \bar{D}^{n-k+1}$.

   Since (iii) cancels (ii) and (iv) cancels (i), $\chi\big(\chi(M, S^{k-1}), S^{n-k}\big) = M$. Therefore, every k-surgery can be reversed by an $(n-k+1)$-surgery.

(d) In Figure 2, we illustrate how to surgically convert a sphere into a torus or a Klein bottle. Combining this with (a)-(c), we see that iterated surgeries can convert any orientable, closed surface into any other.

Now suppose $S^{k-1} \subset \partial M$ is a framed sphere, contained in the boundary of an n-manifold M. Since $\dim(\partial M) = n - 1$, this framing defines a closed tubuluar neighborhood $S^{k-1} \times \bar{D}^{n-k} \subset \partial M$. Similar to the above, note that

$$\partial\big(\bar{D}^k \times \bar{D}^{n-k}\big) = S^{k-1} \times \bar{D}^{n-k} \sqcup_{S^{k-1} \times S^{n-k-1}} \bar{D}^k \times S^{n-k-1}.$$

[4]This may not seem hugely important, but it is! If M or N possesses a diffeomorphism that reverses orientation, then the orientations don't matter in defining $M \# N$. However, this is not always the case. For example, $\mathbb{CP}^2$ does not possess such a diffeomorphism (how would this map act on the cohomology ring?) and it can be shown that $\mathbb{CP}^2 \# \mathbb{CP}^2$ and $\mathbb{CP}^2 \# (-\mathbb{CP}^2)$ are not even homotopy equivalent (they have different signatures).

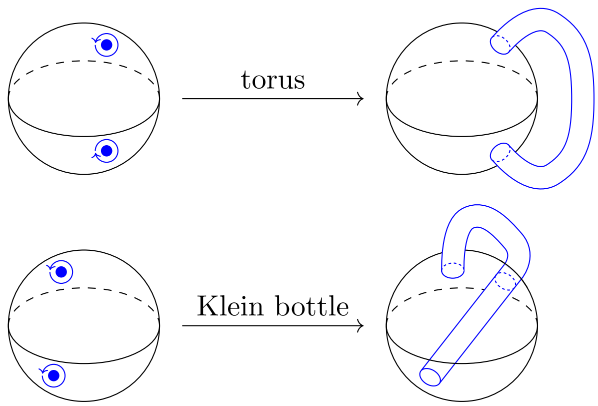


Figure 2: Surgeries on a sphere

Thus we can glue M and $\bar{D}^k \times \bar{D}^{n-k}$ along $S^{k-1} \times \bar{D}^{n-k}$, yielding a manifold

$$M' = M \sqcup_{S^{k-1} \times \bar{D}^{n-k}} (\bar{D}^k \times \bar{D}^{n-k}).$$

Since $\bar{D}^k \times \bar{D}^{n-k}$ has corners, it might seem that the resulting manifold M' only possesses a natural smooth structure on $M' \setminus (S^{k-1} \times S^{n-k-1})$. However, this smooth structure extends uniquely to M'. There are various approaches to this "smoothing" process, but we will not dwell on this irksome detail.[5]

Definition 2.1.3. This smooth manifold M' is said to be formed from M by attaching a smooth k-handle $(\bar{D}^k \times \bar{D}^{n-k})$. We call $\bar{D}^k \times 0$ the core disk and $0 \times \bar{D}^{n-k}$ the belt disk. Similarly, we call $S^{k-1} \times 0$ the attaching sphere and $0 \times S^{n-k-1}$ the belt sphere. (These pieces are all shown in Figure 3.)

Up to homotopy equivalence, attaching a handle is the same as gluing $\bar{D}^k$ to M along the attaching sphere S^{k-1} (deformation retract the handle onto the core disk). Thus attaching a k-handle is the same (up to homotopy) as attaching a k-cell along the same attaching sphere S^{k-1}. In particular, the framing of S^{k-1} does not affect the homotopy type of the manifold M'.

Example 2.1.4. In the following examples, let M denote an n-manifold.

(a) 0-handles: Similarly to Example 2.1.2(a), the boundary ∂M obviously contains a framed $S^{-1} = \emptyset \subset \partial M$. Adding a 0-handle yields $M \sqcup \bar{D}^n$.

[5] The classical approach to smoothing corners is described in the appendix of [Mil59]. A more advanced and comprehensive account can be found in [KS77]. The book [Kos93] presents a nice alternative approach, which never leaves the realm of smooth manifolds.

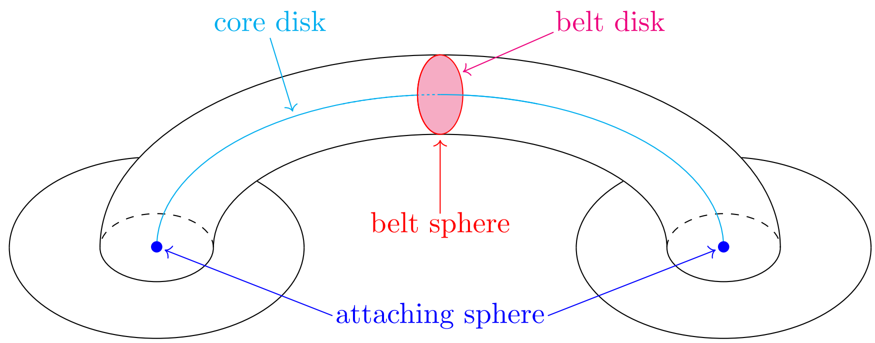


Figure 3: Various parts of a handle

(b) 1-handles: Let M and N be n-manifolds with connected boundaries. As in Example 2.1.2(b), we take $S^0 \subset \partial(M \sqcup N) = \partial M \sqcup \partial N$ to consist of two points $p \in \partial M$ and $q \in \partial N$ (with frames on $T_p\partial M$ and $T_q\partial N$). Attaching a 1-handle along $S^0 = \{p, q\}$ yields the boundary sum $M \natural N$ (as before, the choices of orientation on $T_p\partial M$ and $T_q\partial N$ matter).

The two constructions of surgery and handle attachment are very similar. Indeed, suppose we attach a k-handle to M along a framed $S^{k-1} \subset \partial M$. After this attachment, the tubular neighborhood $S^{k-1} \times D^{n-k}$ disappears from ∂M, because it is now in the interior of the manifold, being glued to

$$S^{k-1} \times D^{n-k} \subset \partial(\bar{D}^k \times \bar{D}^{n-k}).$$

However, the handle attachment also gives a new piece of the boundary:

$$\bar{D}^k \times S^{n-k-1} \subset \partial(\bar{D}^k \times \bar{D}^{n-k}).$$

This piece of boundary is glued to $\partial M \setminus (S^{k-1} \times D^{n-k})$ along $S^{k-1} \times S^{n-k-1}$, so our new manifold M' has boundary $\partial M' = \chi(\partial M, S^{k-1})$ (i.e. attaching a k-handle to a manifold M performs the corresponding k-surgery on ∂M).

Definition 2.1.5. Consider a framed sphere $S^{k-1} \subset M$. Attach a handle to $M \times I$ along the framed sphere $S^{k-1} \times 1 \subset \partial(M \times I)$. The resulting manifold W can be viewed as a triad $\big(W; M, \chi(M, S^{k-1})\big)$, called the trace of the surgery $M \mapsto \chi(M, S^{k-1})$. A simple example is depicted in Figure 4.

The notion of trace helps formalize the assertion that "handle attachment is just surgery on the boundary." Definition 2.1.5 clearly describes k-surgery on M in terms of a k-handle attachment (to the product cobordism $M \times I$).

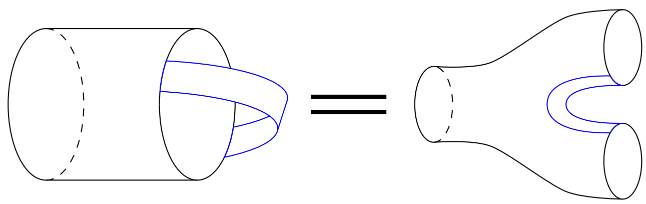


Figure 4: Pair of pants as a trace

Conversely, consider a framed $S^{k-1} \subset \partial N$. The trace $\big(W; \partial N, \chi(\partial N, S^{k-1})\big)$ can be glued to N by identifying the two copies of ∂N. This has the same result as attaching a k-handle along S^{k-1} (since $\partial N \times I \subset W$ can be viewed as a collar of ∂N, to which the handle is attached).[6] Hence, we can view surgery and handle attachment as two sides of the same coin. But there are still some important distinctions, which are not to be haphazardly conflated:

- Any compact manifold can be formed by successively attaching handles to $\emptyset$ (we will prove this shortly). But if M is surgically equivalent to $\emptyset$, then the traces of these surgeries compose to form a triad $(W; \emptyset, M)$. Therefore, if M can be formed by applying successive surgeries to $\emptyset$, then $M = \partial W$ (this is certainly not always possible).

- In Example 2.1.2(c), we noted that surgeries can always be reversed. This is definitely not the case for handles. (What could we possibly glue to a 0-handle $\bar{D}^n$ to get back to $\emptyset$?) Propositions 2.3.1 and 2.3.2 imply that a trace is never an h-cobordism and thus never invertible [HJ18].

  But if a surgery has trace $(W; V_0, V_1)$, its reverse has trace $(W; V_1, V_0)$. Hence, while W can be viewed as $V_0 \times I$ with a k-handle attached, "reversed" W can be viewed as $V_1 \times I$ with an $(n-k)$-handle attached. These facts will follows from Proposition 2.3.1.

- For surgery, we saw in Example 2.1.2(d) (and in the footnote to (b)) that the framing of S^k does impact the homotopy type of $\chi(M, S^k)$. But for handle attachment, we noted that the framing of the attaching sphere does not affect the homotopy type of the resulting manifold. The issue is that, for handle attachment, we must consider $(M, \partial M)$. The homotopy type of this pair *does* depend on the framing of S^k.

[6] Of course, it seems a little nonsensical to say that traces can be used to define handles, when handles have been used to define traces. But in [Mil65a], traces of surgeries are given an independent (entirely smooth) definition, while handles receive no explicit mention.

Exercise 2.1.6. Find an example to demonstrate that the homotopy type of a pair $(M, \partial M)$ resulting from handle attachment does depend on the framing of the attaching sphere.

Example 2.1.7. Figure 5 shows one way to build a torus out of handles, starting with $\emptyset$. We begin with a 0-handle, successively attach two 1-handles, then add a 2-handle. In the four corresponding surgeries on the boundary, the latter two simply reverse the former two. However, the torus resulting as the composition of their traces is certainly non-trivial.

2.2 Morse functions

We will now describe Morse theory, which uses "nice" functions $f : M \to \mathbb{R}$ on a manifold to understand its topology. Originated by Marston Morse, this theory captures the topology of the manifold in the critical points of f, which are finite in number when M is compact. Ultimately, this will help elucidate the relationship between cobordism and surgery/handles.

Let $f : M \to \mathbb{R}$ be a smooth function and let $p \in M$ be a critical point. Then $df : M \to T^*M$ is a section of the cotangent bundle, vanishing at p. We can identify $T_{(p,0)}T^*M = T_pM \oplus T_p^*M$, so the derivative at p is a map

$$D_p(df) : T_pM \to T_pM \oplus T_p^*M.$$

Projection onto the second summand defines the Hessian $H : T_pM \to T_p^*M$ (equivalently, we can view this as a bilinear form $H : T_pM \otimes T_pM \to \mathbb{R}$).

If $(x_1, \ldots, x_n)$ are local coordinates near $p \in M$, then we have local coordinates $(x_1, \ldots, x_n, dx_1, \ldots, dx_n)$ near $(p, 0) \in T^*M$. Then we can write

$$df(a_1, \ldots, a_n) = \left(a_1, \ldots, a_n, \frac{\partial f}{\partial x_1}(a_1), \ldots, \frac{\partial f}{\partial x_n}(a_n)\right).$$

Its derivative is just $D_p(df) = (I_n \mid \mathcal{H})$, where I_n is the identity matrix and

$$\mathcal{H} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{1 \le i,j \le n.} \tag{2}$$

Projecting onto the last n coordinates, we see that the matrix $\mathcal{H}$ locally defines the form $H : T_pM \otimes T_pM \to \mathbb{R}$. In particular, this H is symmetric.

There are (at least) two other coordinate-free definitions of the Hessian: by $H(v, w) = w(Vf)$, where V is any vector field extending $v \in T_pM$ [Mil63], or by $H(v, w) = \langle w, \nabla_v(df)\rangle$ where ∇ is any connection on T^*M [Hut02].

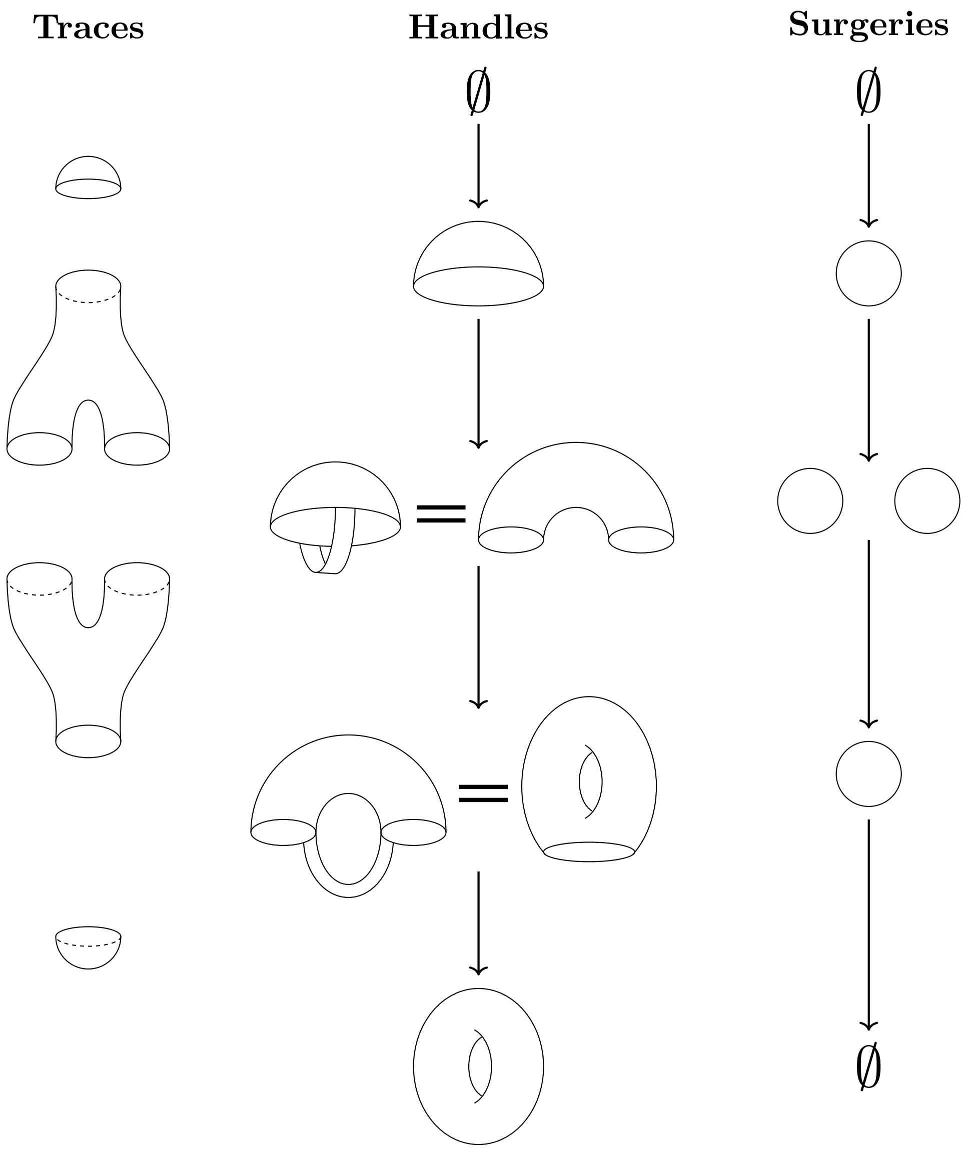


Figure 5: Building a torus from handles

Exercise 2.2.1. In local coordinates, prove that the latter two definitions of H are both described by (2). Prove directly that all three definitions are independent of arbitrary choices made, and that they are symmetric forms. (This all depends notably on p being a critical point.[7])

For a symmetric bilinear form H on a finite-dimensional vector space V:

- The nullity of H is the dimension of the largest subspace of V on which H is identically zero. If we view H as a linear transformation $V \to V^*$, then the nullity of H is just the dimension of the kernel.
- H is non-degenerate if one of the following, equivalent conditions hold: H has nullity zero; for any $v \in V$, there exists $w \in V$ with $H(v, w) \neq 0$; the linear transformation $V \to V^*$ defined by H is an isomorphism; the graph of $H : V \to V^*$ is transverse to the domain $V \times 0 \subset V \times V^*$.
- The index of H is the dimension of the largest subspace of V on which H is negative definite.

With an appropriate choice of basis on V, the form H will be represented by a diagonal matrix with entries in $\{-1, 0, 1\}$. Sylvester's law of inertia states that the number of times each value occurs on the diagonal is independent of the choice of basis. We can read the nullity and index from this matrix:

Exercise 2.2.2. Show that the nullity (resp. index) is the number of times that 0 (resp. -1) occurs on the diagonal. Use this to prove Sylvester's law.

Definition 2.2.3. Consider a function $f : M \to \mathbb{R}$. A critical point p of f is said to be non-degenerate if the Hessian $H : T_pM \otimes T_pM \to \mathbb{R}$ of f at p is non-degenerate. The index $\operatorname{ind}(p)$ of the point p is just the index of H. Also note that we will write $\operatorname{Crit}(f)$ for the set of all critical points of f, and $\operatorname{Crit}_k(f)$ for the set of all (non-degenerate) critical points of index k.

In first defining the Hessian, we described $D_p(df) : T_pM \to T_pM \oplus T_p^*M$ by $v \mapsto \big(v, H(v)\big)$, so the image of $D_p(df)$ is the graph of $H : T_pM \to T_p^*M$. The form H is non-degenerate if and only if its graph is transverse to T_pM,

[7] We only need the first description of H, but the definition in terms of connections provides useful context. The form H should capture some sort of total second derivative, but the differential df is a section of the cotangent bundle T^*M, so differentiation at p requires a horizontal subspace complementary to the vertical space $T_p^*M \subset T_{(p,df_p)}T^*M$. This subspace can be defined canonically (by the zero section M) if and only if $df_p = 0$, which is why H is only defined at critical points. However, if we have a given connection (e.g. the Levi-Civita connection for some metric), then the Hessian is defined everywhere.

and thus the critical point p is non-degenerate if and only if the differential $df : M \to T^*M$ is transverse to the zero section at p.

Our use of the Hessian is entirely captured in the following lemma.

Lemma 2.2.4 (Morse). *Suppose that $f : \mathbb{R}^m \to \mathbb{R}$ is a smooth function, and 0 is a non-degenerate critical point of index k. Then there is a smooth chart about 0, on which $f(x) = f(0) - x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_m^2$.*

Proof. This proof is adapted from [Mil63]. Consider a smooth $\varphi : U \to \mathbb{R}^n$ with $\varphi(0) = 0$, where $U \subset \mathbb{R}^m$ is an open set. We write φ_i for i^{th} component of the function φ, so that we may define smooth functions $a_{ij} : \mathbb{R}^m \to \mathbb{R}$ by

$$a_{ij}(x) = \int_0^1 \frac{\partial \varphi_i}{\partial x_j}(tx)dt.$$

Then $a_{ij}(0) = \frac{\partial \varphi_i}{\partial x_j}(0)$. Using the smooth, matrix-valued map $A(x) = [a_{ij}(x)]$, we can write this as $A(0) = D_0\varphi$. Moreover, the chain rule implies that

$$\varphi_i(x) = \int_0^1 \frac{d}{dt}\varphi_i(tx)dt = \int_0^1 \left(\sum_{j=0}^m \frac{\partial \varphi_i}{\partial x_j}(tx) \cdot x_i \right) dt = \sum_{j=0}^m a_{ij}(x)x_i. \quad (3)$$

Applying (3) for each $i = 1, \ldots, n$, we have $\varphi(x) = A(x) \cdot x$ (i.e. evaluation of the matrix $A(x)$ on the vector x). This is called the Hadamard lemma.

Clearly, we may assume that $f(0) = 0$. Applying the Hadamard lemma, there is a smooth $g : \mathbb{R}^m \to \mathbb{R}^m$ with $g(0) = D_0 f = 0$ and $f(x) = g(x) \cdot x$. Applying the Hadamard lemma again to g, there is a smooth, matrix-valued function $H(x) = [h_{ij}(x)]$ with $g(x) = H(x) \cdot x$. Therefore, we can write

$$f(x) = x^T \cdot H(x) \cdot x = \sum_{1 \le i,j \le m} x_i x_j h_{ij}(x).$$

Possibly replacing H by $\frac{1}{2}(H^T + H)$, we may assume that H is symmetric. Applying the product rule, we can see that $2H(0)$ is the Hessian of f at 0, which is non-singular by assumption. Up to a linear change of coordinates, we may assume that $h_{11}(0) \neq 0$ ($H(0)$ is symmetric and thus diagonalizable). Hence, there is a smaller neighborhood $V \ni 0$ on which $h_{11}(x) \neq 0$. Define

$$y_1(x) = \frac{\sqrt{|h_{11}(x)|}}{h_{11}(x)} \sum_{i=1}^m x_i h_{i1}(x)$$

on this set V. Using the product rule, we calculate $\frac{\partial y_1}{\partial x_1}(0) = \sqrt{|h_{11}(0)|} \neq 0$. Hence, $(y_1, x_2, \ldots, x_m)$ form smooth coordinates about 0, in some smaller neighborhood W. Let $k_{ij} = h_{i1}h_{j1}/h_{11}$ and $\tilde{h}_{ij} = h_{ij} - k_{ij}$. Then we have

$$\begin{aligned}
\operatorname{sgn}\big(h_{11}(x)\big) y_1(x)^2 = \frac{\operatorname{sgn}\big(h_{11}(x)\big)}{|h_{11}(x)|}\left(\sum_{i=1}^{m} x_i h_{i1}(x)\right)^2 &= \sum_{1 \leq i,j \leq m} \frac{x_i x_j h_{i1}(x) h_{j1}(x)}{h_{11}(x)} \\
&= \sum_{1 \leq i,j \leq m} x_i x_j k_{ij}(x) = f(x) - \sum_{1 \leq i,j \leq m} x_i x_j \tilde{h}_{ij}(x).
\end{aligned}$$

But if $i = 1$ or $j = 1$, then $\tilde{h}_{ij} = 0$. Therefore, on the chart W, we have

$$f(y_1, x_2, \ldots, x_m) = \pm\, y_1^2 + \sum_{2 \leq i,j \leq m} x_i x_j \tilde{h}_{ij}(y_1, x_2, \ldots, x_m).$$

Continuing in this fashion gives a coordinate system $(y_1, \ldots, y_m)$ in some neighborhood of 0, such that $f(y_1, \ldots, y_m) = \pm y_1^2 \pm \cdots \pm y_m^2$. By calculating the Hessian in these coordinates, we can easily see that $\operatorname{ind}(0) = k$ is precisely the number of minuses in the expression for f, as desired. $\square$

Corollary 2.2.5. *Consider a smooth manifold M and a smooth $f : M \to \mathbb{R}$. If p is a non-degenerate critical point of f, then p is an isolated critical point.*

The local picture in Lemma 2.2.4 is the model for our "nice" functions.

Definition 2.2.6. Fix a triad $(W; V_0, V_1)$. A smooth function $f : W \to [a, b]$ is called Morse if its critical points are non-degenerate, $\operatorname{Crit}(f) \cap \partial W = \emptyset$, $f^{-1}(a) = V_0$ and $f^{-1}(b) = V_1$. By Corollary 2.2.5, $\operatorname{Crit}(f)$ is a finite set (because triads are compact), so we may define the Morse number:

$$\mu(W; V_0, V_1) = \min\{\#\operatorname{Crit}(f) \mid f : W \to \mathbb{R} \text{ is Morse}\}.$$

In particular, we say that:

- a triad with $\mu(W; V_0, V_1) = 0$ is a product cobordism;
- a triad with $\mu(W; V_0, V_1) = 1$ is an elementary cobordism.

Of course, for the Morse number to be defined at all, we must ensure that Morse functions exist. This is proved below, following Chapter 2 of [Mil65a].

Proposition 2.2.7. *Fix a triad $(W; V_0, V_1)$. Let $\mathcal{F}$ consist of all smooth $f : W \to I$ with no critical points on ∂W, $f^{-1}(0) = V_0$ and $f^{-1}(1) = V_1$. Morse functions form an open, dense subset of $\mathcal{F}$. In particular, there exists a Morse function with no two critical points lying on the same level set.*

Proof. On manifolds without boundary, the existence of Morse functions follows from Sard's lemma and the (weakest) Whitney embedding theorem. But the conditions $f^{-1}(0) = V_0$ and $f^{-1}(1) = V_1$ require a little more work. Hence, we will have to treat some lemmas. We say that a smooth function f is "good" (on a set S), if f has no degenerate critical points (on S).

(a) Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}$ be smooth. For almost all linear $L : \mathbb{R}^n \to \mathbb{R}$, the function $f - L : U \to \mathbb{R}^n$ is good (on its domain U).

   Proof. Consider a smooth map $g : U \to \mathbb{R}$ and the projection map $\pi : T^*U \to (\mathbb{R}^n)^*$ from the standard trivialization $T^*U \cong U \times (\mathbb{R}^n)^*$. Any $p \in U$ is a regular point of $\pi \circ dg : U \to \mathbb{R}^n$ if and only if the image $D_p(dg)(T_pU) \subset T_{(p,dg_p)}(T^*U) = \mathbb{R}^n \times (\mathbb{R}^n)^*$ projects onto the second factor isomorphically. But if p is a critical point of g (i.e. $dg_p = 0$), this is equivalent to the Hessian being non-degenerate at p. Therefore, 0 is a regular value of $\pi \circ dg$ if and only if g is good. Taking $g = f - L$, note that $\pi(dg_p) = \pi\big(d(f - L)_p\big) = \pi(df_p) - L$, so L is a regular value of $\pi \circ df$ if and only if g is good. By applying Sard's lemma to $\pi \circ df$, we see that $g = f - L$ is good for almost all $L \in (\mathbb{R}^n)^*$. $\square$

To speak of an open, dense set in $\mathcal{F}$, we need to give $\mathcal{F}$ a topology. Consider an open set $U \subset \mathbb{R}^n$ and any compact subset $K \subset U$. We define the semi-norm $||f||_\infty = \sup\{|f(x)| : x \in K\}$ on $C^0(U)$. Then, we may define

$$||f|| = ||f||_\infty + \sum_{i=1}^{n} \left|\left|\frac{\partial f}{\partial x_i}\right|\right|_\infty + \sum_{1 \le i,j \le n} \left|\left|\frac{\partial^2 f}{\partial x_i \partial x_j}\right|\right|_\infty$$

This is a semi-norm on $C^2(U)$. Moreover, if $U' \subset \mathbb{R}^n$ is open and $\varphi : U \to U'$ is a diffeomorphism with $\varphi(K) = K'$, then the pairs (U, K) and (U', K') define equivalent semi-norms (we omit the proof, which is not too hard; see [Mil65a] for a quick proof or [Hir76] for a more detailed account).

Let $\{U_i\}$ be a finite cover of W by charts, and let $\{C_i\}$ be a compact refinement. Let $||\cdot||_i$ be the semi-norm on $C^2(U_i)$ defined above and define

$$||f|| = \sum \big|\big|f|_{U_i}\big|\big|_i$$

This defines a norm on $C^2(W)$ and the resulting topology does not depend on the cover and refinement, by the above equivalence of semi-norms. Thus, we get a well-defined "C^2 topology" on $\mathcal{F}$, independent of $\{U_i\}$ and $\{C_i\}$.

(b) Let $U \subset W$ be a coordinate chart and consider a compact $K \subset U$. Then $N(K) = \{f \in C^2(W) : f \text{ is good on } K\}$ is open in $C^2(W)$.

Proof. Let $H_f(p)$ denote the Hessian of f at $p \in U$ (in the coordinates of the chart U, the Hessian is defined for all p). Then p is a degenerate critical point of f if and only if $||df_p|| + |\det H_f(p)| = 0$. Now, define

$$\mu(f) = \min\left\{||df_p|| + |\det H_f(p)| : p \in K\right\}.$$

Note that f is good on K if and only if $\mu(f) > 0$. We will show that $\mu : C^2(W) \to \mathbb{R}$ is a continuous map, from which it follows that $N(K) = \mu^{-1}(0, \infty) \subset C^2(W)$ is open.

Firstly, restriction of functions $C^2(W) \to C^2(U)$ is a continuous map. The following maps $C^2(U) \to C^0(U)$ are also clearly continuous:

$$f \mapsto \frac{\partial f}{\partial x_i} \qquad \text{and} \qquad f \mapsto \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Thus, the maps $C^2(U) \to C^0(U)$ given by $f \mapsto ||df||$ and $f \mapsto |\det H_f|$ are continuous (They are absolute values of polynomials in the first and second partial derivatives). Moreover, the function $C^0(U) \to \mathbb{R}$ given by $f \mapsto \min\{f(p) : p \in K\}$ is clearly continuous. Combining all of these, we see that $\mu : C^2(W) \to \mathbb{R}$ is continous, as desired. □

Consider a compact $K \subset W$. Let $\{U_i\}$ be a finite cover of K by charts with a compact refinement $\{C_i\}$. Each $N(C_i \cap K) \subset C^2(W)$ is open by (b), so their (finite) intersection $N(K)$ is also open. This implies that $N(K) \cap \mathcal{F}$ is open in $\mathcal{F}$. In particular, the Morse functions $N(W) \cap \mathcal{F}$ are open in $\mathcal{F}$.

(c) Let $U \subset W$ be an interior chart and consider a compact set $K \subset U$. For any non-empty, open set $N \subset \mathcal{F}$, we have $N \cap N(K) \neq \emptyset$.

Proof. Choose some $g \in N$. Choose a smooth function $\chi : W \to I$, compactly support in U, with $\chi \equiv 1$ on a neighborhood of K. Viewing the chart U as an open subset of $\mathbb{R}^n$, we may define $g_L : W \to \mathbb{R}$ by

$$g_L(x) = \begin{cases} g(x) - \chi(x)L(x), & x \in U \\ g(x), & x \notin \operatorname{supp}(\chi) \end{cases}$$

for any $L \in (\mathbb{R}^n)^*$. Since the piecewise definitions agree on $U \backslash \operatorname{supp}(\chi)$, this function g_L is smooth. Since $\chi \equiv 1$ on K, we see that $g_L \in N(K)$

if and only if $g|_U - L$ is good on K. We will show that $g_L \in N$ for L sufficiently close to 0. By (a), we can choose L sufficiently close to 0 such that $g|_U - L$ is good on K, which will give us $g_L \in N \cap N(K)$.

Let $N = N' \cap \mathcal{F}$, where $N' \subset C^2(W)$ is open. Note that we can write

$$N = N' \cap \mathcal{F} = \{f \in N' \cap N(\partial W) : f^{-1}(i) = V_i \text{ for } i = 0, 1\}.$$

The function $G : (\mathbb{R}^n)^* \to C^2(W)$ given by $G(L) = g_L$ is continuous (because if $L \in (\mathbb{R}^n)^*$ is small, then $||g - g_L||_\infty = ||\chi L||_\infty$ is also small, as are all first and second partial derivatives), so $G^{-1}\big(N' \cap N(\partial W)\big)$ is an open set containing 0. Moreover, the function g is bounded away from 0 and 1 on $\operatorname{supp}(\chi)$ (since U is an interior chart), so g_L is bounded away from 0 and 1 on U when L is sufficiently small. Thus we have

$$g_L \in N' \cap N(\partial W) \cap \{f \in C^2(M) : f^{-1}(i) = V_i \text{ for } i = 0, 1\} = N$$

for $L \in (\mathbb{R}^n)^*$ sufficiently close to 0, as desired. □

Consider $g \in \mathcal{F}$ (this is vacuous if $\mathcal{F} = \emptyset$) and a neighborhood $N \ni g$. Since g has no critical points on ∂W, there is an open set $X \supset \partial W$ on which g has no critical points. Next, choose open an set $Y \supset \partial W$ with $\overline{Y} \subset X$. Since g is good on Y, we can take the smaller neighborhood $N_0 = N \cap N(\overline{Y})$. Let $\{U_1, \dots, U_n\}$ be a cover of $W \setminus Y$ by interior charts and let $\{C_1, \dots, C_n\}$ be a compact refinement. Applying (c) repeatedly, we see that

$$N \cap N(W) = N_0 \cap N(C_1) \cap \cdots \cap N(C_n) \neq \emptyset,$$

so there is a Morse function in N. Hence, Morse functions are dense in $\mathcal{F}$. But this doesn't prove that Morse functions exist, for we still need $\mathcal{F} \neq \emptyset$.

(d) There is a smooth function $f : W \to I$ with no critical points on ∂W, $f^{-1}(0) = V_0$ and $f^{-1}(1) = V_1$ (i.e. there exists some $f \in \mathcal{F}$).

Proof. Pick a monotone, smooth function $\chi : I \to I$ with $\chi(t) = 1 - t$ for $t < 1/3$ and $\chi(t) = 0$ for $t > 2/3$. Now, choose disjoint collars $\varphi : V_0 \times I \to W$ and $\psi : V_1 \times I \to W$. For $i = 0, 1$, let $\pi_i : V_i \times I \to I$ denote projection onto the second factor. Define $g : W \to [-1, 1]$ by

$$g(x) = \begin{cases} -\chi \circ \pi_0 \circ \varphi^{-1}(x), & x \in \varphi(V_0 \times I) \\ \chi \circ \pi_1 \circ \psi^{-1}(x), & x \in \psi(V_1 \times I) \\ 0, & \text{otherwise.} \end{cases}$$

On the smaller collar neighborhoods $V_0 \times [0, 1/3)$ and $V_1 \times [0, 1/3)$, this function has the forms $g \circ \varphi(x, t) = t - 1$ and $g \circ \psi(x, t) = 1 - t$. Thus g has no critical points on $\partial W = V_0 \sqcup V_1$. Note that $|g| \leq 2/3$ outside of these collars. Therefore, we can take $f = \frac{1}{2}(g + 1)$. □

This completes the proof that Morse functions $f : W \to I$ exist. To show that f can be chosen so that no two critical points lie on the same level set, it suffices to show that f can be modified in a neighborhood of a critical point p, to change $f(p)$ without creating new critical points. We postpone the proof to §3.2, where such adjustments occur in greater generality. □

The most intuitive Morse functions are just the "height" of an embedded submanifold, as in Figure 6. There are four critical points: the saddles have index 1, while the maximum has index 2 and the minimum has index 0. This notion of "height" is general enough, since we can take an embedding $\varphi : M \to \mathbb{R}^n$ and a Morse function $f : M \to \mathbb{R}$ and form a new embedding $\varphi \times f : M \to \mathbb{R}^{n+1}$, for which projection onto the last coordinate is just f. And while there is no reason for an arbitrary "height" function to have only non-degenerate critical points, this is true for a generic direction. Therefore, if M is closed, then this (combined with the Whitney embedding theorem) gives another way to find Morse functions on M.

Exercise 2.2.8. If $M \subset \mathbb{R}^n$ is an embedded, closed manifold, prove that

$$\{v \in S^{n-1} : x \mapsto \langle x, v\rangle \text{ is a Morse function on } M\}$$

is a dense subset of S^{n-1}.

We now address Morse functions with no critical points.

Proposition 2.2.9. *A product cobordism* $(W; V_0, V_1)$ *satisfies* $W \cong V_1 \times I$ *(hence the name) and therefore* $V_0 \cong V_1$.[8]

Proof. Fix an arbitrary metric on W and a Morse function $f : W \to [0, 1]$ with no critical points. Let $\Phi : \mathcal{D} \to W$ denote the flow of the vector field $\xi = -\nabla f/||\nabla f||^2$, where $\mathcal{D} \subset W \times \mathbb{R}$ denotes the domain on which this flow is well-defined. Letting $y = \Phi(x, t)$ for some $(x, t) \in \mathcal{D}$, we can calculate

$$\left.\frac{d}{du}\right|_{u=t} f \circ \Phi(x, u) = (\mathcal{L}_\xi f)(y) = \langle \xi, df\rangle(y) = \frac{-\langle \nabla_y f, df_y\rangle}{||\nabla_y f||^2} = -1.$$

[8] This relies critically on W being compact. The norm function $||\cdot|| : (D^2 \setminus 0) \to [0, 1]$ is a "Morse function" on the "triad" $(D^2 \setminus 0; \emptyset, S^1)$ with no critical points, but $\emptyset \ncong S^1$.

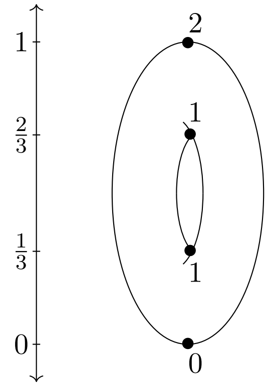


Figure 6: Morse function on a torus

Thus we have $f \circ \Phi(x,t) = f(x) - t$ for any $(x,t) \in \mathcal{D}$. Since W is compact, an integral curve can only terminate on ∂W. Therefore, we have

$$\mathcal{D} = \{(x,t) \in W \times \mathbb{R} : 0 \leq f(x) - t \leq 1\},$$

i.e. every integral curve starts on V_1 and ends on V_0. In particular, we have a well-defined restriction $\tilde{\Phi} : V_1 \times [0,1] \to W$, which is bijective because every integral curve intersects V_1 exactly once. Since $V_1 \times [0,1]$ is compact, this map $\tilde{\Phi}$ is a homeomorphism. Since f has no critical points, the vector field $-\nabla f$ is non-vanishing and thus $\tilde{\Phi}$ is a diffeomorphism. $\square$

We similarly show that Morse functions respect composition of triads. Rather than just using the gradient vector field (as in the previous proof), we treat more general vector fields that will be useful in §3.2.

Proposition 2.2.10. *For two triads $(W; V_0, V_1)$ and $(W'; V_1, V_2)$, we have*

$$\mu(W \sqcup_{V_1} W'; V_0, V_2) \leq \mu(W; V_0, V_1) + \mu(W'; V_1, V_2).$$

More generally, suppose that f and g are Morse functions on $(W; V_0, V_1)$ and $(W'; V_1, V_2)$ with $f(V_1) = g(V_1)$. Then $f \cup g : W \sqcup_{V_1} W' \to \mathbb{R}$ is Morse.

Proof. For simplicity, we suppose that f and g have images $[0,1]$ and $[1,2]$, respectively. Let U be a neighborhood of $V_1 \subset W$. Let ξ be a vector field defined on U, which is outward-pointing along V_1. We also let $\chi : U \to [0,1]$ be compactly supported in $\{x \in U : \xi(x) \neq 0\}$ with $\chi \equiv 1$ near V_1. Define

$$\tilde{\xi} = (1-\chi)\xi + \frac{\chi\xi}{\xi(f)}. \tag{4}$$

Note that $\tilde{\xi}$ is defined wherever ξ is defined (if ξ is defined outside of U, then $\tilde{\xi} = \xi$ on the rest of its domain) and outward-pointing along V_1. Also,

$$\mathcal{L}_{\tilde{\xi}} f = \left\langle (1-\chi)\xi + \frac{\chi\xi}{\xi(f)}, f \right\rangle = (1-\chi)\xi(f) + \chi$$

is identically 1 near V_1. Hence, the flowout of V_1 along $-\tilde{\xi}$ defines a collar $\Phi : V_1 \times [0, \delta) \to W$ with $f \circ \Phi(x,t) = 1 - t$ (for δ very small). Similarly, we may consider a neighborhood U' of $V_1 \subset W'$ and a vector field η on U', which is inward-pointing along V_1. In the same manner, we can then find $\tilde{\eta}$, which is equal to η outside of U' and satisfies $\mathcal{L}_{\tilde{\eta}} g = 1$. As above, the flowout of V_1 along $\tilde{\eta}$ defines a collar $\Psi : V_1 \times [0, \epsilon) \to W'$ with $g \circ \Psi(x,t) = 1 + t$ (for ϵ very small). Gluing these collars gives $W \sqcup_{V_1} W'$ a smooth structure. Since $f(V_1) = \{1\} = g(V_1)$, we get a continuous map $f \cup g : W \sqcup_{V_1} W' \to \mathbb{R}$. We consider the embedded neighborhood $\Phi \cup \Psi : V_1 \times (-\delta, \epsilon) \to W \sqcup_{V_1} W'$ (where we "turned around" the time-interval for Φ). By the above, we have

$$(f \cup g) \circ (\Phi \cup \Psi)(x,t) = 1 + t.$$

Thus $f \cup g$ is smooth and has no critical points on V_1. Moreover, the fields $\tilde{\xi}$ and $\tilde{\eta}$ are given by $\partial/\partial t$ in their respective collars, so they combine to form a smooth field on $W \sqcup_{V_1} W'$ (which agrees with ξ and η outside of $U \cup U'$). For later use, we note the following: if $\xi(f) > 0$ on U, then (4) clearly shows that $\tilde{\xi}(f) > 0$ on U; analogously, if $\eta(g) > 0$ on U', then $\tilde{\eta}(g) > 0$ on U'.

Taking f and g to minimize the number of critical points on each triad, the desired inequality for Morse numbers follows immediately. □

2.3 Elementary cobordisms

We will now describe the most important question of classical Morse theory: if $f : W \to \mathbb{R}$ is a Morse function, how does the sub-level set $f^{-1}(-\infty, a]$ change when a passes a critical point? The usual approach, found in [Mil63], focuses purely on the topological structure. Since we care a great deal about diffeomorphism type, we will give a slightly different approach, as in [Mil65a]. This approach focuses more on traces (and thus smooth handle attachment). However, both approaches use the local picture provided by Lemma 2.2.4, as we will see in the proof below, which follows Chapter 3 of [Mil65a].

Proposition 2.3.1. *Every trace is an elementary cobordism and vice versa.*

Proof. First, consider a closed manifold V_0 and an framed sphere $S^{k-1} \subset V_0$. We start by constructing a manifold $\omega(V_0, S^{k-1})$, which will eventually turn

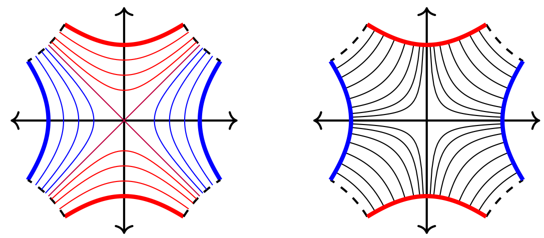


Figure 7: Morse function on a trace

out to be the trace of the surgery $V_0 \mapsto \chi(V_0, S^{k-1})$. First, we define the set

$$L = \{(x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : -1 \leq ||y||^2 - ||x||^2 \leq 1 \text{ and } |xy| < 1\}.$$

This region is illustrated in Figure 7. Notice that ∂L is just the union of

$$S^{k-1} \times D^{n-k} = \{(x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : ||y||^2 - ||x||^2 = -1 \text{ and } |xy| < 1\},$$

$$D^k \times S^{n-k-1} = \{(x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : ||y||^2 - ||x||^2 = 1 \text{ and } |xy| < 1\},$$

which are shown in blue and red, respectively, in Figure 7(a). We have also drawn various level sets of the smooth function $f(x, y) = ||y||^2 - ||x||^2$ on L. We also consider orthogonal trajectories, i.e. integral curves of $\nabla f/||\nabla f||^2$. This is illustrated in Figure 7(b). We will temporarily ignore the two "axes," which comprise all points flowing into or out of the origin. The remaining region $L \setminus (\mathbb{R}^k \cup \mathbb{R}^{n-k})$ is swept out by integral curves starting on the blue boundary $S^{k-1} \times (D^{n-k} \setminus 0)$ (where the axis $\mathbb{R}^k$ is cut out). We now define

$$\Phi : S^{k-1} \times (D^{n-k} \setminus 0) \times [-1, 1] \to L$$

to be the flowout of this portion of the boundary, which starts at time -1, so that $f \circ \Phi(x, y, t) = t$. This is a diffeomorphism onto $L \setminus (\mathbb{R}^k \cup \mathbb{R}^{n-k})$. Returning to V_0, we choose a tubular neighborhood $\Psi : S^{k-1} \times D^{n-k} \to V_0$ of the framed sphere. Then we can glue L to $(V_0 \setminus S^{k-1}) \times [-1, 1]$ via maps:

$$\begin{tikzcd}
 & S^{k-1} \times (D^{n-k} \setminus 0) \times [-1,1] \arrow[dl, "\Phi"'] \arrow[dr, "\Psi \times \mathrm{Id}_{[-1,1]}"] & \\
L & & (V_0 \setminus S^{k-1}) \times [-1,1]
\end{tikzcd}$$

Since Φ and Ψ are smooth, it is straightforward to check that the resulting space $\omega(V_0, S^{k-1})$ is a compact, smooth manifold. Since $f \circ \Phi(x, y, t) = t$, the gluing occurs in a way so that $f : L \to [-1, 1]$ agrees with the projection $(V_0 \setminus S^{k-1}) \times [-1, 1] \to [-1, 1]$ on the intersection of their domains. As such, we get a smooth function $f : \omega(V_0, S^{k-1}) \to [-1, 1]$. The only critical point of this function is the original critical point $0 \in L$. Moreover, we have

$$\partial\omega(V_0, S^{k-1}) = \partial L \cup \big((V_0 \setminus S^{k-1}) \times \{-1, 1\}\big) = f^{-1}(-1) \sqcup f^{-1}(1).$$

Therefore, f is a Morse function on the triad $\big(\omega(V_0, S^{k-1}); f^{-1}(-1), f^{-1}(1)\big)$. Note that $f^{-1}(-1)$ is the union of $(V_0 \setminus S^{k-1})$ and the blue boundary of L, i.e. $f^{-1}(-1) = (V_0 \setminus S^{k-1}) \cup (S^{k-1} \times D^{n-k}) = V_0$ (we glue back in S^{k-1} along its tubular neighborhood $S^{k-1} \times D^{n-k}$). Similarly, we can see that

$$f^{-1}(1) = (V_0 \setminus S^{k-1}) \cup (D^k \times S^{n-k-1}) = \chi(V_0, S^{k-1}),$$

(rather than cutting out the tubular neighborhood to perform the surgery, we are gluing smoothly along the whole tubular neighborhood, as in [Kos93]). Thus we have constructed a triad $\big(\omega(V_0, S^{k-1}); V_0, \chi(V_0, S^{k-1})\big)$ (a candidate for the trace) and a Morse function f with one critical point of index k.

Now consider an elementary cobordism $(W; V_0, V_1)$ and a Morse function $g : W \to [-1, 1]$ with one critical point p of index k (we can take $g(p) = 0$). By Lemma 2.2.4, there is an open set $U \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$ containing the origin and an interior chart $U \hookrightarrow W$ such that $0 \mapsto p$ and $g(x, y) = ||y||^2 - ||x||^2$. We may choose a vector field ξ on W with $\xi = (-x, y)$ on U and $\xi(g) > 0$ away from p (this is proven in Proposition 3.1.2 below). Choose $0 < \epsilon < 1$ small enough that $2\epsilon D^n \subset U$. Define $W_\epsilon = g^{-1}[-\epsilon^2, \epsilon^2]$. Since $g^{-1}[-1, -\epsilon^2]$ and $g^{-1}[\epsilon^2, 1]$ are product cobordisms, Proposition 2.2.9 lets us to replace $(W; V_0, V_1)$ by $\big(W_\epsilon; V_0, V_1\big)$, identifying $V_0 \cong g^{-1}(-\epsilon^2)$ and $V_1 \cong g^{-1}(\epsilon^2)$. Then $U \cap W_\epsilon$ resembles Figure 7; in particular, the $\mathbb{R}^k$ "axis" intersects V_0 in a framed sphere S^{k-1}. Points in $V_0 \setminus S^{k-1}$ miss p and flow along ξ to V_1, so there is a unique embedding $\psi : (V_0 \setminus S^{k-1}) \times [-1, 1] \to W_\epsilon$ satisfying:

- $\psi(x, -1) = x$ for any $x \in V_0 \setminus S^{k-1}$;
- $t \mapsto \psi(x, t)$ is an integral curve of ξ, for any $x \in V_0 \setminus S^{k-1}$;
- $g \circ \psi(x, t) = t\epsilon^2$ for any $x \in V_0 \setminus S^{k-1}$ and $t \in [-1, 1]$.

Define an embedding $\varphi : L \to U \cap W_\epsilon$ by $\varphi(x, y) = (\epsilon x, \epsilon y)$. The two maps φ and ψ agree on the intersection of their domains, so they define a map $\omega(V_0, S^{k-1}) \to W_\epsilon$. The reader may check that this is a diffeomorphism,

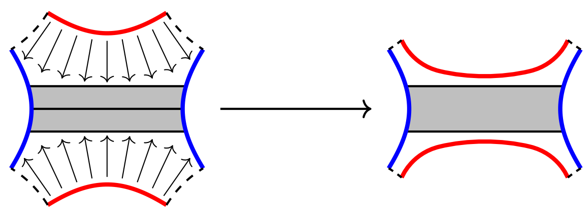


Figure 8: Handle in an elementary cobordism

simply by showing that it is bijective. Since φ identifies the local pictures of L and U (up to scaling), this map also identifies the framed spheres.

At this point: we have constructed a triad $\big(\omega(V_0, S^{k-1}); V_0, \chi(V_0, S^{k-1})\big)$, which is an elementary cobordism associated to a surgery; we have shown that any elementary cobordism is of this form. It remains to show that this is indeed the desired trace. We will give a brief, pictorial sketch[9] (full details can be found in [Kos93]). First, define $H = \{(x, y) \in U \cap W_\epsilon : ||y|| \leq \epsilon/2\}$. This is the gray region in Figure 8; it is essentially a tubuluar neighborhood of the disk $\bar{D}^k = \mathbb{R}^k \cap W_\epsilon$. Then we can see that $H \cong \bar{D}^k \times \bar{D}^{n-k}$ intersects V_0 in the tubular neighborhood $S^{k-1} \times \bar{D}^{n-k}$ of the above framed sphere. Flowing backwards along ξ, every point outside of $V_0 \cup H$ either flows back to V_0 or enters H. Slowing the rate of flow near $V_0 \cup H$, we are able to define a homotopy $\Theta : W_\epsilon \times I \to W_\epsilon$ such that:

- $\Theta|_{(V_0 \cup H) \times \{t\}} = \mathrm{Id}_{V_0 \cup H}$ for all $t \in I$;
- $\Theta|_{W_\epsilon \times \{t\}}$ is a smooth emebedding for all $t < 1$;
- $\Theta(x, t) \in V_0 \cup H$ if and only if $x \in V_0 \cup H$ or $t = 1$.

Then Θ shrinks W_ϵ down towards $V_0 \cup H$, but only collapses W_ϵ onto $V_0 \cup H$ at time $t = 1$ (where we forfeit smoothness). For $0 < \delta < 1$ small enough, $\Theta|_{W_\epsilon \times \{1-\delta\}}$ maps W_ϵ diffeomorphically onto a "slightly thickened $V_0 \cup H$." We can view this as a cylinder $V_0 \times I$ with a smooth handle $\widetilde{H}$ attached. Since H was attached to V_0 along the framed sphere S^{k-1}, the attaching sphere of $\widetilde{H}$ is $S^{k-1} \times 1 \subset V_0 \times I$. Thus $(V_0 \times I) \cup \widetilde{H}$ is the desired trace. $\square$

Corollary 2.3.2. *If $f : W \to \mathbb{R}$ is a Morse function on the triad $(W; V_0, V_1)$ with exactly one critical point, then the Morse number is $\mu(W; V_0, V_1) = 1$, and the index of the critical point of f is independent of the choice of f.*

[9] In what follows, we will only need that $V_0 \sqcup_{S^{k-1}} D^k \subset W_\epsilon$ is a homotopy equivalence. But it is still useful to see how the intuitive Definition 2.1.5 relates to the less transparent definition of $\omega(V_0, S^{k-1})$. The minimal details required can be found in [Mil63] or [Mil65a].

Proof. This result and its proof follow a remark in [Mil65a], almost verbatim. Let X be any space and let $\varphi : S^{k-1} \to X$ be continuous. Then we have

$$H_i(X \sqcup_\varphi D^k, X) = H_i(D^k, S^k) = \begin{cases} \mathbb{Z}, & \text{if } i = k \\ 0, & \text{if } i \neq k \end{cases}$$

by excision. Thus, attaching a k-cell always changes the homotopy type.

Suppose that the critical point of f has index k. We wish to show that k is independent of f. Since Proposition 2.3.1 gives a homotopy equivalence $(W, V_0) \sim (V_0 \sqcup_\varphi D^k, V_0)$, where φ is an attaching sphere, we conclude that

$$H_i(W, V_0) = \begin{cases} \mathbb{Z}, & \text{if } i = k \\ 0, & \text{if } i \neq k \end{cases}$$

This describes k uniquely in terms of $(W; V_0, V_1)$. If we had $\mu(W; V_0, V_1) = 0$, then Proposition 2.2.9 would show that $V_0 \hookrightarrow W$ is a homotopy equivalence, which would contradict $H_k(W, V_0) = \mathbb{Z}$. Therefore $\mu(W; V_0, V_1) = 1$. □

Corollary 2.3.3. *The following results relate cobordism to surgery:*

(a) *Any cobordism $(W; V_0, V_1)$ is a composition of elementary cobordisms. The minimal number of elementary cobordisms needed is $\mu(W; V_0, V_1)$.*

(b) *Closed manifolds are surgically equivalent if and only if they cobordant.*

Proof. Proposition 2.2.7 shows that any $(W; V_0, V_1)$ admits a Morse function $f : W \to \mathbb{R}$ with critical points $p_0, \ldots, p_n$ such that $f(p_0) < \cdots < f(p_n)$. Choose $a_i \in \big(f(p_{i-1}), f(p_i)\big)$ for all $i = 1, \ldots, n$. Let $a_0 = 0$ and $a_{n+1} = 1$. We factor W into the elementary cobordisms $f^{-1}[a_i, a_{i+1}]$ for $i = 0, \ldots, n$ and note that $\mu(W; V_0, V_1) \leq n + 1$ by Proposition 2.2.10. This proves (a).

Suppose that V_1 can be formed by applying successive surgeries to V_0. The composition of their traces is a triad $(W; V_0, V_1)$. Conversely, any triad $(W; V_0, V_1)$ is a composition of traces, by (a) and Proposition 2.3.1. □

We now use Corollary 2.3.3 to compare Figures 5 and 6. Corresponding to the Morse function f in Figure 6, the torus is depicted as a composition of traces in Figure 5. Since attaching a handle is the same thing as gluing on a trace, the sub-level[10] set $f^{-1}(-\infty, a]$ changes by the addition of a handle whenever a passes a critical value (and the level set $f^{-1}(a)$ correspondingly changes by surgery). In the next section, we will see why this presentation of the torus via handles is, in fact, minimal (i.e. $\mu(S^1 \times S^1; \emptyset, \emptyset) = 4$).

[10]In Figure 5, everything is actually upside-down with respect to Figure 6. This is just so that Figure 5 can be written top-to-bottom, with the traces lined up for composition.

3 Morse Homology

3.1 Morse-Smale pairs

In this section, we expand upon the dynamical approach used in the proofs of Propositions 2.2.9 and 2.3.1. The main tool developed here will be vital in §4 and is also the starting point for the topic of Morse homology. We will not discuss the latter topic in full, but its broader relevance is worth noting: in [Bot88], Bott frames this subject as a natural destination for Morse theory and gives an account of how the ideas were developed by Thom and Smale, before being made fully explicit by Witten. Today, most research interests in "Morse theory" are infinite-dimensional (Floer theory), where the classic cell-attachment approach fails, but the homology approach is very fruitful.

Definition 3.1.1. Let $f : W \to [a, b]$ be a Morse function on $(W; V_0, V_1)$. We will say that a vector field ξ on W is a pseudo-gradient of f if $\xi(f) < 0$ on all of $W \setminus \mathrm{Crit}(f)$, and each critical point p admits a local coordinate system $(x_1, \dots, x_n)$, such that $f(x) = f(p) - x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_m^2$ and $\xi(x) = (x_1, \dots, x_k, -x_{k+1}, \dots, -x_n)$ (we will call such a chart Morse).

In particular, since the critical points of any Morse f cannot lie on ∂W, a pseudo-gradient ξ must be inward-pointing on V_1 and outward-pointing on V_0. Also, we clearly have $k = \mathrm{ind}(p)$ in the definition of a Morse chart. The first step is to confirm that pseudo-gradient vector fields exist.

Proposition 3.1.2. *For any Morse function $f : W \to [a, b]$ on $(W; V_0, V_1)$, we indeed have a pseudo-gradient vector field ξ.*[11]

Proof. Let $p_1, \dots, p_\ell$ be the critical points of f. Lemma 2.2.4 gives charts $U_1, \dots, U_\ell$ about $p_1, \dots, p_\ell$ (which we choose to be disjoint from each other and the boundary ∂W), in which $f(x) = f(p_i) - x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_n^2$. Let Y_i be a neighborhood of p_i with $\overline{Y}_i \subset U_i$. Choose smooth functions $\chi_i : W \to [0, 1]$ with $\chi_i \equiv 1$ on Y_i and $\mathrm{supp}(\chi_i) \subset U_i$. Let $\chi = \chi_1 + \cdots + \chi_\ell$. Then we can form a vector field $\xi_i(x) = \chi_i(x)(-x_1, \dots, -x_k, x_{k+1}, \dots, x_n)$ on the chart U_i, which extends by 0 to all of W. Then Y_i is a Morse chart for the pair $(f, -\xi_i)$. Given $x \in W$ with $\chi_i(x) > 0$, we have $x \in U_i$ and thus

$$\xi_i(f)(x) = \chi_i(x)(2x_1 + \cdots + 2x_n) \geq 0,$$

with equality if and only if $x = p_i$. Choose a metric on W and note that

$$\big((\chi - 1)\nabla f\big)(f)(x) = \langle(\chi - 1)\nabla f, df\rangle(x) = (\chi(x) - 1)||\nabla_x f||^2 \leq 0, \quad (5)$$

[11] We already used this fact in the proof of Proposition 2.3.1. No circular logic occurs.

with equality if and only if $\chi(x) = 1$ or $\nabla_x f = 0$ (i.e. x is a critical point). In particular, $\chi_1(x) = \cdots = \chi_\ell(x) = 0$ gives $\chi(x) = 0$ and $x \notin Y_1 \cup \cdots \cup Y_\ell$, so x is not a critical point of f and the inequality in (5) is strict. Hence,

$$\xi = (\chi - 1)\nabla f - \xi_1 - \cdots - \xi_\ell$$

is a pseudo-gradient of f, because at least one summand of $\xi(f)$ is non-zero at any point, and $\xi|_{Y_i} = \xi_i|_{Y_i}$ has the same Morse chart Y_i as above. $\square$

Now, we consider a Morse function $f : W \to \mathbb{R}$ and a pseudo-gradient ξ. Note that ξ does not have any closed orbits: since $\xi(f) < 0$ on $W \setminus \mathrm{Crit}(f)$, the value of f decreases along non-stationary flow lines. Thus a non-critical point $x \in W$ either flows forwards (resp. backwards) along ξ to V_0 (resp. V_1) or approaches a critical point as $t \to \infty$ (resp. $t \to -\infty$).

Definition 3.1.3. Fix a critical point p of a Morse function $f : W \to \mathbb{R}$. Let Φ denote the flow a pseudo-gradient ξ. The descending manifold of p is

$$\mathscr{D}(p) = \left\{ x \in W : \lim_{t \to -\infty} \Phi(x,t) = p \right\}.$$

The ascending manifold $\mathscr{A}(p)$ is defined analogously:

$$\mathscr{A}(p) = \left\{ x \in W : \lim_{t \to \infty} \Phi(x,t) = p \right\}.$$

These should be thought of as points "coming out of" and "going into" p. Note that $f|_{\mathscr{D}(p)}$ $\big($resp. $f|_{\mathscr{A}(p)}\big)$ is uniquely maximized (resp. minimized) at p. It is clear that different descending (resp. ascending) manifolds are disjoint, but descending and ascending manifolds may intersect. We say that the pair (f, ξ) is Morse-Smale if all such intersections are transverse.

Figure 9 depicts descending/ascending manifolds, drawn in blue/red, respectively. In any Morse chart U about the critical point p, the ascending and descending manifolds are just the "coordinate axes" (see Figure 9(a), which represents $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$, where $k = \mathrm{ind}(p)$). Hence, we can see that $\mathscr{D}(p) \cap U$ is a k-dimensional manifold, which is clearly tangent to ξ. Therefore, flowing along ξ preserves this structure, i.e. $\mathscr{D}(p)$ is a k-manifold (because, by definition, any $x \in \mathscr{D}(p)$ is on an integral curve starting in U). Similarly, the ascending manifold $\mathscr{A}(p)$ is an $(n-k)$-dimensional manifold. This shows that the definition of a Morse-Smale pair makes sense.

Exercise 3.1.4. Prove the following about ascending/descending manifolds:

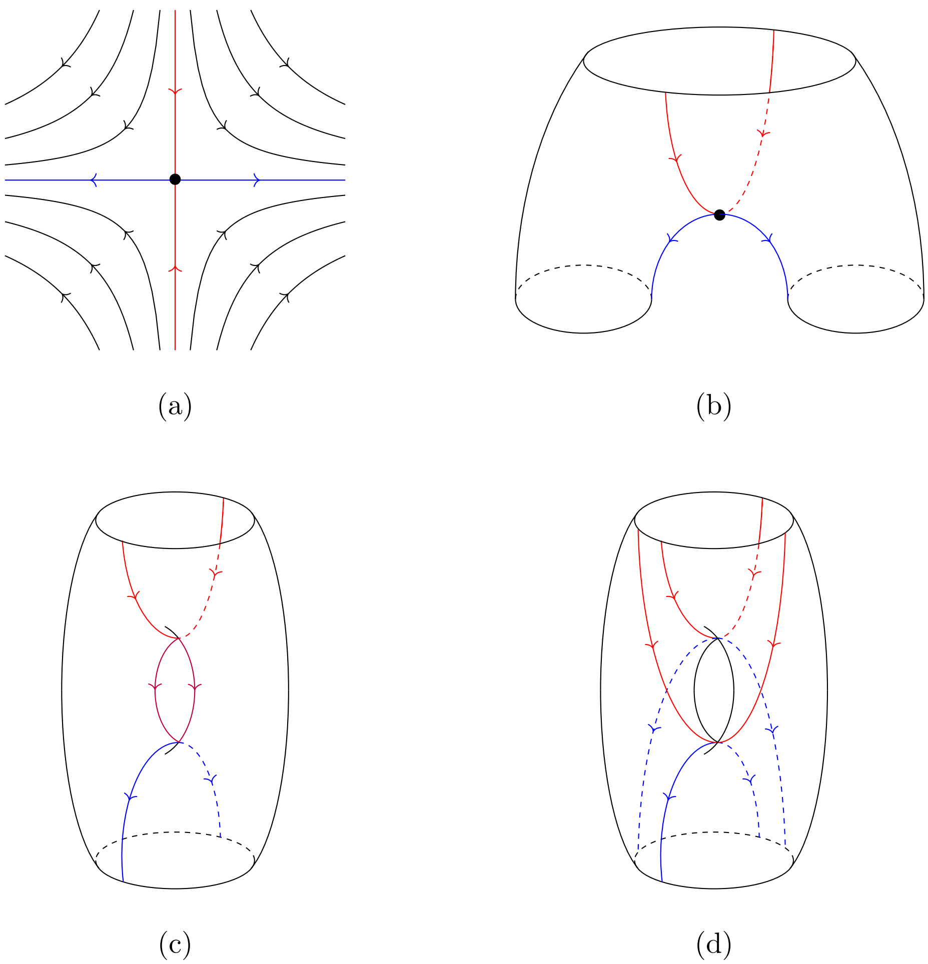


Figure 9: Ascending and descending manifolds

(a) If $V_0 = \emptyset$ and f has critical points $p_1, \ldots, p_\ell$, then $W = \bigsqcup_{i=1}^{\ell} \mathscr{A}(p_i)$.

(b) If $\mathscr{D}(p) \cap V_0 = \emptyset$ and $k = \mathrm{ind}(p)$, we get a diffeomorphism $\mathscr{D}(p) \cong D^k$. (Hint: construct a diffeomorphism $\varphi : D^k \to \mathscr{D}(p)$ such that $\varphi(0) = p$ and $(0,1)v$ maps onto some integral curve of ξ for every $v \in S^{n-1}$.)

Turning everything upside-down, the analogous results clearly hold as well.

When checking that a pair is Morse-Smale, we always have $\mathscr{D}(p) \pitchfork \mathscr{A}(p)$, since they only intersect at p, which sits in a Morse chart. Thus, for a Morse function $f : W \to [-1,1]$ with only one critical point p, any pseudo-gradient forms a Morse-Smale pair (see Figure 9(b)). Fix a Morse chart U about p. We assume that $f(p) = 0$ and restrict to an equivalent cobordism $f^{-1}[-\epsilon, 1]$, for $0 < \epsilon < 1$ small enough that $\mathscr{D}(p) \cap f^{-1}[-\epsilon, 1] \subset U$. In this local picture, we see that $\mathscr{D}(p) \cap f^{-1}[-\epsilon, 1] \cong \bar{D}^k$. As in the proof of Proposition 2.3.1, this $\bar{D}^k$ is the core disk of the handle attachment which forms the trace. Then $\mathscr{D}(p) \cap f^{-1}(-\epsilon)$ is a $(k-1)$-sphere, which flows along ξ to V_0 through the product cobordism $f^{-1}[-1, -\epsilon]$. Thus $\mathscr{D}(p) \cong \bar{D}^k$ is neatly embedded in W with $\partial \mathscr{D}(p) \subset V_0$. Similarly, $\mathscr{A}(p) \cong \bar{D}^{n-k}$ is neatly embedded in W with $\partial \mathscr{A}(p) \subset V_1$. Hence, we view $\mathscr{D}(p)$ and $\mathscr{A}(p)$ as the core and belt disks of the trace W. In fact, this does not require that f has one critical point; we only need $\mathscr{D}(p)$ to "miss" other critical points, so that it flows to V_0.

Definition 3.1.5. Consider a triad $(W; V_0, V_1)$, Morse function $f : W \to \mathbb{R}$ and pseudo-gradient ξ. Consider some $p \in \mathrm{Crit}_k(f)$. If $\mathscr{D}(p) \setminus \{p\}$ does not intersect any ascending manifolds, we may define the lower sphere to be

$$S^{k-1}_{\mathscr{D}}(p) = \partial \mathscr{D}(p) = \mathscr{D}(p) \cap V_0.$$

Similarly, if $\mathscr{A}(p) \setminus \{p\}$ does not intersect any descending manifolds, then

$$S^{n-k-1}_{\mathscr{A}}(p) = \partial \mathscr{A}(p) = \mathscr{A}(p) \cap V_1$$

will be called the upper sphere of p. Note that these spheres inherit framings from the local picture (just as any attaching and belt spheres are framed).

We will also need some upper/lower spheres ($\mathscr{D}(p) \cap V$ and $\mathscr{A}(p) \cap V$) in other non-critical level sets V, but we will specify these as they arise.

We will prove the existence of Morse-Smale pairs below, but it is also instructive to think about when this fails. On a torus, we take the "height" Morse function (as in Figure 6) and construct a pseudo-gradient by which points follow the steepest path of descent. This is illustrated in Figure 9(c) (where we have chopped off the ends the torus, so as to focus on the middle).

We have two integral curves (drawn in purple) flowing between the critical points of index 1, which we write as p and q (with p being the higher point). Then $\mathscr{D}(p) \setminus \{p\} = \mathscr{A}(q) \setminus \{q\}$, so the intersection $\mathscr{D}(p) \cap \mathscr{A}(q)$ is certainly not transverse. However, by altering the pseudo-gradient slightly, we ensure that $\mathscr{D}(p)$ and $\mathscr{A}(q)$ miss each other entirely (see Figure 9(d)). In keeping with the notions of height and steepest descent, this is usually interpreted as tipping/shearing the torus (with the top half going further into the page). The resulting pair is Morse-Smale on the torus (with the ends re-attached).

3.2 Rearrangement of critical points

In this section, we will prove that Morse-Smale pairs always exist and that, in nice situations, the critical values of a Morse function can be moved past each other. Given triads $(W; V_0, V_1)$ and $(W'; V_1, V_2)$ with Morse functions $f : W \to [0, 1]$ and $g : W' \to [1, 2]$, Proposition 2.2.10 showed that they glue to give a Morse function $f \cup g : W \sqcup_{V_1} W' \to [0, 2]$ on the composed triad. In that proof, we further showed that pseudo-gradients of f and g can glue to form a pseudo-gradient of $f \cup g$ (they may be rescaled in a neighborhood of V_1, but this does not alter the integral curves or Morse charts). Hence, we can factor/compose cobordisms in a way that respects pseudo-gradients. With this cut-and-paste approach, we can make some useful modifications.

Lemma 3.2.1. *Let $f : W \to [a, b]$ be a Morse function with pseudo-gradient field ξ. Suppose $a \leq c < d \leq b$ are such that $[c, d]$ does not contain any critical values, so the flow of ξ gives a diffeomorphism $\varphi : f^{-1}(c) \to f^{-1}(d)$. If $h : f^{-1}(d) \to f^{-1}(d)$ is a diffeomorphism isotopic to the identity $Id_{f^{-1}(d)}$, then we can find a new pseudo-gradient $\tilde{\xi}$, agreeing with ξ outside of $f^{-1}(c, d)$ and whose flow induces the diffeomorphism $h \circ \varphi : f^{-1}(c) \to f^{-1}(d)$.*

Proof. For simplicity, we assume that $c = 0$ and $d = 1$. Let $V = f^{-1}(d)$. Because $f^{-1}[c, d]$ is a product cobordism, we may identify it with $V \times I$ (where we identify $f^{-1}(c) \cong V$ via φ). Under this identification, f is just the projection $\pi : V \times I \to I$. Let $h_t : V \to V$ be an isotopy with $h_0 = \mathrm{Id}_V$ and $h_1 = h$. Then the map $H : V \times I \to V \times I$ defined by $H(x, t) = \big(h_t(x), t\big)$ is clearly a diffeomorphism. Take the vector field $\partial/\partial t$ on $V \times I$ and push it forward to define $\eta = H_*(\partial/\partial t)$. This η is still positive in the t-direction, so $\eta(\pi) > 0$ on $V \times I$. If Φ denotes the flow of η, then we clearly have

$$\Phi_t(v, s) = (h_{s+t} \circ h_s^{-1}(x), s + t).$$

In particular, flowing from $V \times 0$ to $V \times 1$ gives the map h. Gluing $f^{-1}[c, d]$ back into W, with the pseudo-gradient $-\eta$, defines the desired vector field $\tilde{\xi}$.

(Technically, we should replace $[c, d]$ by some interval in (c, d) at the outset, so we can be sure that all changes occur strictly within $f^{-1}(c, d)$.) $\square$

To use this lemma, we must have some useful isotopies to apply:

Lemma 3.2.2. *Let V be a closed manifold, with submanifolds M and N. If M is framed (this is not necessary, but it is easier to prove), then there exists an isotopy $h_t : V \to V$ with $h_0 = Id_V$ and $h_1(M)$ transverse to N.*

Proof. Let $M \times D^k \subset V$ be a tubular neighborhood (given by the framing) and let $N' = N \cap (M \times D^k)$. The projection map $M \times D^k \to D^k$ restricts to a smooth map $\pi : N' \to D^k$. For any $v \in D^k$, the manifold N is transverse to $M \times v \subset M \times D^k$ if and only if v is a regular value of π. By Sard's lemma, such a regular value v exists. The desired isotopy can now be defined as

$$h_t(p) = \begin{cases} p, & p \notin M \times D^k \\ \big(x, \varphi_t(v)\big), & p = (x, v) \in M \times D^k, \end{cases}$$

where $\varphi_t : D^k \to D^k$ is an isotopy supported (i.e. different from the identity) in a compact subset of D^k, with $\varphi_0 = \mathrm{Id}_{D^k}$ and $\varphi_1(0) = v$. Existence of such an isotopy is a standard application of bump functions (see [Mil65b]). $\square$

These lemmas give us the tools to prove that Morse-Smale pairs exist. First, we note an important detail. Suppose that we have two critical points p and q of a Morse function $f : W \to \mathbb{R}$. Consider some $x \in \mathscr{A}(p) \cap \mathscr{D}(q)$ (with respect to some pseudo-gradient ξ). Since $\mathscr{A}(p)$ and $\mathscr{D}(q)$ are both preserved under the flow of ξ, their tangent spaces contain the ξ-direction. Hence, transversality depends only on a complementary direction, e.g. T_xV, where $V = f^{-1}\big(f(x)\big)$ is a level set. Therefore $\mathscr{A}(p)$ and $\mathscr{D}(q)$ are transverse in W at x if and only if $\mathscr{A}(p) \cap V$ and $\mathscr{D}(q) \cap V$ are transverse in V at x.

Proposition 3.2.3. *If $f : W \to \mathbb{R}$ is Morse, then there is a pseudo-gradient ξ of f such that (f, ξ) is a Morse-Smale pair.*

Proof. Let ξ be a pseudo-gradient of f and suppose f has critical values $c_0 < c_1 < \cdots < c_k$. For each $i = 1, \ldots, k$, choose some $a_i \in (c_{i-1}, c_i)$. Let M (resp. N) denote the intersection of all ascending (resp. descending) manifolds with $V_1 = f^{-1}(a_1)$. Note that M is framed, because it is a union of upper spheres. By Lemma 3.2.2, we can find an isotopy $h_t : V_1 \to V_1$ with $h_0 = \mathrm{Id}_{V_1}$ and $h_1(M) \pitchfork N$. Using Lemma 3.2.1, we can then modify ξ in $f^{-1}(a_1 - \epsilon, a_1)$ for some small $\epsilon > 0$, so that the new ascending manifolds intersect $f^{-1}(a_1)$ in $h_1(M)$. Thus, the ascending manifolds of critical points in $f^{-1}(c_0)$ intersect the descending manifolds of critical points in $f^{-1}(c_1)$ transversely. Continuing in this fashion,[12] we successively modify ξ near each level set $f^{-1}(a_i)$. After these modification are done, consider an arbitrary intersection $x \in \mathscr{A}(p) \cap \mathscr{D}(q)$. Suppose that $q \in f^{-1}(c_i)$. Then we can find some $\tilde{x} \in f^{-1}(a_i)$ on the same integral curve as x. The modification near $f^{-1}(a_i)$ ensures that $\mathscr{A}(p)$ and $\mathscr{D}(q)$ intersect transversely at $\tilde{x}$ (all later modifications are near some $f^{-1}(a_j)$ with $j > i$, so they do not change ξ on $f^{-1}(-\infty, c_i])$. Flowing back to x, we see that $\mathscr{A}(p)$ and $\mathscr{D}(q)$ intersect transversely at x as well (because $\mathscr{A}(p)$ and $\mathscr{D}(q)$ are both tangent to ξ). $\square$

We now describe how critical values can be moved around:

Proposition 3.2.4. *Let $f : W \to [a, b]$ be a Morse function with critical points $p_0, \dots, p_k, q_0, \dots, q_\ell$. Suppose that $c, d \in (a, b)$ are such that $c \leq d$, $p_0, \dots, p_k \in f^{-1}(c)$ and $q_0, \dots, q_\ell \in f^{-1}(d)$. Fix a pseudo-gradient ξ of f. If*

$$\big(\mathscr{A}(p_0) \cup \cdots \cup \mathscr{A}(p_k)\big) \cap \big(\mathscr{D}(q_0) \cup \cdots \cup \mathscr{D}(q_\ell)\big) = \emptyset,$$

then for any $c', d' \in (a, b)$, there is a function $g : W \to [a, b]$ such that:

- *ξ is a pseudo-gradient of the Morse function g and Crit$(g) =$ Crit(f);*
- *$g - f$ is zero near ∂W and constant near critical points of the functions;*
- *$p_0, \dots, p_k \in g^{-1}(c')$ and $q_0, \dots, q_\ell \in g^{-1}(d')$.*

Proof. This follows [Mil65a] almost verbatim. Let $(W; V_0, V_1)$ be the triad. For ease, take $a = 0$ and $b = 1$. We assumed that the below sets are disjoint:

$$K_c = \mathscr{A}(p_0) \cup \mathscr{D}(p_0) \cup \cdots \cup \mathscr{A}(p_k) \cup \mathscr{D}(p_k),$$

$$K_d = \mathscr{A}(q_0) \cup \mathscr{D}(q_0) \cup \cdots \cup \mathscr{A}(q_\ell) \cup \mathscr{D}(q_\ell).$$

As noted in §3.1, it follows that the ascending and descending manifolds are all closed disks, so K_c and K_d are compact. Define $\pi : W \setminus (K_c \cup K_d) \to V_0$ to be the smooth map taking each point to the endpoint of its integral curve. For any critical point x, the mapping π extends continuously to $\mathscr{D}(x) \setminus \{x\}$, so points near $\mathscr{D}(x) \setminus \{x\}$ map to points near $\pi\big(\mathscr{D}(x) \setminus \{x\}\big) = \mathscr{D}(x) \cap V_0$. But points near $\mathscr{A}(x)$ flow into a Morse chart about x, so they become close

[12] For $i > 1$, the manifold M is no longer a union of upper spheres, because $\mathscr{A}(p)$ may now "get caught on" critical points with values between $f(p)$ and a_i. But M is still framed, as can be seen by flowing out from a Morse chart. Alternatively, we can apply the version of Lemma 3.2.2 without the framing condition, which we have mentioned but not proven.

to $\mathscr{D}(x)$ and therefore also map to points near $\mathscr{D}(x) \cap V_0$. Thus π takes points near K_c (resp. K_d) to points near $K_c \cap V_0$ (resp. $K_d \cap V_0$). We choose a smooth map $\mu : V_0 \to I$, which is 0 near K_c and 1 near K_d. Note that $\mu \circ \pi$ is a smooth function on $W \setminus (K_c \cup K_d)$, which is constant along integral curves of ξ. By the above discussion of the flow, $\mu \circ \pi$ extends to $\bar{\mu} : W \to I$, which is smooth, constant along integral curves, 0 near K_c and 1 near K_d.

Choose a smooth function $G : I^2 \to I$ (illustrated in Figure 10) with:

(i) $\frac{\partial G}{\partial s}(s,t) > 0$ for all $(s,t) \in I^2$;

(ii) $G(s,t) = s$ when s is near 0 or 1;

(iii) $G(s,0) = s + c' - c$ for s near c;

(iv) $G(s,1) = s + d' - d$ for s near d.

We now define $g : W \to I$ by $g(x) = G\big(f(q), \bar{\mu}(q)\big)$. Then (ii)-(iv) imply that $g - f$ is zero near ∂W and constant near $\operatorname{Crit}(f)$, so $\operatorname{Crit}(f) \subset \operatorname{Crit}(g)$. Moreover, every point in $\operatorname{Crit}(f)$ admits a Morse chart for the pair (g, ξ). By the chain rule and the fact that $\bar{\mu}$ is constant along integral curves:

$$\begin{aligned}\xi(g)(q) = \langle dg, \xi\rangle(q) &= \frac{\partial G}{\partial s}\big(f(q), \bar{\mu}(q)\big)\langle df, \xi\rangle(q) + \frac{\partial G}{\partial t}\big(f(q), \bar{\mu}(q)\big)\langle d\bar{\mu}, \xi\rangle(q)\\ &= \frac{\partial G}{\partial s}\big(f(q), \bar{\mu}(q)\big)\langle df, \xi\rangle(q) = \frac{\partial G}{\partial s}\big(f(q), \bar{\mu}(q)\big)\xi(f)(q).\end{aligned}$$

Applying (i), we can see that $\xi(g)$ is a positive (function) multiple of $\xi(f)$, so $\xi(g) < 0$ on $W \setminus \operatorname{Crit}(f)$. This gives $\operatorname{Crit}(g) \subset \operatorname{Crit}(f)$. Thus g is Morse, with pseudo-gradient ξ and the same critical points as f. By (iii) and (iv), we can see that $p_0, \ldots, p_k \in g^{-1}(c')$ and $q_0, \ldots, q_\ell \in g^{-1}(d')$. □

If $c = d$, then there is only one critical level, so ascending and descending manifolds do not intersect (except at the critical points). Hence, the above proposition is applicable, so we can shift the value on any subset of $\operatorname{Crit}(f)$.

For an arbitrary Morse function f, we can cut up the triad into factors which each have one critical level. On each factor, we can "nudge" the value of a critical point, by the above comment. Since Proposition 3.2.4 does not change the Morse function near the boundary, we can glue the factors back together to get an f modified by these "nudges," but otherwise the same.[13] Iterating this makes f injective on $\operatorname{Crit}(f)$, as promised in Proposition 2.2.7.

We actually prefer another sort of nice Morse function. If $\dim W = n$, we say that a Morse function $f : W \to [-1, n+1]$ is self-indexing if we have $\operatorname{Crit}_k(f) \subset f^{-1}(k)$ for all k (i.e. any critical point goes to its index under f).

[13] In particular, the vector field ξ is still a pseudo-gradient of the new Morse function.

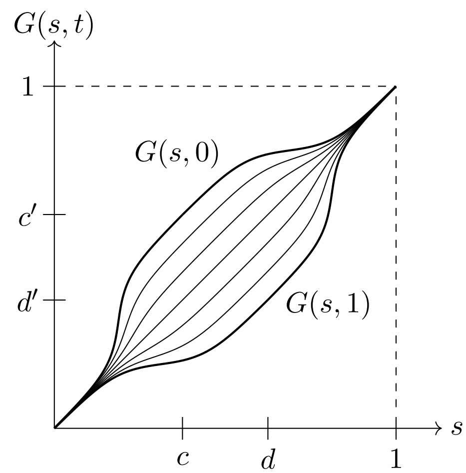


Figure 10: Graphs of the functions $s \mapsto G(s,t)$

Corollary 3.2.5. *If $f : W \to \mathbb{R}$ is Morse, then there exists a self-indexing Morse function g on W with the same critical points (of the same indices). Moreover, if (f, ξ) is a Morse-Smale pair, then (g, ξ) is Morse-Smale.*

Proof. The important part of this result is the order in which critical points are sorted; the actual critical values are immaterial (and can be prescribed after rearrangement). Form a Morse-Smale pair (f, ξ) by Proposition 3.2.3. By the above remark, we can modify f to be injective on its critical set.

Fix distinct critical points p and q such that $f(p) \leq f(q)$, $\operatorname{ind}(p) \geq \operatorname{ind}(q)$, and $\big[f(p), f(q)\big]$ contains no other critical values. If $n = \dim W$, then

$$\begin{aligned} \dim\big(\mathscr{A}(p) \cap \mathscr{D}(q)\big) &= n - \dim \mathscr{A}(p) - \dim \mathscr{D}(q) \\ &= n - \big(n - \operatorname{ind}(p)\big) - \operatorname{ind}(q) \leq 0. \end{aligned}$$

But if $x \in \mathscr{A}(p) \cap \mathscr{D}(q)$, then so is the entire integral curve containing x. Thus $\mathscr{A}(p) \cap \mathscr{D}(q) = \emptyset$. Therefore, we can apply Proposition 3.2.4 to move the values $f(p)$ and $f(q)$ past each other (or move them to the same value). The pseudo-gradient ξ is unchanged by the modification in Proposition 3.2.4, so the Morse-Smale property is preserved as well. Applying this repeatedly, we clearly may arrange the values at critical points in the desired way.[14] □

[14]We use a bubble sort, and the proof shows that necessary swaps can be performed. Once critical points of the same index are all adjacent ($\operatorname{Crit}(f)$ only has a total *pre*order), they can be moved to the same level set, because our proof allowed for $\operatorname{ind}(p) = \operatorname{ind}(q)$.

3.3 Morse homology

Using self-indexing Morse functions, we can give an interesting presentation of homology in terms of critical points and ascending/descending manifolds. The key to the h-cobordism theorem lies in understanding this presentation of homology and why, in nice enough situations, a Morse function exists with the minimal number of critical points necessary to generate the homology. After defining the presentation here, we will pursue this minimality in §4.

First, we recall the definition of intersection numbers of submanifolds. Let W be an oriented, closed n-manifold and let $M_0, M_1 \subset W$ be oriented, closed submanifolds with $\dim M_0 + \dim M_1 = n$. We denote Poincaré duality in W by $D : H^k(W) \to H_{n-k}(W)$. Let $\imath : M_0 \hookrightarrow W$ and $\jmath : M_1 \hookrightarrow W$ denote inclusion. Define the intersection number $I(M_0, M_1) \in \mathbb{Z}$ by the following:

$$I(M_0, M_1)[W] = D^{-1}\imath_*[M_0] \smile D^{-1}\jmath_*[M_1].$$

Alternatively,[15] we can use (the strengthened) Lemma 3.2.2 to deform M_1, so that it is transverse to M_0. Then $I(M_0, M_1)$ is a signed count of the points in $M_0 \cap M_1$, where M_0 and M_1 intersect positively at $p \in W$ if and only if

$$T_pM_0 \oplus T_pM_1 = T_pW \text{ respects orientation.}$$

Notice that the sign of an intersection depends on the order of M_0 and M_1.

Exercise 3.3.1. Let $k = \dim M_0$. In either (or both) of the formulations, prove that $I(M_1, M_0) = (-1)^{k(n-k)} I(M_0, M_1)$.

Throughout the rest of this section, any mention of a triad $(W; V, V')$ assumes that W is oriented. If ξ is a pseudo-gradient of a Morse function f, we may choose arbitrary orientations on all descending manifolds and orient the ascending manifolds so that $\mathscr{A}(p)$ and $\mathscr{D}(p)$ intersect positively at p. We take the induced boundary orientations on all upper and lower spheres. If $a > \min f$, we orient $f^{-1}(a)$ as part of the boundary of $f^{-1}(-\infty, a]$.

Lemma 3.3.2. *Let $f : W \to [a, b]$ be a Morse function on $(W; V, V')$ with critical points $p_1, \dots, p_\ell$, all having index k and the same value under f. Choose a pseudo-gradient ξ. Then W deformation retracts, fixing V, onto*

$$V \cup \mathscr{D}(p_1) \cup \cdots \cup \mathscr{D}(p_\ell) \subset W.$$

[15] The equivalence of these two definitions follows from the Thom isomorphism theorem.

Therefore $H_k(W,V) \cong \mathbb{Z}^{\oplus \ell}$ *with a basis given by* $\{[\mathscr{D}(p_i)] : i = 1, \ldots, \ell\}$. *For any closed* k*-manifold* $M \subset V_1$*, the class* $[M] \in H_k(W,V)$ *is given by*

$$[M] = \sum_{i=1}^{\ell} I\big(S_{\mathscr{A}}^{n-k-1}(p_i), M\big)[\mathscr{D}(p_i)],$$

where the intersection number is calculated in V'.

Proof. All but the last assertion was proven for one critical point in §2.3. The generalization to several critical points (of the same index and the same value under f) is straightforward. Our proof of the last assertion closely follows [Mil65a]. Define the sets $C = \{p_1, \ldots, p_\ell\}$, $\mathscr{D} = \mathscr{D}(p_1) \cup \cdots \cup \mathscr{D}(p_\ell)$ and $S_{\mathscr{A}} = S_{\mathscr{A}}^{n-k-1}(p_1) \cup \cdots \cup S^{n-k-1}\mathscr{A}(p_\ell)$. We will write $r : W \to V \cup \mathscr{D}$ for the deformation retract. Now recall the proof-sketch of Proposition 2.3.1 (modified for several critical points): we shrunk W onto $V \cup H_1 \cup \cdots \cup H_\ell$, where H_i is the handle associated to p_i; the deformation retract is completed by shrinking the handles down onto their core disks. This process is shown in Figure 11, where the marked points denote p_i and $S_{\mathscr{A}}^{n-k-1}(p_i)$. The first retraction just follows flow lines, so all points in $S_{\mathscr{A}}^{n-k-1}(p_i) = \mathbb{R}^{n-k} \cap V'$ stay on the $\mathbb{R}^{n-k}$ axis. Away from the "lower boundary" $S^{k-1} \times D^{n-k} \subset H_i$, the second retraction agrees with the projection map $D^k \times D^{n-k} \to D^k \times 0$. Therefore, the overall deformation retract takes $S_{\mathscr{A}}^{n-k-1}(p_i)$ to p_i for each i, so $r(S_{\mathscr{A}}) = C$. Thus we have the following commutative diagram:

$$\begin{tikzcd}
H_k(M) \arrow[r] & H_k(V') \arrow[r] \arrow[d, "r_*"'] & H_k(V', V' \setminus S_{\mathscr{A}}) \arrow[d, "r_*"] \\
& H_k(W) \arrow[dddr, bend right] & H_k(V \cup \mathscr{D}, V \cup \mathscr{D} \setminus C) \arrow[d, "h"] \\
& & H_k(V \cup \mathscr{D}, V) \arrow[d] \\
& & H_k(W, V)
\end{tikzcd}$$

Here h is an isomorphism, induced by the apparent deformation retract $V \cup \mathscr{D} \setminus C \to V$ (each punctured disk $\mathscr{D}(p_i) \setminus p_i$ deformation retracts onto its boundary $S_{\mathscr{D}}^{k-1}(p_i) \subset V$). All unlabelled maps are induced by inclusion. Since $r : V' \to W$ is homotopic to the inclusion $V' \subset W$, the left-hand maps compose to give the natural inclusion of $[M] \in H_k(W,V)$. The vertical maps on the right-hand side are all isomorphisms, due to the deformation retracts.

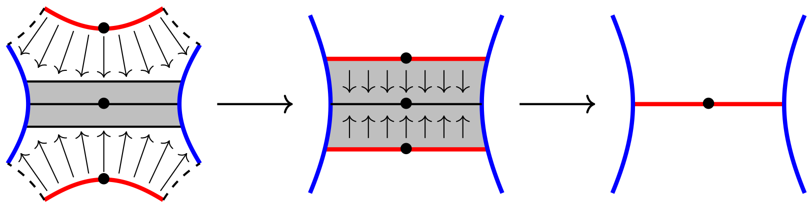


Figure 11: Deformation retract onto on $V \cup \mathscr{D}$

We wish to find the coefficients of $[M]$ with respect to the basis

$$\{[\mathscr{D}(p_i)] \in H_k(V \cup \mathscr{D}, V) : i = 1, \ldots, \ell\}.$$

If $D_i^k \subset \mathscr{D}(p_i)$ is a small disk about p_i (with the same orientation as $\mathscr{D}(p_i)$), then $h[D_i^k] = [\mathscr{D}(p_i)]$. Since $\mathscr{A}(p_i)$ is oriented so that $I(\mathscr{A}(p_i), \mathscr{D}(p_i)) = 1$, we can see that $I(\mathscr{A}(p_i), D_i^k) = 1$. Now, suppose that $\tilde{D}_i^k \subset V'$ is a small disk that intersects $\mathscr{A}(p_i)$ transversely in a single point. Then the two stages of the deformation retract (flowing along ξ and projection onto $\mathscr{D}$) preserve both the transversality and the sign of the intersection with $\mathscr{A}(p_i)$. Hence, if we orient $\tilde{D}_i^k$ such that $I(\mathscr{A}(p_i), \tilde{D}_i^k) = 1$, then we have $r_*[\tilde{D}_i^k] = [D_i^k]$. Since $\tilde{D}_i^k$ lies in the boundary component V', we have

$$I(\mathscr{A}(p_i), \tilde{D}_i^k) = 1 \text{ in } W \quad \iff \quad I(S_{\mathscr{A}}^{n-k-1}(p_i), \tilde{D}_i^k) = 1 \text{ in } V.$$

It remains to find the coefficients of $[M]$ with respect to the basis

$$\{[\tilde{D}_i^k] \in H_k(V', V' \setminus S_{\mathscr{A}}) : i = 1 \ldots, \ell\}.$$

The proof now utilizes the Thom isomorphism theorem, which goes beyond the prerequisites I initially set forth, so I will summarize the reasoning here. First consider disjoint tubular neighborhoods $T_i \subset V'$ of each $S_{\mathscr{A}}^{n-k-1}(p_i)$. Let $T = T_1 \cup \cdots \cup T_\ell$. By excision, we have

$$\begin{aligned} H_k(V', V' \setminus S_{\mathscr{A}}) = H_k(T, T \setminus S_{\mathscr{A}}) &= H_k\left(\bigsqcup_{i=1}^{\ell} T_i, \bigsqcup_{i=1}^{\ell} T_i \setminus S_{\mathscr{A}}^{n-k-1}(p_i)\right) \\ &= \bigoplus_{i=1}^{\ell} H_k\big(T_i, T_i \setminus S_{\mathscr{A}}^{n-k-1}(p_i)\big) \end{aligned}$$

and the i^{th} summand is generated by $[\tilde{D}_i^k]$. Since each $S_{\mathscr{A}}^{n-k-1}(p_i)$ is framed, we may apply Lemma 3.2.2 to perturb M, so that it is transverse to $S_{\mathscr{A}}$.

Moreover, framed manifolds have trivial normal bundles, so we may identify T_i with $\tilde{D}_i^k \times S_{\mathscr{A}}^{n-k-1}(p_i)$. If we choose the neighborhood T_i small enough, we may ensure that $M \cap T_i$ is a collection of disks, each homotopic to a fiber of $\tilde{D}_i^k \times S_{\mathscr{A}}^{n-k-1}(p_i)$. Hence, for each intersection point in $S_{\mathscr{A}}^{n-k-1}(p_i) \cap M$ having sign $\imath = \pm 1$, we get a disk in T_i whose fundamental class is just

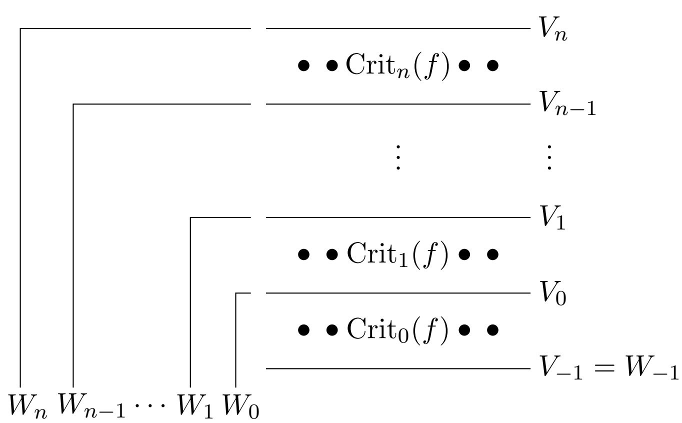
$$\imath[\tilde{D}_i^k] \in H_k\big(T_i, T_i \setminus S_{\mathscr{A}}^{n-k-1}(p_i)\big).$$

Thus a signed count of the disks in $M \cap T_i$ (with sign according to whether their orientation agrees with $\tilde{D}_i^k$) equals a signed count of the intersection points of $S_{\mathscr{A}}^{n-k-1}(p_i)$ and M (with the usual sign convention). This shows (modulo some details) that the coefficient of $[M]$ in $H_k\big(T_i, T_i \setminus S_{\mathscr{A}}^{n-k-1}(p_i)\big)$ is the intersection number $I\big(S_{\mathscr{A}}^{n-k-1}(p_i), M\big)$, as desired. $\square$

Next, suppose that we are given a triad $(W; V, V')$ and a self-indexing Morse function $f : W \to [-1, n+1]$. For all $i = -1, 0, 1, \ldots, n$, we define

$$V_i = f^{-1}\left(i + \tfrac{1}{2}\right) \text{ and } W_i = f^{-1}\left[-\tfrac{1}{2}, i + \tfrac{1}{2}\right].$$

These submanifolds can be schematically represented as follows:



Critical points of index k occur between V_{k-1} and V_k. Chopping a product cobordism off each end of W, we may replace $(W; V, V')$ by $(W_n; V_{-1}, V_n)$.

Now choose a pseudo-gradient ξ. Rather than considering the boundary of the big manifold W, we define upper and lower spheres in each factor, i.e. $S_{\mathscr{D}}^{k-1}(p) = \mathscr{D}(p) \cap V_{k-1}$ and $S_{\mathscr{A}}^{n-k-1}(p) = \mathscr{A}(p) \cap V_k$, where $k = \mathrm{ind}(p)$.

From here, we can give the promised presentation of homology. For each $k = 0, \ldots, n$, let $\mathrm{Crit}_k(f) = \{p_1^k, \ldots, p_{\ell_k}^k\}$. By excision and Lemma 3.3.2,

$$H_j(W_k, W_{k-1}) = H_j\big(f^{-1}[k - \tfrac{1}{2}, k + \tfrac{1}{2}], V_{k-1}\big) = \begin{cases} \mathbb{Z}^{\oplus \ell_k}, & j = k \\ 0, & j \neq k \end{cases}$$

Thus $H_\bullet(W_k, W_{k-1})$ is concentrated in degree k, where it has the basis

$$\{[\mathscr{D}(p)] : p \in \mathrm{Crit}_k(f)\} = \{[\mathscr{D}(p_i^k)] : i = 1, \dots, \ell_k\}.$$

Define $C_k = H_k(W_k, W_{k-1})$ and let $\partial_k : C_k \to C_{k-1}$ denote the boundary map in the homology long exact sequence of the triple (W_k, W_{k-1}, W_{k-2}).

Proposition 3.3.3. *The above-defined $(C_\bullet, \partial)$ is indeed a chain complex, with homology isomorphic to $H_\bullet(W, V)$. In the bases given by descending manifolds, the boundary map $\partial_k : C_k \to C_{k-1}$ has the following matrix:*

$$\Big[I\big(S_{\mathscr{A}}^{n-k}(p_i^{k-1}), S_{\mathscr{D}}^{k-1}(p_j^k)\big)\Big]_{\substack{i=1,\dots,\ell_{k-1}\\ j=1,\dots,\ell_k}}$$

Proof. This proof closely follows [Mil65a]. We have a commutative diagram:

$$\begin{tikzcd}
 & 0 \arrow[d] & & \\
C_{k+1} \arrow[r, "g_1"] \arrow[dr, "\partial_{k+1}"'] & H_k(W_k, W_{k-2}) \arrow[r] \arrow[d, "g_2"] & H_k(W_{k+1}, W_{k-2}) \arrow[r] & 0 \\
 & C_k \arrow[d, "\partial_k"'] & & \\
 & C_{k-1} & &
\end{tikzcd}$$

Here, the horizontal and vertical maps come from the long exact sequence of the triples (W_{k+1}, W_k, W_{k-2}) and (W_k, W_{k-1}, W_{k-2}), respectively (above, we noted that $H_k(W_{k-1}, W_{k-2}) = 0$ and $H_k(W_{k+1}, W_k) = 0$). Therefore,

$$\partial_k \circ \partial_{k+1} = \partial_k \circ g_2 \circ g_1 = 0,$$

by exactness. Thus $(C_\bullet, \partial)$ is a chain complex. Its homology is given by

$$\frac{\ker \partial_k}{\mathrm{Im}\, \partial_{k+1}} = \frac{\mathrm{Im}\, g_2}{\mathrm{Im}\, \partial_{k+1}} = \frac{H_k(W_k, W_{k-2})}{\mathrm{Im}\, g_1} = H_k(W_{k+1}, W_{k-2}),$$

again by exactness. Consider the exact sequence of the triple (W, W_s, W_{s-1}):

$$H_r(W_s, W_{s-1}) \to H_r(W, W_{s-1}) \to H_r(W, W_s) \to H_{r-1}(W_s, W_{s-1}).$$

If $r \neq s, s+1$, then the outer groups are zero, so $H_r(W, W_s) \cong H_r(W, W_{s-1})$. Applying this inductively, we find that:

$$r \leq s \Longrightarrow H_r(W, W_s) \cong H_r(W, W_{s+1}) \cong \cdots \cong H_r(W, W_n) = 0;$$

$$r \geq s+2 \Longrightarrow H_r(W, W_s) \cong H_r(W, W_{s-1}) \cong \cdots \cong H_r(W, W_{-1}) = H_r(W, V).$$

Consider the long exact sequence of the triple (W, W_{k+1}, W_{k-2}):

$$H_{k+1}(W, W_{k+1}) \to H_k(W_{k+1}, W_{k-2}) \to H_k(W, W_{k-2}) \to H_k(W, W_{k+1}).$$

The outer groups are zero, so $H_k(W_{k+1}, W_{k-2}) \cong H_k(W, W_{k-2}) \cong H_k(W, V)$, proving that $(C_\bullet, \partial)$ has k^{th} homology $H_k(W, V)$.

It remains to show that ∂_k has the matrix described above. This follows by applying Lemma 3.3.2 with $M = S_{\mathscr{D}}^{k-1}(p_j^k) \subset V_{j-1}$ for each $j = 1, \ldots, \ell_k$:

$$\begin{aligned}\partial_k[\mathscr{D}(p_j^k)] &= [\partial \mathscr{D}(p_j^k)] = [S_{\mathscr{D}}^{k-1}(p_j^k)] \\ &= \sum_{i=1}^{\ell_{k-1}} I\big(S_{\mathscr{A}}^{n-k}(p_i^{k-1}), S_{\mathscr{D}}^{k-1}(p_j^k)\big)[\mathscr{D}(p_i^{k-1})].\end{aligned}$$ □

It is interesting to see some immediate, powerful consequences of this presentation of homology. The first of these will not be needed in the sequel.

Exercise 3.3.4. Consider some (self-indexing) Morse function $f : W \to \mathbb{R}$. Prove that $\ell_k \geq b_k$ for all k, where $b_k = \dim H_k(W, V)$ and $\ell_k = \#\mathrm{Crit}_k(f)$. Further prove that $\sum_k (-1)^k \ell_k = \sum_k (-1)^k b_k$. With an algebraic lemma described in [Mil63], the stronger "Morse inequalities" follows for all $k \in \mathbb{N}$:

$$\ell_k - \ell_{k-1} + \cdots \pm \ell_1 \mp \ell_0 \geq b_k - b_{k-1} + \cdots \pm b_1 \mp b_0.$$

Another consequence (one that we will actually use) is Lefschetz duality, a generalization of Poincaré duality to compact manifolds with boundary. When considering multiple Morse functions (and their respective indices, descending manifolds, etc.), we will specify the function with a subscript. An omitted subscript always refers to the "original" Morse function f.

Corollary 3.3.5. *Given any triad $(W; V, V')$ with W oriented, there exists an isomorphism $H_k(W, V) \cong H^{n-k}(W, V')$, where $n = \dim W$ and $k \in \mathbb{Z}$.*

Proof. As above, choose a self-indexing Morse function $f : W \to [-1, n+1]$ and a pseudo-gradient ξ making (f, ξ) into a Morse-Smale pair. Then $n - f$ is also a self-indexing Morse-function and $(n - f, -\xi)$ is a Morse-Smale pair. Let $(C_\bullet, \partial)$ and $(\tilde{C}_\bullet, \tilde{\partial})$ denote the chain complexes given by these two pairs. Let $(\tilde{C}^\bullet, \delta)$ be the dual cochain complex to $(\tilde{C}_\bullet, \tilde{\partial})$. We have an isomorphism

$$H^{n-k}(\tilde{C}^\bullet, \delta) \cong H^{n-k}(W, V'),$$

by Proposition 3.3.3.[16] Note that $(\tilde{C}_\bullet, \tilde{\partial})$ looks like $(C_\bullet, \partial)$ with everything turned upside-down: $\mathrm{Crit}_k(f) = \mathrm{Crit}_{n-k}(n-f)$, $\mathscr{D}_f(p) = \mathscr{A}_{n-f}(p)$, et cetera. Thus $\tilde{C}_{n-k}$ has the basis $\{[\mathscr{A}(p)] : p \in \mathrm{Crit}_k(f)\}$, so we get a dual basis $\{[\mathscr{A}(p)]^* : p \in \mathrm{Crit}_k(f)\}$ for $\tilde{C}^{n-k}$. Then, we can define an isomorphism $\varphi_k : C_k \to \tilde{C}^{n-k}$ by $\varphi_k[\mathscr{D}(p)] = [\mathscr{A}(p)]^*$ for all $p \in \mathrm{Crit}_k(f)$.

Upper spheres of (f, ξ) are lower spheres of $(n - f, -\xi)$ and vice versa, so we can apply Proposition 3.3.3 to see that $\tilde{\partial}_{n-k+1}$ has matrix

$$\begin{aligned} X &= \Big[I\big(S^{k-1}_{\mathscr{A},n-f}(p^k_i), S^{n-k}_{\mathscr{D},n-f}(p^{k-1}_j)\big)\Big]_{\substack{i=1,\dots,\ell_k \\ j=1,\dots,\ell_{k-1}}} \\ &= \Big[I\big(S^{k-1}_{\mathscr{D}}(p^k_i), S^{n-k}_{\mathscr{A}}(p^{k-1}_j)\big)\Big]_{\substack{i=1,\dots,\ell_k \\ j=1,\dots,\ell_{k-1}}} \end{aligned}$$

Thus $\delta_{n-k} : \tilde{C}^{n-k} \to \tilde{C}^{n-k+1}$ is given by the matrix X^T. Now, we note that

$$\begin{aligned} (-1)^{(k-1)(n-k)} X^T &= \Big[(-1)^{(k-1)(n-k)} I\big(S^{k-1}_{\mathscr{D}}(p^k_j), S^{n-k}_{\mathscr{A}}(p^{k-1}_i)\big)\Big]_{\substack{i=1,\dots,\ell_{k-1} \\ j=1,\dots,\ell_k}} \\ &= \Big[I\big(S^{n-k}_{\mathscr{A}}(p^{k-1}_i), S^{k-1}_{\mathscr{D}}(p^k_j)\big)\Big]_{\substack{i=1,\dots,\ell_{k-1} \\ j=1,\dots,\ell_k}} \end{aligned}$$

is the matrix for ∂_k in Proposition 3.3.3. But φ_k identifies the bases in which the matrices were computed, so we get a diagram that commutes *up to sign*:

$$\begin{tikzcd} \cdots \arrow[r] & C_{k+1} \arrow[r, "\partial_{k+1}"] \arrow[d, "\varphi_{k+1}"] & C_k \arrow[r, "\partial_k"] \arrow[d, "\varphi_k"] & C_{k-1} \arrow[r] \arrow[d, "\varphi_{k-1}"] & \cdots \\ \cdots \arrow[r] & \tilde{C}^{n-k-1} \arrow[r, "\delta_{n-k-1}"] & \tilde{C}^{n-k} \arrow[r, "\delta_{n-k}"] & \tilde{C}^{n-k+1} \arrow[r] & \cdots \end{tikzcd}$$

Thus $\{\varphi_k : k \in \mathbb{N}\}$ induces an isomorphism on the (co)homology groups:

$$H_k(W, V) \cong H_k(C_\bullet, \partial) \cong H^{n-k}(\tilde{C}^\bullet, \delta) \cong H^{n-k}(W, V'). \qquad \square$$

While our presentation strikes a balance between the handle-attachment and gradient flow-line approaches, the modern literature is (for good reason) much more focused on the latter. As such, we will give a terse description of Morse homology (not needed below) placing more emphasis on flow-lines.

Let (f, ξ) be a Morse-Smale pair. Given distinct critical points p and q, the transverse intersection $\mathscr{D}(p) \cap \mathscr{A}(q)$ is a manifold of dimension

$$n - \big(\mathrm{ind}(p) + n - \mathrm{ind}(q)\big) = \mathrm{ind}(p) - \mathrm{ind}(q).$$

[16] By the universal coefficients theorem, two chain complexes with isomorphic homology induce dual cochain complexes with isomorphic cohomology, without need for a morphism of chain complexes (although the isomorphism will not be canonical).

Since $\mathscr{D}(p)$ and $\mathscr{A}(q)$ are both closed under the flow of ξ, this flow induces a free $\mathbb{R}$-action on $\mathscr{D}(p) \cap \mathscr{A}(q)$ (the intersection contains no critical points). Thus we define a manifold $\mathscr{M}(p,q) = \big(\mathscr{D}(p) \cap \mathscr{A}(q)\big)/\mathbb{R}$, the moduli space of integral curves going from p to q. Notice that $\mathscr{M}(p,q)$ has dimension $\operatorname{ind}(p) - \operatorname{ind}(q) - 1$. We orient $\mathscr{M}(p,q)$ by the convention in [Hut02]. Namely, consider $\gamma \in \mathscr{M}(p,q)$ and $x \in \gamma$. We have (canonical) exact sequences

$$\begin{tikzcd}
 & & & 0 \arrow[d] & \\
 & 0 \arrow[d] & & T_x\mathscr{A}(q) \arrow[d] & \\
 & T_x\gamma \arrow[d] & & T_xW \arrow[d] & \\
0 \arrow[r] & T_x\big(\mathscr{D}(p) \cap \mathscr{A}(q)\big) \arrow[r] \arrow[d] & T_x\mathscr{D}(p) \arrow[r] & \dfrac{T_xW}{T_x\mathscr{A}(q)} \arrow[r] \arrow[d] & 0 \\
 & T_\gamma\mathscr{M}(p,q) \arrow[d] & & 0 & \\
 & 0 & & &
\end{tikzcd}$$

In an exact sequence $0 \to A \to B \to C \to 0$ of vector spaces, orientations on two of the three spaces induce an orientation on the third, so that any splitting $B \cong A \oplus C$ is orientation-preserving. Notice that γ is oriented by ξ, W and $\mathscr{D}(p)$ have arbitrary orientations, and $\mathscr{A}(q)$ inherits an orientation from $\mathscr{D}(q)$ and W. Thus, the diagram defines an orientation on $T_\gamma\mathscr{M}(p,q)$. Flowing along γ shows that this is independent of x. A local trivialization of the flow near x shows that this orientation is a continuous function of γ. Therefore, if $\operatorname{ind}(p) = \operatorname{ind}(q) + 1$, we have a sign on each point in the finite set $\mathscr{M}(p,q)$, so we get a signed count $\#\mathscr{M}(p,q)$ of flow lines from p to q.

Let $\mathscr{C}_k = \mathbb{Z}\big[\operatorname{Crit}_k(f)\big]$ (free abelian group). Define $\partial_{k+1} : \mathscr{C}_{k+1} \to \mathscr{C}_k$ by

$$\partial(p) = \sum_{q \in \operatorname{Crit}_k(f)} \#\mathscr{M}(p,q) \cdot q.$$

Proposition 3.3.6. *The above-defined $(\mathscr{C}_\bullet, \partial)$ is indeed a chain complex, with homology isomorphic to $H_\bullet(W,V)$.*

Proof. By Corollary 3.2.5, we can replace f by a self-indexing function g, where (g, ξ) is still Morse-Smale. But $(\mathscr{C}_\bullet, \partial)$ only depends on ξ, so we may assume that f was self-indexing to begin with, and relate the groups $(\mathscr{C}_\bullet, \partial)$ to the above chain complex $(C_\bullet, \partial)$, defined in terms of sub-levels of f.

For all $k \in \mathbb{N}$, we define an isomorphism $\varphi_k : \mathscr{C}_k \to C_k$ by $\varphi_k(p) = [\mathscr{D}(p)]$ for every $p \in \mathrm{Crit}_k(f)$. Similarly to the previous proof, we will show that

$$\varphi_k \circ \partial_{k+1} = (-1)^{n+k} \partial_{k+1} \circ \varphi_{k+1}$$

for all k. This shows that $(\mathscr{C}_\bullet, \partial)$ is a chain complex and each φ_k induces

$$H_k(\mathscr{C}_\bullet, \partial) \cong H_k(C_\bullet, \partial) \cong H_k(W, V).$$

In Proposition 3.3.3, we described the matrix of $\partial_k : C_k \to C_{k-1}$ in terms of intersection numbers in $V_k = f^{-1}(k + \frac{1}{2})$. For any critical points p and q of indices $k+1$ and k, respectively, we will show that

$$\#\mathscr{M}(p, q) = (-1)^{n+k} I\big(S^{n-k-1}_{\mathscr{A}}(q), S^k_{\mathscr{D}}(p)\big). \tag{6}$$

From this, it follows at once that, in the bases identified under $\{\varphi_k : k \in \mathbb{N}\}$, the matrices of the two boundary maps ∂_{k+1} just differ by a sign of $(-1)^{n+k}$.

Since there are no critical values between k and $k+1$, every (non-critical) point in $\mathscr{D}(p) \cap \mathscr{A}(q)$ lies on an integral curve that intersects V_k exactly once. Therefore, we have an embedding $g : \mathscr{M}(p, q) \to V_k$ given by $g(\gamma) = \gamma \cap V_k$. Consider $x \in \operatorname{Im} g$. Let the intersection $x \in S^{n-k-1}_{\mathscr{A}}(q) \cap S^k_{\mathscr{D}}(p)$ have sign $\imath$; let $\gamma = g^{-1}(x)$ have sign $\jmath$. Then (6) follows immediately from $\jmath = (-1)^{n+k}\imath$ (for any $x \in \operatorname{Im} g$), which is left as the following exercise. □

Exercise 3.3.7. Fix a basis $v_1, \ldots, v_n$ for $T_{g(x)}W$, such that:

- v_{k+1} is a positive multiple of $\xi_{g(x)}$ and $v_i \in T_{g(x)}V_k$ for all $i \neq k$;
- $T_{g(x)}\mathscr{D}(p) = \mathrm{span}(v_1, \ldots, v_{k+1})$, so $T_{g(x)}S^k_{\mathscr{D}}(p) = \mathrm{span}(v_1, \ldots, v_k)$;
- $T_{g(x)}\mathscr{A}(p) = \mathrm{span}(v_{k+1}, \ldots, v_n)$, so $T_{g(x)}S^{n-k-1}_{\mathscr{A}}(q) = \mathrm{span}(v_{k+2}, \ldots, v_n)$.

Use this basis to wade through the definitions and show that $\jmath = (-1)^{n+k}\imath$. (Depending on your sign conventions, your final answer may actually differ. The only thing important in the above proof is that the sign is uniquely determined by n and k, for it still ensures that $\varphi_k \circ \partial_{k+1} = \pm\, \partial_{k+1} \circ \varphi_{k+1}$.)

Exercise 3.3.8. Returning to Figure 9(d), use these techniques to calculate the homology of the torus, via the Morse-Smale pseudo-gradient therein.

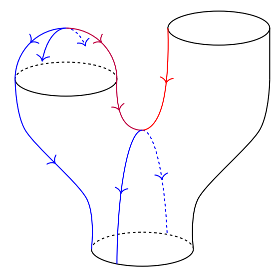


Figure 12: One integral curve between two critical points

4 Eliminating Critical Points

4.1 Cancellation of neighboring indices

In this section, we develop a tool for reducing the number of critical points of a Morse function, which will eventually result in the promised minimal presentation of homology. This is the most technical part of the entire proof, and as such, will be the sketchiest. In particular, we will assume two lemmas, which are plausible but tricky to prove. This all closely follows [Mil65a].

Suppose we have a triad $(W; V_0, V_1)$ and a Morse function $f : W \to \mathbb{R}$, with exactly two critical points p and q of indices $k+1$ and k, respectively. Furthermore, we suppose that ξ is a pseudo-gradient such that $\mathscr{D}(p) \cap \mathscr{A}(q)$ consists of one integral curve γ going from p to q ($\mathscr{M}(p,q)$ is a singleton). We will prove that these two critical points "cancel" with each other.

Example 4.1.1. In previous examples, we have already seen this in action! In Figure 4, we saw that the pair of pants is a cylinder $S^1 \times I$ with a smooth 1-handle attached; in Figure 1, we saw that the cylinder can be recovered from a pair of pants by "capping a leg" with a 2-handle. The Morse-theoretic version is depicted in Figure 12, where γ is illustrated in purple (as usual, the Morse function is height and the pseudo-gradient is the actual gradient).

The first lemma provides the local model where our modifications occur.

Lemma 4.1.2. *There exists a smooth function $v : \mathbb{R} \to \mathbb{R}$ and an interior chart $\Phi : \mathbb{R}^n \to W$ containing γ and the critical points p and q, such that:*

- $\left|\frac{dv}{dt}\right| = 1$ *outside of* $(1/4, 3/4)$ *and* $v(0) = v(1) = 0$*;*
- $\Phi(0) = p$ *and* $\Phi(e_1) = q$*, where* $e_1 = (1, 0, \ldots, 0)$*;*
- Φ *pulls back* ξ *to* $\zeta(x) = \big(v(x_1), x_2, \ldots, x_k, x_{k+1}, -x_{k+2}, \ldots, -x_n\big)$.

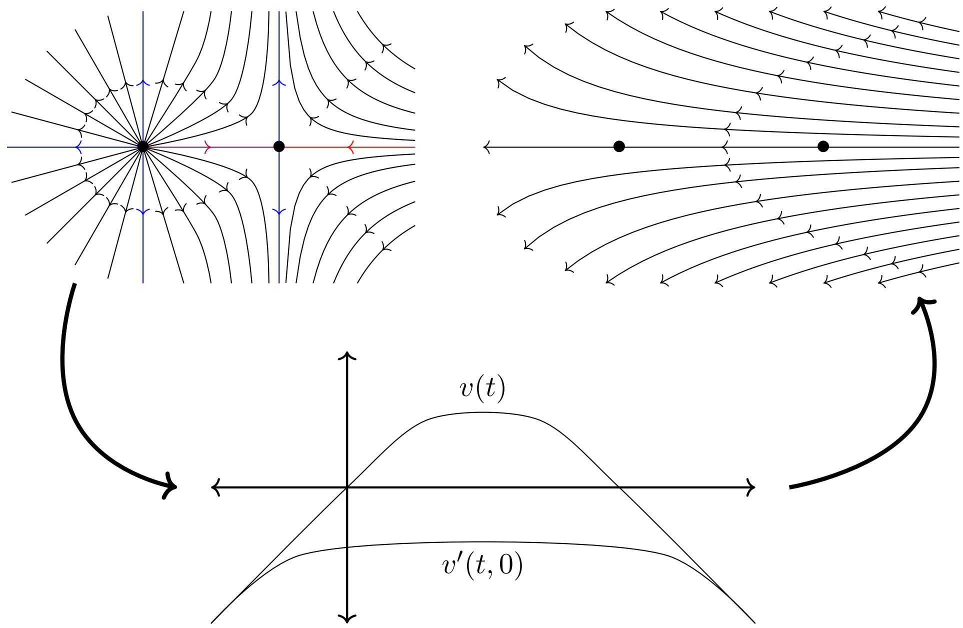


Figure 13: Modifying the pseudo-gradient near γ

The chart Φ (for $n = 2$) is drawn in the upper-left corner of Figure 13. The upper-right shows the modified version with critical points eliminated. The function v lies on the graph below, along with the modified function used to define this new vector field.

Proposition 4.1.3. *Under the above assumptions (f has two critical points p and q, of respective indices $k+1$ and k, which are connected by a single flow line going from p to q), we have $\mu(W; V_0, V_1) = 0$.*

Proof. Following [Mil65a], this proof proceeds from the lemma in four steps:

(a) There are neighborhoods $U' \subset U$ of γ, with closures contained in $\operatorname{Im}\Phi$, such that no point in U' flows out of U and then flows back into U'.

We can choose U to be any neighborhood of γ with closure $\overline{U} \subset \operatorname{Im}\Phi$. Choose a metric on W and let Ψ_t denote the flow of ξ. We define

$$\delta_0(x) = \inf\{\operatorname{dist}(\Psi_t(x), q) : t \geq 0 \text{ and } \Psi_t(x) \text{ is well-defined}\},$$
$$\delta_1(x) = \inf\{\operatorname{dist}(\Psi_t(x), p) : t \leq 0 \text{ and } \Psi_t(x) \text{ is well-defined}\}.$$

These are upper semi-continuous functions $\delta_i : W \to [0, \infty)$, such that

$$\delta_0(x) = 0 \iff x \in \mathscr{A}(q) \qquad \text{and} \qquad \delta_1(x) = 0 \iff x \in \mathscr{D}(p).$$

Thus $\delta = \delta_0 + \delta_1$ is an upper semi-continuous function $W \to [0, \infty)$, which vanishes precisely along γ. Since U is open and W is compact, there is some $2\epsilon > 0$ bounding δ from below on $W \setminus U$. Notice that δ_0 (resp. δ_1) is decreasing (resp. increasing) along any integral curve.

Define $U' = \delta^{-1}[0, \epsilon) \cap U$. Suppose that η is an integral curve of ξ containing points x, y, z (which occur in that order), where $x, z \in U'$ and $y \notin U$. This yields a contradiction, since $\delta(y) \geq 2\epsilon$, but we have

$$\delta_0(y) \leq \delta_0(x) \leq \delta(x) < \epsilon \qquad \text{and} \qquad \delta_1(y) \leq \delta_1(z) \leq \delta(z) < \epsilon.$$

(b) We can modify ξ on a compact subset of U', so that ξ is non-vanishing and every point $x \in U$ flows both forwards and backwards out of U.

Let $K \subset U'$ be a compact neighborhood of γ. We choose $v' : \mathbb{R}^2 \to \mathbb{R}$ such that $v(s, 0) < 0$ for every $s \in \mathbb{R}$ and if $\Phi(x_1, \ldots, x_n) \notin K$, then

$$v'\big(x_1, \sqrt{x_2^2 + \cdots + x_n^2}\big) = v(x_1).$$

This is shown in Figure 13. We modify ξ by replacing ζ with

$$\zeta'(x) = \Big((v'\big(x_1, \sqrt{x_2^2 + \cdots + x_n^2}\big), x_2, \ldots, x_{k+1}, -x_{k+2}, \ldots, -x_n\Big).$$

Since $\zeta \equiv \zeta'$ outside of $\Phi^{-1}(K)$, this defines a smooth vector field ξ', which agrees with ξ outside of K. Note that ξ' cannot vanish, since

$$x_2 = x_3 = \cdots = x_n = 0 \Longrightarrow v'\big(x_1, \sqrt{x_2^2 + \cdots + x_n^2}\big) < 0.$$

Let η be an integral curve of ξ' starting at some $\Phi(a_1, \ldots, a_n) \in U$. Then the component of $\eta \cap \operatorname{Im} \Phi$ containing $\Phi(a_1, \ldots, a_n)$ is given by

$$\eta_i(0) = a_i \qquad \text{and} \qquad \frac{d\eta_i}{dt}(t) = \zeta_i'\big(\eta(t)\big),$$

for all $i = 1, \ldots, n$ (for the rest of (b), the locality allows us to make all necessary identifications under Φ; if we leave $\operatorname{Im} \Phi$, then we have left U as well and thus our work is done). In particular, it follows that

$$\eta_i(t) = \begin{cases} a_i e^t, & 2 \leq i \leq k+1 \\ a_i e^{-t}, & k+2 \leq i \leq n \end{cases}$$

If $a_i \neq 0$ with $2 \leq i \leq k+1$, then η_i is unbounded (in forward time), so η leaves the compact $\overline{U}$. Assume that $a_2 = \cdots = a_{k+1} = 0$. Then

$$\sqrt{\eta_2(t)^2 + \cdots + \eta_n(t)^2} = \frac{\sqrt{a_{k+2}^2 + \cdots + a_n^2}}{e^{-t}}$$

approaches zero as t grows (it will equal zero iff $a_{k+2} = \cdots = a_n = 0$). Thus for t very large, there is some $\epsilon < 0$ which is an upper bound for

$$\frac{d\eta_1}{dt}(t) = \zeta_1\big(\eta(t)\big) = v'\Big(\eta_1(t), \sqrt{\eta_2(t)^2 + \cdots + \eta_n(t)^2}\Big),$$

by the properties of v'. For T very large and any $t \geq 0$, we then have

$$\eta_1(T+t) \leq \eta_1(T) - \epsilon t.$$

Thus η_1 is unbounded (in forward time), so η leaves the compact $\overline{U}$. The result for flowing backwards follows by analogous casework.

(c) If ξ' is the modifed field, then every flow-line of ξ' goes from V_1 to V_0.

Let η be any integral curve of ξ'. Suppose that we can find $y \in \eta$, which never flows forwards into U'. Then at points past y, the curve η is just an integral curve of ξ, since $\xi \equiv \xi'$ outside of U'. But if a point x does not flow forward along the pseudo-gradient ξ towards p or q, then x must flow forward to V_0. Since $p, q \in U'$, this is the case for y.

Thus, it suffices to find some such y. If $\eta \cap U' = \emptyset$, then we can pick any point along η. If $x \in \eta \cap U'$, then (b) shows that x flows along ξ' to some $y \in \eta \setminus U$. Then (a) shows that y never flows back into U' (again using the fact that $\xi \equiv \xi'$ outside of U'). This proves that η terminates on V_0. An analogous argument shows that η starts on V_1.

(d) The cobordism $(W; V_0, V_1)$ is trivial (i.e. has Morse number zero).

Let Ψ_t denote the flow of ξ'. We can uniquely define a smooth function $\tau : V_1 \to (0, \infty)$ by the condition that $\Psi_{\tau(x)}(x) \in V_0$. Then the map

$$(x, t) \mapsto \Psi_{\tau(x)t}(x)$$

is an equivalence $(V_0 \times I; V_0 \times 0, V_0 \times 1) \cong (W; V_0, V_1)$. $\square$

For convenience, we rewrite this result in the following handy corollary. We say the critical points p and q therein are "cancelled against each other."

Corollary 4.1.4. *Fix a Morse function $f : W \to \mathbb{R}$ and pseudo-gradient ξ. Let f have critical values $c < d$ and no critical values in the interval (c, d). If $p \in f^{-1}(d)$ and $q \in f^{-1}(c)$ satisfy all the hypotheses of Proposition 4.1.3 (but f may have more critical points), then there exists a Morse function $g : W \to \mathbb{R}$, with $Crit(g) = Crit(f) \setminus \{p, q\}$ and $g \equiv f$ near $Crit(g) \cup \partial W$.*

Proof. Let $5\epsilon = d - c$. By the comment following Proposition 3.2.4, we can modify f in neighborhoods of p and q, so that $f(p) = d - 2\epsilon$ and $f(q) = c + 2\epsilon$, while ξ remains a pseudo-gradient. Then the hypotheses of Proposition 4.1.3 are satisfied on $c_1 = \big(f^{-1}[c+\epsilon, d-\epsilon], f^{-1}(c+\epsilon), f^{-1}(d-\epsilon)\big)$. Define triads

$$c_0 = \big(f^{-1}(-\infty, c+\epsilon]; V_0, f^{-1}(c+\epsilon)\big), \qquad c_2 = \big(f^{-1}[d-\epsilon, \infty); f^{-1}(d-\epsilon), V_1\big).$$

Then $(W; V_0, V_1) = c_0 c_1 c_2$ and $\mu(c_1) = 0$, so we replace $f|_{c_1}$ by projection onto the second factor of $f^{-1}(d-\epsilon) \times [c+\epsilon, d-\epsilon]$. By Proposition 2.2.10, this modification glues in smoothly, yielding the desired function g. $\square$

Next, we state the second lemma, which is known as the "Whitney trick."

Lemma 4.1.5. *Consider a simply-connected, n-dimensional manifold V, with $\partial V = \emptyset$. Let $M, N \subset V$ be closed submanifolds of dimensions r and s, respectively. Consider intersection points $x, y \in M \cap N$. Suppose that:*

- *$r \geq 2$, $s \geq 3$, $r + s = n$ and $M \pitchfork N$ (in particular, $M \cap N$ is finite);*
- *M and N are connected, and if $r = 2$, then $\pi_1(V \setminus N) = 1$;*
- *M and N are both orientable and the two intersection points x and y have opposite signs (V is orientable, because it is simply connected).*

Then there is an isotopy $h_t : V \to V$ with $h_1(M) \cap N = M \cap N \setminus \{x, y\}$, $h_0 = Id_V$ and $h(z,t) = z$ for z near $M \cap N \setminus \{x, y\}$ (which gives $h_1(M) \pitchfork N$).

Using this lemma, we can strengthen Proposition 4.1.3 in certain cases. This result, due to Smale, is known as the "Handle cancellation theorem." Consider a triad $(W; V_0, V_1)$, where W, V_0 and V_1 are simply-connected. Suppose that $f : W \to \mathbb{R}$ is a Morse function with exactly two critical points p and q of indices $k+1$ and k, respectively. Assume that $f(q) < f(p)$ and consider a pseudo-gradient ξ. For any $f(q) < a < f(p)$, we may consider the spheres $S^{n-k}_{\mathscr{A}}(q)$ and $S^{k-1}_{\mathscr{D}}(p)$ in the level $V = f^{-1}(a)$, where $n = \dim W$.

Theorem 4.1.6. *Suppose that $n \geq 6$ and $I\big(S^{n-k-1}_{\mathscr{A}}(q), S^{k}_{\mathscr{D}}(p)\big) = \pm 1$ in V (if (f, ξ) is Morse-Smale, then the latter is equivalent to $\#\mathscr{M}(p,q) = \pm 1$). Then we have $\mu(W; V_0, V_1) = 0$.*

Proof. By possibly modifying ξ, we may assume that (f, ξ) is Morse-Smale. Because Lemma 3.2.1 just alters upper/lower spheres by an ambient isotopy, the intersection numbers are preserved. We split up the proof into cases:[17]

[17]We omit the case of $k = 1$, because it requires a more detailed version of the Whitney trick than the one given above. This case will not be needed in anything that follows.

- Suppose that $k = 0$. Then $S^0_{\mathscr{D}}(p)$ is just two points, so the condition $I\big(S^{n-1}_{\mathscr{A}}(q), S^0_{\mathscr{D}}(p)\big) = \pm 1$ ensures that $S^{n-1}_{\mathscr{A}}(q) \cap S^0_{\mathscr{D}}(p)$ is a single point. Since any flow line in $\mathscr{A}(q) \cap \mathscr{D}(p)$ must intersect V, we see that there is only such flow line. Thus Proposition 4.1.3 gives $\mu(W; V_0, V_1) = 0$.

- Suppose that $2 \leq k \leq n-4$. Assume that $S^{n-k-1}_{\mathscr{A}}(q) \cap S^k_{\mathscr{D}}(p)$ contains more than two points. Then $I\big(S^{n-k-1}_{\mathscr{A}}(q), S^k_{\mathscr{D}}(p)\big) = \pm 1$ implies that there are intersection points $x, y \in S^{n-k-1}_{\mathscr{A}}(q) \cap S^k_{\mathscr{D}}(p)$ of opposite sign. For $M = S^k_{\mathscr{D}}(p)$ and $N = S^{n-k-1}_{\mathscr{A}}(q)$, we will show that Lemma 4.1.5 gives some $h_t : V \to V$ as described therein. In particular, $h_0 = \mathrm{Id}_V$ implies that we can modify ξ on $f^{-1}(a-\epsilon, a]$ (for some very small ϵ), so that the new upper sphere of q is $h_1(M)$. Therefore, $h_1(M) \pitchfork N$ and there are two fewer intersection points. Continuing in this fashion, we can modify ξ so that $S^{n-k-1}_{\mathscr{A}}(q) \cap S^k_{\mathscr{D}}(p)$ has at most two points, and (f, ξ) is still Morse-Smale. Then we can proceed as when $k = 0$.

  It remains to show that the Whitney trick applies. We have $r = k \geq 2$ and $s = n-k-1 \geq 3$. The spheres in question are certainly connected and orientable. We also have $r + s = n - 1 = \dim V$. Recall that W has the homotopy type of V with a $(k+1)$-cell and an $(n-k)$-cell attached. Each of these cells has dimension ≥ 3, so attaching them does not change the fundamental group. Hence $\pi_1(V) = \pi_1(W) = 1$. If $k \geq 3$, we are done, but if $k = 2$, we need to show $\pi_1(V \setminus N) = 1$.

  Suppose that $k = 2$. Flowing along ξ gives a diffeomorphism

  $$V_0 \setminus S^1_{\mathscr{D}}(q) \cong V \setminus S^{n-3}_{\mathscr{A}}(q).$$

  Let $T \subset V_0$ be a tubular neighborhood of the framed sphere $S^1_{\mathscr{D}}(q)$. Then we have $T \cong D^{n-k-1} \times S^1_{\mathscr{D}}(q)$ and $n-k-1 \geq 3$, so $\pi_1(T) \cong \mathbb{Z}$ and $\pi_1\big(T \setminus S^1_{\mathscr{D}}(q)\big) \cong \pi_1(T)$, where the latter is induced by inclusion. Notice that $\big(V_0 \setminus S^1_{\mathscr{D}}(q)\big) \cup T = V_0$ and $\big(V_0 \setminus S^1_{\mathscr{D}}(q)\big) \cap T = T \setminus S^1_{\mathscr{D}}(q)$, so we may apply van Kampen's theorem to get a pushout diagram:

$$\begin{tikzcd}
 & \pi_1\big(T \setminus S^1_{\mathscr{D}}(q)\big) = \mathbb{Z} \arrow[dl] \arrow[dr, "\cong"] & \\
\pi_1\big(V_0 \setminus S^1_{\mathscr{D}}(q)\big) \arrow[dr] & & \pi_1(T) = \mathbb{Z} \arrow[dl] \\
 & \pi_1(V_0) = 1 &
\end{tikzcd}$$

Since pushouts preserve isomorphisms, this gives the desired result:

$$\pi_1\big(V \setminus S^{n-3}_{\mathscr{A}}(q)\big) \cong \pi_1\big(V_0 \setminus S^1_{\mathscr{D}}(q)\big) \cong \pi_1(V_0) = 1.$$

- Suppose that $3 \leq k \leq n-1$. The pair $(-f, -\xi)$ is still Morse-Smale. Note that $p \in \mathrm{Crit}_{n-k-1}(-f)$ and $q \in \mathrm{Crit}_{n-k}(-f)$. Note also that

  $$\begin{aligned} I\big(S^k_{\mathscr{A},-\xi}(p), S^{n-k-1}_{\mathscr{D},-\xi}(q)\big) &= I\big(S^k_{\mathscr{D}}(p), S^{n-k-1}_{\mathscr{A}}(q)\big) \\ &= (-1)^{k(n-k-1)}\, I\big(S^{n-k-1}_{\mathscr{A}}(q), S^k_{\mathscr{D}}(p)\big) = \pm 1. \end{aligned}$$

  Since $n-k-1 \leq n-4$, we then have $\mu(W; V_0, V_1) = \mu(W; V_1, V_0) = 0$ by one of the previous cases. (We will not need the case of $k = n-2$, which corresponds to the unproven case of $k = 1$ above.)

Since we must have $0 \leq k < k+1 \leq n$, this covers all possible cases. □

4.2 Eliminating extremal indices

Our first use of §4.1 is the elimination of critical points in the "extremal" index range, i.e. 0, 1, $n-1$ and n. Suppose that $(W; V, V')$ is a triad with $\dim W = n$ and that $f : W \to [-1, n+1]$ is a self-indexing Morse function. Recall the notation of $V_i = f^{-1}(i + \frac{1}{2})$ and $W_i = \big[-\frac{1}{2}, i + \frac{1}{2}\big]$ used in §3.3. As usual, we will also consider a pseudo-gradient ξ and its flow.

Proposition 4.2.1. *We will prove the following about extremal indices:*

(a) If $H_0(W, V) = 0$, then any $p \in Crit_0(f)$ can be cancelled against some critical point of index 1 (in particular, we have $Crit_1(f) \neq \emptyset$).

(b) If $H_0(W, V') = 0$, then any $p \in Crit_n(f)$ can be cancelled against some critical point of index $n-1$ (in particular, we have $Crit_{n-1}(f) \neq \emptyset$).

Now suppose that W is simply-connected and $n \geq 5$.

(c) Suppose that V is connected and $Crit_0(f) = \emptyset$. For any $p \in Crit_1(f)$, the function f can be modified to introduce two critical points q and r of indices 2 and 3, such that p and q cancel.

(d) Suppose that V' is connected and $Crit_n(f) = \emptyset$. For any $p \in Crit_{n-1}(f)$, the function f can be modified to introduce two critical points q and r of indices $n-2$ and $n-3$, such that p and q cancel.

In (c) and (d), we say that the critical point r is "traded" for the point p.

Proof. Note that (a) $\iff$ (b) and (c) $\iff$ (d), by reversing the cobordism (and replacing (f, ξ) by $(n - f, -\xi)$). Thus it suffices to prove (a) and (c).

Consider the boundary map $\partial_1 : C_1 \to C_0$ defined in §3.3. The condition $H_0(W, V) = 0$ is equivalent to ∂_1 being surjective. Fix some $q \in \mathrm{Crit}_0(f)$. For any $p \in \mathrm{Crit}_1(f)$, the spheres $S^{n-1}_{\mathscr{A}}(q), S^0_{\mathscr{D}}(p) \subset V_0$ intersect in at most the two points of $S^0_{\mathscr{D}}(p)$. In order to have $[\mathscr{D}(q)] \in \operatorname{Im} \partial_1$, there must be some $p \in \mathrm{Crit}_1(f)$ with $I\big(S^{n-1}_{\mathscr{A}}(q), S^0_{\mathscr{D}}(p)\big) = \pm 1$, by Proposition 3.3.3 (the only other options are 0 and ± 2, which are divisible by 2). Then $S^{n-1}_{\mathscr{A}}(q) \cap S^0_{\mathscr{D}}(p)$ is a single point, so there is one flow line from p to q. Thus p and q cancel by Corollary 4.1.4, proving (a). The proof of (c) is a little more involved, and the argument we give is a sketch of the full proof in [Mil65a].

Fix some $p \in \mathrm{Crit}_1(f)$. The idea is as follows: we find a sphere $S^1_0 \subset V_1$ which intersects $S^{n-2}_{\mathscr{A}}(p)$ once; then we generate a cancelling pair of critical points q and r (by a sort of Proposition 4.1.3 in reverse), such that $S^1_{\mathscr{D}}(q)$ is the prescribed sphere, so p and q satisfy Corollary 4.1.4. We depict this process in Figure 14, where the rainbow reflects chronology. We can easily find an embedded curve $\varphi : I \to V_1$ that intersects $S^{n-2}_{\mathscr{A}}(p)$ once at $\varphi(1/2)$, transversely, and avoids every other upper sphere in V_1. Then $\varphi(0)$ and $\varphi(1)$ are disjoint from all upper spheres, so they flow along ξ to some $x, y \in V_0$. But $V_0 \cong V$ is connected (because $\mathrm{Crit}_0(f) = \emptyset$) with dimension $n - 1 \geq 2$, so we can find an embedded path from x to y that misses the lower spheres $S^0_{\mathscr{D}}(p')$ for all $p' \in \mathrm{Crit}_1(f)$. Then this path flows backwards along ξ to V_1, where it defines an embedded curve $\psi : I \to V_1$ that misses all upper spheres and satisfies $\psi(i) = \varphi(i)$ for $i = 0, 1$. Gluing ψ and φ yields a map $S^1 \to V_1$, which intersects $S^{n-2}_{\mathscr{A}}(p)$ once, transversely, and is smooth except at the two points of gluing. This can be approximated by a self-transverse map $S^1 \to V_1$ which agrees with the original along $\varphi|_{[1/4,3/4]}$. But $\dim V_1 \geq 4 > 2 \cdot \dim S^1$, so this new map is an embedding. If this approximation is sufficiently close, then this gives the desired $S^1_0 \subset V_1$ intersecting $S^{n-2}_{\mathscr{A}}(p)$ once, transversely.

Using Lemmas 3.2.1 and 3.2.2, we can modify ξ so that S^1_0 misses $S^1_{\mathscr{D}}(q')$ for every $q' \in \mathrm{Crit}_2(f)$. Thus S^1_0 flows backwards along ξ to a sphere $S^1_1 \subset V_2$. By construction, the sphere S^1_1 is disjoint from $S^{n-3}_{\mathscr{A}}(q')$ for all $q' \in \mathrm{Crit}_2(f)$. By an isotopy theorem of Whitney,[18] because $\dim V_2 \geq 4 = 2 + 2 \cdot \dim S^1$, any two embeddings of $S^1 \to V_2$ are ambient-isotopic if and only if they are homotopic. But W is formed from V_2 by successively attaching cells of dimension $n - 1, n - 2 \geq 3$ on one side and $3, 4, \ldots, n - 1, n$ on the other, so $\pi_1(V_2) = \pi_1(W) = 1$. Thus any two 1-spheres in V_2 are ambient-isotopic. By Lemma 3.2.1, we can modify ξ on $f^{-1}[2\frac{1}{4}, 2\frac{1}{2}]$ so that S^1_1 becomes any

[18] As well as the isotopy extension theorem of Thom, Palais and Cerf [Kos93].

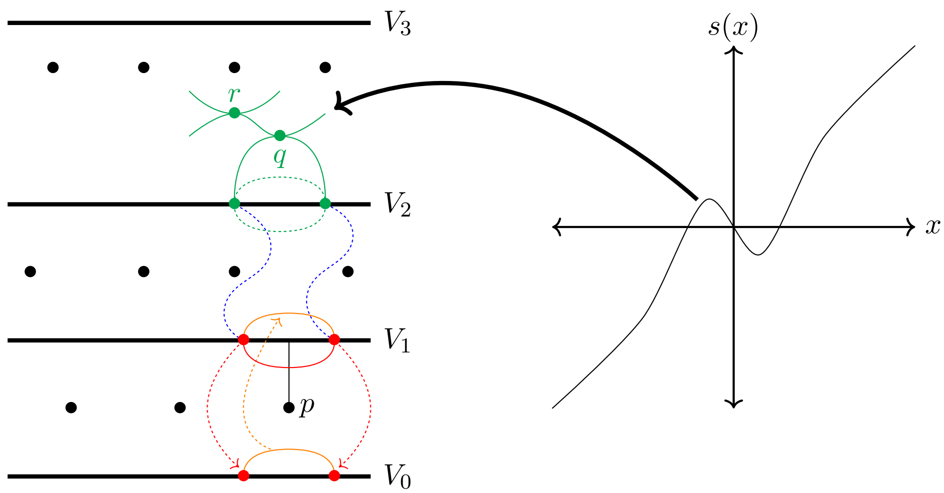


Figure 14: Trading critical points of index 1 and 3

embedded $S^1 \subset V_2$ of our choosing. The upper spheres in V_2 are transformed by the same modification to ξ, so they remain disjoint from the modified S^1_1.

We wish to generate critical points of index 2 and 3. Towards this end, we pick some small chart $\varphi : \mathbb{R}^n \to f^{-1}(2\frac{1}{2}, 3)$, such that $f \circ \varphi(x) = x_1 + 2\frac{3}{4}$. For any $0 \leq k < n$, it is possible to find a function $g : \mathbb{R}^n \to \mathbb{R}$ such that:

- $g(x) = x_1$ outside of a compact set;
- g has two critical points $\tilde{q}$ and $\tilde{r}$, both non-degenerate;
- $\operatorname{ind}(\tilde{q}) = k$, $\operatorname{ind}(\tilde{r}) = k + 1$ and $g(\tilde{q}) < g(\tilde{r})$.

For $n = 1$, this is straightforward and illustrated in Figure 14 (in this case, call the function s). For $n > 1$, we can carefully interpolate $x \mapsto x_1$ with

$$x \mapsto s(x_1) - x_2^2 - \cdots - x_{k+1}^2 + x_{k+2}^2 + \cdots + x_n^2, \tag{7}$$

so that the only critical points are the two arising from (7). Refer to [Mil65a] for the full details. Taking $k = 2$, we can define a new Morse function

$$\tilde{f}(x) = \begin{cases} f(x), & x \notin \operatorname{Im} \varphi \\ g(v), & x = \varphi(v) \text{ for some } v \in \mathbb{R}^n \end{cases}$$

This gives two new critical points $q = \varphi(\tilde{q})$ and $r = \varphi(\tilde{r})$ of indices 2 and 3, respectively. We now make the following modifications to the vector field ξ:

- on $\operatorname{Im}\varphi$, modify ξ to ensure that it remains a pseudo-gradient;
- on $f^{-1}[2\frac{3}{4}, 3]$, modify ξ so that $\mathscr{A}(r) \cap \mathscr{D}(r') = \emptyset$ for all $r' \in \mathrm{Crit}_3(f)$;
- on $f^{-1}[2\frac{1}{4}, 2\frac{1}{2}]$, modify ξ so that $S_1^1 = S_{\mathscr{D}}^1(q)$ (as discussed above).

Since $S_1^1 \cap S_{\mathscr{A}}^{n-3}(q')$ for all $q' \in \mathrm{Crit}_2(f)$ (with whatever modifications were made to ξ), we can modify $\tilde{f}$ on a neighborhood of q and r (e.g. on $\operatorname{Im}\varphi$) so that ξ is still a pseudo-gradient, $\tilde{f}(q) = 2$ and $\tilde{f}(r) = 3$. Since ξ has not been modified on $f^{-1}[1, 2]$, we see that $\mathscr{D}(q) \cap V_2 = S_1^1$ gives $\mathscr{D}(q) \cap V_1 = S_0^1$, by flowing along ξ from V_2 to V_1. Thus $S_{\mathscr{A}}^{n-2}(p), S_{\mathscr{D}}^1(q) \subset V_1$ intersect in just one point, so there is one flow line from q to p. Therefore, p and q cancel by Corollary 4.1.4, proving (c). □

Corollary 4.2.2. *Suppose that $(W; V, V')$ is a triad of dimension $n \geq 5$, with W simply-connected and $H_\bullet(W, V) = 0$. Then this triad admits some Morse function without any critical points of 0, 1, $n-1$ or n.*

Proof. By the universal coefficients theorem, we can see that $H^\bullet(W, V) = 0$. Then $H_k(W, V') \cong H^{n-k}(W, V) = 0$ for any $k \in \mathbb{Z}$, by Proposition 3.3.5 (since simply-connected manifolds are orientable). In particular, we have $\tilde{H}_0(V) = \tilde{H}_0(W) = 0$ and $\tilde{H}_0(V') = \tilde{H}_0(W) = 0$, so V and V' are connected. Choose a self-indexing Morse function $f : W \to \mathbb{R}$. By Proposition 4.2.1 (a) and (b), we can cancel all critical points of index 0 or n against critical points of index 1 or $n-1$. Then $\mathrm{Crit}_0(f) = \mathrm{Crit}_n(f) = \emptyset$. Using (c), we can now trade all critical points of index 1 for points of index 3. Because $3 < n$, we still have $\mathrm{Crit}_n(f) = \emptyset$, so we use (d) to trade all critical points of index $n-1$ for critical points of index $n-3$. Since $n-3 > 1$, the resulting function has no critical points of index 0, 1, $n-1$ or n. □

4.3 Eliminating middle indices

In this section, which is independent of §4.2, we will prove the simplest case of minimality for Morse homology: this is the case where the homology vanishes and it is possible to find a Morse function with no critical points. While it is also useful to have minimality in other cases (i.e. to find a Morse function with a minimal number of critical points necessary to generate some non-zero homology), we will not need this. Such a result occurs in [Kos93].

We start with the "Basis Theorem" of [Mil65a], which translates linear algebra on homology groups into alterations of the descending manifolds.

Proposition 4.3.1. *Let $(W; V_0, V_1)$ be a triad with $\dim W = n$. Suppose that $f : W \to \mathbb{R}$ is a Morse function, whose critical points all have index k and all lie on one level $f^{-1}(a)$. Fix a pseudo-gradient ξ. If $2 \leq k \leq n-2$ and W is connected, then for any basis of $H_k(W, V_0)$, there exists a Morse function g, which a pseudo-gradient ζ, such that:*

- *$\mathrm{Crit}(f) = \mathrm{Crit}(g) \subset g^{-1}(a)$;*
- *$g \equiv f$ and $\zeta \equiv \xi$ on some neighborhood of ∂W;*
- *the descending manifolds of ζ determine the chosen basis of $H_k(W, V_0)$.*

Proof. Let $\mathrm{Crit}(f) = \{p_1, \ldots, p_\ell\}$. To go between any two bases, it suffices to repeatedly perform the three elementary operations: permuting the basis $\{[\mathscr{D}(p)] : p \in \mathrm{Crit}(f)\}$ corresponds to permuting $\mathrm{Crit}(f)$; rescaling $[\mathscr{D}(p_i)]$ by -1 corresponds to switching the chosen orientation on $\mathscr{D}(p_i)$; it remains to show that $[\mathscr{D}(p_1)]$ can be replaced by $[\mathscr{D}(p_1)] + [\mathscr{D}(p_2)]$, without changing the other basis elements. We present a sketch of the proof given in [Mil65a].

Consider some neatly embedded $M \subset W$ with $\partial M \subset V_0$ and $\dim M = k$. By a variant of Lemma 3.3.2, the class $[M] \in H^k(W, V_0)$ is given by

$$[M] = \sum_{i=1}^{\ell} I\big(\mathscr{A}_\xi(p_i), M\big)[\mathscr{D}_\xi(p_i)]. \tag{8}$$

A proper proof utilizes the Thom isomorphism theorem, so we will take this formula as a given (the reader may show how the formula in Lemma 3.3.2 follows from this one). Then we wish to find g and ζ satisfying the following:

$$I\big(\mathscr{A}_\xi(p_i), \mathscr{D}_\zeta(p_j)\big) = \begin{cases} 1, & \text{if } i = j \text{ or } i-1 = j = 1 \\ 0, & \text{otherwise} \end{cases}$$

The idea is as follows: we raise $f(p_1)$ slightly, then find an isotopy that pulls $S^{k-1}_{\mathscr{D}}(p_1)$ across $S^{n-k-1}_{\mathscr{A}}(p_2)$, with a single moment of transverse intersection; if ξ is transformed into ζ by Proposition 3.2.1 with this isotopy, then $\mathscr{D}_\zeta(p_1)$ will intersect $\mathscr{A}_\xi(p_2)$ in just a single point (we will also ensure that $\mathscr{D}_\zeta(p_1)$ intersects $A_\xi(p_j)$ in δ_{1j} points for all $j \neq 2$, where δ_{ij} is the Kronecker delta). Then we will have prescribed not only the intersection number, but the total number of intersections! We need not worry about the sign of intersections, because we have already shown how to change signs in $\{[\mathscr{D}(p)] : p \in \mathrm{Crit}(f)\}$. We depict this process in Figure 15, with the local picture that we will use.

By Proposition 3.2.4, we can modify f in some small neighborhood of p_1, so that $f(p_1) > a$. Choose $b \in (a, f(p_1))$ and define the level set $V = f^{-1}(b)$.

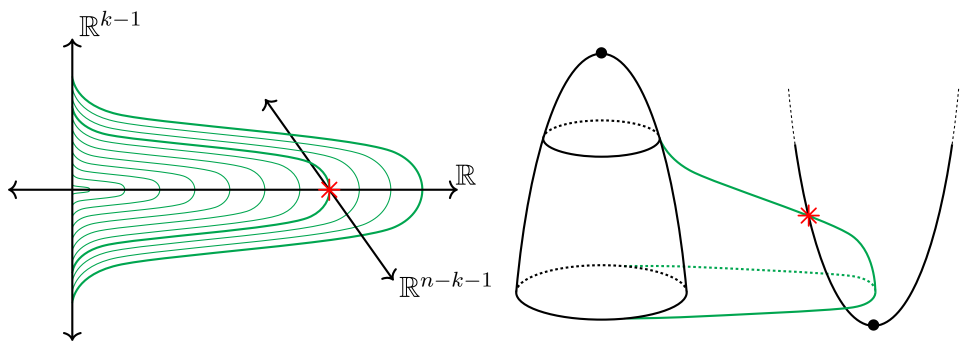


Figure 15: Creating an intersection for an elementary operation

Then we may consider arbitrary points in the upper and lower spheres:

$$x \in S^{k-1}_{\mathscr{D},\xi}(p_1) = \mathscr{D}_\xi(p_1) \cap V \qquad \text{and} \qquad y \in S^{n-k-1}_{\mathscr{A},\xi}(p_2) = \mathscr{A}_\xi(p_2) \cap V.$$

Since p_1 and p_2 originally lay on the same level, these spheres are disjoint. Note that W is formed from V by attaching a k-cell on one side and $\ell - 1$ $(n-k)$-cells on the other side. Since $k \geq 2$ and $n - k \geq 2$, we see that $\pi_0(W) = \pi_0(V)$. But we assumed that W is connected, so V is connected. We have $k - 1 \leq n - 3$ and $n - k - 1 \leq n - 3$, so $S^{k-1}_{\mathscr{D}}(p_1)$ and $S^{n-k-1}_{\mathscr{A}}(p_i)$ have codimension ≥ 2 in V ($i = 2, \ldots, \ell$). Thus, we can find a smoothly embedded curve $\varphi : I \to V$, which avoids these upper/lower spheres except at the endpoints $\varphi(0) = x$ and $\varphi(1) = y$. With the support of this curve φ, we can define[19] an embedding $\Phi : \mathbb{R} \times \mathbb{R}^{k-1} \times \mathbb{R}^{n-k-1} \to V$ such that:

- $\Phi(0,0,0) = x$ and $\Phi(1,0,0) = y$;
- $\Phi^{-1}\big(S^{k-1}_{\mathscr{D}}(p_1)\big) = 0 \times \mathbb{R}^{k-1} \times 0$;
- $\Phi^{-1}\big(S^{n-k-1}_{\mathscr{A}}(p_2)\big) = 1 \times 0 \times \mathbb{R}^{n-k-1}$;
- $\operatorname{Im}\Phi$ is otherwise disjoint from the upper and lower spheres.

With the use of bump functions, we also define a function $\alpha : \mathbb{R} \to [1, \infty)$ and a compactly supported isotopy H_t on $\mathbb{R} \times \mathbb{R}^{k-1} \times \mathbb{R}^{n-k-1}$, such that:

- $\alpha(x) = 1$ for $x \geq 2$, while $\alpha(x) > 2$ for $x \leq 1$;

[19] See [Mil65a] for the details of this construction, which requires Riemannian geometry but is not particular difficult. Essentially, this a tubular neighborhood, with a "twist" whose existence is guaranteed by the connectivity of Stiefel manifolds $V_k(\mathbb{R}^n)$ for $k < n$.

- $H_t(0, x, 0) = \Big(t\alpha\big(||x||^2\big), x, 0\Big)$ for all $x \in \mathbb{R}^{k-1}$.

The images $H_t(0 \times \mathbb{R}^{k-1} \times 0)$ for various $t \in [0,1]$ are shown in Figure 15. Since $H_t = \mathrm{Id}$ outside of a compact set, it extends via Φ to an isotopy of V. As described in Lemma 3.2.1, we can now use H_t to modify ξ on $f^{-1}(b, b+\epsilon)$ for an arbitrary $0 < \epsilon < a - b$. Call the resulting vector field ζ. It remains to check that $\mathscr{D}_\zeta(p_1)$ has the desired intersection numbers with each $\mathscr{A}_\xi(p_i)$.

Since $\mathscr{A}_\xi(p_1) \subset f^{-1}[a, \infty)$ and $\mathscr{D}_\zeta(p_1) \subset f^{-1}(-\infty, a]$, they only intersect at the point p_1. Since $\xi = \zeta$ outside of $f^{-1}[b, b+\epsilon]$, any intersection point in

$$\mathscr{D}_\zeta(p_1) \cap \mathscr{A}_\xi(p_i) \cap f^{-1}\big((-\infty, b] \cup [b+\epsilon, \infty)\big)$$

must flow back to $f^{-1}(b)$ or $f^{-1}(b+\epsilon)$ (for any $i = 2, \ldots, \ell$). Thus, we only need to investigate $f^{-1}[b, b+\epsilon] \cong V \times I$. Since $p_1, \ldots, p_\ell$ were originally all on the same level, we have $\mathscr{D}_\xi(p_1) \cap \mathscr{A}_\xi(p_i) = \emptyset$. But $\mathscr{D}_\xi(p_1)$ and $\mathscr{D}_\zeta(p_1)$ agree outside of $\mathrm{Im}\,\Phi \times I$, so we can focus just on the region parametrized by $\Phi \times I$. Here, our construction makes it clear that $D_\zeta(p_1)$ intersects $\mathscr{A}_\xi(p_i)$ exactly δ_{2i} times for each $i = 2, \ldots, \ell$, as desired (see Figure 15 again).

Lastly, we move $f(p_1)$ back to a. If there were an intersection between $\mathscr{D}_\zeta(p_1)$ and $\mathscr{A}_\zeta(p_i)$ for some $i = 2, \ldots, \ell$, then it would flow to an intersection point in $S_{\mathscr{D},\zeta}(p_1) \cap S_{\mathscr{A},\zeta}(p_i) \subset V$. But $\xi = \zeta$ below V, so $S_{\mathscr{A},\zeta}(p_i) = S_{\mathscr{A},\xi}(p_i)$. We know that $\mathscr{D}_\zeta(p_1)$ does not intersect $\mathscr{A}_\xi(p_i)$ at any point in V and thus

$$S_{\mathscr{D},\zeta}(p_1) \cap S_{\mathscr{A},\zeta}(p_i) = S_{\mathscr{D},\zeta}(p_1) \cap S_{\mathscr{A},\xi}(p_i) = \emptyset.$$

Therefore $\mathscr{D}_\zeta(p_1) \cap \mathscr{A}_\zeta(p_i) = \emptyset$ for all $i = 1, \ldots, \ell$, so we can move $f(p_1)$ back down to a by Proposition 3.2.4. We have now modified f twice (near p_1). The proof is completed by calling the resulting function g. $\square$

Corollary 4.3.2. *Suppose that $(W; V, V')$ is a triad of dimension $n \geq 6$, with W, V and V' simply-connected and $H_\bullet(W, V) = 0$. If this triad admits some Morse function without any critical points of index 0, 1, $n-1$ or n, then $\mu(W; V, V') = 0$.*

Proof. Let $f : W \to \mathbb{R}$ be Morse, with the minimal number of critical points among all Morse functions without critical points of index 0, 1, $n-1$ or n. By Corollary 3.2.5, we find a self-indexing Morse function g with the same critical points as f, having the same indices. Choose a pseudo-gradient ξ. Since $H_\bullet(W, V) = 0$, the chain complex $(C_\bullet, \partial)$ of Proposition 3.3.3 is exact. If $\mathrm{Crit}(g) \neq \emptyset$, then there must be a non-zero group in the exact sequence

$$0 \longrightarrow C_{n-2} \xrightarrow{\partial_{n-2}} C_{n-3} \longrightarrow \cdots \longrightarrow C_3 \xrightarrow{\partial_3} C_2 \longrightarrow 0$$

Therefore, there exists some $k \in \mathbb{N}$ with $\ker(\partial_k) \neq 0$. Let $z_1, \ldots, z_\ell \in C_k$ be a basis for $\ker \partial_k$. Then there is exists basis $a_1, \ldots, a_m, b_1, \ldots, b_\ell \in C_{k+1}$ with $\partial_{k+1}(b_i) = z_i$ for all $i = 1, \ldots, \ell$. By Proposition 4.3.1, we can modify g and ξ on $g^{-1}[k - \frac{1}{2}, k + \frac{1}{2}]$ so that g is still self-indexing and $z_1 = [\mathscr{D}(q)]$ for some $q \in \mathrm{Crit}_k(f)$. Similarly, we can modify g and ξ on $g^{-1}[k + \frac{1}{2}, k + \frac{3}{2}]$ so that $b_1 = [\mathscr{D}(p)]$ for some $p \in \mathrm{Crit}_{k+1}(f)$. Now, we recall the description of the matrix in Proposition 3.3.3. Because $\partial[\mathscr{D}(p)] = \partial b_1 = z_1 = [\mathscr{D}(q)]$, taking the entry corresponding to p and q yields $I\big(S^{n-k-1}_{\mathscr{A}}(q), S^k_{\mathscr{D}}(p)\big) = 1$. By the comment following Proposition 3.2.4, we see that we can modify g in small neighborhoods of p and q, so that $g(p) = k + \frac{3}{4}$ and $g(p) = k + \frac{1}{4}$, and ξ is still a pseudo-gradient. Now, we define some manifolds and triads:

$$\tilde{V} = g^{-1}(k + \tfrac{1}{8}), \qquad \tilde{V}' = g^{-1}(k + \tfrac{7}{8}), \qquad c_0 = \big(g^{-1}[-1, k + \tfrac{1}{8}]; V, \tilde{V}\big),$$

$$c_1 = \big(g^{-1}[k + \tfrac{1}{8}, k + \tfrac{7}{8}]; \tilde{V}, \tilde{V}'\big), \qquad c_2 = \big(g^{-1}[k + \tfrac{7}{8}, n + 1]; \tilde{V}', V'\big).$$

Then we have the composition $(W; V, V') = c_0 c_1 c_2$. We will write c_i to refer to both the cobordism and the total space of the triad. Notice the following:

(a) Since g restricts to a Morse function on c_2, we see that $c_1 c_2$ is formed by the successive attachment of cells of dimension $\geq k + 1 \geq 3$ to c_1. Hence, the inclusion $c_1 \subset c_1 c_2$ induces an isomorphism on π_1.

(b) Similarly, $-g$ restricts to a Morse function on $\bar{c}_0$, with critical points of index $\geq n - k \geq 3$ (since $k + 1 \leq n - 2$). As above, this shows that the inclusion $\bar{c}_2 \bar{c}_1 \subset \bar{c}_2 \bar{c}_1 \bar{c}_0$ induces an isomorphism on π_1.

(c) Now $\pi_1(c_1) = \pi_1(c_1 c_2) = \pi_1(\overline{c_2 c_1}) = \pi_1(\overline{c_2 c_1 c_0}) = \pi_1(c_0 c_1 c_2) = 1$.

(d) Because g only has critical points of index ≥ 2, the inclusion $V \subset c_0$ induces a surjection on π_1. But $\pi_1(V) = 1$, so we also have $\pi_1(c_0) = 1$.

(e) As in (b), the Morse function $-g|_{\bar{c}_0}$ shows that the inclusion $\tilde{V} \subset \bar{c}_0$ induces an isomorphism $\pi_1(\tilde{V}) = \pi_1(\bar{c}_0) = \pi_1(c_0) = 1$.

(f) Similarly to (d) and (e), we can see that $\pi_1(\tilde{V}') = 1$.

We have proven that the triad c_1 satisfies the condition of Theorem 4.1.6, so $\mu(c_1) = 0$. Thus we can modify g to eliminate the critical points on c_1 (see the proof of Corollary 4.1.4). But this contradicts the minimality of f, so our assumption that $\mathrm{Crit}(g) \neq \emptyset$ must have been wrong. $\square$

5 The H-Cobordism theorem

We are now at the point where we can prove some truly amazing results! We begin with the h-cobordism theorem, from which much else will follow.

Theorem 5.0.1 (Smale)**.** *Suppose that W, V_0 and V_1 are simply-connected. If $H_\bullet(W, V_0) = 0$ and $n \geq 6$ (where $n = \dim W$), then $\mu(W; V_0, V_1) = 0$.*

Proof. This follows immediately from Corollaries 4.2.2 and 4.3.2. □

An immediate corollary gives the h-cobordism theorem its name:

Corollary 5.0.2. *For $n \geq 6$, an n-dimensional, simply-connected cobordism is an h-cobordism if and only if it is a product cobordism. Therefore closed, simply-connected manifolds of dimension ≥ 5 are h-cobordant if and only if they are diffeomorphic.*

Proof. Let $(W; V_0, V_1)$ be a simply-connected h-cobordism with $\dim W \geq 6$. Then we have homotopy equivalences $V_0 \sim W \sim V_1$, so all three manifolds are simply connected. Moreover, $H_\bullet(W, V) = 0$ by the long exact sequence in homology. Therefore, the hypotheses of Theorem 5.0.1 are all satisfied. All of the other implications follow at once from Proposition 2.2.9. □

In fact, the seemingly weaker conditions of Theorem 5.0.1 imply that $(W; V_0, V_1)$ is an h-cobordism. Since W and V_0 are both simply-connected, the assumption that $H_\bullet(W, V_0) = 0$ implies that $V_0 \subset W$ is a homotopy equivalence, by the relative Hurewicz theorem. The proof of Corollary 4.2.2 used Lefschetz duality to show that $H_\bullet(W, V_0) = 0 \implies H_\bullet(W, V_1) = 0$ (whenever W is orientable), so the same line of reasoning shows that $V_1 \subset W$ is a homotopy equivalence. Note that this did not use $n \geq 6$.

As another, fairly simple consequence, we can classify the n-disk:

Corollary 5.0.3. *If M is a contractible, n-dimensional manifold with $n \geq 6$ and ∂M simply connected, then $M \cong \bar{D}^n$.*

Proof. Pick some interior point $p \in M$ and an interior chart $\varphi : \mathbb{R}^n \to M$ centered at p. Let $M' = M \setminus D^n$, where D^n is the unit disk in the chart φ. Then $(M'; S^{n-1}, \partial M)$ is a triad with M', S^{n-1} and ∂M simply connected (since a contraction of a loop in M can be deformed to avoid D^n). Since M is contractible, the inclusion $D^n \subset M$ is a homotopy equivalence and thus

$$H_\bullet(M', S^{n-1}) \cong H_\bullet(M, D^n) = 0,$$

by excision. Thus, we may apply Theorem 5.0.1 to the triad $(M'; S^{n-1}, \partial M)$. Therefore, the cobordism $(M; \emptyset, \partial M)$ is the composition of $(D^n; \emptyset, S^{n-1})$ with the product cobordism $(M'; S^{n-1}, \partial M)$, so we see that $M \cong D^n$. $\square$

We are now ready to prove the generalized Poincaré conjecture in higher dimensions! First, some definitions and a brief history of the conjecture:

- A homology (resp. homotopy) n-sphere is a closed n-manifold, which has the homology of S^n (resp. which is homotopy equivalent[20] to S^n). Depending on context, the manifold may be smooth or just topological.
- A twisted n-sphere is a closed n-manifold M with $\mu(M; \emptyset, \emptyset) = 2$.
- An exotic n-sphere is homeomorphic, but not diffeomorphic, to S^n.

Originally, Poincaré conjectured that any (topological) homology 3-sphere was indeed homeomorphic to the 3-sphere. He soon found a counterexample and conjectured instead that any (topological) 3-manifold M that is closed and simply-connected must be a 3-sphere. Since simply-connected manifolds are orientable, such an M must be a homotopy 3-sphere by Poincaré duality. The topological/smooth Poincaré conjecture in dimension n asks if every homotopy n-sphere is homeomorphic/diffeomorphic to S^n. This claim is:

- Trivially true in the topological and smooth cases for $n = 0, 1$ [Mil65b].
- Long known to be true in the topological and smooth cases for $n = 2$. The classification of surfaces was well-known by the time of Poincaré, although it was only in the piecewise-linear (PL) case, until Radó proved that topological, PL and smooth surfaces are equivalent [Moi52].
- Only recently proven true for $n = 3$. Moise proved that topological, PL and smooth 3-manifolds are equivalent in [Moi52], so the Poincaré conjecture is equivalent in these three categories. Perelman proved it in the early 2000's, completing a program of Hamilton and Thurston.
- Known to be true for $n = 4$ in the topological case. It is not known whether there exist any exotic 4-spheres (the smooth and PL cases are equivalent for $n = 4$). There are many candidates for exotic 4-spheres, but a proof of either exotic-ness or non-existence remains elusive.

[20] By the Hurewicz theorem, a simply-connected n-manifold M ($n > 1$) is a homology n-sphere if and only if $\pi_i(M) = \pi_i(S^n)$ for all $i = 1, \ldots, n$. If so, we choose a generator $\varphi : S^n \to M$ of $\pi_n(M) = \mathbb{Z}$. By the Whitehead theorem and relative Hurewicz theorem, φ is a homotopy equivalence. Thus for any $n > 1$, a homotopy n-sphere is the same thing as a simply-connected homology n-sphere.

- Generally false in the higher-dimensional smooth case. Milnor found exotic 7-spheres and in later joint work with Kervaire, described exotic n-spheres for $n \geq 5$ via stable homotopy groups of spheres. The full description requires a solution to the "Kervaire invariant problem," which (after much hard work) only remains open in dimension 126.
- Known to be true in the topological case (and the PL case) for $n \geq 5$. The first step in this direction was Smale's use of the h-cobordism theorem to prove that any smooth homotopy n-sphere (where $n \geq 5$) is homeomorphic to S^n. It is this spectacular result that we will prove.

The first step is a fairly elementary result of Morse theory:

Lemma 5.0.4 (Reeb)**.** *Every twisted n-sphere M is homeomorphic to S^n.*

Proof. Let $f : M \to \mathbb{R}$ be Morse, with $\mathrm{Crit}(f) = \{p, q\}$ and $f(q) > f(p)$. Extreme points of f are critical, so f is maximized at q and minimized at p. Therefore $\mathrm{ind}(q) = n$ and $\mathrm{ind}(p) = 0$. Now choose a pseudo-gradient ξ for f. Exercise 3.1.4(b) shows that we can define a diffeomorphism $\varphi : D^n \to \mathscr{D}(q)$, with $\varphi(0) = q$ and $(0, 1)v$ mapping to an integral curve of ξ for all $v \in S^{n-1}$. Exercise 3.1.4(a) shows that:

$$M = \mathscr{A}(p) \sqcup \mathscr{A}(q) = \mathscr{A}(p) \sqcup \{q\},$$

$$M = \mathscr{D}(p) \sqcup \mathscr{D}(q) = \{p\} \sqcup \mathscr{D}(q). \tag{9}$$

Thus $\varphi(D^n \setminus 0) \subset M \setminus \{q\} = \mathscr{A}(p)$. As $t \to \infty$, all integral curves in $\mathscr{A}(p)$ tend to p, so φ extends to a continuous $\bar{\varphi} : \bar{D}^n \to M$ with $\bar{\varphi}(S^{n-1}) = \{p\}$. Since $\bar{\varphi}$ is surjective by (9) and these two spaces are both compact Hausdorff, $\bar{\varphi}$ descends to a homeomorphism $\bar{D}^n/\bar{\varphi} \to M$. Note that $\bar{D}^n/\bar{\varphi} = \bar{D}^n/S^{n-1}$ is obviously homeomorphic to S^n, as desired. $\square$

The desired result for $n \geq 6$ now clearly follows from:

Theorem 5.0.5 (Smale)**.** *If $n \geq 6$, then every smooth homotopy n-sphere is a twisted n-sphere and thus is homeomorphic to S^n.*

Proof. Let M be a smooth homotopy n-sphere. A smooth function $M \to \mathbb{R}$ has at least two critical points, so it suffices to show that $\mu(M; \emptyset, \emptyset) \leq 2$. Choose a small embedded disk $\bar{D}^n \subset M$. By Lefschetz duality and excision:

$$H_i(M \setminus D^n) = H^{n-i}(M \setminus D^n, S^n) = H^{n-i}(M, \bar{D}^n) = \widetilde{H}^{n-i}(M).$$

Hence, $M \setminus D^n$ is simply-connected and has the same homology as a point, so $M \setminus D^n \cong \bar{D}^n$ by the Hurewicz theorem and Corollary 5.0.3. Therefore,

$$\mu(M;\emptyset,\emptyset) \leq \mu(\bar{D}^n;\emptyset,S^n) + \mu(M \setminus D^n;S^n,\emptyset) = 2$$

by Proposition 2.2.10, completing the proof. $\square$

For $n = 5$, the result requires another theorem of Milnor and Kervaire, which we will not prove: if M is a homotopy n-sphere with $n = 4, 5$ or 6, then M bounds a contractible $(n+1)$-manifold N [Mil65a]. If $n = 5$ or 6, then $N \cong \bar{D}^{n+1}$ by Corollary 5.0.3, so $M \cong S^n$. (this actually shows that the smooth Poincaré conjecture holds true in dimensions 5 and 6).

In conclusion, we will briefly discuss Theorem 5.0.1 in lower dimensions. For $n = 0, 1, 2$, it is trivial. For $n = 3$, it follows from Poincaré's conjecture.

Proof. Let $(W; V_0, V_1)$ denote a triad as in Theorem 5.0.1 with $\dim W = 3$. Then $V_0 \cong V_1 \cong S^2$, so we may form a new manifold M by gluing disks $\bar{D}_0^3$ and $\bar{D}_1^3$ to W along V_0 and V_1, respectively. Then M is simply-connected, so $M \cong S^3$ by the Poincaré conjecture. Any two orientation-preserving embeddings $\bar{D}^3 \subset D^3$ are ambient-isotopic (due to Cerf and Palais [Mil65a]), so removing any disk from the interior of $\bar{D}^3$ will yield a product cobordism. In particular, removing disks U_0 and U_1 from the interiors of $\bar{D}_0^3$ and $\bar{D}_1^3$ transforms M back into the cobordism $(W; V_0, V_1)$. But we can choose these disks so that $M \setminus (U_0 \cup U_1) \cong S^2 \times I$ is obvious. It follows that $(W; V_0, V_1)$ is a product cobordism. This proves Theorem 5.0.1 for $n = 3$. $\square$

For $n = 5$, Theorem 5.0.1 is false smoothly (Donaldson), while it is true topologically (Freedman). This topological h-cobordism theorem was used to prove the topological Poincaré conjecture in dimension 4. An accessible, wonderfully illustrated account of these results can be found in [Sco05].

For $n = 4$, the problem of Theorem 5.0.1 is similar to the case of $n = 3$, but complicated by the fact that the smooth Poincaré conjecture is still open in dimension 4. In fact, the following five conjectures are all equivalent:

(a) The smooth Poincaré conjecture in dimension 4.

(b) Theorem 5.0.1 for $n = 4$ (the h-cobordism theorem).

(c) Corollary 5.0.3 for $n = 4$ (classification of the 4-disk).

(d) Theorem 5.0.5 for $n = 4$ (every homotopy 4-sphere is twisted).

(e) Corollary 5.0.3 for $n = 5$ (classification of the 5-disk).

Proof. (a) $\Longrightarrow$ (b) is proved exactly as above (for $n = 3$), while (b) $\Longrightarrow$ (c) and (c) $\Longrightarrow$ (d) follow as in the proofs of Corollary 5.0.3 and Theorem 5.0.5, respectively. "A difficult theorem of Cerf" states that any twisted 4-sphere is actually diffeomorphic to S^4 [Mil65a], which clearly gives (d) $\Longrightarrow$ (a).

Lastly, we prove (a) $\Longleftrightarrow$ (e). As mentioned above, a theorem of Milnor and Kervaire shows that any homotopy 4-sphere M bounds a contractible manifold N. If (e) is true, then we have $N \cong \bar{D}^5$ and thus $M = \partial N \cong S^4$, proving (a). Conversely, we now suppose that W is a contractible 5-manifold with a simply-connected boundary. Just as in the proof of Corollary 5.0.3, we fix a small disk $\bar{D}^5$ in the interior of W and note that $(W \setminus D^5; S^4, \partial W)$ satisfies the conditions of Theorem 5.0.1 (for $n = 5$, where it is *not* true). But by the comment following Corollary 5.0.2, we see that $(W \setminus D^5; S^4, \partial W)$ is an h-cobordism, so $S^4 \sim W \setminus D^5 \sim \partial W$. Thus ∂W is a homotopy 4-sphere, so if (a) is true, then $\partial W \cong S^4$. We form a closed manifold $M = W \sqcup_{S^4} \bar{D}^5$. Because W is contractible, we get homotopy equivalences $M \sim M/W \sim S^5$. Thus M is a homotopy 5-sphere, so $M \cong S^5$. Any two orientation-preserving embeddings $\bar{D}^5 \subset S^5$ are ambient-isotopic (due to Cerf and Palais [Mil65a]), so we may assume that $\bar{D}^5 \subset M \cong S^5$ is just identified with a hemisphere. Then $W = M \setminus D^5 \cong S^5 \setminus D^5 = \bar{D}^5$ (the other hemisphere), proving (e). $\square$

References

[Bot88] Raoul Bott. Morse theory indomitable. *Pubs. Maths. de l'IHÉS*, 68:99–114, 1988.

[Hir76] Morris Hirsch. *Differential Topology*, volume 33 of *Graduate Texts in Mathematics*. Spring-Verlag, 1st edition, 1976.

[HJ18] J. Hausmann and B. Jahren. A simplification problem in manifold theory. *Enseign. Math.*, 64:207–248, 2018.

[Hut02] Michael Hutchings. Lecture notes on morse homology (with an eye towards floer theory and pseudoholomorphic curves). Available at https://math.berkeley.edu/∼hutching/, 2002.

[Kos93] Antoni Kosinski. *Differential Manifolds*. Academic Press, 1993. Republished by Dover Books in 2007 with a new appendix.

[KS77] R. Kirby and L. Siebenmannn. *Foundational Essays on Topological Manifolds, Smoothings, and Triangulations*. Number 88 in Annals of Mathematics Studies. Princeton University Press, 1977.

[Lee12] John Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, 2nd edition, 2012.

[Mil59] John Milnor. Differentiable manifolds which are homotopy spheres, 1959. Unpublished manuscript (mimeographed). Can be found in *Collected Papers of John Milnor, Vol. III: Differential Topology*.

[Mil61] John Milnor. A procedure for killing homotopy groups of differentiable manifolds. *Proceedings of Symposia in Pure Math.*, 1961.

[Mil63] John Milnor. *Morse Theory*. Number 51 in Annals of Mathematics Studies. Princeton University Press, 1963.

[Mil65a] John Milnor. *Lectures on the H-Cobordism Theorem*. Princeton University Press, 1965. Printed through Princeton Legacy Library.

[Mil65b] John Milnor. *Topology from the Differentiable Viewpoint*. Princeton Landmarks in Mathematics. Princeton University Press, 1965.

[Moi52] Edwin Moise. The triangulation theorem and hauptvermutung. *Annals of Mathematics (2)*, 56(1):96–114, 1952.

[Sco05] Alexandru Scorpan. *The Wild World of 4-Manifolds*. American Mathematical Society, 2005. Reprinted in 2011.

[Sto68] Robert Stong. *Notes on Cobordism Theory*. Princeton University Press, 1968. Printed through Princeton Legacy Library.

[VK91] O. Viro and V. Kharlamov. On the work of Vladimir Abramovich Rokhlin in topology. *Leningrad Mathematical Journal*, 2, 1991.