Summary of Morse Theory

One central theme in geometric topology is the classification of selected classes of smooth manifolds up to diffeomorphism. Complete information on this problem is known for compact **1 –** dimensional and **2 –** dimensional smooth manifolds, and an extremely good understanding of the **3 –** dimensional case now exists after more than a century of work. A closely related theme is to describe certain families of smooth manifolds in terms of relatively simple decompositions into smaller pieces. The following quote from **http://en.wikipedia.org/wiki/Morse_theory** states things very briefly but clearly:

> In differential topology, the techniques of Morse theory give a very direct way of analyzing the topology of a manifold by studying differentiable functions on that manifold. According to the basic insights of Marston Morse, a differentiable function on a manifold will, in a ***typical*** case, reflect the topology quite directly. Morse theory allows one to find CW [cell complex] structures and handle decompositions on manifolds and to obtain substantial information about their homology.

Morse's approach to studying the structure of manifolds was a crucial idea behind the following breakthrough result which S. Smale obtained about 50 years ago: If a compact smooth manifold M^n (without boundary) is homotopy equivalent to the sphere S^n, where $n \geq 5$, then M^n is homeomorphic to S^n. — Earlier results of J. Milnor constructed a smooth **7 –** manifold which is homeomorphic but not diffeomorphic to S^n, so one cannot strengthen the conclusion to say that M^n is diffeomorphic to S^n.

We shall use Morse's approach to retrieve some low – dimensional classification and decomposition results which were obtained before his theory was developed. The two classic references are books by Milnor:

> **J. Milnor.** ***Morse theory*** (Based on lecture notes by M. Spivak and R. Wells). Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, 1963.
>
> **J. Milnor.** Lectures on the ***h*** **–** cobordism theorem (Notes by L. Siebenmann and J. Sondow, Princeton Mathematical Notes No. 1). Princeton University Press, Princeton, 1965.

The latter is out of print but available online:

http://www.maths.ed.ac.uk/~aar/surgery/hcobord.pdf

There is also a brief but informative set of slides (with accompanying) notes that gives a nice overview of Morse Theory**:**

http://maths.dept.shef.ac.uk/magic/course_files/52/notes.pdf

http://maths.dept.shef.ac.uk/magic/course_files/52/lecture07_handout.pdf

http://maths.dept.shef.ac.uk/magic/course_files/52/lecture08_handout.pdf

http://maths.dept.shef.ac.uk/magic/course_files/52/lecture09_handout.pdf

http://maths.dept.shef.ac.uk/magic/course_files/52/lecture10_handout.pdf

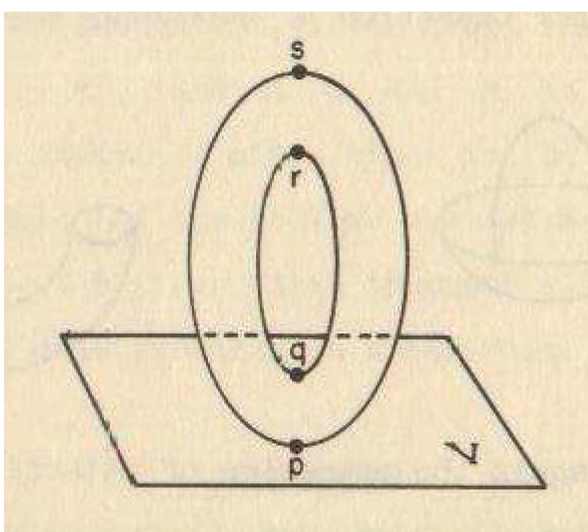
http://maths.dept.shef.ac.uk/magic/course_files/52/lecture11_handout.pdf

http://maths.dept.shef.ac.uk/magic/course_files/52/lecture12_handout.pdf

Underlying concepts of Morse Theory

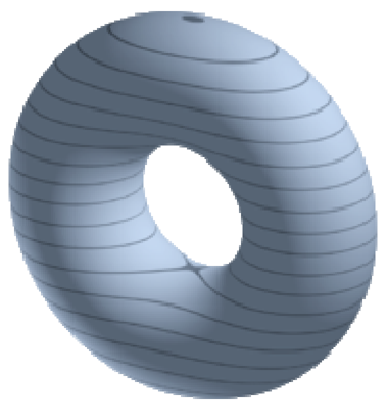
The usual approach is to start with the **2 –** dimensional torus presented as a surface ***F*** of revolution of the circle $x^2 + (y-2)^2 = 1$ about the x – axis**.** The drawings in Milnor's book on Morse Theory are excellent and hard to improve upon**,** but we shall also use some similar illustrations in color from the ***Wikipedia*** article on Morse Theory**;** unless stated otherwise**,** the illustrations below come from one of these sources**.**

Let h be the height function on F defined by taking the z – coordinate**.** Then h has exactly three ***critical points*** (where the derivative vanishes)**;** specifically, an absolute minimum value of -3 at the point $(0, 0, -3)$**,** a relative maximum value of 3 at the point $(0, 0, 3)$**,** and two saddle points at $(0, 0, -1)$ and $(0, 0, 1)$**.**



We note that the critical points of this function are ***non – degenerate***; namely, the matrices of second partial derivatives at these points (in some coordinate system) are invertible symmetric matrices.

The basic idea of Morse Theory is to see how the sublevel sets F^a (= all x such that $h(x) \leq a$) change as a increases from the minimum to the maximum value. In the picture below, the level curve L^a (= all x such that $h(x) = a$) are drawn for several values of a, and of course the sets F^a consist of all on or below the curves L^a. Observe that if a is not a critical value of h then L^a is a bicollared smooth submanifold; in fact, this is true for every smooth function on a manifold without boundary.



A close inspection of this picture suggests that if the function h has no critical values between a and b, the set $V[a, b]$ of all points x such that $a \leq h(x) \leq b$ is diffeomorphic to a product of either L^a or L^a with the closed interval $[a, b]$. In fact, the following generalization of this observation is a cornerstone of Morse Theory:

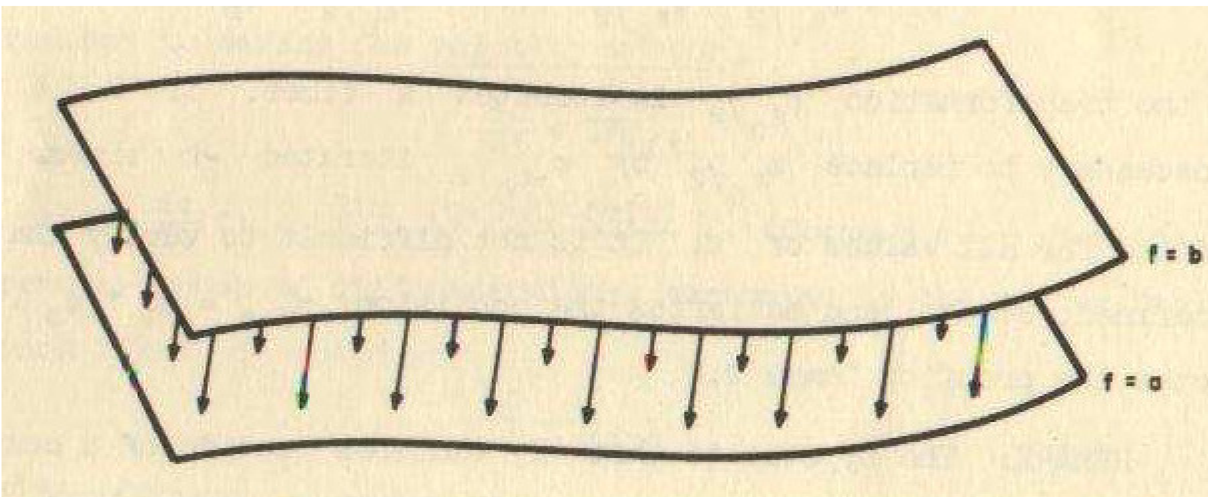
Product Theorem. Let M^n be a compact smooth manifold with boundary ∂M such that ∂M is the disjoint union of two submanifolds $\partial_0 M$ and $\partial_1 M$, and let f be a smooth real valued function on M^n which takes values in the closed interval $[0, 1]$, mapping $\partial_0 M$, $\partial_1 M$ and the interior of M^n to $\{0\}$, $\{1\}$, and $(0, 1)$ respectively. Suppose we are given real numbers $0 < a < b < 1$ such that f has no critical values on the interval $[a, b]$. Then the following hold:

(1) The level sets L^a or L^b, given by points where the function's value is a or b, are bicollared smooth submanifolds, and for $t = a, b$ the sublevel sets F^t are smooth submanifolds whose boundaries are the disjoint unions of $\partial_0 M$ and L^t.

(2) The set $V[a, b]$ of all x such that $a \leq h(x) \leq b$ is a submanifold which is diffeomorphic to both $L^a \times [a, b]$ and $L^b \times [a, b]$.

It follows immediately that ***the manifolds*** F^a ***and*** F^b ***are diffeomorphic.***

The idea of the proof is to take a smooth Riemannian metric on M which looks like a product metric near manifolds L^a and L^b, and to construct the gradient of h with respect to this metric. Then the diffeomorphism is defined using the integral flow of this vector field. In the drawing below, the arrows point in the **opposite** direction of the gradient in our discussion.



We shall now apply this to the function which was defined on the torus. The Product Theorem verifies that the sublevel sets F^t are all diffeomorphic in the following ranges:

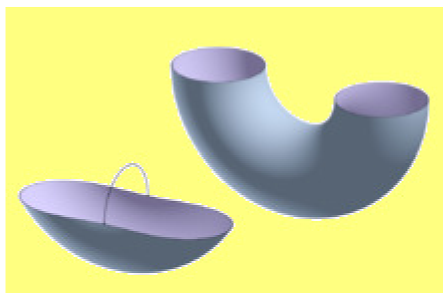
$$-3 < t < -1, \qquad -1 < t < 1, \qquad 1 < t < 3$$

In view of this, it is natural to ask the following question:

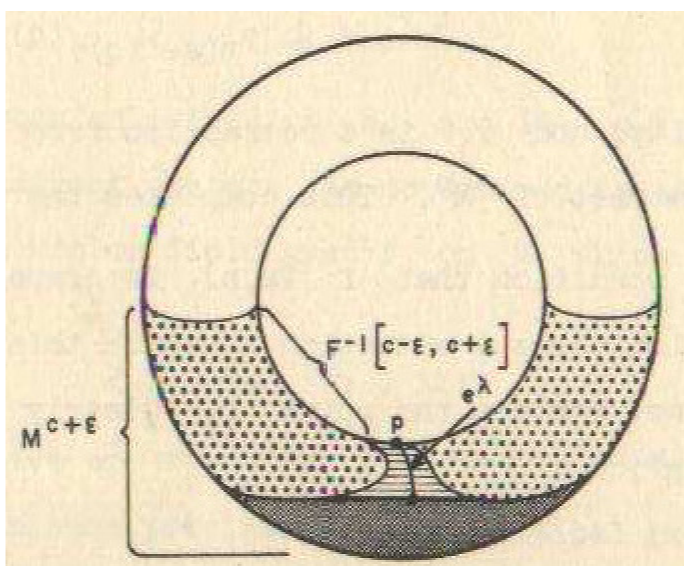
> Suppose that c is a critical value of h and there is an interval $[a, b]$ containing c in its interior such that no other value in the interval is a critical value. How are the manifolds F^a and F^b related? In the example, there is only one critical point between L^a and L^b, and hence there is only one critical value in this range.

For our basic example, a complete representative list of sublevel sets is given by F^{-4}, F^{-2}, F^0, F^2 and F^4. Note that the first of these is empty and the second is the entire torus. Passage from F^{-4} to F^{-2} amounts to taking a **2 –** disk at the bottom of the torus.

The next passage is from F^{-2} to F^{0}. From a homotopy – theoretic viewpoint we are adding a **1 –** cell along the boundary of the disk (see the color picture below). Topologically, we are actually adjoining a thickened **1 –** cell called a ***1 – handle*** or a ***handle of index* 1.** Roughly speaking, we first thicken the copy of S^{0} in the boundary of F^{-2} to a smooth embedding of $S^{0} \times D^{2}$, and then we attach a copy of $D^{1} \times D^{2}$ along $S^{0} \times D^{2}$ (see the second picture below; the original disk is the dark region and the handle is indicated by horizontal stripes). Finally, we add a collar along the boundary to get the entire sublevel set (the dotted region); note that if we have M such that ∂M is a disjoint union of pieces $\partial_0 M$ and $\partial_1 M$, and we glue a collar $\partial_1 M \times [0, 1]$ onto M along $\partial_1 M$, then the manifold we obtain is diffeomorphic to M, so the final step does not change the diffeomorphism type.

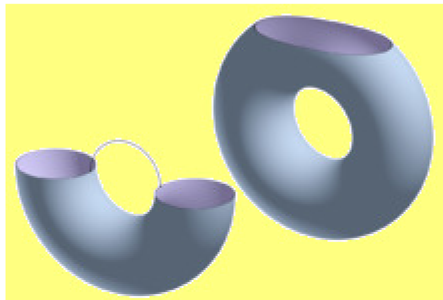


A homotopy model for F^{0}



The thickening process – adding a handle and then a collar

Similar considerations hold for passage from F^2 to F^4, but now the $1-$cell's end points lie on different components of the level set L^0 (see the picture below). Note that the level set L^2 is a circle.



Finally, we can now pass from F^2 to F^4 by gluing on a $2-$disk along L^2.

Generalizing the basic example

In Morse Theory it is customary to formulate everything in terms of **cobordisms**, which are merely manifolds M with boundaries ∂M that are split into two pieces denoted by $\partial_0 M$ and $\partial_1 M$; such objects contain all manifolds with our without boundaries, for we can let either one or both boundary pieces by empty. We shall be interested in smooth functions f from M to $[0, 1]$ such that (as before) f maps $\partial_0 M$, $\partial_1 M$ and the interior of M^n to $\{0\}$, $\{1\}$, and $(0, 1)$ respectively.

If we are given a smooth manifold M without boundary, we shall often consider Morse functions on the manifold N formed from M by choosing two disjointly embedded closed $n-$disks A_0 and A_1 (where n is the dimension of M as usual) and removing their interiors; in this case we arbitrarily set $\partial A_0 = \partial_0 N$ and $\partial A_1 = \partial_1 N$. Note that if M is oriented and we choose orientation – preserving embeddings g_i from D^n onto A_i, then the induced orientations on ∂A_1 is the usual one and the induced orientation on ∂A_0 will be its opposite (by definition, if P is an oriented manifold with boundary, then the orientation on ∂P is obtained by first pulling back the orientation form to an orientation ω on $\partial P \times (0, 1)$ by a collar neighborhood, then taking the interior product or contraction $-(d/dt) \lrcorner\, \omega$, and finally restricting the latter to a slice $\partial P \times \{t\}$; as in multivariable calculus, we want to use the ***outward*** pointing normal to the boundary).

We need to assume some regularity property for the behavior of the function f near the boundary components. One simple way of doing this is to assume that there are smooth collar neighborhoods

$$c_0 : \partial_0 M \times [0, \varepsilon) \to M \quad \text{and} \quad c_1 : \partial_1 M \times (1 - \varepsilon, 1] \to M$$

such that $f c_i (y, t) = t$ for all i, y, t.

In the example of a function on the torus, the critical points were ***nondegenerate*** in the sense that the Hessian matrices of second partial derivatives were invertible. The natural generalization of this property is the defining condition for a Morse function. More precisely, we need the following local result in order to proceed:

Lemma. Suppose that U is an open subset of $\mathbb{R}^n$ and we are given a smooth function $f : U \to \mathbb{R}$ and a point p in U such that p is a critical point of f and the Hessian matrix $H(f; p)$ of second partial derivatives at p is invertible. Let h be a diffeomorphism from an open subset V to U, and choose q such that $h(q) = p$. Then $f h$ has a critical point at q and $\mathrm{H}(f h; q) = {}^{\mathrm{T}}A\, \mathrm{H}(f; p) A$, where $A = Dh(q)$ and ${}^{\mathrm{T}}A$ denotes the transpose of A.

This is discussed in the top half of page 144 of Hirsch, and it has two important consequences. First of all, $H(f; p)$ is invertible if and only if $\mathrm{H}(f h; q)$ is (since A is invertible). Second, if we define the ***Morse index*** of a symmetric matrix to be the number of negative eigenvalues counted with multiplicities (recall that the eigenvalues of a real symmetric matrices are always real), then the Morse indices of the two Hessians are equal by standard results on symmetric real bilinear forms.

Definition. Let f be a smooth real valued function on a smooth manifold W without boundary, and let p be a point of W. Then f is said to be a nondegenerate critical point if there is a coordinate chart $h : U \to W$ at p, with $h(q) = p$ for some (unique) q, such that $f h$ is a critical point of q and the Hessian of $f h$ at q is invertible. The (Morse) index of the critical point p is given by the number of negative eigenvalues for this Hessian and is often denoted by λ_p or simply λ.

The preceding discussion implies that the definitions of nondegenerate critical points and their indices do not depend upon the choice of smooth coordinate charts.

We can now define a **Morse function** on a cobordism M to be a smooth function f which satisfies the previously stated conditions and also has only nondegenerate critical points. Note that if M is not a cylinder $\partial_0 M \times [0, 1]$, then f must have at least one critical point; in particular, if N is a compact unbounded manifold and M is formed by deleting the interior of two closed disks, then there must be at least one critical point on M unless N is homeomorphic to a sphere.

A fundamental but fairly straightforward result (the ***Morse Lemma***; see pages 145 147 in Hirsch) states that we can always find a smooth coordinate chart φ at p of the form $\varphi: V_0 \times V_1 \to W$, where V_0 is an open neighborhood of $\mathbf{0}$ in $\mathbb{R}^{\lambda}$ and V_1 is an open neighborhood of $\mathbf{0}$ in $\mathbb{R}^{n-\lambda}$, such that $\varphi(0,0) = p$ and $f\varphi(x,y) = |y|^2 - |x|^2 + C$, where C is the critical value and $|v|$ denotes the length of the vector v.

The Morse Lemma has many far – reaching consequences for Morse functions, beginning with the following:

Corollary. The critical points of a Morse function are isolated.

Obviously, the usefulness of Morse functions depends upon our ability to construct them, but fortunately this is no problem (see pages 147 – 148 of Hirsch to see how this falls out of basic general position theorems or pages 16 – 18 of Milnor, ***Lectures on the h– cobordism Theorem***, for a more self – contained proof that one can construct Morse function approximations to arbitrary smooth functions on a cobordism which send the pieces to the appropriate subsets of the unit interval). Furthermore, we can do this so that no two critical points map to the same critical value (Milnor works this out explicitly).

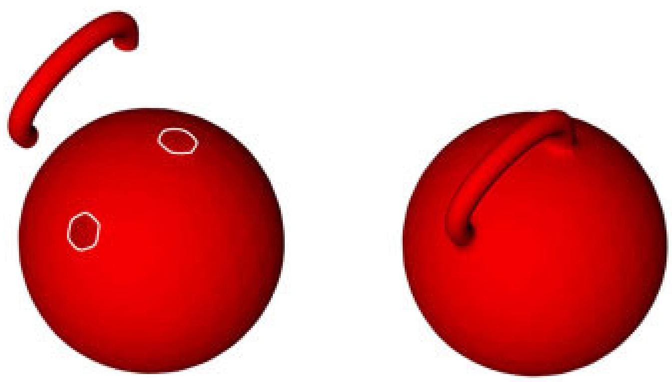
This puts us into a situation very much like the one we had on the torus. There is a finite number of critical points, and we can label them in sequence as $p_1, \dots, p_m$ such that $0 < f(p_1) < \dots < f(p_m) < 1$. We know that the sublevel sets F^t are cylinders for all positive $t < f(p_1)$, and once again we want to know what happens when we pass from a sublevel set $F^{p_k - \varepsilon}$ to $F^{p_k + \varepsilon}$, where ε is positive and small enough that $p_{k-1} < p_k - \varepsilon < p_k < p_k + \varepsilon < p_{k+1}$. Fortunately, we can do this fairly explicitly.

Theorem. Suppose that we are given a Morse function h on an n – dimensional cobordism M with exactly one critical point, and the index of the critical point is equal to λ. Then the following hold:

(1) There is a smooth embedding $g: S^{\lambda-1} \times D^{n-\lambda} \to \partial_0 M$ such that M is homotopy equivalent to $\partial_0 M$ with a λ – cell attached along the restriction of g to $S^{\lambda-1} \times \{0\}$.

(2) If i_1 denotes the standard inclusion of X in $X \times [0, 1]$ as the slice $X \times \{i\}$, then M is homeomorphic to the union of $\partial_0 M \times [0, 1]$ and $D^{\lambda} \times D^{n-\lambda}$ with the image of $i_1 g$ in the former identified with $S^{\lambda-1} \times D^{n-\lambda}$ in the latter (there is an illustration below). Furthermore, there is a canonical smooth structure on M which depends only on the embedding h.

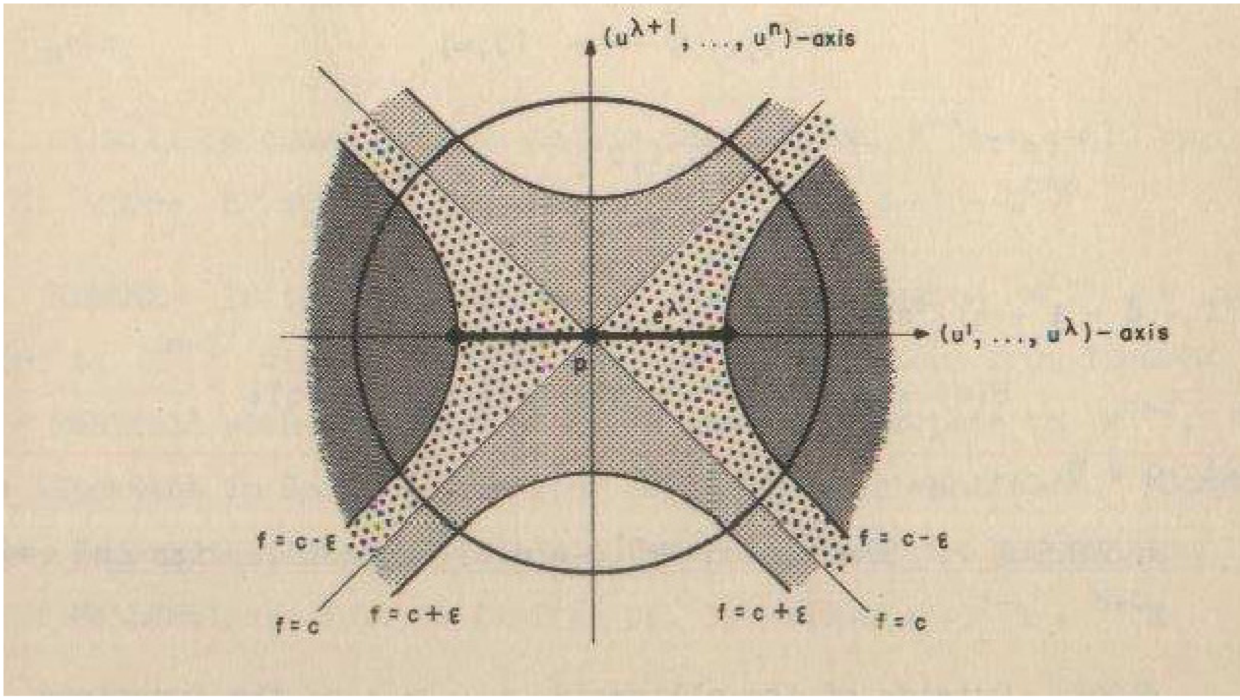
Special cases of this principle are implicit in the torus example (including pictures) which was discussed earlier. The picture below illustrates what happens for $\partial_0 M = S^2$ and $\lambda = 1$, and it suggests the reason why the construction is often called ***handle attachment***.



(Source: http://www.uff.br/cdme/pdp/pdp-html/handle-01.jpg)

The relationship between $\partial_0 M$ and $\partial_1 M$ is that the second is obtained from the first by the following process, which is called a **surgery** (or spherical modification) of type $(\lambda, n - \lambda)$: One removes the interior of $h(S^{\lambda-1} \times D^{n-\lambda})$ from M and glues in a copy of $D^{\lambda} \times S^{n-\lambda-1}$ using the obvious identification of **hbdy** with $\partial(D^{\lambda} \times S^{n-\lambda-1}) = S^{\lambda-1} \times S^{n-\lambda-1}$. In the example depicted above, the surgery relates S^2 with a manifold which is homeomorphic (in fact, diffeomorphic) to the torus T^2.

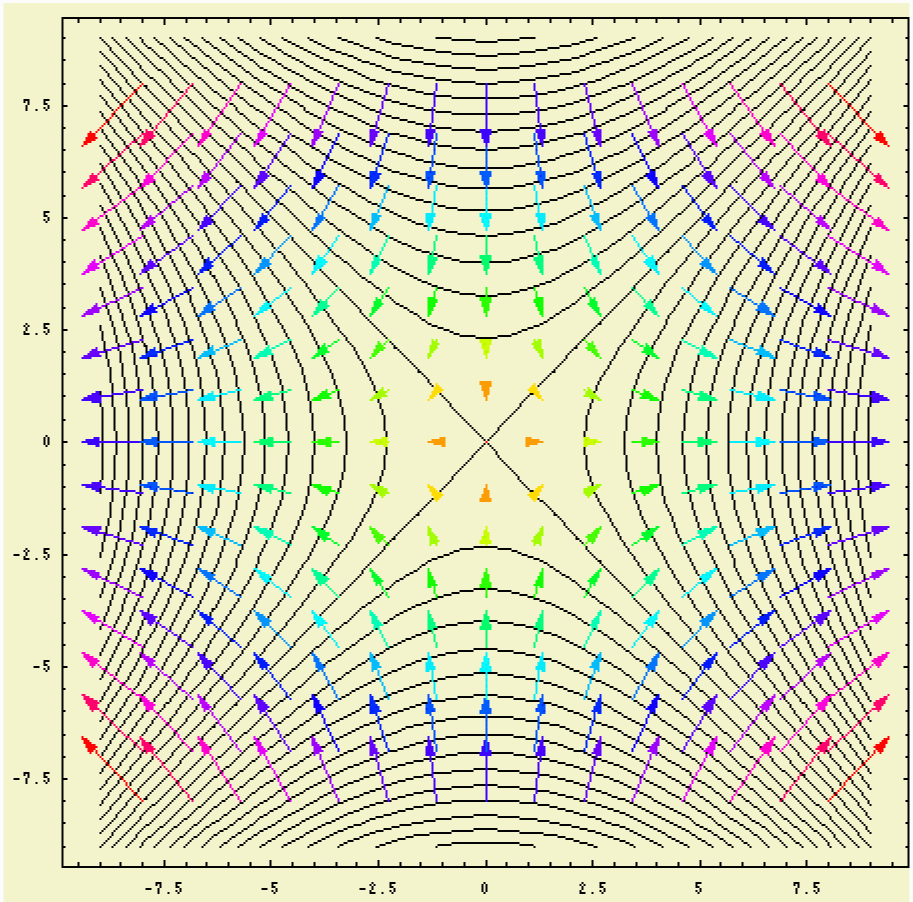
Given a Morse function f on an n – dimensional cobordism M with exactly one critical point, let c denote the associated critical value. As in the case of no critical points, we shall use a suitably defined gradient – like vector field on the cobordism to analyze the topological structure of M, and not surprisingly we shall also use the Morse Lemma which describes the behavior of f near the isolated fixed point. The latter yields the following picture of the situation near the single critical point. Note that if we remove the points inside the circle, then the closed region bounded by $L^{c-\varepsilon}$ and $L^{c+\varepsilon}$ is a product of the relevant part of either with the closed interval $[c - \varepsilon, c + \varepsilon]$.



In order to proceed, we need to define a gradient – like vector field X for the Morse function h explicitly. Such a vector field should have the following properties:

1. The directional derivative Xh is positive away from the critical points.
2. On collar neighborhoods of the boundary pieces $c_0 : \partial_0 M \times [0, \varepsilon) \to M$ and $c_1 : \partial_1 M \times (1 - \varepsilon, 1] \to M$ the vector field X corresponds to the standard vector field $\partial/\partial t$.
3. Near a critical point p, for a suitably chosen smooth chart φ at p so that the conclusion of the Morse Lemma is true, the vector field X corresponds to the gradient of $h\varphi(x, y) = |y|^2 - |x|^2 + C$.

It will be extremely important to understand how such a vector field behaves near critical points, so we shall elaborate on the consequences of the third condition using the following picture (corresponding to the case $\lambda = 1$ and $n = 2$).



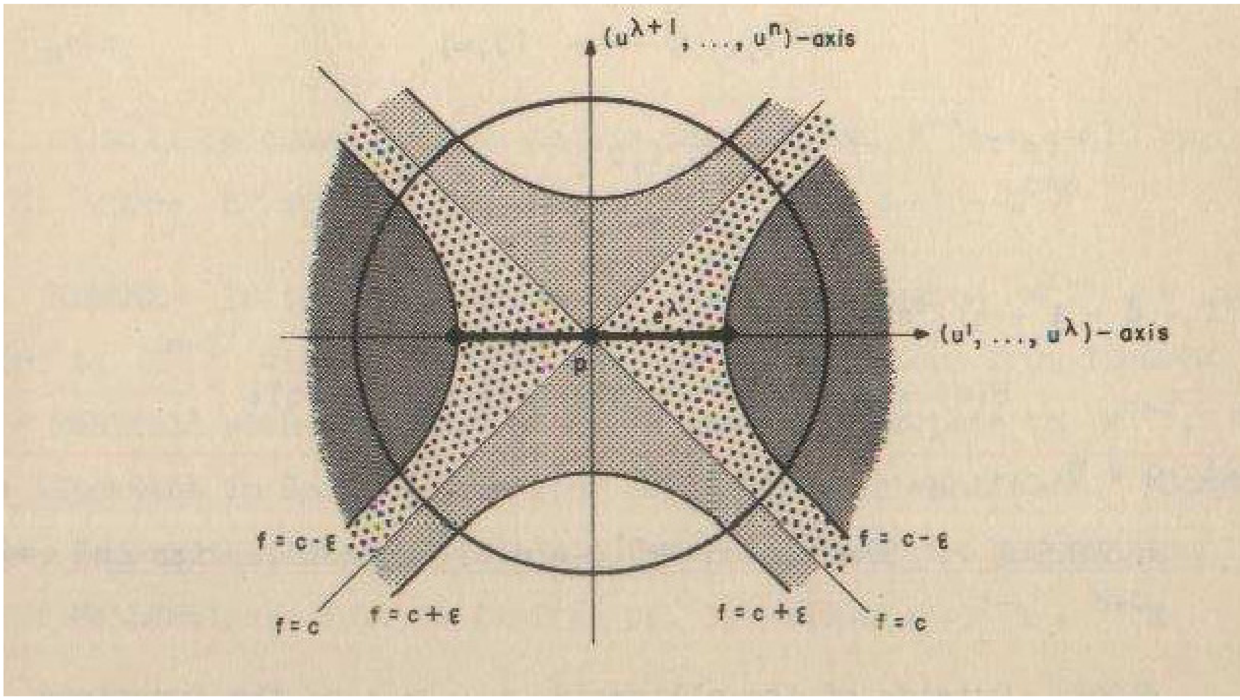
(Source: http://www.math.ou.edu/~amiller/math/images/vf3.gif)

In this drawing the critical point is the origin, the critical value is $\mathbf{0}$, and the level curves correspond to the hyperbolas drawn in black. The colored arrows describe the gradient vector field, and the two coordinate axes represent $\boldsymbol{\lambda}$ – and $(\boldsymbol{n} - \boldsymbol{\lambda})$ – dimensional vector spaces. For points not on these axes the integral curves of the vector field are given by branches of the hyperbolas $|\boldsymbol{x}|\cdot|\boldsymbol{y}| = \boldsymbol{constant}$, which are orthogonal to the level sets for $|\boldsymbol{y}|^2 - |\boldsymbol{x}|^2$. For nonzero points on the axes, the integral curves either move radially towards the origin or radially away from it. Of course, at the critical point the vector field vanishes and the integral curve is constant.

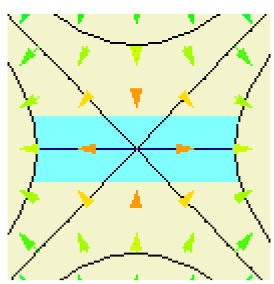
To construct a gradient – like vector field, it will suffice to take a Riemannian metric which is built locally out of pieces and has the desired properties (these clearly exist) and to piece them together using a smooth partition of unity; one can then take $\boldsymbol{X}$ to be the gradient of $\boldsymbol{h}$ with respect to this metric.

<u>Sketch of proof of the theorem.</u> We shall derive the first conclusion from the second, so we begin with that part. Our discussion will be based on the argument given on pages 33 – 34 of ***Lectures on the h– cobordism Theorem*****,** and we shall only give the main points here and explain why the methods yield the stronger conclusion which we have stated**.**

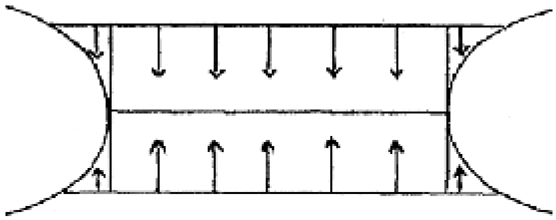
We need to examine the picture at the top of page 10 more closely.

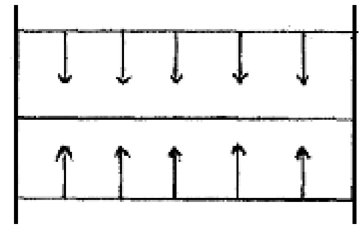


As before**,** in this picture the function is given by $|y|^2 - |x|^2 + C$**.** By definition, the disk $\mathbf{e}^\lambda$ is the disk of radius ε on the horizontal axis**.** Let K be the set of (x, y) such that $f(\mathrm{x}, \mathrm{y})$ lies in $[C - \varepsilon, C + \varepsilon]$ and $|y| < \varepsilon/10$**.** In the drawing below (which is not to scale), the set K is shaded in turquoise.

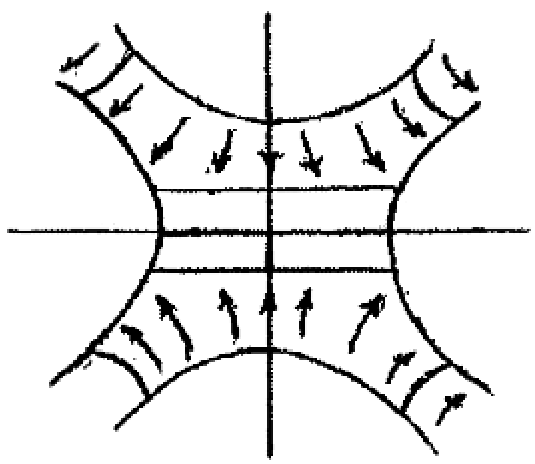


We claim that K is homeomorphic to $D^\lambda \times D^{n-\lambda}$ such that $K \cap L^{c-\varepsilon}$ corresponds to $S^{\lambda-1} \times D^{n-\lambda}$**;** we can do this using homeomorphism which flattens out the vertical part of the boundary of K by radial shrinking in the horizontal direction, as indicated in the drawings below**:**





It follows that $\boldsymbol{F}^{c\,-\,\varepsilon} \cup \boldsymbol{K}$ has the structure specified in part **(2)** of the theorem. We now need to show that $\boldsymbol{F}^{c\,-\,\varepsilon} \cup \boldsymbol{K}$ is homeomorphic to the original cobordism, and this is where we need to invoke the argument from ***Lectures on the h–cobordism Theorem*.** The idea is to push everything from $\boldsymbol{F}^{c\,+\,\varepsilon}$ into $\boldsymbol{F}^{c\,-\,\varepsilon} \cup \boldsymbol{K}$ using the integral curves of the gradient – like vector field, as suggested by the drawing below. Since the Morse function has no critical points besides the given one, so that $\boldsymbol{M}$ and $\boldsymbol{F}^{c\,+\,\varepsilon}$ are diffeomorphic, a modification of this construction yields a deformation retraction from $\boldsymbol{M}$ back to $\boldsymbol{F}^{c\,-\,\varepsilon} \cup \boldsymbol{K}$.



Since $\boldsymbol{M}$ and $\boldsymbol{F}^{c\,+\,\varepsilon}$ are diffeomorphic, the assertion on the topological structure of $\boldsymbol{M}$ reduces to showing that $\boldsymbol{F}^{c\,+\,\varepsilon}$ is homeomorphic to $\boldsymbol{F}^{c\,-\,\varepsilon} \cup \boldsymbol{K}$ with a collar attached to the upper component of the boundary of the latter. In other words, we need to show that the closure of the set

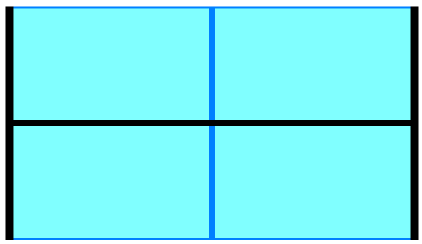
$$\boldsymbol{F}^{c\,+\,\varepsilon} \;-\; (\boldsymbol{F}^{c\,-\,\varepsilon} \cup \boldsymbol{K})$$

is homeomorphic to the product of $\partial_1 \boldsymbol{M}$ with a closed interval. In fact, one can also use the gradient vector field to define this homeomorphism; the definition requires two cases, depending upon whether or not the integral curve through a point meets the set $|\boldsymbol{y}| = \boldsymbol{\varepsilon/10}$ before or exactly when it meets $\boldsymbol{L}^{c\,-\,\varepsilon}$, but direct computation shows that this is equivalent to the condition

$$|x| \cdot |y| \;\leq\; \varepsilon^2 \, \frac{\sqrt{101}}{100}.$$

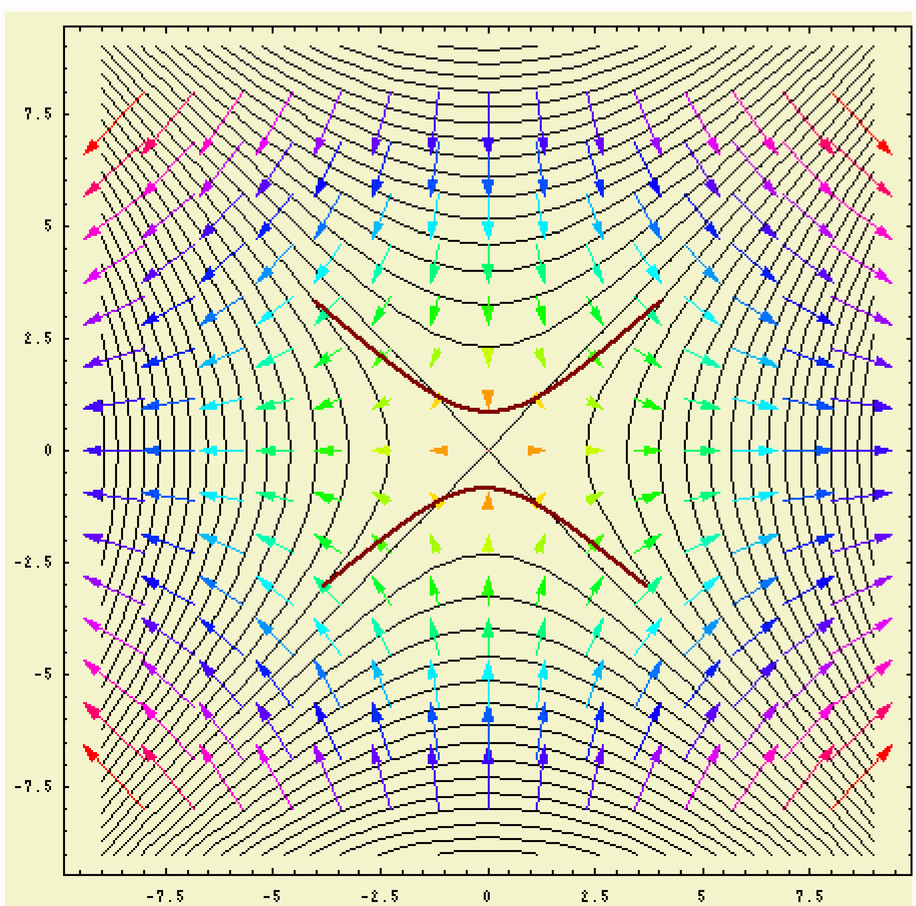
We now have to show the first part of the result. However, given the second part and the local picture, this reduces immediately to the following statement:

CLAIM. $S^{\lambda-1} \times \{0\}$ is a strong deformation retract of $D^{\lambda} \times D^{n-\lambda}$**.**



The proof is fairly elementary and left as an exercise (one can use the picture as a basis for an argument)**.**

The following picture suggests how one might perform a smooth version of the handle attachment described above**,** with suitably rounded edges**:**



As indicated in Section 3 of ***Lectures on the h– cobordism Theorem*****,** there is a converse of sorts to the preceding result. Given an arbitrary topological embedding of $S^{\lambda-1} \times D^{n-\lambda}$ in a manifold without boundary N^{n-1}, one can carry out the topological handle attachment construction described above to obtain a topological cobordism whose **0 –** component is N. By Theorem 3.12 on page 30 of that reference**,** if the embedding is smooth, then one has a canonical way of making this cobordism into a smooth manifold and defining a Morse function on it which has exactly one critical point**,** whose index is equal to λ.

The preceding discussion has important implications for the homotopy structure of the cobordism M. If we take a Morse function such that different critical points assume different critical values, it follows that M can be obtained from $\partial_0 M$ inductively by attaching cells of dimensions λ_1, *etc.* given by the indices of the associated critical points**.** In particular, if we start with a manifold P without boundary and form M by deleting the interiors of two closed disks**,** then it will follow that P has the homotopy type of a finite cell complex which is given in terms of the Morse function**.**