§2. Definitions and Lemmas.

The words "smooth" and "differentiable" will be used interchangeably to mean differentiable of class C^∞. The tangent space of a smooth manifold M at a point p will be denoted by TM_p. If $g\colon M \to N$ is a smooth map with $g(p) = q$, then the induced linear map of tangent spaces will be denoted by $g_*\colon TM_p \to TN_q$.

Now let f be a smooth real valued function on a manifold M. A point $p \in M$ is called a critical point of f if the induced map $f_*\colon TM_p \to T\mathbf{R}_{f(p)}$ is zero. If we choose a local coordinate system $(x^1,\ldots,x^n)$ in a neighborhood U of p this means that

$$\frac{\partial f}{\partial x_1}(p) = \ldots = \frac{\partial f}{\partial x^n}(p) = 0 .$$

The real number $f(p)$ is called a critical value of f.

We denote by M^a the set of all points $x \in M$ such that $f(x) \le a$. If a is not a critical value of f then it follows from the implicit function theorem that M^a is a smooth manifold-with-boundary. The boundary $f^{-1}(a)$ is a smooth submanifold of M.

A critical point p is called non-degenerate if and only if the matrix

$$\left(\frac{\partial^2 f}{\partial x^i \partial x^j}(p)\right)$$

is non-singular. It can be checked directly that non-degeneracy does not depend on the coordinate system. This will follow also from the following intrinsic definition.

If p is a critical point of f we define a symmetric bilinear functional f_{**} on TM_p, called the Hessian of f at p. If $v, w \in TM_p$ then v and w have extensions $\tilde{v}$ and $\tilde{w}$ to vector fields. We let* $f_{**}(v,w) = \tilde{v}_p(\tilde{w}(f))$, where $\tilde{v}_p$ is, of course, just v. We must show that this is symmetric and well-defined. It is symmetric because

$$\tilde{v}_p(\tilde{w}(f)) - \tilde{w}_p(\tilde{v}(f)) = [\tilde{v},\tilde{w}]_p(f) = 0$$

where $[\tilde{v},\tilde{w}]$ is the Poisson bracket of $\tilde{v}$ and $\tilde{w}$, and where $[\tilde{v},\tilde{w}]_p(f) = 0$

* Here $\tilde{w}(f)$ denotes the directional derivative of f in the direction $\tilde{w}$.

since f has p as a critical point.

Therefore f_{**} is symmetric. It is now clearly well-defined since $\tilde{v}_p(\tilde{w}(f)) = v(\tilde{w}(f))$ is independent of the extension $\tilde{v}$ of v, while $\tilde{w}_p(\tilde{v}(f))$ is independent of $\tilde{w}$.

If $(x^1,\ldots,x^n)$ is a local coordinate system and $v = \sum a_i \left.\frac{\partial}{\partial x^i}\right|_p$, $w = \sum b_j \left.\frac{\partial}{\partial x^j}\right|_p$ we can take $\tilde{w} = \sum b_j \frac{\partial}{\partial x^j}$ where b_j now denotes a constant function. Then

$$f_{**}(v,w) = v(\tilde{w}(f))(p) = v\left(\sum b_j \frac{\partial f}{\partial x^j}\right) = \sum_{i,j} a_i b_j \frac{\partial^2 f}{\partial x^i \partial x^j}(p) ;$$

so the matrix $\left(\frac{\partial^2 f}{\partial x^i \partial x^j}(p)\right)$ represents the bilinear function f_{**} with respect to the basis $\left.\frac{\partial}{\partial x^1}\right|_p, \ldots, \left.\frac{\partial}{\partial x^n}\right|_p$.

We can now talk about the index and the nullity of the bilinear functional f_{**} on TM_p. The index of a bilinear functional H, on a vector space V, is defined to be the maximal dimension of a subspace of V on which H is negative definite; the nullity is the dimension of the null-space, i.e., the subspace consisting of all $v \in V$ such that $H(v,w) = 0$ for every $w \in V$. The point p is obviously a non-degenerate critical point of f if and only if f_{**} on TM_p has nullity equal to 0. The index of f_{**} on TM_p will be referred to simply as the index of f at p. The Lemma of Morse shows that the behaviour of f at p can be completely described by this index. Before stating this lemma we first prove the following:

LEMMA 2.1. Let f be a C^∞ function in a convex neighborhood V of 0 in $\mathbf{R}^n$, with $f(0) = 0$. Then

$$f(x_1,\ldots,x_n) = \sum_{i=1}^{n} x_i g_i(x_1,\ldots,x_n)$$

for some suitable C^∞ functions g_i defined in V, with $g_i(0) = \frac{\partial f}{\partial x_i}(0)$.

PROOF:

$$f(x_1,\ldots,x_n) = \int_0^1 \frac{df(tx_1,\ldots,tx_n)}{dt}\,dt = \int_0^1 \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(tx_1,\ldots,tx_n)\cdot x_i\,dt .$$

Therefore we can let $g_i(x_1,\ldots,x_n) = \int_0^1 \frac{\partial f}{\partial x_i}(tx_1,\ldots,tx_n)\,dt$.

LEMMA 2.2 (Lemma of Morse). Let p be a non-degenerate critical point for f. Then there is a local coordinate system $(y^1,\ldots,y^n)$ in a neighborhood U of p with $y^i(p) = 0$ for all i and such that the identity

$$f = f(p) - (y^1)^2 - \cdots - (y^\lambda)^2 + (y^{\lambda+1})^2 + \cdots + (y^n)^2$$

holds throughout U, where λ is the index of f at p.

PROOF: We first show that if there is any such expression for f, then λ must be the index of f at p. For any coordinate system $(z^1,\ldots,z^n)$, if

$$f(q) = f(p) - (z^1(q))^2 - \cdots - (z^\lambda(q))^2 + (z^{\lambda+1}(q))^2 + \cdots + (z^n(q))^2$$

then we have

$$\frac{\partial^2 f}{\partial z^i\,\partial z^j}(p) = \begin{cases} -2 & \text{if } i = j \leq \lambda\,, \\ 2 & \text{if } i = j > \lambda\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

which shows that the matrix representing f_{**} with respect to the basis $\left.\frac{\partial}{\partial z^1}\right|_p,\ldots,\left.\frac{\partial}{\partial z^n}\right|_p$ is

$$\begin{pmatrix} -2 & & & & & \\ & \ddots & & & & \\ & & -2 & & & \\ & & & 2 & & \\ & & & & \ddots & \\ & & & & & 2 \end{pmatrix}.$$

Therefore there is a subspace of TM_p of dimension λ where f_{**} is negative definite, and a subspace V of dimension $n-\lambda$ where f_{**} is positive definite. If there were a subspace of TM_p of dimension greater than λ on which f_{**} were negative definite then this subspace would intersect V, which is clearly impossible. Therefore λ is the index of f_{**}.

We now show that a suitable coordinate system $(y^1,\ldots,y^n)$ exists. Obviously we can assume that p is the origin of $\mathbf{R}^n$ and that $f(p) = f(0) = 0$. By 2.1 we can write

$$f(x_1,\ldots,x_n) = \sum_{j=1}^{n} x_j g_j(x_1,\ldots,x_n)$$

for $(x_1,\ldots,x_n)$ in some neighborhood of 0. Since 0 is assumed to be a critical point:

$$g_j(0) = \frac{\partial f}{\partial x^j}(0) = 0\,.$$

Therefore, applying 2.1 to the g_j we have

$$g_j(x_1,\ldots,x_n) = \sum_{i=1}^{n} x_i h_{ij}(x_1,\ldots,x_n)$$

for certain smooth functions h_{ij}. It follows that

$$f(x_1,\ldots,x_n) = \sum_{i,j=1}^{n} x_i x_j h_{ij}(x_1,\ldots,x_n)\,.$$

We can assume that $h_{ij} = h_{ji}$, since we can write $\bar{h}_{ij} = \frac{1}{2}(h_{ij} + h_{ji})$, and then have $\bar{h}_{ij} = \bar{h}_{ji}$ and $f = \sum x_i x_j \bar{h}_{ij}$. Moreover the matrix $(\bar{h}_{ij}(0))$ is equal to $\left(\frac{1}{2}\frac{\partial^2 f}{\partial x^i\,\partial x^j}(0)\right)$, and hence is non-singular.

There is a non-singular transformation of the coordinate functions which gives us the desired expression for f, in a perhaps smaller neighborhood of 0. To see this we just imitate the usual diagonalization proof for quadratic forms. (See for example, Birkhoff and MacLane, "A survey of modern algebra," p. 271.) The key step can be described as follows.

Suppose by induction that there exist coordinates $u_1,\ldots,u_n$ in a neighborhood U_1 of 0 so that

$$f = \pm(u_1)^2 \pm \cdots \pm (u_{r-1})^2 + \sum_{i,j \geq r} u_i u_j H_{ij}(u_1,\ldots,u_n)$$

throughout U_1; where the matrices $(H_{ij}(u_1,\ldots,u_n))$ are symmetric. After a linear change in the last $n-r+1$ coordinates we may assume that $H_{rr}(0) \neq 0$. Let $g(u_1,\ldots,u_n)$ denote the square root of $|H_{rr}(u_1,\ldots,u_n)|$. This will be a smooth, non-zero function of $u_1,\ldots,u_n$ throughout some smaller neighborhood $U_2 \subset U_1$ of 0. Now introduce new coordinates $v_1,\ldots,v_n$ by

$$v_i = u_i \qquad \text{for } i \neq r$$

$$v_r(u_1,\ldots,u_n) = g(u_1,\ldots,u_n)\Big[u_r + \sum_{i>r} u_i H_{ir}(u_1,\ldots,u_n)/H_{rr}(u_1,\ldots,u_n)\Big].$$

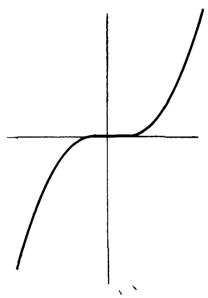
It follows from the inverse function theorem that $v_1,\ldots,v_n$ will serve as coordinate functions within some sufficiently small neighborhood U_3 of 0. It is easily verified that f can be expressed as

$$f = \sum_{i \leq r} \pm(v_i)^2 + \sum_{i,j>r} v_i v_j H'_{ij}(v_1,\ldots,v_n)$$

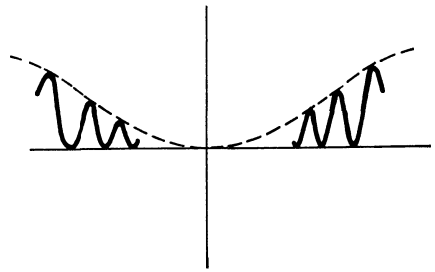
throughout U_3. This completes the induction; and proves Lemma 2.2.

COROLLARY 2.3 Non-degenerate critical points are isolated.

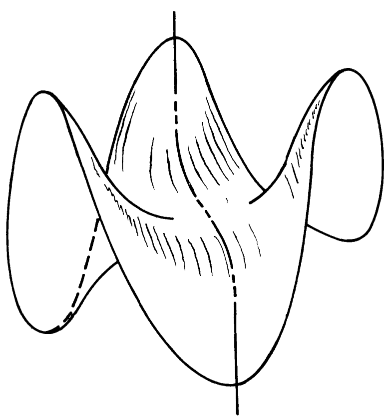
Examples of degenerate critical points (for functions on **R** and $\mathbf{R}^2$) are given below, together with pictures of their graphs.



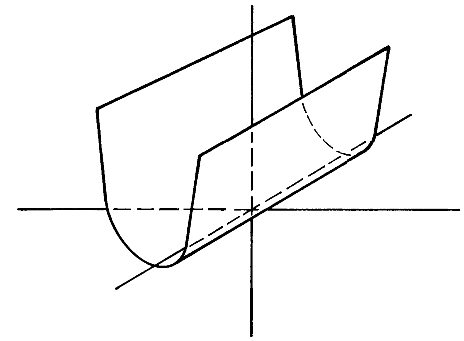
(a) $f(x) = x^3$. The origin is a degenerate critical point.



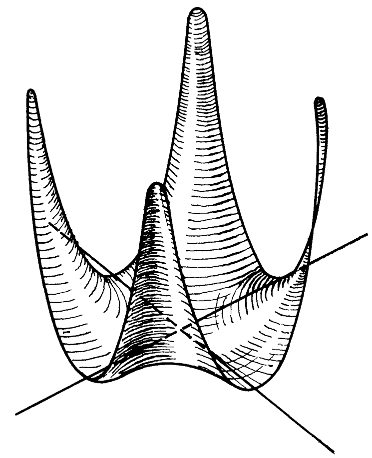
(b) $F(x) = e^{-1/x^2}\sin^2(1/x)$. The origin is a degenerate, and non-isolated, critical point.



(c) $f(x,y) = x^3 - 3xy^2$ = Real part of $(x + iy)^3$. (0.0) is a degenerate critical point (a "monkey saddle").



(d) $f(x,y) = x^2$. The set of critical points, all of which are degenerate, is the x axis, which is a sub-manifold of $\mathbf{R}^2$.



(e) $f(x,y) = x^2y^2$. The set of critical points, all of which are degenerate, consists of the union of the x and y axis, which is not even a sub-manifold of $\mathbf{R}^2$.

We conclude this section with a discussion of 1-parameter groups of diffeomorphisms. The reader is referred to K. Nomizu, "Lie Groups and Differential Geometry," for more details.

A <u>1-parameter group of diffeomorphisms</u> of a manifold M is a C^∞ map

$$\varphi : \mathbf{R} \times M \to M$$

such that

1) for each $t \in \mathbf{R}$ the map $\varphi_t: M \to M$ defined by $\varphi_t(q) = \varphi(t,q)$ is a diffeomorphism of M onto itself,

2) for all $t,s \in \mathbf{R}$ we have $\varphi_{t+s} = \varphi_t \circ \varphi_s$

Given a 1-parameter group φ of diffeomorphisms of M we define a vector field X on M as follows. For every smooth real valued function f let

$$X_q(f) = \lim_{h \to 0} \frac{f(\varphi_h(q)) - f(q)}{h} .$$

This vector field X is said to <u>generate</u> the group φ.

LEMMA 2.4. A smooth vector field on M which vanishes outside of a compact set $K \subset M$ generates a unique 1-parameter group of diffeomorphisms of M.

PROOF: Given any smooth curve

$$t \to c(t) \in M$$

it is convenient to define the <u>velocity vector</u>

$$\frac{dc}{dt} \in TM_{c(t)}$$

by the identity $\frac{dc}{dt}(f) = \lim_{h \to 0} \frac{fc(t+h) - fc(t)}{h}$. (Compare §8.) Now let φ be a 1-parameter group of diffeomorphisms, generated by the vector field X. Then for each fixed q the curve

$$t \to \varphi_t(q)$$

satisfies the differential equation

$$\frac{d\varphi_t(q)}{dt} = X_{\varphi_t(q)} ,$$

with initial condition $\varphi_0(q) = q$. This is true since

$$\frac{d\varphi_t(q)}{dt}(f) = \lim_{h \to 0} \frac{f(\varphi_{t+h}(q)) - f(\varphi_t(q))}{h} = \lim_{h \to 0} \frac{f(\varphi_h(p)) - f(p)}{h} = X_p(f) ,$$

where $p = \varphi_t(q)$. But it is well known that such a differential equation, locally, has a unique solution which depends smoothly on the initial condition. (Compare Graves, "The Theory of Functions of Real Variables," p. 166. Note that, in terms of local coordinates $u^1, \dots, u^n$, the differential equation takes on the more familiar form: $\frac{du^i}{dt} = x^i(u^1, \dots, u^n)$, $i = 1, \dots, n$.)

Thus for each point of M there exists a neighborhood U and a number $\varepsilon > 0$ so that the differential equation

$$\frac{d\varphi_t(q)}{dt} = X_{\varphi_t(q)}, \qquad \varphi_0(q) = q$$

has a unique smooth solution for $q \in U$, $|t| < \varepsilon$.

The compact set K can be covered by a finite number of such neighborhoods U. Let $\varepsilon_0 > 0$ denote the smallest of the corresponding numbers ε. Setting $\varphi_t(q) = q$ for $q \notin K$, it follows that this differential equation has a unique solution $\varphi_t(q)$ for $|t| < \varepsilon_0$ and for all $q \in M$. This solution is smooth as a function of both variables. Furthermore, it is clear that $\varphi_{t+s} = \varphi_t \circ \varphi_s$ providing that $|t|, |s|, |t+s| < \varepsilon_0$. Therefore each such φ_t is a diffeomorphism.

It only remains to define φ_t for $|t| \geq \varepsilon_0$. Any number t can be expressed as a multiple of $\varepsilon_0/2$ plus a remainder r with $|r| < \varepsilon_0/2$. If $t = k(\varepsilon_0/2) + r$ with $k \geq 0$, set

$$\varphi_t = \varphi_{\varepsilon_0/2} \circ \varphi_{\varepsilon_0/2} \circ \cdots \circ \varphi_{\varepsilon_0/2} \circ \varphi_r$$

where the transformation $\varphi_{\varepsilon_0/2}$ is iterated k times. If $k < 0$ it is only necessary to replace $\varphi_{\varepsilon_0/2}$ by $\varphi_{-\varepsilon_0/2}$ iterated $-k$ times. Thus φ_t is defined for all values of t. It is not difficult to verify that φ_t is well defined, smooth, and satisfies the condition $\varphi_{t+s} = \varphi_t \circ \varphi_s$. This completes the proof of Lemma 2.4

REMARK: The hypothesis that X vanishes outside of a compact set cannot be omitted. For example let M be the open unit interval $(0,1) \subset \mathbf{R}$, and let X be the standard vector field $\frac{d}{dt}$ on M. Then X does not generate any 1-parameter group of diffeomorphisms of M.

§3. Homotopy Type in Terms of Critical Values.

Throughout this section, if f is a real valued function on a manifold M, we let

$$M^a = f^{-1}(-\infty,a] = \{p \in M : f(p) \leq a\} .$$

THEOREM 3.1. Let f be a smooth real valued function on a manifold M. Let $a < b$ and suppose that the set $f^{-1}[a,b]$, consisting of all $p \in M$ with $a \leq f(p) \leq b$, is compact, and contains no critical points of f. Then M^a is diffeomorphic to M^b. Furthermore, M^a is a deformation retract of M^b, so that the inclusion map $M^a \to M^b$ is a homotopy equivalence.

The idea of the proof is to push M^b down to M^a along the orthogonal trajectories of the hypersurfaces f = constant. (Compare Diagram 2.)

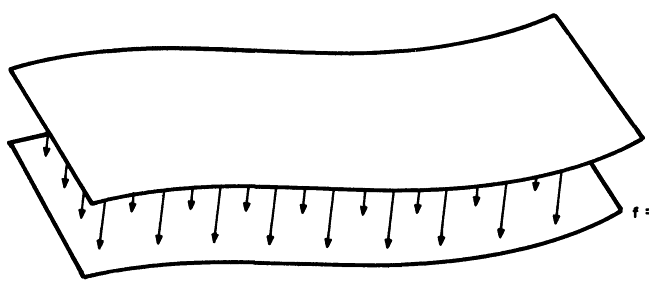


Diagram 2.

Choose a Riemannian metric on M; and let $\langle X,Y\rangle$ denote the inner product of two tangent vectors, as determined by this metric. The gradient of f is the vector field grad f on M which is characterized by the identity*

$$\langle X, \text{grad } f\rangle = X(f)$$

(= directional derivative of f along X) for any vector field X. This vector field grad f vanishes precisely at the critical points of f. If

* In classical notation, in terms of local coordinates $u^1,\ldots,u^n$, the gradient has components $\sum_j g^{ij} \frac{\partial f}{\partial u^j}$.

$c: \mathbf{R} \to M$ is a curve with velocity vector $\frac{dc}{dt}$ note the identity

$$\langle \frac{dc}{dt}, \text{grad } f\rangle = \frac{d(f\circ c)}{dt} .$$

Let $\rho: M \to \mathbf{R}$ be a smooth function which is equal to $1/\langle \text{grad } f, \text{grad } f\rangle$ throughout the compact set $f^{-1}[a,b]$; and which vanishes outside of a compact neighborhood of this set. Then the vector field X, defined by

$$X_q = \rho(q)\ (\text{grad } f)_q$$

satisfies the conditions of Lemma 2.4. Hence X generates a 1-parameter group of diffeomorphisms

$$\varphi_t: M \to M.$$

For fixed $q \in M$ consider the function $t \to f(\varphi_t(q))$. If $\varphi_t(q)$ lies in the set $f^{-1}[a,b]$, then

$$\frac{df(\varphi_t(q))}{dt} = \langle \frac{d\varphi_t(q)}{dt}, \text{grad } f\rangle = \langle X, \text{grad } f\rangle = +1 .$$

Thus the correspondence

$$t \to f(\varphi_t(q))$$

is linear with derivative +1 as long as $f(\varphi_t(q))$ lies between a and b.

Now consider the diffeomorphism $\varphi_{b-a}: M \to M$. Clearly this carries M^a diffeomorphically onto M^b. This proves the first half of 3.1.

Define a 1-parameter family of maps

$$r_t: M^b \to M^b$$

by

$$r_t(q) = \begin{cases} q & \text{if } f(q) \leq a \\ \varphi_{t(a-f(q))}(q) & \text{if } a \leq f(q) \leq b . \end{cases}$$

Then r_0 is the identity, and r_1 is a retraction from M^b to M^a. Hence M^a is a deformation retract of M^b. This completes the proof.

REMARK: The condition that $f^{-1}[a,b]$ is compact cannot be omitted. For example Diagram 3 indicates a situation in which this set is not compact. The manifold M does not contain the point p. Clearly M^a is not a deformation retract of M^b.

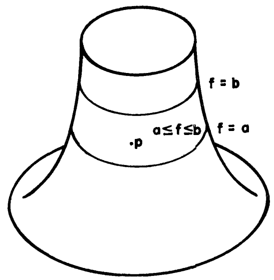


Diagram 3.

THEOREM 3.2. Let $f\colon M \to \mathbf{R}$ be a smooth function, and let p be a non-degenerate critical point with index λ. Setting $f(p) = c$, suppose that $f^{-1}[c-\varepsilon, c+\varepsilon]$ is compact, and contains no critical point of f other then p, for some $\varepsilon > 0$. Then, for all sufficiently small ε, the set $M^{c+\varepsilon}$ has the homotopy type of $M^{c-\varepsilon}$ with a λ-cell attached.

The idea of the proof of this theorem is indicated in Diagram 4, for the special case of the height function on a torus. The region

$$M^{c-\varepsilon} = f^{-1}(-\infty, c-\varepsilon]$$

is heavily shaded. We will introduce a new function $F\colon M \to \mathbf{R}$ which coincides with the height function f except that $F < f$ in a small neighborhood of p. Thus the region $F^{-1}(-\infty, c-\varepsilon]$ will consist of $M^{c-\varepsilon}$ together with a region H near p. In Diagram 4, H is the horizontally shaded region.

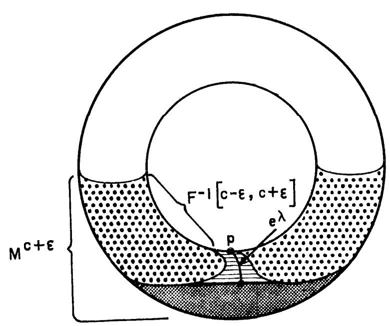


Diagram 4.

Choosing a suitable cell $e^\lambda \subset H$, a direct argument (i.e., pushing in along the horizontal lines) will show that $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $M^{c-\varepsilon} \cup H$. Finally, by applying 3.1 to the function F and the region $F^{-1}[c-\varepsilon, c+\varepsilon]$ we will see that $M^{c-\varepsilon} \cup H$ is a deformation retract of $M^{c+\varepsilon}$. This will complete the proof.

Choose a coordinate system $u^1, \ldots, u^n$ in a neighborhood U of p so that the identity

$$f = c - (u^1)^2 - \cdots - (u^\lambda)^2 + (u^{\lambda+1})^2 + \cdots + (u^n)^2$$

holds throughout U. Thus the critical point p will have coordinates

$$u^1(p) = \cdots = u^n(p) = 0 .$$

Choose $\varepsilon > 0$ sufficiently small so that

(1) The region $f^{-1}[c-\varepsilon, c+\varepsilon]$ is compact and contains no critical points other than p.

(2) The image of U under the diffeomorphic imbedding

$$(u^1, \ldots, u^n)\colon U \longrightarrow \mathbf{R}^n$$

contains the closed ball.

$$\{(u^1, \ldots, u^n)\colon \ \sum (u^i)^2 \leq 2\varepsilon\} .$$

Now define e^λ to be the set of points in U with

$$(u^1)^2 + \cdots + (u^\lambda)^2 \leq \varepsilon \quad \text{and} \quad u^{\lambda+1} = \cdots = u^n = 0.$$

The resulting situation is illustrated schematically in Diagram 5.

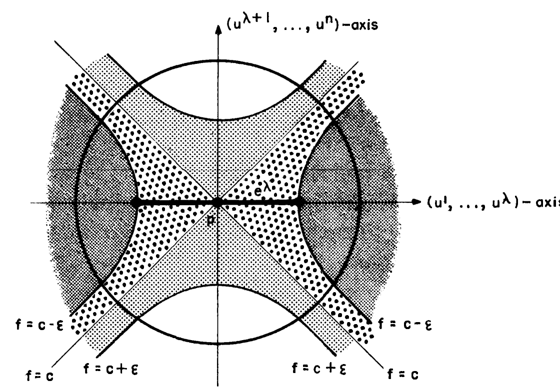


Diagram 5.

The coordinate lines represent the planes $u^{\lambda+1} = \ldots = u^n = 0$ and $u^1 = \ldots = u^\lambda = 0$ respectively; the circle represents the boundary of the ball of radius $\sqrt{2\varepsilon}$; and the hyperbolas represent the hypersurfaces $f^{-1}(c-\varepsilon)$ and $f^{-1}(c+\varepsilon)$. The region $M^{c-\varepsilon}$ is heavily shaded; the region $f^{-1}[c-\varepsilon, c]$ is heavily dotted; and the region $f^{-1}[c, c+\varepsilon]$ is lightly dotted. The horizontal dark line through p represents the cell e^λ.

Note that $e^\lambda \cap M^{c-\varepsilon}$ is precisely the boundary $\dot{e}^\lambda$, so that e^λ is attached to $M^{c-\varepsilon}$ as required. We must prove that $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $M^{c+\varepsilon}$.

Construct a new smooth function $F: M \longrightarrow \mathbf{R}$ as follows. Let

$$\mu : \mathbf{R} \longrightarrow \mathbf{R}$$

be a C^∞ function satisfying the conditions

$$\mu(0) > \varepsilon$$
$$\mu(r) = 0 \quad \text{for } r \geq 2\varepsilon$$
$$-1 < \mu'(r) \leq 0 \quad \text{for all } r,$$

where $\mu'(r) = \frac{d\mu}{dr}$. Now let F coincide with f outside of the coordinate neighborhood U, and let

$$F = f - \mu((u^1)^2 + \ldots + (u^\lambda)^2 + 2(u^{\lambda+1})^2 + \ldots + 2(u^n)^2)$$

within this coordinate neighborhood. It is easily verified that F is a well defined smooth function throughout M.

It is convenient to define two functions

$$\xi, \eta : U \longrightarrow [0, \infty)$$

by

$$\xi = (u^1)^2 + \ldots + (u^\lambda)^2$$
$$\eta = (u^{\lambda+1})^2 + \ldots + (u^n)^2$$

Then $f = c - \xi + \eta$; so that:

$$F(q) = c - \xi(q) + \eta(q) - \mu(\xi(q) + 2\eta(q))$$

for all $q \in U$.

ASSERTION 1. The region $F^{-1}(-\infty, c+\varepsilon]$ coincides with the region $M^{c+\varepsilon} = f^{-1}(-\infty, c+\varepsilon]$.

PROOF: Outside of the ellipsoid $\xi + 2\eta \leq 2\varepsilon$ the functions f and F coincide. Within this ellipsoid we have

$$F \leq f = c - \xi + \eta \leq c + \tfrac{1}{2}\xi + \eta \leq c + \varepsilon .$$

This completes the proof.

ASSERTION 2. The critical points of F are the same as those of f.

PROOF: Note that

$$\frac{\partial F}{\partial \xi} = -1 - \mu'(\xi + 2\eta) < 0$$
$$\frac{\partial F}{\partial \eta} = 1 - 2\mu'(\xi + 2\eta) \geq 1 .$$

Since

$$dF = \frac{\partial F}{\partial \xi} d\xi + \frac{\partial F}{\partial \eta} d\eta$$

where the covectors $d\xi$ and $d\eta$ are simultaneously zero only at the origin, it follows that F has no critical points in U other than the origin.

Now consider the region $F^{-1}[c-\varepsilon, c+\varepsilon]$. By Assertion 1 together with the inequality $F \leq f$ we see that

$$F^{-1}[c-\varepsilon, c+\varepsilon] \subset f^{-1}[c-\varepsilon, c+\varepsilon] .$$

Therefore this region is compact. It can contain no critical points of F except possibly p. But

$$F(p) = c - \mu(0) < c - \varepsilon .$$

Hence $F^{-1}[c-\varepsilon, c+\varepsilon]$ contains no critical points. Together with 3.1 this proves the following.

ASSERTION 3. The region $F^{-1}(-\infty, c-\varepsilon]$ is a deformation retract of $M^{c+\varepsilon}$.

It will be convenient to denote this region $F^{-1}(-\infty, c-\varepsilon]$ by $M^{c-\varepsilon} \cup H$; where H denotes the closure of $F^{-1}(-\infty, c-\varepsilon] - M^{c-\varepsilon}$.

REMARK: In the terminology of Smale, the region $M^{c-\varepsilon} \cup H$ is described as $M^{c-\varepsilon}$ with a "handle" attached. It follows from Theorem 3.1 that the manifold-with-boundary $M^{c-\varepsilon} \cup H$ is diffeomorphic to $M^{c+\varepsilon}$. This fact is important in Smale's theory of differentiable manifolds. (Compare S. Smale, Generalized Poincaré's conjecture in dimensions greater than four, Annals of Mathematics, Vol. 74 (1961), pp. 391-406.)

Now consider the cell e^λ consisting of all points q with

$$\xi(q) \leq \varepsilon, \quad \eta(q) = 0.$$

Note that e^λ is contained in the "handle" H. In fact, since $\frac{\partial F}{\partial \xi} < 0$, we have

$$F(q) \leq F(p) < c-\varepsilon$$

but $f(q) \geq c-\varepsilon$ for $q \in e^\lambda$.

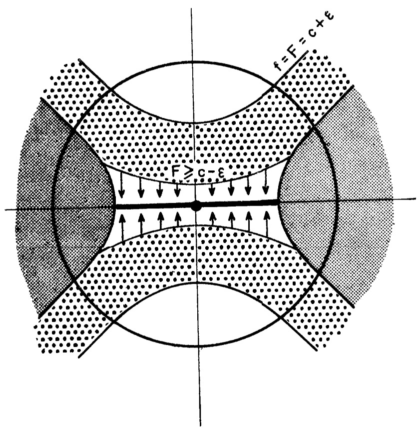


Diagram 6.

The present situation is illustrated in Diagram 6. The region $M^{c-\varepsilon}$ is heavily shaded; the handle H is shaded with vertical arrows; and the region $F^{-1}[c-\varepsilon,c+\varepsilon]$ is dotted.

ASSERTION 4. $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $M^{c-\varepsilon} \cup H$.

PROOF: A deformation retraction $r_t: M^{c-\varepsilon} \cup H \to M^{c-\varepsilon} \cup H$ is indicated schematically by the vertical arrows in Diagram 6. More precisely let r_t be the identity outside of U; and define r_t within U as follows. It in necessary to distinguish three cases as indicated in Diagram 7.

CASE 1. Within the region $\xi \leq \varepsilon$ let r_t correspond to the transformation

$$(u^1,\ldots,u^n) \to (u^1,\ldots,u^\lambda,tu^{\lambda+1},\ldots,tu^n) .$$

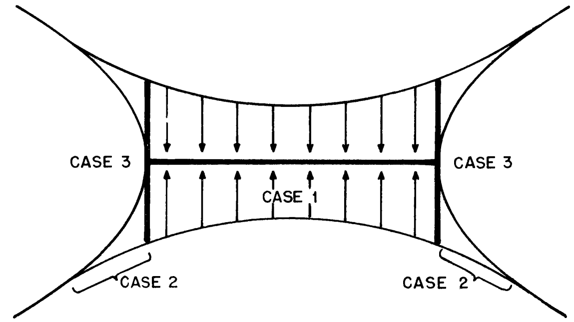


Diagram 7.

Thus r_1 is the identity and r_0 maps the entire region into e^λ. The fact that each r_t maps $F^{-1}(-\infty,c-\varepsilon]$ into itself, follows from the inequality $\frac{\partial F}{\partial \eta} > 0$.

CASE 2. Within the region $\varepsilon \leq \xi \leq \eta + \varepsilon$ let r_t correspond to the transformation

$$(u^1,\ldots,u^n) \to (u^1,\ldots,u^\lambda,s_t u^{\lambda+1},\ldots,s_t u^n)$$

where the number $s_t \in [0,1]$ is defined by

$$s_t = t + (1-t)((\xi-\varepsilon)/\eta)^{1/2} .$$

Thus r_1 is again the identity, and r_0 maps the entire region into the hypersurface $f^{-1}(c-\varepsilon)$. The reader should verify that the functions $s_t u^i$ remain continuous as $\xi \to \varepsilon$, $\eta \to 0$. Note that this definition coincides with that of Case 1 when $\xi = \varepsilon$.

CASE 3. Within the region $\eta + \varepsilon \leq \xi$ (i.e., within $M^{c-\varepsilon}$) let r_t be the identity. This coincides with the preceeding definition when $\xi = \eta + \varepsilon$.

This completes the proof that $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $F^{-1}(-\infty,c+\varepsilon]$. Together with Assertion 3, it completes the proof of Theorem 3.2.

REMARK 3.3. More generally suppose that there are k non-degenerate critical points $p_1,\ldots,p_k$ with indices $\lambda_1,\ldots,\lambda_k$ in $f^{-1}(c)$. Then a similar proof shows that $M^{c+\varepsilon}$ has the homotopy type of $M^{c-\varepsilon} \cup e^{\lambda_1} \cup \ldots \cup e^{\lambda_k}$.