A reprint from

PROCEEDINGS OF SYMPOSIA
IN PURE MATHEMATICS

Volume XV

GLOBAL ANALYSIS

AMERICAN MATHEMATICAL SOCIETY
P. O. Box 6248
Providence, Rhode Island 02904

CRITICAL POINT THEORY AND THE MINIMAX PRINCIPLE

RICHARD S. PALAIS[1]

1. **Introduction.** Since the goal of this paper is to present an exposition of a fairly general method of attack on a certain class of problems in analysis, it is perhaps in order to begin with a discussion of the domain of applicability of the concepts and techniques we are going to describe, and to illustrate them in some simple cases.

In a typical problem in analysis, both linear and nonlinear, we are given a space X and a set of "equations" defined on X and are asked to describe the set S of solutions of these equations.

There are really two quite separate types of description, depending on whether one is interested in the properties of the elements of S on the one hand or in describing the nature of the set S on the other.

Typical of the first type of description is classical "complex variable theory." Here we may take for X the set of say C^1 complex valued functions defined in some open set in the complex plane and for S the set of solutions of the Cauchy-Riemann equations. The emphasis is placed on determining the properties that elements of S have as distinguished from the general element of X (e.g. the open mapping property, the maximum modulus property, complex analyticity etc.).

It is, however, the second side of analysis that will engage us in this paper. Here the elements of X are considered to be propertyless "points" and the emphasis is placed on describing the character of S, either considered intrinsically or as a subset of X. For example a uniqueness theorem is a statement to the effect that S contains at most one point and a local uniqueness theorem is a statement that S is a discrete subset of X in some topology. An existence theorem takes the form that S is not empty or that S has cardinality (or dimension, or Lusternik-Schirelman category) greater than some given positive integer. On a more sophisticated level, when X is a smooth Banach manifold, transversality theorems are designed to give the result that S is a smooth submanifold of X of a certain given codimension.

One of the most powerful methods of attack on this type of problem consists in setting up a bijective correspondence between the set S and the set of fixed points of some self-mapping (or set of self-mappings) and then analyzing the nature of these fixed point sets by means of one of the many fixed point theorems, such as the Banach contraction principle or one of the forms of the Brouwer-Leray-Schauder-Lefschetz theorem.

A second and seemingly distinct method of attack is to set up a bijective cor-

[1] Research supported in part by Air Force Grant No. AFOSR 68-1403.

respondence between S and the set of critical points of some set of smooth real valued functions on differentiable manifolds and then to use critical point theory, and in particular the so-called "Minimax Principle," to analyze these sets of critical points. It is this second technique that will be the basic concern of this paper. As we shall see it is closely connected to the fixed point approach via the method of "steepest descent" or "gradient-like flows." Before discussing this connection we recall a well-known example of the critical point method to clarify precisely how it works in practice.

Let $A: \boldsymbol{R}^n \to \boldsymbol{R}^n$ be a C^1 map and let us write

$$A(x) = (A_1(x_1, \ldots, x_n), \ldots, A_n(x_1, \ldots, x_n)).$$

We look for the set S of solutions of the equations $A(x) = \lambda x$ where λ is any real number. In particular if A is linear, say $A_i(x) = \sum_j a_{ij} x_j$, then this is just the problem of finding the eigenvectors of A. To apply the critical point method we assume $\partial A_i/\partial x_j = \partial A_j/\partial x_i$ (in case A is linear this is just the condition that A be selfadjoint, i.e. $a_{ij} = a_{ji}$). Then there exists a C^2-function $f: \boldsymbol{R}^n \to \boldsymbol{R}$ (unique up to an additive constant) such that $A_i = \partial f/\partial x_i$. In the linear selfadjoint case we can take $f(x) = \frac{1}{2}\langle Ax, x\rangle = \frac{1}{2}\sum_{ij} a_{ij} x_i x_j$. Then the Lagrange-multiplier theorem tells us that if $\|x_0\| = r > 0$ then $x_0 \in S$ if and only if x_0 is a critical point of the restriction of f to the sphere $\Sigma(r) = \{x \in \boldsymbol{R}^n \mid \|x\|^2 = r^2\}$. In particular we have the theorem that for each $r > 0$ there are at least two points of S on the sphere $\Sigma(r)$, namely points where f assumes a minimum or maximum on $\Sigma(r)$. Note that everything works equally well if we replace $\boldsymbol{R}^n$ by a hilbert space H, except that since $\Sigma(r)$ is no longer compact when H is infinite-dimensional we can no longer assert the existence of a maximum or minimum of f on $\Sigma(r)$ in this case. However, going back to the case where A is linear, it is easily seen that if A is a compact operator then f must indeed assume a minimum and maximum value on each $\Sigma(r)$. This suggests that in proving existence theorems for critical points of a function f, compactness or finite-dimensionality assumptions on the manifold on which f is defined can be replaced by some sort of "compactness" assumption on f itself. Such a condition has been introduced by the author [7] and S. Smale [12] and is referred to as Condition (C) below. It will be formulated explicitly in §4.

Returning to the case of $\boldsymbol{R}^n$ for the moment, it might seem that there is a weakness in the critical point method. For if A is linear then there will not be just two critical points of f on $\Sigma(r)$ but always at least $2n$ (and "generically" exactly $2n$). However, if the full power of the "minimax principle" is used this result follows. Indeed as was proved by Lusternik and Schnirelman [5] if f is a smooth real-valued function on a compact manifold X then it follows from the minimax principle that f has at least cat(X) critical points on X (where cat(X), the Lusternik-Schnirelman category of X, is the least number of closed sets, each contractible in X, needed to cover X). Since the function $f(x) = \langle Ax, x\rangle$ satisfies $f(x) = f(-x)$, its restriction to the sphere $\Sigma(r)$ can be considered as a function on the real projective $(n-1)$-space $P(r) = \Sigma(r)/x \sim (-x)$, and f will have twice as many critical points on $\Sigma(r)$ as it does on $P(r)$, so it will suffice to prove that cat$(P(r)) = n$. But it is well known ([**8**] and [**2**]) that for connected X, $\text{cat}(X) \leq \dim(X) + 1$ and $\text{cat}(X) \geq \text{cuplong}(X) + 1$, where $\text{cuplong}(X)$ is the largest integer k such that there exist k cohomology classes of X of dimension greater than zero (with coefficients in any ring) whose cup product is nonzero. Since if x is the generator of $H^1(P(r), Z_2)$ then x^{n-1} is the generator of $H^{n-1}(P(r), Z_2)$, $\text{cuplong}(P(r)) = n - 1 = \dim(X)$ so $\text{cat}(X) = n$. We shall discuss the Lusternik-Schnirelman theory in detail later and we shall see that it too extends to the infinite-dimensional setting under the assumption of Condition (C).

If M is a smooth finite-dimensional manifold and $g: M \to \boldsymbol{R}$ is a C^2 function, then at a critical point p of g the second differential of g is a well-defined symmetric bilinear form on $T(M)_p$, (called the Hessian of g at p) and if it is nonsingular, p is called a nondegenerate critical point and its index (i.e., the number of minus signs when the corresponding quadratic form is written as a sum and difference of squares, or better the dimension of a maximal subspace on which the Hessian is negative definite) is called the index of the critical point p. Clearly the index measures the dimension of the "space of directions at p in which the function g is decreasing" and is equal to 0 at a minimum and $\dim M$ at a maximum. It can be shown by the Sard-Brown theorem that, except for a set of g of the first category in $C^2(M, \boldsymbol{R})$, g has only nondegenerate critical points. Such a g is called a Morse function and for such functions on compact manifolds Marston Morse [**6**] established considerably more delicate results than follow from the Lusternik-Schnirelman theory (and indeed the latter results were motivated by the attempt to see what could be said when the nondegeneracy condition was violated). In particular Morse showed that there must be at least as many critical points of index k as the kth Betti number of M (relative to any coefficient field). For example if we go back to the function g induced on $P(r)$ by the function $f(x) = \langle Ax, x\rangle$ on $\boldsymbol{R}^n$ then it is easily seen that g is a Morse function if and only if A has no nondegenerate eigenvalues. Since the mod 2 Betti numbers of $P(r)$ in dimension k, $0 \leq k \leq n - 1$ are all one it follows that in this range there will be at least one critical point of index k. In fact there is exactly one, namely the eigenspace corresponding to the $(k + 1)$st eigenvalue (arranged in increasing order). Once again all this goes through in the infinite dimensional setting if we assume Condition (C), except that we have no analogue to the theorem that "most" functions are Morse functions in this case.

It is perhaps in order to clarify at this point, by another example, why we put so much emphasis on extending results to the "infinite dimensional setting." We start with an apparently finite-dimensional situation, namely a complete C^r n-dimensional Riemannian manifold M and a C^2 path $\sigma:[a, b] \to M$ joining say p and q. If $\lambda: [a, b] \to T(M)$ $(\lambda(t) \in T(M)_{\sigma(t)})$ is a C^1 vector field along σ then the condition that λ be "auto-parallel" along σ is that $\delta\lambda/\delta t = 0$ where $\delta/\delta t$ denotes the covariant derivative along σ. If σ' denotes the tangent vector field of σ then σ is a "geodesic" if σ' is auto-parallel, i.e. if $\delta\sigma'/\delta t = 0$. We consider the problem of finding the set S of all solutions of this second order ordinary differential equation in the set $C^2([a, b], M; p, q)$ of all C^2 maps of $[a, b]$ into M starting at

p and ending at q. In particular we would like to prove that S is not empty. For ease of exposition we assume M is isometrically embedded in some R^N (by a theorem of J. Nash this is always possible) and we assume p is the origin of $\boldsymbol{R}^N$. Let H denote the set of all absolutely continuous maps $\sigma:[a,b] \to \boldsymbol{R}^N$ such that $\sigma(a) = 0$ and $\int_a^b \|\sigma'(t)\|^2 dt < \infty$. Then H is a Hilbert space under the inner product $\langle \sigma, \tau \rangle = \int_a^b \langle \sigma'(t), \tau'(t) \rangle$ dt and $\Omega = \{\sigma \in H | \text{im } \sigma \subset M \text{ and } \sigma(b) = q\}$ is a C^∞ closed submanifold of H. We define the "energy function" $J:\Omega \to \boldsymbol{R}$ by $J(\sigma) = \int_a^b \|\sigma'(t)\|^2 dt$, i.e., the square of the norm of H restricted to Ω. Then it can be shown [**7**] that J satisfies Condition (C) and that if q is chosen outside a set of measure zero in M, all the critical points of M are nondegenerate. Moreover the critical points of J are precisely the geodesics of M joining p to q, all of which are C^∞. It follows that the number of geodesics from p to q of index k is at least as large as the kth Betti number of Ω. In particular since it follows from theorems of Serre that Ω has infinitely many nonvanishing Betti numbers, there are infinitely many geodesics joining p to q. This is the simplest of a whole class of calculus of variations theorems for which one can obtain strong existence type theorems using this approach. Actually in the geodesic example above Morse obtained the basic results by "approximating" in an ingenious way the manifold Ω by certain compact finite-dimensional submanifolds [**6**]. However, in calculus of variations problems involving several independent variables these approximation techniques seem not to be possible and one must really work in the infinite-dimensional setting.

The basic technique in critical point theory is that of gradient (or gradient-like) flows, also called "the method of steepest descent." As remarked above it establishes a connection with fixed point theory and also leads directly to the general "minimax principle." For simplicity of description we shall consider the case of a C^∞ real-valued function f on a compact C^∞ Riemannian manifold M. Before preceding with the technical aspects of steepest descents we describe an extremely valuable heuristic idea which not only explains the name of the game but also accounts for most of the geometric intuition behind many of the theorems. In the product manifold $M \times \boldsymbol{R}$ we regard the M direction as "horizontal" and the positive $\boldsymbol{R}$ direction as "vertical." We consider $M \times \{0\}$ as representing "sea-level" and we identify M with the graph of f under the usual "graph embedding" $x \mapsto (x, f(x))$. Thus the value of f at a point of M is just its projection on the vertical axis, i.e., the height above sea level of the point, and the level surfaces $f =$ constant are "isoclines." If we place a "particle" at $p \in M$ the "gravitational force" it will experience will be zero at a critical point of f and elsewhere its component tangent to M will be orthogonal to the isoclines, in the direction of $-\nabla f_p$, where ∇f_p, the gradient of f at p, is the tangent vector at p dual to df_p (i.e., satisfying $\langle \nabla f_p, X \rangle = df_p(X) = Xf$ for all $X \in T(M)_p$) so that by Schwartz's inequality $-\nabla f_p$ has the direction in which df_p is least, i.e., in which f is decreasing most rapidly. We release our particle at time $t = 0$ and let it move under Aristotelian (rather than Newtonian) dynamics; to be precise so that the horizontal component of its *velocity* at any point is just the value of $-\nabla f$ at that point. Let $\phi_t(p)$ denote the position of the particle at time t, so that ϕ_t is the "flow" (i.e., the one-parameter group of diffeomorphisms of M) generated by the C^r vector field $-\nabla f$. For what should now be obvious reasons $\phi_t(p)$ is called the path of steepest descent of the point p. Now note that

$$\frac{d}{dt}(f(\phi_t(p))) = -\nabla f_{\phi_t(p)}(f) = \langle \nabla f_{\phi_t(p)}, -\nabla f_{\phi_t(p)} \rangle = -\|\nabla f_{\phi_t(p)}\|^2.$$

Thus there are exactly two possibilities: either $\nabla f_p = 0$, i.e., p is a critical point of f, in which case $\phi_t(p) = p$ for all t, or else $\nabla f_p \neq 0$ in which case $f(\phi_t(p))$ is less than (greater than) $f(p)$ for t greater than (less than) zero, so that $\phi_t(p) \neq p$ for $t \neq 0$. Thus *for any $t \neq 0$ the set of critical points of f is the same as the set of fixed points of the self-mapping $\phi_t: M \to M$.* This is the sense in which the critical point method is a special case of the fixed point method. There is more to the connection than this of course. For example a critical point p of f is nondegenerate if and only if it is a nondegenerate fixed point of ϕ_t, and then the index of p as a fixed point of ϕ_t is $(-1)^k$ where k is the index of p as a critical point of f, and it is also the degree of the zero of ∇f at p. Thus if all the critical points of f are nondegenerate the Lefschetz fixed point formula, the Hopf formula for the Euler characteristic χ of M in terms of the degrees of the zeros of ∇f, and the "Morse equality" (namely that $\chi = \Sigma(-1)^k M_k$ where M_k is the number of critical points of f of index k) all reduce to the same thing.

Let $K = K(f)$ denote the set of critical points of f, clearly a closed and hence compact subset of M. The compact set of real numbers $f(K)$ is called the set of *critical values* of f and the complementary open set of real numbers is called the set of *regular values* of f. Thus $c \in \boldsymbol{R}$ is a critical value of f if the "level" $f^{-1}(c)$ contains at least one critical point of f and is a regular value if $f^{-1}(c)$ is empty or contains only regular points of f (so a nonvalue of f is a regular value of f). The minimax principle is a very general method for locating critical values of f. It is a consequence of the deformation theorem which we shall consider next.

We shall denote by f^c the part of M "below the level c", i.e., $f^c = f^{-1}((-\infty, c])$, and K_c will denote the set of critical points of f at the level c, i.e., $K_c = K \cap f^{-1}(c)$ (so $K_c \neq \varnothing$ if and only if c is a critical value of f).

LEMMA. *Let $x_0 \in M$ be a regular point of f and let $f(x_0) = c$. Then there is an $\varepsilon > 0$ and a neighborhood V of x_0 such that $\phi_1(V) \subseteq f^{c-\varepsilon}$.*

PROOF. $df(\phi_t(x_0))/dt = -\|\nabla f_{\phi_t(x_0)}\|^2$ so $f(\phi_t(x_0))$ is monotone nonincreasing. Moreover at $t = 0$ the derivative is $-\|\nabla f_{x_0}\|^2$ which is strictly negative, and hence $f(\phi_1(x_0))$ is less than $f(\phi_0(x_0)) = f(x_0) = c$, so for some $\varepsilon > 0$, $f(\phi_1(x_0)) < c - \varepsilon$, and hence for x in a neighborhood V of x_0, $f(\phi_1(x)) < c - \varepsilon$. Q.E.D.

DEFORMATION THEOREM. *Given $c \in \boldsymbol{R}$ let U be a neighborhood of K_c in M. Then there is an $\varepsilon > 0$ such that $\phi_1(f^{c+\varepsilon} - U) \subseteq f^{c-\varepsilon}$. In particular if c is a regular value of f then there is an $\varepsilon > 0$ such that $\phi_1(f^{c+\varepsilon}) \subseteq f^{c-\varepsilon}$.*

PROOF. For each x in the compact set $X = f^{-1}(c) - U$ choose a neighborhood V_x of x in M and a $\delta_x > 0$ so that $\phi_1(V_x) \subseteq f^{c-\delta_x}$. Let $V_{x_1} \cup \ldots \cup V_{x_m}$ cover X and let $\delta = \min(\delta_{x_1}, \ldots, \delta_{x_m})$, so that if $\varepsilon < \delta$ then $\phi_1(V_{x_1} \cup \ldots \cup V_{x_m}) \subseteq f^{c-\varepsilon}$. Since M is compact and $\mathcal{O} = U \cup V_{x_1} \cup \ldots \cup V_{x_m}$ is a neighborhood of $f^{-1}(c)$, there is an $\varepsilon > 0$, which we can assume less than δ, so that $f^{-1}([c-\varepsilon, c+\varepsilon]) \subseteq \mathcal{O}$. Then since $f^{c+\varepsilon} \subseteq f^{c-\varepsilon} \cup f^{-1}([c-\varepsilon, c+\varepsilon])$ and $\mathcal{O} - U \subseteq V_{x_1} \cup \ldots \cup V_{x_m}$, $f^{c+\varepsilon} - U \subseteq f^{c-\varepsilon} \cup (V_{x_1} \cup \ldots \cup V_{x_m})$. But both $\phi_1(f^{c-\varepsilon})$ and $\phi_1(V_{x_1} \cup \ldots \cup V_{x_m})$ are included in $f^{c-\varepsilon}$. Q.E.D.

We are now in a position to formulate and prove the minimax principle. Let $\mathcal{F}$ be a family of subsets of M. We define the minimax of f over $\mathcal{F}$ by

$$\text{Minimax}(f, \mathcal{F}) = \operatorname*{Inf}_{F \in \mathcal{F}} \text{Sup} \{ f(x) | x \in F \}.$$

It is easily seen that an equivalent definition is:

$$\text{Minimax}(f, \mathcal{F}) = \text{Inf} \{ c \in \boldsymbol{R} | \exists F \in \mathcal{F} \text{ with } F \subseteq f^c \}.$$

The family $\mathcal{F}$ is called *ambient isotopy invariant* if given an isotopy g_t of M (i.e., a C^r map $(t, x) \to g_t(x)$ of $[0, 1] \times M$ into M with each g_t a diffeomorphism of M onto M and g_0 the identity) and $F \in \mathcal{F}$ it follows that $g_1(F) \in \mathcal{F}$.

MINIMAX PRINCIPLE. *If $\mathcal{F}$ is an ambient isotopy invariant family of subsets of M then Minimax$(f, \mathcal{F})$ is a critical value of f.*

PROOF. Suppose $c = \text{Minimax}(f, \mathcal{F})$ is a regular value of f and let $F \in \mathcal{F}$ with $F \subseteq f^{c+\varepsilon}$, where $\varepsilon > 0$ is chosen as in the deformation theorem. Since $\{\phi_t\}_{0 \le t \le 1}$ is an isotopy of M it follows that $\phi_1(F) \in \mathcal{F}$. But $\phi_1(F) \subseteq \phi_1(f^{c+\varepsilon}) \subseteq f^{c-\varepsilon}$ so $\text{Minimax}(f, \mathcal{F}) \le c - \varepsilon$, a contradiction. Q.E.D.

One of the most important applications of the minimax principle is the derivation of Lusternik-Schnirelman theory. However, before taking up that topic, we will consider some more elementary applications.

EXAMPLES.

(a) Let $\mathcal{F} = \{M\}$. Then $\text{Minimax}(f, \mathcal{F}) = \text{Max}_{x \in M} f(x)$.

(b) Let $\mathcal{F}$ be the family of all one point subsets of M. Then $\text{Minimax}(f, \mathcal{F}) = \text{Min}_{x \in M} f(x)$.

(c) Let X be any space and let $[X, M]$ denote the set of homotopy classes of maps of X into M. Given $\gamma \in [X, M]$ let $\mathcal{F} = \mathcal{F}(\gamma) = \{g(X) | g \in \gamma\}$. Then $\text{Minimax}(f, \mathcal{F}(\gamma))$ is a critical value of f. In particular (if $X = S^k$) this associates critical values of f to each element of $\pi_k(M)$.

(d) Let X be a smooth manifold of dimension less than dim M. If γ is an isotopy class of embeddings of X into M or a regular homotopy class of immersions of X in M, then clearly $\mathcal{F} = \mathcal{F}(\gamma) = \{g(X) | g \in \gamma\}$ is ambient isotopy invariant so that $\text{Minimax}(f, \mathcal{F})$ is a critical value of f.

(e) Let H_k denote the k dimensional homology functor (with arbitrary coefficients). Given $\gamma \in H_k(M)$, $\gamma \neq 0$, let $\mathcal{F}$ denote the set of subsets $F \subseteq M$ such that γ is in the image of $H_k(i_F): H_k(F) \to H_k(M)$. In particular if M is connected and we take γ to be the generator of $H_k(M)$ where $k = \dim M$ or zero, we get back to Examples (a) and (b) respectively.

(f) Let H^k denote a k-dimensional cohomology functor, $\gamma \neq 0$ an element of $H^k(M)$ and let $\mathscr{F}$ denote the family of subsets F of M such that γ is not annihilated by the restriction map $H^k(i_F): H^k(M) \to H^k(F)$. Again by taking the generators of $H^k(M)$ when M is connected and $k = 0$ or $\dim M$ we get back to Examples (a) and (b).

All of these examples beyond (a) and (b) can be used for finding critical points beyond the obvious minimum and maximum. However, we will only mention one example of such an application – the same tired old example as in every other discussion of critical point theory. Namely we let M be the two-dimensional torus represented as an inner tube standing in ready-to-roll position on the (x, y)-plane and we let f denote the projection of M on the vertical z axis. Aside from the minimum and maximum of f there are two nondegenerate critical points of index one which can be located by any of the methods of (c), (d), (e) or (f) [in (c) or (d) we let $X = S^1$ and take γ to be the usual generators of $\pi_1(M)$, which can be represented by embeddings, and in (e) or (f) we take $k = 1$ and γ the usual generators of H_1 or H^1].

Let us now take up the basic Lusternik-Schnirelman results on critical point theory. A subset A of M is said to have Lusternik-Schnirelman category m in M (and we write $\mathrm{cat}(A; M) = m$) if A can be covered by m (but not fewer) closed subsets of M each of which is "contractible to a point in M," the latter meaning that the inclusion map into M is homotopic, as a map into M, to a constant map. We define $\mathrm{cat}(M) = \mathrm{cat}(M; M)$.

The following are some obvious properties of the nonnegative integer valued set function $\mathrm{cat}(\cdot\,; M)$:

(i) $\mathrm{cat}(A; M) = 0$ if and only if $A = \varnothing$.

(ii) $\mathrm{cat}(A; M) = 1$ if and only if $\bar{A}$ is contractible in M.

(iii) $\mathrm{cat}(A; M) = \mathrm{cat}(\bar{A}; M)$.

(iv) If A is closed in M then $\mathrm{cat}(A, M) \leq m$ if and only if A is the union of m closed sets each contractible in M.

(v) $\mathrm{cat}(\cdot\,; M)$ is monotone, i.e., if $A \subseteq B \subseteq M$ then $\mathrm{cat}(A; M) \leq \mathrm{cat}(B; M)$.

(vi) $\mathrm{cat}(\cdot\,; M)$ is subadditive, i.e., $\mathrm{cat}(A \cup B; M) \leq \mathrm{cat}(A; M) + \mathrm{cat}(B; M)$.

(vii) If A is closed and deformable through M into B (i.e., if the inclusion map of A into M is homotopic, as a map into M, with a map of A into B), then $\mathrm{cat}(A; M) \leq \mathrm{cat}(B; M)$.

PROOF. Letting $h_t: A \to M$ be the homotopy, if $B \subseteq F_1 \cup \ldots \cup F_m$ with each F_i closed and contractible in M, then $A = G_1 \cup \ldots \cup G_m$ where $G_i = h_1^{-1}(F_i)$ is closed in M. Since $h_t|G_i$ is a homotopy of the inclusion of G_i with a map into F_i, G_i is contractible in M.

(viii) If h is a homeomorphism of M onto itself then $\mathrm{cat}(h(A); M) = \mathrm{cat}(A; M)$.

Given a positive integer m less than or equal to $\mathrm{cat}(M)$ let us define $\mathscr{F}_m$ to be the family of subsets F of M such that $\mathrm{cat}(F; M) \geq m$. It is immediate from property

(viii) that $\mathscr{F}_m$ is ambient isotopy invariant. It follows that if we define $c_m = c_m(f)$ by

$$\begin{aligned} c_m(f) &= \operatorname{Minimax}(f, \mathscr{F}_m) \\ &= \operatorname*{Inf}_{\operatorname{cat}(A\,:\,M) \geq m} \operatorname{Sup}\{f(x) | x \in A\} \\ &= \operatorname{Inf}\{c \in \boldsymbol{R} | \exists A \subseteq f^c \text{ with } \operatorname{cat}(A : M) \geq m\} \end{aligned}$$

then by the minimax principle for each positive integer $m \leq \operatorname{cat}(M)$, $c_m(f)$ is a critical value of f. We note that by the monotonicity of $\operatorname{cat}(\cdot\,; M)$ an equivalent definition of $c_m(f)$ is

$$c_m(f) = \operatorname{Inf}\{c \in \boldsymbol{R} | \operatorname{cat}(f^c, M) \geq m\}.$$

Since $\operatorname{cat}(f^c, M) \geq m + 1$ implies $\operatorname{cat}(f^c, M) \geq m$ (or equivalently since $\mathscr{F}_{m+1} \subset \mathscr{F}_m$) it is clear that $c_m(f) \leq c_{m+1}(f)$. It can of course happen that equality occurs. For example if f is constant all the $c_m(f)$ are equal. However, if equality should occur then it is made up for by there being more than one critical point on that critical level. In fact we have the following remarkable fact.

LUSTERNIK-SCHNIRELMAN MULTIPLICITY THEOREM. *If* $c = c_{n+1}(f) = c_{n+2}(f) = \ldots = c_{n+k}(f)$, *then f has at least k critical points on the critical level c. Hence if* $1 \leq m \leq \operatorname{cat}(M)$ *then f has at least m critical points at or below the level* $c_m(f)$ (i.e., in f^{c_m}) *and in particular f has at least* $\operatorname{cat}(M)$ *critical points altogether.*

PROOF. We can assume that there are only a finite number of critical points on the level c, say $x_1, \ldots, x_r$ and we must prove that $r \geq k$. Choose open neighborhoods V_i of x_i whose closures are disjoint closed balls, so that if $\mathscr{O} = V_1 \cup \ldots \cup V_r$ then clearly $\operatorname{cat}(\mathscr{O}, M) \leq r$. By the deformation theorem, for some $\varepsilon > 0$, $f^{c+\varepsilon} - \mathscr{O}$ can be deformed into $f^{c-\varepsilon}$ and since

$$c - \varepsilon < c_{n+1}(f) = \operatorname{Inf}\{a \in \boldsymbol{R} | \operatorname{cat}(f^a : M) \geq n + 1\}$$

it follows from property (vii) of $\operatorname{cat}(\cdot, M)$ that $\operatorname{cat}(f^{c+\varepsilon} - \mathscr{O}, M) \leq n$. By the monotonicity and subadditivity of $\operatorname{cat}(\cdot, M)$,

$$\operatorname{cat}(f^{c+\varepsilon}, M) \leq \operatorname{cat}((f^{c+\varepsilon} - \mathscr{O}) \cup \mathscr{O} : M) \leq n + r$$

and hence

$$c < c + \varepsilon \leq \operatorname{Inf}\{a \in \boldsymbol{R} | \operatorname{cat}(f^a : M) \geq n + r + 1\} = c_{n+r+1}$$

so that $n + r + 1 > n + k$ and hence $r \geq k$. Q.E.D.

Our final topic in this introductory section is a discussion of Morse theory in the finite-dimensional setting.

Recall that at a critical point p of f the second differential of f at p is a symmetric bilinear form on the tangent space to M at p which is independent of the coordinate system at p used in computing this second differential. This form, denoted by $H(f)_p$, is called the Hessian of f at p and if it is nondegenerate, p is called a nondegenerate critical point and the dimension of a maximal subspace of $T(M)_p$ on which $H(f)_p$ is negative definite is called the index of f at p. Let $\mathscr{O}$ be a neighborhood of a point p of M and let $\phi : \mathcal{O} \to T(M)_p$ be a chart at p mapping p to zero. We shall say that ϕ is a *Morse chart for f at p* if there is an orthogonal projection P of $T(M)_p$ onto a subspace V such that for $x \in \mathcal{O}$

$$f(x) - f(p) = \|P\phi(x)\|^2 - \|(I - P)\phi(x)\|^2.$$

Choose an orthonormal basis $e_1, \ldots, e_{k+l}$ for $T(M)_p$ with $e_1, \ldots, e_l$ in V and $e_{l+1}, \ldots, e_{k+l}$ in $V^\perp$. Then we have coordinates $x_1, \ldots, x_l, y_1, \ldots, y_k$ in $\mathcal{O}$ defined by $x_i(q) = \langle e_i, \phi(q)\rangle$ and $y_j(q) = \langle e_{l+j}, \phi(q)\rangle$ and in terms of these coordinates f is given by

$$f|\mathcal{O} = f(p) + \sum_{i=1}^{l} x_i^2 - \sum_{j=1}^{k} y_j^2.$$

Note that if $q \in \mathcal{O}$ and we put $\bar{q} = \phi(q)$ and identify $T(M)_q$ with $T(M)_p$ by the isomorphism $d\phi_q$ then df_q is given by

$$df_q(v) = 2\langle \bar{q}, Pv\rangle - 2\langle \bar{q}, (I - P)v\rangle$$

so that $df_q = 0$ if and only if $q = 0$, i.e., if and only if $q = p$. Thus p is the unique critical point of f in $\mathcal{O}$. Moreover the Hessian of f at p is clearly

$$H(f)_p(v, \omega) = 2\langle Pv, \omega\rangle - 2\langle (I - P)v, \omega\rangle$$

which is a nondegenerate and has $V^\perp$ as a maximal subspace on which it is negative definite. Thus p is a nondegenerate critical point and the index of p is rank$(I - P)$ $= \dim V^\perp = k$. The starting point of "Morse Theory" is the following crucial lemma.

MORSE LEMMA. *If p is a nondegenerate critical point of f, then there is a Morse chart for f at p.*

We shall not repeat here a proof of this famous result of Marston Morse. An elementary proof valid in the finite dimensional case we are considering at present can be found in many places, in particular on p. 6 of [3]. The Morse Lemma was extended to the infinite-dimensional case by the author in [7] (the proof appears on p. 307).

As an immediate consequence of the Morse Lemma and the remark above (that p is the only critical point in the domain of a Morse chart at p) it follows that a nondegenerate critical point of f is isolated in the set K of critical points of f. Recalling that f is called a Morse function if all the critical points of f are nondegenerate it follows that for a Morse function K is discrete. Since K is always closed in M, and hence compact, a Morse function has only a finite number of critical points. More generally if we define a critical value c of f to be a *nondegenerate critical level of f* if $f^{-1}(c)$ contains only nondegenerate critical points, then a nondegenerate critical level c of f contains only a finite number of critical points and there exists an $\varepsilon > 0$ such that c is the only critical value of f in $[c - \varepsilon, c + \varepsilon]$.

The importance of the Morse Lemma lies in the fact that it provides the basic technical tool for solving what might well be called "The Main Problem of Morse

Theory" namely: given regular values $a < b$ of a Morse function f reconstruct f^b from a knowledge of f^a and of the critical point structure of f in $f^{-1}([a, b])$. This reconstruction of f^b can be either in the strictest sense of its diffeomorphism type (and in this sense it is called "handlebody theory" and was first carried out in full detail by S. Smale [**12**], [**4**] and lies at the heart of his h-cobordism theorem) or in some weaker sense such as its homotopy type or "homology type".

We first consider the easy special case when $[a, b]$ is a "noncritical interval," that is when all $c \in [a, b]$ are regular values of f or equivalently when there are no critical points of f in $f^{-1}([a, b])$. In this case f is actually diffeomorphic (and in fact isotopic) to f^a and indeed a little more is true, namely:

NONCRITICAL INTERVAL THEOREM. *If f has no critical values in the interval $[a, b]$ then there is a C^∞ flow ψ_t on M such that*

(1) *f is monotonically nonincreasing along each orbit of ψ_t,*

(2) $\psi_1(f^b) = f^a$,

(3) *given $\varepsilon > 0$ we can suppose ψ_t leaves $f^{a-\varepsilon}$ pointwise fixed.*

PROOF. The function $g : f^{-1}([a, b]) \to \boldsymbol{R}$ defined by $g(x) = (a - b)/\|\nabla f_x\|^2$ is well defined, C^∞, and strictly negative and can be extended to be a C^∞ nonpositive function on M which is identically zero on $f^{a-\varepsilon}$ if $\varepsilon > 0$. Let ψ_t be the C^∞ flow on M generated by the C^∞ vector field $g \nabla f$. Then clearly ψ_t leaves $f^{a-\varepsilon}$ pointwise fixed and since if $p = \psi_t(p_0)$ then

$$\frac{d}{dt} f(\psi_t(p_0)) = g(p)\nabla f_p(f) = g(p)\|\nabla f\|_p^2 \leq 0,$$

f is monotonically nondecreasing along each orbit. If $p_0 \in f^b$ and $\psi_t(p_0) \notin f^a$ then $p = \psi_t(p_0) \in f^{-1}([a, b])$ so $g(p) = (a - b)/\|\nabla f_p\|^2$ and hence $df(\psi_t(p_0))/dt = a - b$, and since the same holds for all smaller positive values of t, $f(\psi_t(p_0)) = f(p_0) + t(a - b)$.

It follows that when $t = (f(p_0) - a)/(b - a) \leq 1$ then $f(\psi_t(p_0)) = a$ and hence $f(\psi_1(p_0)) \leq a$ i.e., $\psi_1(f^b) \subseteq f^a$. A similar argument shows that $\psi_{-1}(f^a) \subseteq f^b$ and since ψ_1 and ψ_{-1} are inverse diffeomorphisms $\psi_1(f^b) = f^a$. Q.E.D.

In view of the above theorem it is clear that the general case of the Main Problem of Morse Theory can be reduced to the problem of reconstructing f^b from f^a when there is only one critical level c in the interval $[a, b]$ and moreover it even suffices to consider the case $a = c - \varepsilon$ and $b = c + \varepsilon$ for $\varepsilon > 0$ arbitrarily small. To describe the situation in this case we need the concept of "attaching a handle" to a manifold with boundary.

Let D^k denote the closed unit ball in a Hilbert space of dimension k $(0 \leq k \leq \infty)$ considered as a C^∞ manifold with boundary S^{k-1} and let $\mathring{D}^k = D^k - S^{k-1}$ denote the open unit ball.

DEFINITION. Let M be a C^r hilbert manifold (possibly with boundary) and N a closed C^∞ submanifold of M which is the closure of an open submanifold. Let α be a homeomorphism of $D^k \times D^l$ onto a closed subset h of M. We shall say that

M arises from N by a C^r attachment of a handle of type (k, l) (or of index k) and we write $M = N \cup_\alpha h$ if

(1) $M = N \cup h$,

(2) $\alpha|(S^{k-1} \times D^l)$ is a C^r isomorphism onto $h \cap \partial N$,

(3) $\alpha|(D^k \times D^l)$ is a C^r isomorphism onto $M - N$.

Here are two examples of the process, which are of low enough dimension to be pictured, and either one of which suggests why the nomenclature is as it is.

In the first M is a C^r submanifold of the sphere S^2 consisting of the lower hemisphere N (think of N as a basket) and a handle h consisting of a tubular neighborhood of that part σ of a great circle lying in the upper hemisphere. If $r > 0$ we must "round off" or smooth the angles where h and N meet so that $M = N \cup h$ is of class C^r. In this case α maps $D^k \times \{0\} = D^1 \times \{0\}$ onto σ and maps $\{0\} \times D^l = \{0\} \times D^1$ onto a little segment of great circle normal to σ at the north pole.

In the second example M is the "solid torus" thought of in the usual way as a solid sphere $D^3 = N$ with a handle $h = D^1 \times D^2$ attached by identifying the two "ends" $\{-1\} \times D^2$ and $\{1\} \times D^2$ with two disjoint discs on $\partial D^3 = S^2$. Of course if $r > 0$ we must again smooth the corners along $N \cap h$.

Suppose we have a sequence of C^r manifolds $N = N_0 \subseteq N_1 \subseteq \ldots \subseteq N_s = M$ such that N_{i+1} arises from N_i by a C^r attachment α_i of a handle of type (k_i, l_i). If the images of the α_i are disjoint subsets of N, then we shall say that M arises from N by disjoint C^r attachments $(\alpha_1, \ldots, \alpha_s)$ of s handles of type $((k_1, l_1), \ldots, (k_s, l_s))$ (or of indices $(k_1, \ldots, k_s)$). With this terminology we are in a position to describe how f^c changes as c "passes through" a nondegenerate critical value.

THEOREM ON PASSING A NONDEGENERATE CRITICAL LEVEL. *Let c be a nondegenerate critical level of f and suppose c is the only critical value of f in $[a, b]$. If $k_1, \ldots, k_s$ are the indices of the critical points of f on the level c then there is an isotopy ψ_t of M such that $\psi_1(f^b)$ arises from f^a by the disjoint C^∞ attachment of s handles of indices $(k_1, \ldots, k_s)$. Given $\varepsilon > 0$ we can assume $f^{a-\varepsilon}$ is pointwise fixed under ψ_t.*

PROOF. The proof of this theorem is given in excruciating detail in §§11 and 12 of [7]. Here we shall only give a sketch, emphasizing the geometric ideas. As has already been remarked it follows from the noncritical interval theorem that we can replace a by $c - \varepsilon$ and b by $c + \varepsilon$ with $\varepsilon > 0$ arbitrarily small. It is also clear that we lose no generality if we assume $c = 0$. Finally it simplifies the writing to assume $s = 1$, the necessary modifications for general s being fairly clear, hence we assume that there is a single critical point p of f on the level zero and that p has index k. We identify a neighborhood $\mathcal{O}$ of p in M with a neighborhood of the origin in $T(M)_p$ via a Morse chart ϕ for f at p. Then f is given in $\mathcal{O}$ by $f(x, y) = \|x\|^2 - \|y\|^2$ where we have written $T(M)_p = H^l \oplus H^k$, H^k being a maximal subspace of $T(M)_p$ on which the Hessian of f at p is negative definite and H^l its orthogonal complement. The part of f^ε inside $\mathcal{O}$ is represented on Figure 1 by the region between the two branches of the hyperbola marked $f = \varepsilon$, while the part of $f^{-\varepsilon}$ in $\mathcal{O}$ is similarly represented by the two regions which are respectively above and below the upper and lower branches of the hyperbola marked $f = -\varepsilon$.

We now explain how to define a flow ψ_t on M such that $\psi_1(f^{\varepsilon})$ arises from $f^{-\varepsilon}$ by the C^r attachment of a handle of index k.

Let $\lambda:\boldsymbol{R} \to \boldsymbol{R}$ be a C^r nonincreasing function with $\lambda(x) = 1$ for $x \leq \frac{1}{2}$, $\lambda(x) > 0$ for $x < 1$ and $\lambda(x) = 0$ for $x \geq 1$ and define $g:\mathcal{O} \to \boldsymbol{R}$ by

$$g(x, y) = f(x, y) - (3\varepsilon/2)\lambda(\|x\|^2/\varepsilon).$$

Suppose $f(x, y) = \|x\|^2 - \|y\|^2 \geq -2\varepsilon$ and that $f(x, y) \neq g(x, y)$. Then clearly $\lambda(\|x\|^2/\varepsilon) \neq 0$ so $\|x\|^2 < \varepsilon$ and then $\|y\|^2 \leq \|x\|^2 + 2\varepsilon < 3\varepsilon$ and $\|x\|^2 + \|y\|^2 \leq 4\varepsilon$. Thus for ε so small that the sphere of radius $2\sqrt{\varepsilon}$ is included in $\mathcal{O}$ we can extend g to be C^r in $f^{-1}([-2\varepsilon, \varepsilon])$ by defining $g = f$ outside $\mathcal{O}$.

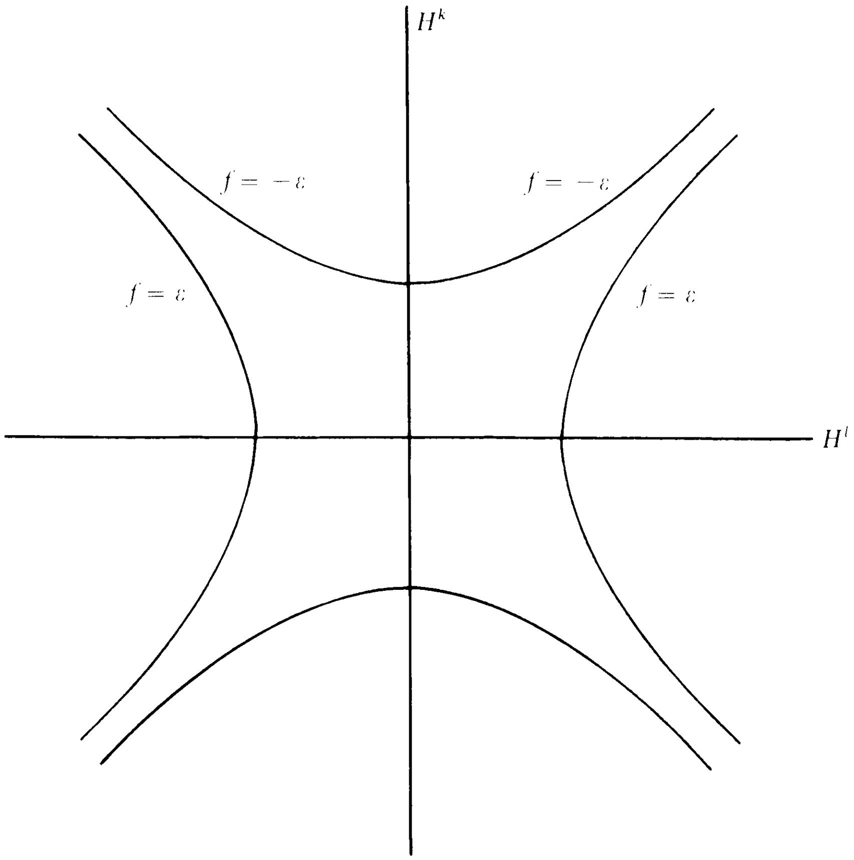

FIGURE 1

Note also that if $f(x, y) \geq \varepsilon$ then $\|x\|^2 \geq \varepsilon$ so $\lambda(\|x\|^2/\varepsilon) = 0$ and $g(x, y) = f(x, y)$. In particular the levels $f = \varepsilon$ and $g = \varepsilon$ coincide and so do the regions f^{ε} and g^{ε}. One next shows that $\nabla f(g) > 0$ in the region $g^{-1}([-\varepsilon, \varepsilon])$. For points outside $\mathcal{O}$, where $g = f$, this is obvious and inside $\mathcal{O}$ it follows from a straightforward computation (Proposition (1) §12, [7]). By an argument essentially identical to the proof of noncritical level theorem (using for ψ_t the flow generated by $-G\nabla f$ where $G: M \to \mathbf{R}$ is a nonnegative C^∞ function equal to $2\varepsilon/\nabla f(g)$ in $g^{-1}([-\varepsilon, \varepsilon])$) it follows that there is a C^r flow ψ_t on M such that ψ_t leaves $f^{-2\varepsilon}$ fixed (because we can assume $G = 0$ there) and such that $\psi_1(f^{\varepsilon}) = \psi_1(g^{\varepsilon}) = g^{-\varepsilon}$. Thus we are reduced to proving that $g^{-\varepsilon}$ arises from $f^{-\varepsilon}$ by the C^r attachment F of a handle of index k. The handle h is defined as the set of $(x, y) \in \mathcal{O}$ such that $g(x, y) \leq -\varepsilon$ and $f(x, y) \geq -\varepsilon$, the hatched region in Figure 2. The homeomorphism F of $D^k \times D^l$ onto h is given explicitly as

$$F(x, y) = (\varepsilon\sigma(\|x\|^2)\|y\|^2 + \varepsilon)^{1/2}x + (\varepsilon\sigma(\|x\|^2))^{1/2}y$$

where $\sigma: [0, 1] \to [0, 1]$ is defined by taking for $\sigma(s)$ the unique solution σ of the equation

$$\lambda(\sigma)/(1 + \sigma) = \tfrac{2}{3}(1 - s).$$

That F indeed maps $D^k \times D^l$ homeomorphically onto h and satisfies the other conditions for attaching a handle onto $f^{-\varepsilon}$ is another straightforward but somewhat messy computation which is carried out in §11 of [7]. Q.E.D.

The following corollary is the homology version of what happens when we pass a nondegenerate critical level.

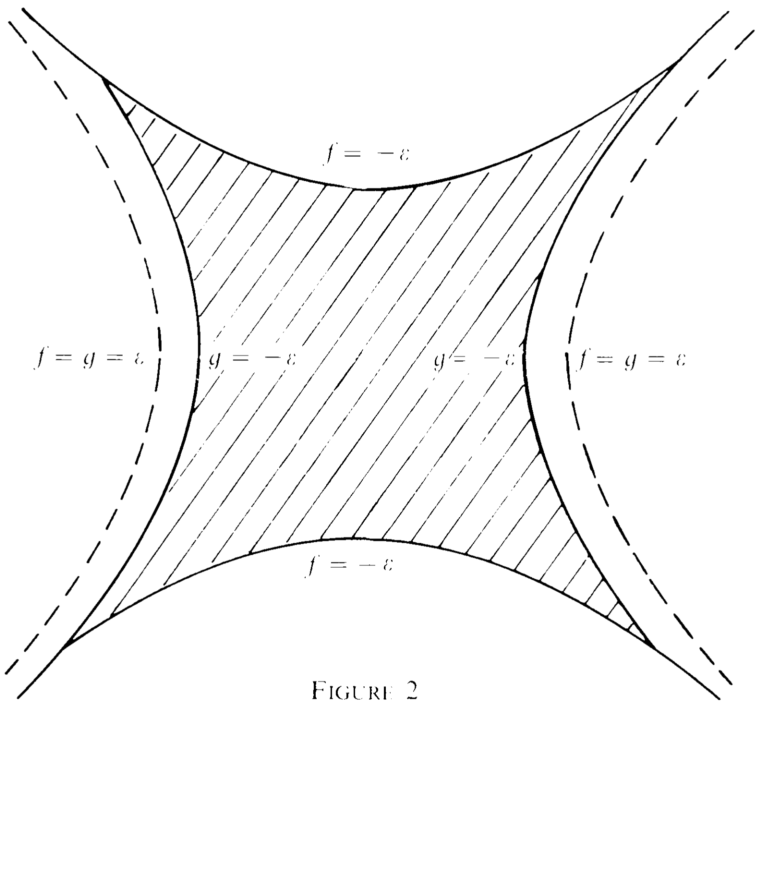

FIGURE 2

COROLLARY. *For each nonnegative integer k let $m_k(c)$ denote the number of critical points of index k on the level c. Then the homology group in dimension k of the pair (f^b, f^a) with coefficients in G is $H_k(f^b, f^a) \simeq G^{m_k(c)}$.*

PROOF. It follows by the homotopy and excision axioms that

$$H_k(f^b, f^a) \simeq H_k\Big(\bigcup_{i=1}^{s} (D^{k_i} \times D^{l_i}), \bigcup_{i=1}^{s} (S^{k_i - 1} \times D^{l_i})\Big)$$

where the unions are disjoint. Since $(D^{k_i}, S^{k_i - 1})$ is a strong deformation retract of $(D^{k_i} \times D^{l_i}, S^{k_i - 1} \times D^{l_i})$ it follows that $H_k(f^b, f^a)$ is the direct sum of the groups $H_k(D^{k_i}; S^{k_i - 1})$ and the latter of course is equal to G if $k = k_i$ and is otherwise the trivial group. Q.E.D.

"MORSE INEQUALITIES" THEOREM. *Let $f\colon M \to \boldsymbol{R}$ be a Morse function and let $M_k = M_k(f)$ denote the total number of critical points of f having index k and $R_k = \dim H_k(M, F)$ the kth Betti number of M relative to some coefficient field F. Then*

$$\sum_{m=0}^{k} (-1)^{k-m} M_m(f) \geq \sum_{m=0}^{k} (-1)^{k-m} R_m$$

and

$$\sum_{m=0}^{\dim M} (-1)^m M_m(f) = \sum_{m=0}^{k} (-1)^m R_m = \chi(M).$$

COROLLARY. $M_k(f) \geq R_k$, $k = 0, 1, \ldots, \dim M$.

PROOF. The corollary follows from the theorem by adding two adjacent inequalities. To prove the theorem let $c_1, \ldots, c_r$ be the critical values of f and let $a_0 < c_1 < a_1 < c_2 < \ldots < a_{p-1} < c_r < a_r$. Define integer valued functions of pairs of spaces (X, Y) of the homotopy type of finite C-W complexes by

$$R_k(X, Y) = \dim H_k(X, Y),$$

$$S_k(X, Y) = \sum_{m \leq k} (-1)^{k-m} R_m(X, Y),$$

$$\chi(X, Y) = \sum_{m=0}^{\infty} (-1)^m R_m(X, Y).$$

By the above Corollary to the theorem on passing a nondegenerate critical level we have $R_l(f^{a_i}, f^{a_{i-1}}) = m_l(c_i)$ so that

$$S_k(f^{a_i}, f^{a_{i-1}}) = \sum_{l \leq k} (-1)^{k-l} m_l(c_i),$$

$$\chi(f^{a_i}, f^{a_{i-1}}) = \sum_{l=0}^{\dim M} (-1)^l m_l(c_i),$$

hence

$$\sum_{i=1}^{r} S_k(f^{a_i}, f^{a_{i-1}}) = \sum_{l \le k} (-1)^{k-l} M_l$$

and

$$\sum_{i=1}^{r} \chi(f^{a_i}, f^{a_{i-1}}) = \sum_{l=0}^{\dim M} (-1)^l M_l.$$

On the other hand,

$$S_k(f^{a_r}, f^{a_0}) = S_k(M, \varnothing) = \sum_{l \le k} (-1)^{k-l} R_l,$$

$$\chi(f^{a_r}, f^{a_0}) = \chi(M, \varnothing) = \sum_{l=0}^{\dim M} (-1)^l R_l.$$

Thus the theorem will follow once it is shown that S_k is "subadditive" and χ "additive" in the sense that for $X \supseteq Y \supseteq Z$

$$S_k(X, Z) \le S_k(X, Y) + S_k(X, Z), \qquad \chi(X, Z) = \chi(X, Y) + \chi(X, Z).$$

Now from the exact homology sequence for the triple (X, Y, Z)

$$\to H_m(Y, Z) \xrightarrow{i_m} H_m(X, Z) \xrightarrow{j_m} H_m(X, Y) \xrightarrow{\partial_m} H_{m-1}(Y, Z) \to$$

we have the three short exact sequences

$$0 \to \operatorname{im}(\partial_{m+1}) \to H_m(Y, Z) \to \operatorname{im}(i_m) \to 0$$

$$0 \to \operatorname{im}(i_m) \to H_m(X, Z) \to \operatorname{im}(j_m) \to 0$$

$$0 \to \operatorname{im}(j_m) \to H_m(X, Y) \to \operatorname{im}(\partial_m) \to 0$$

from which follows

$$R_m(Y, Z) = \dim \operatorname{im}(\partial_{m+1}) + \dim \operatorname{im}(i_m),$$

$$R_m(X, Z) = \dim \operatorname{im}(i_m) + \dim \operatorname{im}(j_m),$$

$$R_m(X, Y) = \dim \operatorname{im}(j_m) + \dim \operatorname{im}(j_m);$$

hence

$$R_m(X, Z) - R_m(X, Y) - R_m(Y, Z) = -(\dim \operatorname{im}(\partial_m) + \dim \operatorname{im}(\partial_{m+1})).$$

If we multiply by $(-1)^{k-m}$ (respectively $(-1)^m$) and sum over $0 \le m \le k$ (respectively $0 \le m \le \dim M$), then we get respectively

$$S_k(X, Z) - S_k(X, Y) - S_k(Y, Z) = (-1)^{k+1} \dim \operatorname{im}(\partial_0) - \dim \operatorname{im}(\partial_{k+1})$$

(which is negative since $\dim \operatorname{im}(\partial_0) = 0$) and $\chi(X, Z) - \chi(X, Y) - \chi(Y, Z) = 0$.

Q.E.D.

This completes our survey of critical point theory for C^∞ real valued functions on compact manifolds. In the remainder of the paper we will be concerned with generalizing the results obtained above to the case of C^r functions on infinite-dimensional Banach manifolds. The key step is to generalize the technique of "steepest descent" which after all gave the deformations that are at the heart of the proofs of the two basic theorems, the "Deformation Theorem" and the "Theorem on passing a nondegenerate critical level." For Hilbert manifolds one can define the gradient vector field of a function (relative to a Riemannian metric) just as in finite dimensions. For arbitrary Banach manifolds there is no real gradient field: however, one can define "pseudo-gradient" fields which have all the essential properties of a gradient field. The real problem in both cases lies in the fact that since the manifold is no longer compact, the flow generated by a vector field is no longer a global flow (i.e., a one parameter group) and it is only by making certain completeness assumptions on the manifold and a kind of compactness assumption relative to the function (Condition (C)) that one can show that flow generated by the pseudo-gradient field really "descends to the critical point set" in a strong enough sense to carry out the proofs of the above theorems.

Not only do these conditions seem to be met in some important calculus of variations problems, but indeed they have recently played the role of suggesting what calculus of variations problems would be interesting to study from a geometrical and topological point of view. The reader can find a discussion of these applications in the author's *Foundations of global nonlinear analysis*, Benjamin, New York, 1968.

2. **Finsler manifolds.** Let E be a Banach space bundle over a space X and let $\| \ \|$ be a continuous real valued function on E such that the restriction to each fiber E_x is an admissible norm, $\| \ \|_x$, for that fiber. If we trivialize E in a neighborhood of x_0, using E_{x_0} as the standard fiber, then for each x near x_0, $\| \ \|_x$ becomes a norm on E_{x_0}. If for each such trivialization in an atlas defining the bundle structure of E and each $k > 1$ we have $(1/k)\| \ \|_x < \| \ \|_{x_0} < k\| \ \|_x$ provided x is sufficiently near x_0, then we say that $\| \ \|$ is a *Finsler structure* for the bundle E. We note that the dual bundle of E has a natural "dual" Finsler structure, also denoted by $\| \ \|$, characterized of course by

$$\|l\| = \operatorname{Sup}\{|l(e)| \mid \|e\|_x = 1\}$$

for $l \in E_x^*$.

A *Finsler manifold* is a regular C^1 Banach manifold M together with a Finsler structure on $T(M)$. As just remarked $T^*(M)$ has a natural Finsler structure, so that if $f: M \to \boldsymbol{R}$ is a C^1 real valued function then $\|df\|$, $x \mapsto \|df_x\|$, is a well-defined nonnegative continuous real-valued function on M.

If M is a regular C^1 Hilbert manifold and $\langle\ ,\ \rangle$ is a C^0 Riemannian structure for M (i.e., for $T(M)$) then $\|e\| = \langle e, e\rangle^{1/2}$ defines a Finsler structure for M, and of course the natural C^0 isomorphism of $T(M)$ with $T^*(M)$, given by the Riemannian structure, is isometric from this Finsler structure to the dual Finsler structure. In particular if f is a C^1 real valued function on M and we define a C^0 vector field ∇f on M by the condition that it be duel to df (i.e., $df_x(e) = \langle e, \nabla f\rangle_x$

for $v \in T(M)_x$) then $\|df\| = \|\nabla f\|$. As usual ∇f is called the gradient of f and satisfies $\nabla f(f) = df(\nabla f) = \|\nabla f\|^2$, so that if $\sigma : [a, b] \to$ is a C^1 path which is an integral curve of $-\nabla f$ (i.e. $\sigma'(t) = \nabla f_{\sigma(t)}$ for all $t \in [a, b]$) then $df(\sigma(t))/dt = -\|\nabla f_{\sigma(t)}\|^2$.

Let M be a Finsler manifold and $\sigma : [a, b] \to M$ a C^1 path in M. We define the *length* of σ to be the nonnegative real number $\int_a^b \|\sigma'(t)\| dt$. For a C^1 path $\sigma : (a, b) \to M$ defined on an open interval we define the length of σ to be

$$\lim_{\alpha \to a;\, \beta \to b} \int_\alpha^\beta \|\sigma'(t)\| dt$$

which of course may be infinite. If p and q are points in the same component of M, then we define the distance $\rho(p, q)$ from p to q to be the infimum of the lengths of all C^1 paths from p to q. It is a triviality that ρ is symmetric and that ρ satisfies the triangle inequality. It is often also incorrectly asserted to be a triviality that $\rho(p, q) > 0$ for $p \neq q$ (so ρ is a metric) and that the topology induced by ρ coincides with the manifold topology. Since I have been guilty of glossing over the difficulty myself[2] and indeed since there seems to be no published complete proof I shall give one here.

THEOREM. *Let M be a regular Banach manifold and $\phi : \mathcal{O} \to V$ a chart for M. Let $\| \quad \|$ be a norm for V and given $x_0 \in \mathcal{O}$ and a positive real number r define*:

$$\begin{aligned} B(x_0, r) &= \{x \in \mathcal{O} \mid \|\phi(x) - \phi(x_0)\| \leq r\}, \\ B(x_0, r) &= \{x \in \mathcal{O} \mid \|\phi(x) - \phi(x_0)\| < r\}, \\ S(x_0, r) &= \{x \in \mathcal{O} \mid \|\phi(x) - \phi(x_0)\| = r\}. \end{aligned}$$

Then if r is sufficiently small $B(x_0, r)$ is a closed neighborhood of x_0 in M, $B(x_0, r)$ is its interior relative to M and $S(x_0, r)$ is its frontier relative to M. For such r $S(x_0, r)$ separates $B(x_0, r)$ from $M - B(x_0, r)$ and in particular if $\sigma : [a, b] \to M$ is a continuous path in M with $\sigma(a) = x_0$ and $\operatorname{im}(\sigma) \not\subseteq B(x_0, r)$ then there is a $c \in [a, b]$ such that $\sigma(c) \in S(x_0, r)$ and $\sigma([a, c)) \subseteq B(x_0, r)$.

PROOF. Since M is regular we can choose a closed neighborhood G of x_0 in M with $G \subseteq \mathcal{O}$. Since ϕ maps $\mathcal{O}$ homeomorphically onto an open set $\phi(\mathcal{O})$ of V, if r is sufficiently small then $F = \{v \in V \mid \|v - \phi(x_0)\| \leq r\}$ is a closed neighborhood of $\phi(x_0)$ in V which is included in $\phi(G)$. Then $B(x_0, r) = \phi^{-1}(F)$ is a closed neighborhood of x_0 in $\mathcal{O}$ and since $\phi^{-1}(F) \subseteq G$ and G is closed in M, $B(x_0, r)$ is a closed neighborhood of x_0 in M. Since ϕ is a homeomorphism it is clear that $B(x_0, r)$ is the interior of $B(x_0, r)$ relative to $\mathcal{O}$; hence relative to M since $\mathcal{O}$ is open in M. The frontier of a closed set is its frontier relative to any open set which includes it, so the frontier of $B(x_0, r)$ relative to M is its frontier relative to $\mathcal{O}$ which, again since ϕ is a homeomorphism, is clearly $S(x_0, r)$. Since $B(x_0, r)$ and $M - B(x_0, r)$ are disjoint open sets whose complement is $S(x_0, r)$, they are separated by $S(x_0, r)$ (i.e., any connected subset of $M - S(x_0, r)$ is included in either $B(x_0, r)$ or

[2] Melvyn Huff pointed out to me that my proof of the above facts in [8] used without proof a non-obvious fact, namely the conclusions of the theorem proved below.

$M - B(x_0, r)$). Finally given σ as in the statement of the theorem $\sigma^{-1}(M - S(x_0, r))$ is an open subset of $[a, b]$ which is therefore the disjoint union of open intervals, whose frontier points are contained in the complementary subset of $[a, b]$ namely $\sigma^{-1}(S(x_0, r))$. In particular if $[a, c]$ is the interval which contains a, then in particular $\sigma(c) \in S(x_0, r)$ and $\sigma([a, c))$, being a connected subset of $M - S(x_0, r)$ which contains $\sigma(a) = x_0 \in B(x_0, r)$, is entirely included in $B(x_0, r)$. Q.E.D.

It is now relatively easy to complete the proof that ρ is a metric for M whose topology is consistent with the manifold topology. Given $x_0 \in M$ choose a chart $\phi : \mathcal{O} \to T(M)_{x_0}$ at x_0 and $r > 0$ satisfying the above theorem. We can assume that $\phi(\mathcal{O}) = \{v \in T(M)_{x_0} | \|v\| < 2r\}$ and also (by definition of a Finsler structure) that using the identification of $T(M)_x$ with $T(M)_{x_0}$ given by $d\phi_x$, for some $k > 1$ we have for all $x \in \mathcal{O}$

$$(1/k)\| \quad \|_{x_0} \leq \| \quad \|_x \leq k\| \quad \|_{x_0}.$$

If $\sigma : [a, b] \to \mathcal{O}$ is C^1 then we have length of σ

$$= \textstyle\int_a^b \|\sigma'(t)\|_{\sigma(t)} dt \geq (1/k)\int_a^b \|\tilde{\sigma}'(t)\|_{x_0} dt \geq (1/k)\|\int_a^b \tilde{\sigma}'(t) dt\|_{x_0} \geq (1/k)\|\tilde{\sigma}(b) - \tilde{\sigma}(a)\|_{x_0},$$

where $\tilde{\sigma} = \phi \circ \sigma$. If $\sigma : [a, b] \to M$ is C^1 with $\sigma(a) = x_0$ and $\mathrm{im}(\sigma) \not\subseteq \mathcal{O}$, then choosing $c \in [a, b]$ as in the above theorem the above inequality gives length of $\sigma \geq$ length of $\sigma|[a, c] \geq (1/k)\|\tilde{\sigma}(c)\| = r/k$. In particular if $q \notin \mathcal{O}$ is in the same component of M as x_0 then $\rho(x_0, q) \geq r/k$. If $q \in \mathcal{O}$ then the length of any path σ from x_0 to q is $\geq r/k$ if $\mathrm{im}(\sigma) \not\subseteq \mathcal{O}$ and is $\geq (1/k)\|\phi(q)\|_{x_0}$ if $\mathrm{im}(\sigma) \subseteq \mathcal{O}$, hence $\rho(x_0, q) \geq (1/k)\min(r, \|\phi(q)\|)$ which is positive if $q \neq x_0$. Finally we prove that the topology induced by ρ coincides with the manifold topology of M. First if $\rho(q_n, x_0) \to 0$ then eventually $\rho(q_n, x_0) < r/k$ so $q_n \in \mathcal{O}$. Moreover we then have $\|\phi(q_n)\|_{x_0} \leq k\rho(q_n, x_0)$ so $\phi(q_n) \to 0$ in $T(M)_{x_0}$; hence $q_n \to x_0$ in $\mathcal{O}$, hence in M. Conversely if $q_n \to x_0$ in M then eventually $q_n \in \mathcal{O}$. Define $\sigma_n : [0, 1] \to \mathcal{O}$, a C^1 path joining x_0 to q_n, by $\tilde{\sigma}_n(t) = \phi(\sigma_n(t)) = t\phi(q_n)$. Then $\rho(x_0, q_n) \leq$ length of σ_n $= \int_0^1 \|\tilde{\sigma}_n'(t)\| dt = \int_0^1 \|\phi(q_n)\|_{t\phi(q_n)} dt \leq k\int_0^1 \|\phi(q_n)\|_{x_0} dt = k\|\phi(q_n)\|$. But $\phi(q_n) \to 0$ (since $q_n \to x_0$) and hence $\rho(x_0, q_n) \to 0$ and we are done.

Doubtless some readers will feel that I have put too much stress on what are mere details. Unfortunately much of the foundational literature of infinite-dimensional manifolds has minor but disconcerting errors caused by ignoring precisely these and related details. In order to put the reader on guard against some of the subtle traps that lie in wait for the unwary I have added an appendix to this section with a couple of amusing counterexamples.

Appendix to § 2. There are two well-known "facts" concerning infinite-dimensional manifolds that have frequently been used without proof and often without even explicit mention.

"FACT" 1. A second countable, Hausdorff, Banach manifold is paracompact.

"FACT" 2. Let M be a paracompact Banach manifold and ϕ a chart for M mapping an open set $\mathcal{O}$ of M homeomorphically onto a Banach space V. If B is a closed ball in V then $\phi^{-1}(B)$ is closed in M.

Sad to say both of these "facts" are false. It is easy to give a counterexample to "Fact" 2. By a well-known theorem of Bessaga [1] if H denotes a separable Hilbert space there is a $C^{\prime}$ diffeomorphism ϕ of $H - \{0\}$ onto H such that $\phi(x) = x$ if $\|x\| \geq 1$. Taking $M = H$ and $\mathcal{O} = H - \{0\}$ gives our counterexample. Indeed, using the notation of the theorem of §2 (i.e., $B(0, r) = \{v \in H \mid \|\phi(v)\| \leq r\}$ etc.) it is clear that letting $r > 1$, $B(0, r) = \{v \in H \mid 0 < \|v\| \leq r\}$ which is not closed in M. To see some interesting pathology that this gives rise to, let $\{x_n\}$ be a sequence in $B(0, r)$ which converges to $0 \notin \mathcal{O}$. Since ϕ is the identity on $S(0, r)$ we can assume $x_n \in B(0, r)$. Let $\sigma_n : [1 - 1/2^n, 1 - 1/2^{n+1}] \to B(0, r)$ be a continuous path from x_n to x_{n+1} and define $\sigma : [0, 1] \to M$ by $\sigma|[1 - 1/2^n, 1 - 1/2^{n+1}] = \sigma_n$ and $\sigma(1) = 0$. Then σ is a continuous path in M starting in $B(0, r)$ and ending up outside the domain $\mathcal{O}$ of the chart ϕ *but never intersecting* $S(0, r)$! Most people are willing to bet this cannot happen, and I believe it justifies the care taken in stating and proving the theorem of §2.

"Fact" 1 was first pointed out to me by Douady. After thinking about it I published a "proof" of a more general incorrect result (Theorem 2 of *Homotopy theory of infinite-dimensional manifolds*, Topology, **5** (1966), 1–16) and restated Fact 1 as a corollary. I am indebted to E. Michael for pointing out both that it is incorrect and also that, using a joint result of his with H. H. Corson, it becomes correct if we replace "Hausdorff" by "regular" in its statement. We now construct a second countable $C^{\prime}$ Hilbert manifold M which is Hausdorff but not regular. The construction is based on the same result of Bessaga used above, however in a slightly different equivalent form. Namely if Σ denotes the unit sphere in a Hilbert space H, $\Sigma = \{v \in H \mid \|v\| = 1\}$, then there is a diffeomorphism of Σ onto H. Let e_0 denote a unit vector of H (the "north pole" of Σ), $\bar{H}^+$ the closed half space $\{v \in H \mid \langle v, e_0 \rangle \geq 0\}$, H^+ its interior $\{v \in H \mid \langle v, e_0 \rangle > 0\}$, and ∂H^+ its boundary $\{v \in H \mid \langle v, e_0 \rangle = 0\}$. As a set we put $M = H^+ \cup S$ where S is some countable dense subset of ∂H^+.

LEMMA. *There is a homeomorphism h of $H^+ \cup \{0\}$ onto H which restricts to a $C^{\prime}$ diffeomorphism of H^+ onto $H - \{0\}$.*

PROOF. Since ∂H^+ is linearly isomorphic to H there is by Bessaga a $C^{\prime}$ diffeomorphism of Σ with ∂H^+. Now ∂H^+ is diffeomorphic to $D = \{v \in \partial H^+ \mid \|v\| < 1\}$ (say by $v \mapsto \lambda(1 + \|v\|^2)v$ where $\lambda : [1, 2] \to [1, \infty]$ is a diffeomorphism which is the identity near 1) and stereographic projection from $-e_0$ maps D diffeomorphically onto $\Sigma \cap H^+$. Hence there is a diffeomorphism $g : \Sigma \approx \Sigma \cap H^+$. Define $h : H \approx H^+ \cup \{0\}$ by $h(x) = \|x\| g(x/\|x\|)$ for $x \neq 0$ and $h(0) = 0$. Then $h^{-1}(x) = \|x\| g^{-1}(x/\|x\|)$ for $x \neq 0$ and $h^{-1}(0) = 0$. Clearly h and h^{-1} are continuous and since $x \mapsto \|x\|$ is a $C^{\prime}$ map of $H - \{0\}$ onto R it follows that h maps $H - \{0\}$ diffeomorphically onto H^+. Q.E.D.

THEOREM. *M admits the structure of a connected, second countable, $C^{\prime}$ Hilbert manifold which is Hausdorff but not regular. H^+ is a dense open submanifold of M having its usual $C^{\prime}$ structure (i.e., as an open submanifold of H) and for each $s \in S$,*

$H^+ \cup \{s\}$ with its usual topology (i.e. as a subspace of H) is an open submanifold of M diffeomorphic to H.

PROOF. For each $s \in S$ we define a chart $\phi_s : H^+ \cup \{s\} \approx H$ by $\phi_s(x) = h(x - s)$, where $h : H^+ \cup \{0\} \approx H$ is the homeomorphism of the lemma. Given s_1 and s_2 in S the intersection of the domains of ϕ_{s_1} and ϕ_{s_2} is H^+ and the map $\phi_{s_1} \circ \phi_{s2}^{-1}$ of $H - \{0\} = \phi_{s_2}(H^+)$ onto $H - \{0\} = \phi_{s_1}(H^+)$ is given by $h \circ T \circ h^{-1}$ where $T: H^+ \to H^+$ is the C^r diffeomorphism $x \mapsto x + s_2 - s_1$. It follows from the lemma that $\phi_{s_1} \circ \phi_{s_2}^{-1}$ is C^r, i.e. that ϕ_{s_1} and ϕ_{s_2} are C^r related charts for M. Hence $\{\phi_s\}_{s \in S}$ is a C^r atlas for M and defines M as a C^r Hilbert manifold, which is second countable since S is countable. That each $H^+ \cup \{s\}$ is an open submanifold of M diffeomorphic to H follows from the fact that each ϕ_s is a chart. Moreover by the lemma ϕ_s maps $H \cup \{s\}$ with its usual topology homeomorphically onto H, so its usual topology agrees with the topology induced by the manifold topology of M. It follows that each point of S is adherent to H^+ in the topology of M, i.e., that H^+ is a dense open submanifold of M (that H^+ has its usual C^r structure as imbedded in M follows from the fact that h maps H^+ diffeomorphically onto $H - \{0\}$ hence so do the charts ϕ_s). Since H^+ is connected it follows that M is connected. The sets S, $H^+ \cup \{s\}$, H^+ are open Hausdorff subspaces of M (S is even discrete) and any pair of points of M is contained in one of them, so M is Hausdorff. To see that M is not regular we shall show that points of S do not have arbitrarily small closed neighborhoods, and in fact that if $s \in S$ and V is any M neighborhood of s in $H^+ \cup \{s\}$ then we can find $\bar{s} \in S$, distinct from s, (hence not in V) in the M-closure of V. Indeed since the open set $H^+ \cup \{s\}$ has its usual topology there is a $\delta > 0$ such that $\tilde{V} = \{x \in H^+ \mid \|x - s\| < \delta\} \subseteq V$, while since S is dense in ∂H^+ there exist $\bar{s} \in S$ such that $0 < \|\bar{s} - s\| < \delta$. Now similarly any M-neighborhood of $\bar{s}$ includes a set of the form $\{x \in H^+ \mid \|x - \bar{s}\| < \varepsilon\}$ for some $\varepsilon > 0$, which meets V (and in fact meets $\tilde{V}$), proving that $\bar{s}$ is in the M-closure of V. This completes the proof of the Theorem. Q.E.D.

We note that unlike points of S, points of H^+ do have arbitrarily small M-neighborhoods which are closed in M, for if $e \in H^+$ has distance δ from ∂H^+, then the closed balls of radius $\varepsilon < \delta$ about e are neighborhoods of e which are closed both in H^+ and in each $H^+ \cup \{s\}$, hence in M. It follows that no homeomorphism of M can map a point of S onto a point of H^+. Thus M is an example of a connected Hausdorff Hilbert manifold which is not topologically homogeneous (it is easy to see that connected, regular Hilbert manifold must be topologically homogeneous).

3. **Pseudogradient vector fields.** Let M be a C^{r+1} Finsler manifold $r \geq 1$ and let $f : M \to \boldsymbol{R}$ be a C^1 function on M. By definition of $\|df_p\|$, the Sup of $Xf = df_p(X)$ as X ranges over the sphere of radius $\|df_p\|$ in $T(M)_p$ is $\|df_p\|^2$. In case M is Riemannian this Sup is attained at a unique point, namely ∇f_p, by the Schwartz inequality. If the model Banach space for M is reflexive the Sup will be attained (by the weak compactness of closed balls) although not generally at a unique point unless $\| \ \|_p$ is uniformly convex. Finally if the model space is not reflexive then in general the Sup will not be attained at all. If one examines the deformations

provided by gradient flows one sees that to carry out the proofs of critical point theory it is fortunately not really essential that the flow curves be paths of steepest descent as long as they are paths of "steep enough" descent. If one looks at this in enough detail one sees that it is possible to get along with flows defined by "pseudogradient" vector fields for f.

3.1 DEFINITION. A vector $X \in T(M)_p$ is called a pseudogradient vector for f at p if $\|X\| \leq 2\|df_p\|$ and $df_p(X) \geq \|df_p\|^2$. A vector field is called a pseudogradient vector field for f if at each point of its domain it is a pseudogradient vector for f.

Let $K = \{p \in M | df_p = 0\}$ denote the closed set of critical points of f and let $M^* = M - K$ denote the set of regular points of f. Given $p \in M^*$ we can find $Y \in T(M)_p$ with $\|Y\| = 1$ such that $df_p(Y)$ is as close as we wish to $\|df_p\|$, say $df_p(Y) > \frac{2}{3}\|df_p\|$. Let $X = \frac{3}{2}\|df_p\| Y$ so that $\|X\| = \frac{3}{2}\|df_p\| < 2\|df_p\|$ and $df_p(X) > \|df_p\|^2$. Extend X to a C^r vector field in a neighborhood of p (say by making it "constant" with respect to a chart at p). The set of points x where $\|X_x\| < 2\|df_x\|$ and where $df_x(X_x) > \|df_x\|^2$ is clearly open. This proves

3.2. LEMMA. *For each $p \in M^*$ there is a C^r pseudogradient vector field for f defined in some open neighborhood of p.*

It is immediate from the definition that the set of pseudogradient vectors for f at any point p is a convex subset of $T(M)_p$. It follows that if M^* admits C^r partitions of unity then we can get a C^r pseudogradient vector field for f in M^* by patching together the local C^r vector fields of the above lemma. Since we are looking for greatest generality, and in particular since we would like our theorems to be applicable to calculus of variations problems, for which the model space is often $L^1(\boldsymbol{R}^n)$ where even C^1 partitions of unity do not exist, we shall not make this assumption. The only smoothness we really need in a pseudogradient vector field is enough to insure that it generates a flow, and it is well known that essentially the minimal hypothesis for this is that it be locally Lipschitz. Fortunately all paracompact C^1 Banach manifolds admit locally Lipschitz partitions of unity. To see this we first note that if $\mathcal{O}$ is an open set in a Banach space V then there is a locally Lipschitz nonnegative real valued function g on V which is positive precisely on $\mathcal{O}$; for example $g(v) =$ distance of v from the complement of $\mathcal{O} = \operatorname{Inf}\{\|v - w\| \,|\, w \notin \mathcal{O}\}$. To see that g is in fact globally Lipschitz, given $v_1, v_2 \in V$ and $\varepsilon > 0$ choose $w \notin \mathcal{O}$ so that $\|v_2 - w\| < g(v_2) + \varepsilon$. Then $g(v_1) \leq \|v_1 - w\| \leq \|v_1 - v_2\| + \|v_2 - w\| \leq \|v_1 - v_2\| + g(v_2) + \varepsilon$ and hence since ε is arbitrary we have $g(v_1) - g(v_2) \leq \|v_1 - v_2\|$. Interchanging the roles of v_1 and v_2 gives $|g(v_1) - g(v_2)| \leq \|v_1 - v_2\|$. Since C^1 coordinate changes preserve the property of being locally Lipschitz it now follows that any sufficiently small open set in a regular C^1 Banach manifold M is the set of points where a nonnegative locally Lipschitz real valued function is positive. Here "sufficiently small" means that the closure is inside the domain of a chart.

Now given an open cover $\{U_\alpha\}$ of a paracompact C^1 Banach manifold M choose a locally finite open refinement $\{\mathcal{O}_\beta\}$ such that $\mathcal{O}_\beta = \{x \in M | g_\beta(x) > 0\}$ where

$g_\beta : M \to \boldsymbol{R}^+$ is locally Lipschitz. Putting $g = \sum_\beta g_\beta$ and $h_\beta = g_\beta/g$, $\{g_\beta\}$ is a locally finite, locally Lipschitz partition of unity subordinate to $\{U_\alpha\}$. We now have the following result:

3.3. There exists a locally Lipschitz pseudogradient vector field for f in M^*.

In what follows X will denote a fixed locally Lipschitz pseudogradient vector field for f in M^* and ϕ_t will denote the maximal flow in M^* generated by $-X$. For each $p \in M^*$ the map $t \mapsto \phi_t(p)$ is a C^1 map of an open interval $(\alpha(p), \omega(p))$ containing zero into M^* such that $\phi_0(p) = p$ and whose tangent vector at any $t \in (\alpha(p), \omega(p))$ is $-X_{\phi_t(p)}$. The maximality of ϕ_t is equivalent to the property that for each $p \in M^*$ either $\alpha(p) = -\infty$ or else $\phi_t(p)$ has no limit point in M^* as $t \to \alpha(p)$ and similarly either $\omega(p) = \infty$ or else $\phi_t(p)$ has no limit point in M^* as $t \to \omega(p)$. The set $D_\phi = \{(p, t) \in \boldsymbol{R} \times M^* | \alpha(p) < t < \omega(p)\}$ is open in $\boldsymbol{R} \times M^*$ and ϕ is a locally Lipschitz map of D_ϕ into M^* such that $\phi_t(\phi_s(p)) = \phi_{t+s}(p)$ whenever the left side is defined. In particular if we put $D_\phi^t = \{p \in M^* | (t, p) \in D_\phi\}$, then D_ϕ^t is open in M^*, ϕ_t is a locally Lipschitz homeomorphism of D_ϕ^t onto D_ϕ^{-t} and ϕ_{-t} is its inverse.

Given $p \in M^*$ and $\alpha(p) < t_1 < t_2 < \omega(p)$ let $l_p(t_1, t_2)$ denote the length of the C^1 curve $t \mapsto \phi_t(p)$ between t_1 and t_2. Since the tangent vector to $t \mapsto \phi_t(p)$ at t is $-X_{\phi_t(p)}$ we have by definition of length and distance in a Finsler manifold that

$$\rho(\phi_{t_1}(p), \phi_{t_2}(p)) \leq l_p(t_1, t_2) = \textstyle\int_{t_2}^{t_1} \|X_{\phi_t(p)}\| dt \leq 2\int_{t_1}^{t_2} \|df_{\phi_t(p)}\| dt,$$

the latter inequality following from Definition 3.1. If g is any C^1 function on M^* then by definition of the tangent vector to a C^1 curve it follows that

$$\frac{d}{dt} g(\phi_t(p)) = -X_{\phi_t(p)}(g) = dg(-X_{\phi_t(p)})$$

and in particular by Definition 3.1 again

$$-\frac{d}{dt} f(\phi_t(p)) = df(X_{\phi_t(p)}) \geq \|df_{\phi_t(p)}\|^2$$

so that

$$\textstyle\int_{t_1}^{t_2} \|df_{\phi_t(p)}\|^2 dt \leq f(\phi_{t_1}(p)) - f(\phi_{t_2}(p)).$$

Now by Schwartz's inequality

$$\textstyle\int_{t_1}^{t_2} \|df_{\phi_t(p)}\| dt \leq (t_2 - t_1)^{1/2} (\int_{t_1}^{t_2} \|df_{\phi_t(p)}\|^2 dt)^{1/2}.$$

We have proved

3.4. THEOREM. *Let* $p \in M^*$ *and* $\alpha(p) < t_1 < t_2 < \omega(p)$. *Then*

(1) $\int_{t_2}^{t_1} \|df_{\phi_t(p)}\|^2 dt \leq f(\phi_{t_1}(p)) - f(\phi_{t_2}(p))$,

(2) $\rho(\phi_{t_1}(p), \phi_{t_2}(p)) \leq l_p(t_1, t_2) \leq 2(t_2 - t_1)^{1/2}(f(\phi_{t_1}(p)) - f(\phi_{t_2}(p)))^{1/2}$.

The above two inequalities are as we shall see basic estimates for dealing with pseudogradient vector fields. Note in particular the interpretation of the second inequality: if in moving along an integral curve during a time interval of length T we go a length or distance D, f must decrease by at least $D^2/4T$.

3.5. COROLLARY. *If $\omega(p) < \infty$ and $f(\phi_t(p))$ is bounded below then $\{\phi_t(p)\}$ is Cauchy as $t \to \omega(p)$.*

PROOF. If $B = \operatorname{Inf} f(\phi_t(p))$ and we put $C = 2(f(p) - B)^{1/2}$, then inequality (2) of Theorem 3.4 gives $\rho(\phi_{t_1}(p), \phi_{t_2}(p)) < C|t_1 - t_2|^{1/2}$ from which the corollary follows. Q.E.D.

We will be interested in asymptotic properties of $\phi_t(p)$ as $t \to \omega(p)$. The following is our first result in this direction.

3.6. THEOREM. *Let $q \in M$ and suppose that for each sufficiently small $\delta > 0$ there is a positive ε such that $\|df_x\| > \varepsilon$ if $\delta \le \rho(q, x) \le 2\delta$. Then for $p \in M^*$ either $\lim_{t\to\omega(p)}\phi_t(p) = q$ (in which case q is an isolated critical point of f) or else $\phi_t(p)$ does not have q as a limit point as $t \to \omega(p)$.*

Suppose q is a limit point of $\phi_t(p)$ as $t \to \omega(p)$. Since $f(\phi_t(p))$ is monotone nonincreasing it follows that $\lim_{t\to\omega(p)} f(\phi_t(p)) = f(q)$ and in particular $f(\phi_t(p))$ is bounded below.

CASE 1. $\omega(p) < \infty$. In this case by Corollary 3.5 $\phi_t(p)$ is Cauchy as $t \to \omega(p)$ so $\lim_{t\to\omega(p)}\phi_t(p) = q$. Since in this case the maximality of ϕ_t precludes even a limit point of $\phi_t(p)$ in M^* as $t \to \omega(p)$ we have $q \in M - M^* = K$ and the assumption of the theorem clearly implies no other critical points of f are near q.

CASE 2. $\omega(p) = \infty$. Let $t_n \to \infty$ with $\phi_{t_n}(p) \to q$. For small $\delta > 0$ it will suffice to show $\rho(q, \phi_t(p)) < 2\delta$ for all large t (and we can suppose $\rho(q, \phi_{t_n}(p)) \le \delta$). If not there is a sequence $s_n \to \infty$ so that $\rho(q, \phi_{s_n}(p)) \ge 2\delta$ and passing to subsequences we can suppose $t_n < s_n < t_{n+1}$. We shall derive a contradiction by showing that $f(\phi_{t_n}(p)) - f(\phi_{t_{n+1}}(p)) \ge \varepsilon\delta/2$ for all n, which implies that $f(\phi_{t_n}(p)) \to -\infty$, contrary to what we observed above. Recalling from (1) of Theorem 3.4 that

$$f(\phi_{t_n}(p)) - f(\phi_{t_{n+1}}(p)) \ge \int_{t_n}^{t_{n+1}} \|df_{\phi_t(p)}\|^2 dt$$

it will suffice to find $t_n \le \alpha_n < \beta_n \le t_{n+1}$ so that $\int_{\alpha_n}^{\beta_n} \|df_{\phi_t(p)}\|^2 \ge \frac{1}{2}\varepsilon\delta$. By continuity of $\rho(q, \phi_t(p))$ as a function of t there is a largest t in the interval $[t_n, s_n]$ with $\rho(q, \phi_t(p)) \le \delta$; we call it α_n. Then there is a smallest t in the interval $[\alpha_n, s_n]$ such that $\rho(q, \phi_t(p)) \ge 2\delta$; we call it β_n and note that $\rho(\phi_{\alpha_n}(p), \phi_{\beta_n}(p)) \ge \delta$. Then $\frac{1}{2}\delta \le \frac{1}{2}\rho(\phi_{\alpha_n}(p), \phi_{\beta_n}(p)) \le \frac{1}{2}\int_{\alpha_n}^{\beta_n} \|X_{\phi_t(p)}\| dt \le \int_{\alpha_n}^{\beta_n} \|df_{\phi_t(p)}\| dt$. On the other hand for $t \in [\alpha_n, \beta_n]$ we have $\delta \le \rho(q, \phi_t(p)) \le 2\delta$ so that $\|df_{\phi_t(p)}\| \ge \varepsilon$ and hence $\frac{1}{2}\varepsilon\delta \le \varepsilon\int_{\alpha_n}^{\beta_n} \|df_{\phi_t(p)}\| dt \le \int_{\alpha_n}^{\beta_n} \|df_{\phi_t(p)}\|^2 dt$ as was to be shown. It remains to show that in Case 2, q must be a critical point of f. If not, that is if $q \in M^*$, then

$$\phi_s(q) = \phi_s(\lim_{t\to\infty} \phi_t(p)) = \lim_{t\to\infty} \phi_s(\phi_t(p)) = \lim_{t\to\infty} \phi_{s+t}(p) = q$$

for all s which implies $-X_q = 0$, and since $\|df_q\|^2 \le df_q(X_q)$ it follows that $df_q = 0$, so q is a critical point of f. Q.E.D.

3.7. COROLLARY. *If q is a limit point of $\phi_t(p)$ as $t \to \omega(p)$, then q is a critical point of f.*

PROOF. If $q \notin K$ then $df_q \neq 0$ so $\|df_q\| > \varepsilon > 0$ and for δ sufficiently small $\|df_x\| > \varepsilon$ if $\rho(q, x) < \delta$. Q.E.D.

4. **Condition (C) and its Consequences.** We will maintain the standing assumptions and notations of the preceding section and in addition we make the following three assumptions.

ASSUMPTION 1. f is bounded below, and we put $B = \operatorname{Inf}\{f(x) | x \in M\}$.

ASSUMPTION 2. For each $c \in \boldsymbol{R}$, $f^c = f^{-1}((-\infty, c])$ is complete in the Finsler metric for M.

ASSUMPTION 3. f satisfies Condition (C), i.e., given a sequence $\{p_n\}$ in M such that $f(p_n)$ is bounded and such that $\|df_{p_n}\| \to 0$, a subsequence of $\{p_n\}$ converges (automatically, by the continuity of $\|df\|$, to a critical point of f).

4.1. THEOREM. *Given $p \in M^*$, $f(\phi_t(p))$ converges monotonically to a limit $a \geq B$ as $t \to \omega(p)$. If $\omega(p) < \infty$ then $\phi_t(p)$ converges to an element of K_a while if $\omega(p) = \infty$ then $\phi_t(p)$ has at least one limit point q as $t \to \infty$. Any such q belongs to K_a and if q is an isolated point of K, then in fact $\phi_t(p) \to q$ as $t \to \infty$. In particular a is always a critical level of f. As $t \to \alpha(p)$ either $f(\phi_t(p))$ diverges monotonically to ∞ or else it converges monotonically to a finite limit b. In the latter case if $\alpha(p) > -\infty$ then $\phi_t(p)$ converges to a point of K_b as $t \to \alpha(p)$ while if $\alpha(p) = -\infty$ then $\phi_t(p)$ has at least one limit point $\bar{q}$ as $t \to -\infty$, any such $\bar{q} \in K_b$, and $\phi_t(p) \to \bar{q}$ as $t \to -\infty$ if $\bar{q}$ is an isolated point of K. In particular b is a critical value of f.*

PROOF. That $f(\phi_t(p))$ is monotonically decreasing we have already seen, and since it is bounded below by B it converges to a finite limit $a \geq B$ as $t \to \omega(p)$ and a finite limit b or $+\infty$ as $t \to \alpha(p)$. In particular $\phi_t(p)$ is in the complete set f^c ($c = f(p)$) for $0 \leq t < \omega(p)$ and (if $b \neq \infty$) in the complete set f^b for $\alpha(p) < t \leq 0$. If $\omega(p) < \infty$ then it is immediate from Corollary 3.5 that $\phi_t(p) \to q$ as $t \to \omega(p)$ and of course $f(q) = \lim_{t \to \omega(p)} f(\phi_t(p)) = a$, so by Corollary 3.7, $q \in K_a$. If $\omega(p) = \infty$ then in (1) of Theorem 3.5 take $t_1 = 0$ and we get for any $t_2 > 0$ that

$$\int_0^{t_2} \|df_{\phi_t(p)}\|^2 dt \leq f(p) - f(\phi_{t_2}(p)) \leq f(p) - B$$

and letting $t_2 \to \infty$ gives $\int_0^\infty \|df_{\phi_t(p)}\|^2 < \infty$ from which it follows that $\|df_{\phi_t(p)}\|$ cannot be bounded away from zero for $0 \leq t < \infty$. By Condition (C) $\phi_t(p)$ has a critical point as a limit point as $t \to \infty$ and in fact by Corollary 3.7 each limit point q of $\phi_t(p)$ as $t \to \infty$ is a critical point of f and, since $f(q) = \lim_{t \to \infty} f(\phi_t(p))$ $= a$, $q \in K_a$. If q is an isolated critical point of f we can choose $\delta > 0$ so small that f is bounded and has no other critical points in the closed ball of radius 2δ about q. If $S = \{x \in M | \delta \leq \rho(x, q) \leq 2\delta\}$, then it follows from Condition (C) that for some $\varepsilon > 0$, $\|df_x\| > \varepsilon$ for all $x \in S$ and by Theorem 3.6 we have $q = \lim_{t \to \infty} \phi_t(p)$. The arguments for $t \to \alpha(p)$ are essentially identical. Q.E.D.

4.2. THEOREM. *Let g be a locally Lipschitz real valued function on M with $0 \leq g \leq 1$ and suppose g vanishes for all x with $f(x)$ larger than some c and also in a neighborhood of K. Define a locally Lipschitz vector field Y on M by $Y_x = 0$ for $x \in K$ and $Y_x = -g(x)X_x$ for $x \in M^*$. Then the maximal flow ψ_t generated by*

Y is a one parameter group of locally Lipschitz homeomorphisms of M, i.e. for each $p \in M$, $\psi_t(p)$ is defined for $-\infty < t < \infty$.

PROOF. We will show that $\psi_t(p)$ is defined for $0 \le t < \infty$, the proof for $-\infty < t \le 0$ being similar. If $Y_p = 0$ then of course $\psi_t(p) = p$ for all $t \in \boldsymbol{R}$. If $Y_p \neq 0$ then since Y is proportional to X with proportionality factor between zero and one it follows that there is a monotone map $\lambda: [0, \bar{\omega}(p)) \to [0, \omega(p)]$ such that $\lambda(t) \le t$ and $\psi_t(p) = \phi_{\lambda(t)}(p)$. If $\bar{\omega}(p) < \infty$ then $\psi_t(p)$ could not have a limit point in M as $t \to \bar{\omega}(p)$, i.e., $\phi_t(p)$ could not have a limit point in M as $t \to \lim_{t \to \bar{\omega}(p)} \lambda(t) \le \omega(p)$. But this contradicts Theorem 4.1 so that $\bar{\omega}(p) = \infty$. Q.E.D.

There is an interesting elementary consequence of Condition (C) which we shall need.

4.3. THEOREM. *$f|K$ is proper. That is, given $-\infty < a < b < \infty$, $K \cap f^{-1}([a, b])$ is compact. In particular for each $c \in \boldsymbol{R}$, $K_c = K \cap f^{-1}(c)$ is compact.*

PROOF. Since $\|df\|$ is identically zero on K it follows from Condition (C) that any sequence $\{p_n\}$ in $K \cap f^{-1}([a, b])$ has a convergent subsequence. Q.E.D.

REMARK. It is worth mentioning that if f itself is proper (in particular if M is compact), then Condition (C) is automatically satisfied relative to any Finsler structure for M and, given Assumption 1, so is Assumption 2. But note that if f is proper then M is locally compact hence finite dimensional, while Condition (C) holds in many interesting infinite-dimensional cases.

4.4. COROLLARY. *The set $f(K)$ of critical values of f is closed and hence the set of regular values of f is open.*

4.5. COROLLARY. *If U is any neighborhood of K_c in M there is a $\delta > 0$ such that $\{x \in M | \rho(x, K_c) \le 2\delta\} \subseteq U$. If c is an isolated critical value of f, then we can assume that for all sufficiently small positive ε and for all x with $\delta \le \rho(x, K_c) \le 2\delta$ we have $\|df_x\| > \varepsilon$.*

PROOF. The first remark is trivial from the compactness of K_c from which it also follows that if $\lambda > 0$ then we will have $|f(x) - c| < \lambda$ for $\rho(x, K_c) \le 2\delta$ provided δ is small enough. If c is a regular value or an isolated critical value of f and we choose $\lambda > 0$ so that there are no other critical values of f in $[c - \lambda, c + \lambda]$, then there are no critical points of f in the closed set $f^{-1}([c - \lambda, c + \lambda]) \cap \{x \in M | \rho(x, K_c) \ge \delta\}$ and hence by Condition (C) we must have $\|df_x\| > \varepsilon > 0$ in this set and a fortiori in $\{x \in M | \delta \le \rho(x, K_c) \le 2\delta\}$. Q.E.D.

We now come to our main result, an analogue of the Deformation Theorem of §1. The assumption that c is an isolated critical value of f is not really necessary, it merely simplifies the proof (rather drastically). The proof in the general case can be found in [**8**]. The argument given there is a modification of a proof given by J. T. Schwartz [**11**] that works in the Riemannian case.

4.6. DEFORMATION THEOREM. *Let c be a regular value of f or an isolated critical value of f. If U is any neighborhood of K_c in M there is a one parameter group ψ_t of Lipschitz homeomorphisms of M and an $\varepsilon > 0$ such that $\psi_1(f^{c+\varepsilon} - U) \subset f^{c-\varepsilon}$.*

PROOF. Using Corollary 4.5 choose $\delta > 0$ so that $\rho(x, K_c) \leq 2\delta \Rightarrow x \in U$, and then choose $\varepsilon > 0$ so that

(a) c is the only critical value of f in $(c - 3\varepsilon, c + 3\varepsilon)$.

(b) $\|X_p\| > (8\varepsilon)^{1/2}$ if $\delta \leq \rho(x, K_c) \leq 2\delta$.

(c) $\varepsilon < \frac{1}{8}\delta^2$.

Using a locally Lipschitz partition of unity we can construct a locally Lipschitz real valued function g on M with $0 \leq g \leq 1$ such that $g(x) = 0$ if $\rho(x, K_c) \leq \frac{1}{2}\delta$ or if $|f(x) - c| \geq 2\varepsilon$ and $g(x) = 1$ if $|f(x) - c| \leq \varepsilon$ and $\rho(x, K_c) \geq \delta$. Let ψ_t be the one parameter group of locally Lipschitz homeomorphisms generated by $-gX = Y$ (Theorem 4.2). We note that if it should happen that $\delta \leq \rho(\psi_t(p), K_c)$ and $|f(\psi_t(p)) - c| \leq \varepsilon$ for $0 \leq t \leq t_0$, then $\psi_{t_0}(p) = \phi_{t_0}(p)$.

Given $p \in f^{c+\varepsilon} - U$ we must show that $\psi_1(p) \in f^{c-\varepsilon}$. Suppose on the contrary that $f(\psi_1(p)) > c - \varepsilon$ so that, since $f(\psi_t(p))$ is monotone nonincreasing, $|f(\psi_t(p)) - c| < \varepsilon$ and $f(p) - f(\psi_t(p)) < 2\varepsilon$ for $0 \leq t \leq 1$.

CASE 1. $\rho(\psi_t(p), K_c) > \delta$, $0 \leq t \leq 1$.

Then as remarked above $\phi_t(p) = \psi_t(p)$, $0 \leq t \leq 1$. Moreover by (c) $\|X_{\phi_t(p)}\| \geq (8\varepsilon)^{1/2}$ for $0 \leq t \leq 1$ and hence $l_p(0, 1) \geq (8\varepsilon)^{1/2}$ so by (2) of Theorem 3.4 $(8\varepsilon)^{1/2} \leq 2(f(p) - f(\psi_1(p)))^{1/2}$ and squaring both sides gives a contradiction, namely $f(p) - f(\psi_1(p)) \geq 2\varepsilon$.

CASE 2. $\rho(\psi_t(p), K_c) \leq \delta$ for some $t \in [0, 1]$.

Let t_0 be the first such t. Then $\psi_t(p) = \phi_t(p)$, $0 \leq t \leq t_0$. Moreover $\rho(p, K_c) \geq 2\delta$ (because $p \notin U$) and since $\rho(p, K_c) \leq \delta$, $\rho(p, \phi_{t_0}(p)) \geq \delta$ and using (2) of Theorem 3.4 once more $\delta \leq 2\sqrt{t_0}(f(p_0) - f(\psi_{t_0}(p)))^{1/2}$. Squaring both sides and remembering $t_0 \leq 1$ gives $f(p) - f(\psi_1(p)) \geq \frac{1}{4}\delta^2$ and since $\frac{1}{8}\delta^2 > \varepsilon$ we have the same contradiction. Q.E.D.

4.7. THEOREM (MINIMAX PRINCIPLE). *Let $\mathscr{F}$ be an ambient isotopy invariant family of subsets of M and assume f is bounded above on at least one $F \in \mathscr{F}$. Then* Minimax $(f, \mathscr{F})$ *is a critical value of f.*

PROOF. The assumption that Minimax $(f, \mathscr{F})$ is a regular value of f leads to a contradiction exactly as in §1.

4.8. COROLLARY. *f assumes its minimum on M and on each component M_0 of M. If f is bounded above it assumes its maximum on M.*

PROOF. Take $\mathscr{F}$ to be respectively the families $\{\{p\}|p \in M\}$, $\{\{p\}|p \in M_0\}$ and $\{M\}$.

4.9. COROLLARY. *Given an integer k with $1 \leq k \leq \operatorname{cat}(M)$ define*

$$c_k = \operatorname{Inf}\{c \in \boldsymbol{R} | \exists A \subseteq f^c \text{ with } \operatorname{cat}(A : M) \geq k\} = \operatorname{Inf}\{c \in \boldsymbol{R} | \operatorname{cat}(f^c, M) \geq k\}.$$

Then either $c_k = \infty$ or else c_k is a critical value of f. Moreover if $c = c_{n+1} = c_{n+2} = \dots = c_{n+k} < \infty$, then either c is an isolated critical value of f, in which case f has at least k critical points on the level c, or else for any $\varepsilon > 0$, f has an infinite number of critical points p satisfying $|f(p) - c| < \varepsilon$.

PROOF. As in the finite-dimensional case.

REMARK. As remarked preceding the proof of the Deformation Theorem (Theorem 4.6) the assumption that c is an isolated critical value of f in that theorem can be dispensed with. Using this it follows that if $c = c_{n+1} = c_{n+2} = \ldots = c_{n+k} < \infty$ then in fact K_c always contains at least k points. In fact with a little more care something stronger can be shown, namely that $\operatorname{cat}(K_c, M) \geq k$. This implies, for example, that if M is connected $\dim K_c \geq k - 1$ (where dim means covering dimension) so that, in this case, when $k \geq 2$, K_c is always infinite. For details we refer to [**8**].

4.10. COROLLARY. *If $c_k < \infty$ the f has at least k critical points in $f^{c_k+\varepsilon}$ for any $\varepsilon > 0$.*

Note that we cannot immediately deduce that f has at least $\operatorname{cat}(M)$ critical points because of the possibility that $c_k = \infty$ for some $k \leq \operatorname{cat}(M)$. However, suppose f had only finitely many critical points. Then for some $c \in \boldsymbol{R}$ all critical values of f would be less than c. In this case it is easy to see that M can be deformed into f^c so that if $k \leq \operatorname{cat}(M)$, then $\operatorname{cat}(f^c, M) \geq k$ and $c_k < c$.

PROOF. Let g be a locally Lipschitz function with $0 \leq g \leq 1$ and $g(x) = 1$ for $f(x) \geq c$ and $g(x) = 0$ for $x \leq c - \varepsilon$, where $c - \varepsilon$ is greater than all critical values of f. If ψ_t is the maximal flow generated by $-gX$ then $\psi_t(p)$ is defined for all $t \geq 0$. Moreover for each $p \in M$ $f(\psi_t(p))$ is eventually less than c (because $\phi_t(p)$ has a limit point in $K \subseteq f^{c-\varepsilon}$) and the first t, say $\lambda(p)$, where $\psi_t(p) \in f^c$ is easily seen to be continuous. Then h_t defined by $h_t(p) = \psi_{\lambda(p)t}(p)$ defines a deformation of M into f^c.

Thus either all $c_k < \infty$ for $k \leq \operatorname{cat}(M)$ in which case Corollary 4.10 implies f has at least $\operatorname{cat}(M)$ critical points, or else f has infinitely many critical points. In any case we have

4.11. THEOREM. *f has at least* $\operatorname{cat}(M)$ *critical points altogether.*

REFERENCES

1. C. Bessaga, *Every infinite dimensional hilbert space is diffeomorphic with its unit sphere*, Bull. Acad. Polon. Sci. **14** (1966), 27–31.

2. I. Bernstein and T. Ganea, Homotopical nilpotency, Illinois. J. Math. **5** (1961), 99–130.

3. J. Milnor, *Morse theory*, Ann. of Math. Studies No. 51, Princeton Univ. Press, Princeton, N.J., 1963.

4. ———, *Lectures on the h-cobordism theorem*, Mathematical Lecture Notes Series, Princeton Univ. Press, Princeton, N.J., 1965.

5. L. Lusternik and L. Schnirelman, *Méthodes topologiques dans les problèmes variationels*, Hermann, Paris, 1934.

6. M. Morse, *Calculus of variations in the large*, Amer. Math. Soc. Colloq. Publ., vol. XVII, Amer. Math. Soc., Providence, R.I., 1934.

7. R. S. Palais, *Morse theory on Hilbert manifolds*, Topology **2** (1963), 299–340.

8. ———, *Lusternik-Schnirelman theory on Banach manifolds*, Topology **5** (1966), 115–132.

9. R. S. Palais and S. Smale, *A generalized Morse theory*, Bull. Amer. Math. Soc. **70** (1964), 165–171.

10. E. H. Rothe, *Critical points and the gradient fields of scalars in Hilbert space*, Acta Math. **85** (1951), 73–98.

11. J. T. Schwartz, *Generalizing the Lusternik-Schnirelman theory of critical points*, Comm. Pure and Appl. Math. **17** (1964), 307–315.

12. S. Smale, *Morse theory and a non-linear generalization of the Dirichlet problem*, Ann. of Math. (2) **80** (1964), 382–396.

13. ———, *Generalized Poincaré's conjecture in dimensions greater than four*, Ann. of Math. (2) **74** (1961), 391–406.

Brandeis University and
Institute for Advanced Study