

INEQUALITIES OF CRITICAL POINT THEORY¹

EVERETT PITCHER

A purpose of critical point theory is the counting of critical points of functions. Principal theorems in the subject state in precise terms that topological complexity of the underlying space is reflected in the existence and nature of critical points of any smooth real-valued function defined on the space.

The initial development of critical point theory is peculiarly the work of one man, Marston Morse. Readers of the fundamental papers of Morse,² particularly his *Calculus of Variations in the Large* [MM1], have found them difficult not only because of the intrinsic difficulties but for another reason. The work was done at a time when the requisite algebraic topology was not so adequately or systematically developed as at present. As a consequence, a substantial part of his exposition is concerned with proof of purely topological results in the special context of critical point theory. Thus part of his exposition deals with various aspects of the exactness of the homology sequence of a pair of spaces and part with the relation between deformations and the homomorphisms of homology theory.

It will be supposed that the reader knows a modest amount of homology theory, which may be summarized in the statement that the axioms of Eilenberg and Steenrod [E-S1] are theorems in singular homology theory.

In this paper an account is given of a specific problem in critical point theory, namely the problem of a smooth function on a Riemannian manifold. The form of statements is chosen in such fashion that they may be extended reasonably to a wider class of problems. Thus this is intended simultaneously as an exposition of a particular useful case and a model. Critical points are defined locally and are classified locally in the neighborhood of separated sets of such points. Theorem 7.3 states that the local classification is possible, Corollary 7.4 permits direct sum decomposition, and Theorem 10.2 details the computation for nondegenerate critical points. The end result of this

The first part of an address delivered before the New York Meeting of the Society, February 25, 1956, by invitation of the Committee to Select Hour Speakers for Eastern Sectional Meetings; received by the editors August 9, 1957.

¹ Some of the work behind this paper was carried out while the writer was a fellow of the John Simon Guggenheim Memorial Foundation. Some of the work was sponsored by the National Science Foundation.

² See [MM1] for references to the earlier papers.

portion of the exposition, which occupies the first eleven sections, is the inequalities of Morse in Theorem 11.1.

Certain desiderata not obtained in the Morse inequalities are described by example in §12, and the related lines of development are discussed briefly. In particular, the need for integer coefficients in the homology groups is noted. The development of strengthened inequalities is the subject of §§13 through 15. The method was suggested by an effort to reformulate for general coefficient groups the group theoretic formulation of the Morse inequalities, in [MM2] Theorem 8.7, which depends on the use of a field of coefficients.

Attention is given to degenerate critical points and to the class of functions without degenerate critical points. There are inequalities which give the lower bound of the number of critical points of a non-degenerate approximating function.

The approach of Deheuvels [D1] through spectral theory, which is of substantial interest, will not be discussed here. The reader who is interested in this point may wish to read the paper [DGB1] of Bourgin.

§§1 through 12 are an exposition of classical material, but §§13–15 contain material previously unpublished except in abstract [P1, 2].

1. The space and the function. Suppose X is a compact n -dimensional Riemannian manifold of class C^3 and $f: X \rightarrow R$, where R is the real numbers, is a function of class C^3 . It will be supposed that X is defined in terms of overlapping local coordinates, and $(x) = (x^1, \dots, x^n)$ will be used for a typical set. The element of arc will be supposed given in local coordinates by

$$ds^2 = g_{ij} dx^i dx^j$$

with the usual convention that i, j are summed from 1 to n and with the understanding that the g_{ij} are functions of x of class C^2 , symmetric in indices i and j , which transform in the covariant manner (to which it will be unnecessary to refer formally again, so that the explicit statement of the well known concept can be omitted).

The element of arc is used to define length of sufficiently smooth curves by integration. The greatest lower bound of lengths of curves joining two points is a metric on the space. It is the topology of this metric which is used throughout.

The gradient of f is the vector whose covariant components are (f_{x^i}) . A point P is a *critical point* if the gradient of f at P is (0) and is an *ordinary point* otherwise. The set of critical points is closed.

Pictorially, one may suppose that X is embedded in a euclidean space of suitable dimension m with f itself as the last coordinate y_m .

One can surely do this by embedding X in a euclidean space of dimension $m-1$ and then introducing f as the m th coordinate, though other embeddings may have the desired property also. Then the critical points of f are the points where the tangent plane to X is perpendicular to the y_m -axis. Easy examples for illustrative purposes are furnished by two dimensional manifolds embedded in 3-space.

Of the intuitively reasonable definitions of critical point of a smooth function, this is probably the most inclusive. In particular, there are points called critical which are trivial in theorems on existence and number of critical points. For easy example, let X be the circle and θ an angular coordinate on it and let f be a function of class C^3 taking on the value θ^3 for θ near 0. (The circle is used only because this is an exposition limited to compact manifolds.) Then the direct summand (Corollary 7.4) of the critical groups at level 0 (§7) associated with $\theta=0$ as a critical point is trivial.

The critical points are the points neighboring which there is no deformation of class C^1 such that along the trajectory of each point in the deformation, f has a negative derivative with respect to the parameter ("time") of the deformation.

The set of points P with $f(P) = c$ will be called the *points at the level* c . The words level and value will be used interchangeably as convenient. A level is *critical* if there is a critical point at the level and is *ordinary* otherwise.

The set of critical levels of f is closed, and of course bounded by the absolute maximum and absolute minimum of f , which are themselves critical levels.

The set σ of critical points at the level c will be called the *critical set at the level* c .

It is possible that the function f be not constant on a connected set of critical points. An example is given by Whitney in [W1]. A theorem of A. P. Morse in [APM1] states that this cannot occur if f is of class C^n , where n is the dimension of X . In preference to assumptions that X and f are highly differentiable, it will be assumed that the critical levels of f are isolated. In consequence, f is constant on connected critical sets.

The notations

$$f_a = \{x \mid f(x) \leq a\},$$

$$f_a^0 = \{x \mid f(x) < a\}$$

will be used.³

³ The notation f_a^- would be used by some in preference to f_a^0 but there is the notational difficulty that A^- is to be used for the closure of the set A .

2. **The deformations Δ and Γ .** The advantage to the particular problem under consideration as a model is the ease with which useful deformations⁴ are constructed. The first is described in the following lemma.

LEMMA 2.1. *The space X admits a deformation*

$$\Delta: X \times [0, 1] \rightarrow X$$

with the properties that

- (a) *The critical points of f are fixed under Δ .*
- (b) *The function f on the path of any ordinary point under the deformation Δ is a decreasing function.*

The deformation $\Delta(P, \cdot)$ is defined in local coordinates (x) in which P has the coordinates (x_0) as the solution of the system

$$\frac{dx^i}{dt} = -g^{ij}f_{x^j}$$

which satisfies the initial conditions

$$x^i(0) = x_0^i.$$

Here, as usual, g^{ij} is defined so that $g_{ik}g^{kj} = \delta_i^j$ (the Kronecker delta) and the quantities $(g^{ij}f_{x^j})$ are the contravariant components of the gradient vector. The definition of Δ in terms of local coordinates is readily seen to be invariant. Along a trajectory under Δ , one finds that

$$\frac{df}{dt} = -f_{x^i}g^{ij}f_{x^j} = -|\text{grad } f|^2$$

so that (a) and (b) of the lemma follow.

It should be remarked that the deformation Δ in fact establishes a homeomorphism of X with itself in which (x) and $\Delta(x, 1)$ correspond. However, this does not remain true in a larger class of problems for which the theorems of this paper might serve as model. Accordingly properties of Δ beyond those stated in Lemma 2.1 are not used except where specifically introduced in §8.

The deformation Γ is associated with a pair of levels, $a < b$, with the property that the interval $[a, b]$ is free of critical levels. It is described in the following lemma.

⁴ In §3, the vocabulary of deformations is reviewed in preparation for more complex statements about deformations which are to follow. It may be read before §2.

LEMMA 2.2. *If $[a, b]$ is free of critical levels, the subspace f_b^0 admits a deformation*

$$\Gamma: f_b^0 \times [0, 1] \rightarrow f_b^0$$

such that

- (a) *The points of f_a remain on f_a during the deformation.*
 (b) $\Gamma(x, 1) \in f_a$.

One does in fact prove more, namely that points of f_a are fixed and that along any trajectory f is a nonincreasing function, decreasing as long as the functional value exceeds a . Neither of these facts is used later, so they are not included in the lemma itself.

The deformation Γ is defined as follows. Through each point of $f_b^0 - f_a^0$, say (x_0) in local coordinates, there is the solution of the differential equation

$$\frac{dx^i}{dt} = -K g^{ij} f_{x^j} / g^{hk} f_{x^k} f_{x^i},$$

$$x^i(0) = x_0^i$$

where K is a constant yet to be specified. Along any solution

$$\frac{df}{dt} = -K.$$

Each point (x_0) of $f_b^0 - f_a^0$ is to be deformed along the solution initiating at that point with $K = f(x_0) - a$ and will be at the level a when $t = 1$. Points at level a remain fixed according to this definition and by further definition all points of f_a are to be fixed under Γ . The definition of Γ in terms of local coordinates is seen to be invariant.

3. Vocabulary of deformations. In this section some remarks on the vocabulary of deformations are recorded for precision and easy reference.

With

$$\begin{array}{c} A \supset C \supset E \\ \cup \quad \cup \quad \cup \\ B \supset D \supset F \end{array}$$

as subsets of a space, one says that δ deforms (C, D) over (A, B) into (E, F) if

$$\delta: (C \times [0, 1], D \times [0, 1]) \rightarrow (A, B)$$

is a continuous function for which $\delta|_{C \times 0}$ is the identity while

$$\delta(C \times 1, D \times 1) \subset (E, F).$$

Empty sets may be suppressed in the notation, so that one may speak of deforming C over A into E . The statement that (C, D) is deformed over (C, D) into (E, F) is shortened to the statement that (C, D) is deformed into (E, F) . The language is applied to collections of more than two sets in the obvious fashion. Thus the statement that (C, D) is deformed into (E, F) and that the trajectories of points of E and F lie on E and F respectively can be shortened to the statement that (C, D, E, F) is deformed into (E, F, E, F) .

With spaces $A \supset B$ and $C \supset D$, two maps

$$\phi_0, \phi_1: (A, B) \rightarrow (C, D)$$

are *homotopic* if there is a map

$$\phi: (A \times [0, 1], B \times [0, 1]) \rightarrow (C, D)$$

such that

$$\phi(P, s) = \phi_s(P) \quad s = 0, 1.$$

The two pairs (A, B) and (C, D) are then said to be of the *same homotopy type* if there are maps

$$\xi: (A, B) \rightarrow (C, D) \quad \eta: (C, D) \rightarrow (A, B)$$

for which $\eta\xi$ and $\xi\eta$ are each homotopic to identity maps. Then ξ and η are said to be *homotopy inverses*. The relation of belonging to the same homotopy type is transitive. The pair (A, B) is said to be of *trivial homotopy type* if it is of the same homotopy type as (B, B) .

If

$$\begin{array}{ccc} A & \supset & B \\ \cup & & \cup \\ C & \supset & D \end{array}$$

and if there is a deformation of (A, B, C, D) into (C, D, C, D) then (A, B) is of the same homotopy type as (C, D) .

4. Invariants of critical levels. Throughout this section it will be supposed that c is a level, which is either ordinary or critical, of special interest. Further it is supposed that $a' < c < b'$ where the levels on $[a', c)$ and $(c, b']$ are ordinary, and that $a' < a, a_1, a_2 < c < b, b_1, b_2 < b'$.

THEOREM 4.1. *The homotopy type of (f_b, f_a) is independent of (a, b) and the homotopy type is the same if f_a or f_b is replaced by f_a^0 and f_b^0 respectively. The homotopy type is trivial if c is ordinary.*

The use of the deformation Δ shows that (f_b^0, f_a) and (f_b, f_a^0) and (f_b^0, f_a^0) have the same homotopy type as (f_b, f_a) .

Now suppose c is ordinary. Then the deformation Δ , iterated sufficiently often, deforms (f_b, f_a) into (f_a, f_a) . For there is a constant $h > 0$ such that $\text{grad } f \geq h$ on $f_b - f_a^0$. Thus Δ deforms (f_b, f_a) onto (f_{b-h}, f_a) and need be repeated at most $1 + (b-a)/h$ times to deform (f_b, f_a) into (f_a, f_a) . This proves the last statement of the theorem.

Suppose

$$(4.1) \quad a_1 \leq a_2 \quad b_1 \leq b_2$$

then the inclusion map of (f_{b_1}, f_{a_1}) into (f_{b_2}, f_{a_2}) admits a homotopy inverse. For, as just shown, sufficient iteration of Δ deforms $(f_{b_2}, f_{a_2}, f_{b_1}, f_{a_1})$ into $(f_{b_1}, f_{a_1}, f_{b_1}, f_{a_1})$ since $[b_1, b_2]$ and $[a_1, a_2]$ are free of critical levels.

Now remove the restriction (4.1). One sets $a = \max(a_1, a_2)$ and $b = \max(b_1, b_2)$ and applies the result of the previous paragraph to (f_b, f_a) and (f_b, f_a) with $s = 1, 2$. Since homotopy equivalence is transitive, the proof is complete.

In the particular problem under discussion the topological type of (f_b, f_a) is independent of a, b , but this fact does not follow in more general problems for which one might wish to use this problem as a model.

The force of Theorem 4.1 is that invariants of homotopy type can be attached to critical levels. In the critical point theory of Morse, it is the relations arising from the use of relative homology groups which are sought.

The following special theorem is useful.

THEOREM 4.2. *If there is no critical level on $[a, c]$, then the homotopy type of (f_c^0, f_a) is trivial.*

This is established through the use of the deformation Γ of §2 associated with the levels a, c .

5. Remarks on the homology theory. The developments to be explained in the following paragraphs will be made in terms of singular homology theory. In the work of Morse a field was used for the coefficient group. In this exposition, the coefficient group G will be a principal ideal domain (see [J1, I, 77 and 121]). The cases of interest appear to be that G be a field or the integers. The integers are useful in the specific problem under discussion because they are a universal coefficient group. Their merits are brought out in §§12-15.

6. A deformation lemma. The following lemma is well known but no reference is at hand, so the proof is given.

LEMMA 6.1. *Suppose $W \supset Z \supset Y \supset A$ is a nest of spaces. Suppose there is a deformation δ of (W, Y, A) over (W, Z, A) into (Y, Y, A) . Then in the sequence of inclusion maps*

$$(Y, A) \xrightarrow{i} (Z, A) \xrightarrow{j} (W, A)$$

the induced map $j_ | i_* H_k(Y, A)$ is an isomorphism onto $H_k(W, A)$.*

For proof consider the diagram

$$\begin{array}{ccccccc} & & & & (Y, A) & & \\ & & & \nearrow \delta_1 & \downarrow i & & \\ (Z, A) & \xrightarrow{i} & (W, A) & \xrightarrow{id_1} & (Z, A) & \xrightarrow{j} & (W, A) \end{array}$$

in which $\delta_1(P) = \delta(P, 1)$. Consider the induced maps

$$\begin{aligned} (id_1)_* &: H_k(W, A) \rightarrow H_k(Z, A), \\ j_* | i_* H_k(Y, A) &: i_* H_k(Y, A) \rightarrow H_k(W, A). \end{aligned}$$

The map $j_* i_* \delta_{1*}$ is the identity map on $H_k(W, A)$ because $ji\delta_1$ is homotopic to the identity map. But

$$j_* i_* \delta_{1*} = (j_* | i_* H_k(Y, A))(id_1)_*$$

so that $j_* | i_* H_k(Y, A)$ is a homomorphism onto $H_k(W, A)$. Again, $i_* \delta_{1*} j_*$ is the identity map on $H_k(Z, A)$ because $id_1 j$ is homotopic to the identity. Thus, cut down in domain and range to $i_* H_k(Y, A)$, it is still an identity map. Thus $j_* | i_* H_k(Y, A)$ is an isomorphism into $H_k(W, A)$. The statement affirmed in the lemma follows.

7. General considerations about critical groups. We continue with the situation of §4 in which c is a level on which attention is focused and $[a, c)$ and $(c, b]$ are free from critical levels. The set of critical points at level c is denoted by σ .

Following Theorem 4.1, the relative homology groups $H_k(f_b, f_a; G)$ will be called the *critical groups at the level c* . The notation for the coefficient group G will usually be suppressed. For convenience in statement and proof specific representations or isomorphic copies of the critical groups will sometimes be used.

Critical groups are discussed in this section and the next. In this section the emphasis is on the conclusions which can be drawn from Lemmas 2.1 and 2.2 (the latter through its single application in

Theorem 4.2) and general topological considerations. The properties considered in this section are capable of extension to a much wider class of problems.

On the one hand, the critical groups will be shown to depend only on the function in an arbitrary neighborhood of the critical set. On the other hand, conditions will be considered under which the critical groups can be determined without the use of points at levels greater than or equal to c except for the points of σ .

Two lemmas are needed.

LEMMA 7.1. *The homomorphism on $H_k(f_b, f_a)$ to $H_k(f_b, f_c^0)$ induced by inclusion is an isomorphism onto the latter.*

In proof one notes that the homology sequence of the triple (f_b, f_c^0, f_a) , namely

$$\cdots \rightarrow H_k(f_c^0, f_a) \rightarrow H_k(f_b, f_a) \rightarrow H_k(f_b, f_c^0) \rightarrow H_{k-1}(f_c^0, f_a) \rightarrow \cdots,$$

is exact. But the pair f_c^0, f_a is homotopically trivial by virtue of Theorem 4.2 so that its homology groups vanish.

The force of Lemma 7.1 is to describe the injection under which the groups $H_k(f_b, f_c^0)$ may be taken to represent the critical groups at the level c .

LEMMA 7.2. *If U is an open set with $\sigma \subset U \subset f_b$ there is an open set V with $\sigma \subset V \subset U$ such that in the sequence of maps*

$$(f_c^0 \cup V, f_c^0) \xrightarrow{i} (f_c^0 \cup U, f_c^0) \xrightarrow{j} (f_b, f_c^0)$$

the induced map

$$j_* \mid i_* H_k(f_c^0 \cup V, f_c^0)$$

is an isomorphism onto $H_k(f_b, f_c^0)$.

Let V_1 denote a neighborhood of σ deformed over U by Δ . Let e with $c < e \leq b$ be such that f_e is deformed by Δ onto $f_c^0 \cup V_1$. That V_1 exists follows from the fact that points of σ are left fixed by Δ . That e exists is seen as follows. The set $f_c \cap V_1'$ is compact. Every point of $f_c \cap V_1'$ at level c is displaced by Δ so that on $f_c \cap V_1'$, $\max f(\Delta(\cdot, 1)) < c$. Thus $f_c \cap V_1'$ has a neighborhood S which is deformed onto f_c^0 . The union $S \cup V_1$ is an open set which contains f_c and therefore contains f_e if e is sufficiently near c .

Set

$$V = f_e^0 \cap V_1, \quad U_1 = f_e^0 \cap U$$

and consider the diagram

$$\begin{array}{ccccc}
 (f_c^0 \cup V, f_c^0) & \xrightarrow{i_1} & (f_c^0 \cup U_1, f_c^0) & \xrightarrow{j_1} & (f_b, f_c^0) \\
 & \searrow i & \downarrow h_1 & & \downarrow h_2 \\
 & & (f_c^0 \cup U, f_c^0) & \xrightarrow{j} & (f_b, f_c^0)
 \end{array}$$

in which all the maps are inclusions and which is commutative. First,

$$f_b \supset f_c^0 \cup U_1 \supset f_c^0 \cup V \supset f_c^0.$$

These sets with the deformation Δ (cut down to f_b) satisfy the hypotheses of Lemma 6.1. Consequently $j_{1*} | i_{1*} H_k(f_c^0 \cup V, f_c^0)$ is an isomorphism onto $H_k(f_b, f_c^0)$. Further h_{2*} is an isomorphism and $h_{2*} j_{1*} = j_* k_{1*}$ so that $j_* h_{1*} | i_{1*} H_k(f_c^0 \cup V, f_c^0)$ is an isomorphism onto $H_k(f_b, f_c^0)$. Thus

$$j_* | h_{1*} i_{1*} H_k(f_c^0 \cup V, f_c^0) = j_* | i_* H_k(f_c^0 \cup V, f_c^0)$$

is an isomorphism onto $H_k(f_b, f_c^0)$ as required.

A basic theorem is the following.

THEOREM 7.3. *If U is an open set with $\sigma \subset U \subset f_b$ there is an open set V with $\sigma \subset V \subset U$ such that in the sequence of inclusion maps*

$$(V, f_c^0 \cap V) \xrightarrow{i} (U, f_c^0 \cap U) \xrightarrow{j} (f_b, f_c^0)$$

the map $j_ | i_* H_k(V, V \cap f_c^0)$ is an isomorphism onto the representative $H_k(f_b, f_c^0)$ of the critical group in dimension k .*

For proof, let U_1 be an open set with $\sigma \subset U_1$ and $U_1^- \subset U$ and let V_1 with U_1 and the sequence of maps

$$(f_c^0 \cup V_1, f_c^0) \xrightarrow{i_1} (f_c^0 \cup U_1, f_c^0) \xrightarrow{j_1} (f_b, f_c^0)$$

satisfy Lemma 7.2. Then the excision of $f_c^0 \cap U'$ from the first two pairs of the sequence is admissible, so that in the sequence of inclusion maps

$$((f_c^0 \cap U) \cup V_1, f_c^0 \cap U) \xrightarrow{i_2} ((f_c^0 \cap U) \cup U_1, f_c^0 \cap U) \xrightarrow{j_2} (f_b, f_c^0)$$

the map $j_{2*} | i_{2*} H_k((f_c^0 \cap U) \cup V_1, f_c^0 \cap U)$ is an isomorphism onto $H_k(f_b, f_c^0)$. If one factors j_2 into the inclusion maps

$$((f_c^0 \cap U) \cup U_1, f_c^0 \cap U) \xrightarrow{i_3} (U, f_c^0 \cap U) \xrightarrow{j} (f_b, f_c^0)$$

and sets $V = (f_c^0 \cap U) \cup V_1$, whence $f_c^0 \cap U = f_c^0 \cap V$, one finds that

$j_*i_{3*}|i_{2*}H_k(V, f_c^0 \cap V)$ is an isomorphism onto $H_k(f_b, f_c^0)$. With the observation that $i_3i_2 = i$, the theorem follows readily.

The force of Theorem 7.3 is that the critical groups at the level c are determined by the function f cut down to any neighborhood of the critical set at the level σ .

If σ is the union of two separated subsets σ_1 and σ_2 , there is a neighborhood U of σ which is the union of two disjoint neighborhoods U_1 of σ_1 and U_2 of σ_2 . Associated with U in Theorem 7.2 is $V \subset U$ which is necessarily the union of neighborhoods V_1 and V_2 with $\sigma_1 \subset V_1 \subset U_1$ and $\sigma_2 \subset V_2 \subset U_2$. Let i_1 and i_2 denote the inclusion maps of V_1 and V_2 on U_1 and U_2 . Then the following corollary to Theorem 7.3 holds.

COROLLARY 7.4.

$$H_k(f_b, f_c) \approx i_{1*}H_k(V_1, V_1 \cap f_c^0) + i_{2*}H_k(V_2, V_2 \cap f_c^0).$$

The direct sum is established by the injections induced by the inclusion maps of U_1 and U_2 into U followed by the inclusion j of Theorem 7.3.

One now considers the exclusion from (f_b, f_c^0) of points at levels c and higher, except for the points of σ .

One considers the possibility that the inclusion map of (f_c, f_c^0) into (f_b, f_c^0) may induce isomorphism

$$H_k(f_c, f_c^0) \approx H_k(f_b, f_c^0).$$

Examination of the exact sequence of the triple (f_b, f_c, f_c^0) shows that this is the case if and only if $H_k(f_b, f_c) = 0$ for all k . It is clear in any event that the inclusion map of $(f_c^0 \cup \sigma, f_c)$ into (f_c, f_c^0) induces isomorphism of the corresponding homology groups, for Δ deforms $(f_c, f_c^0 \cup \sigma, f_c^0)$ into $(f_c^0 \cup \sigma, f_c^0 \cup \sigma, f_c^0)$.

The critical set σ will be called *regularly situated* if there is a neighborhood U of σ , say on f_b , such that $(U, U \cap f_c)$ admits a deformation A over (f_b, f_c) into (f_c, f_c) .

THEOREM 7.5. *If σ is regularly situated, the inclusion map of $(f_c^0 \cup \sigma, f_c^0)$ into (f_b, f_c^0) induces isomorphism between the homology groups of $(f_c^0 \cup \sigma, f_c^0)$ and the critical groups at level c .*

The proof that (f_b, f_c) can be deformed into (f_c, f_c) will not be given in detail. It can be accomplished through use of Δ and A .

8. Special considerations about critical groups. The gradient curves of f permit the definition of deformations with properties beyond those stated in Lemma 2.1. These will be exploited here to obtain a

localization of critical groups which is simpler than Theorem 7.1 and to obtain an instance of regularly situated critical sets. These theorems presumably do not admit such wide generalization as those of §7.

The notation

$$\Delta(S, t) = \{ \Delta(x, t) \mid x \in S \}$$

will be used.

Special properties of the deformation Δ beyond those stated in Lemma 2.1 are the following.

LEMMA 8.1. *The deformation Δ has the properties*

- (a) $\Delta(\cdot, 1)$ is a homeomorphism on X to itself.
- (b) $\Delta(\Delta(f_e, 1), t) \subset \Delta(f_e, 1)$ for all e .

Both properties follow from uniqueness properties of the solutions of ordinary differential equations. Property (a) is used only to imply $\Delta(\cdot, 1)$ is open.

The following "single neighborhood" localization of the critical groups is now possible. The symbols a, c, b, σ continue to be used as described at the beginning of §7.

THEOREM 8.2. *The critical set σ has arbitrarily small neighborhoods $W \subset f_b$ with the property that the inclusion map*

$$i: (W, W \cap f_e^0) \rightarrow (f_b, f_e^0)$$

induces maps i_ which are isomorphisms onto the representatives $H_k(f_b, f_e^0)$ of the critical groups of the level c .*

Let U denote a neighborhood of σ with $U \subset f_b$. The neighborhood W is to lie on U . Suppose U_1 is a neighborhood of σ with $U_1 \subset U$. Suppose as in the proof of Lemma 7.2 that $c < e < b$ and that f_e is deformed by Δ into $f_e^0 \cup U_1$. Let $S = \Delta(f_e^0, 1)$. It is open by virtue of Lemma 8.1. Under Δ , points of $(S, S \cap f_e^0)$ are deformed over $(S, S \cap f_e^0)$, again by virtue of Lemma 8.1. Thus the inclusion

$$i_1: (S, S \cap f_e^0) \rightarrow (f_e, f_e^0)$$

induces isomorphisms of the corresponding homology groups. Now the excision of $S \cap U'$ from $(S, S \cap f_e^0)$ is admissible. Let $W = S \cap U$ and map (f_e, f_e^0) by inclusion into (f_b, f_e^0) . Thus the inclusion

$$i: (W, W \cap f_e) \rightarrow (f_b, f_e)$$

induces isomorphism on the corresponding homology groups as required.

It is easy to prove Theorem 7.3 from Theorem 8.2 but the proof

will not serve as a model in similar situations where Theorem 7.3 is valid but presumably Theorem 8.2 is not.

Next the consideration of regularly situated critical sets is resumed.

LEMMA 8.3. *If the critical set σ consists of isolated points the pair (f_b, f_c) is homotopically trivial.*

For proof, a deformation B of (f_b, f_c) is defined as follows. Under B , points of f_c remain fixed. For $0 \leq t < 1$ the point with local coordinates (x_0) lying on $f_b - f_c^0$ is deformed into the point $B(x_0, t)$ at the point t on the solution of the system

$$\frac{dx^i}{dt} = - \frac{K g^{ij} f_{x^j}}{g^{hk} f_{x^k} f_{x^k}},$$

$$x^i(0) = x_0^i$$

with $K = f(x_0) - c$. Further, one defines

$$B(x_0, 1) = \lim_{t \rightarrow 1} B(x_0, t).$$

The limit can be shown to exist because σ is a finite set. The function B is seen to be continuous and to have the required properties.

From Lemma 8.3 and either Theorem 7.5 or direct observation, the following theorem is deduced.

THEOREM 8.4. *If the critical set σ consists of isolated points the inclusion map on $(f_c^0 \cup \sigma, f_c^0)$ to (f_b, f_c^0) induces isomorphism between the homology groups of $(f_c^0 \cup \sigma, f_c^0)$ and the critical groups at level c .*

By use of excision, the critical groups can be determined as the homology groups of the pair $((f_c^0 \cap U) \cup \sigma, f_c^0 \cap U)$ where U is a neighborhood of σ . Further, taking U as the union of separated neighborhoods of the points of σ shows that the critical groups can be written as a direct sum with a summand $H_k((f_c^0 \cap V) \cup P, f_c^0 \cap V)$ for each point P of σ , where V is a neighborhood of P .

A specific instance of a connected set of degenerate critical points which is regularly situated is considered by Bott in [RB1].

9. Local representation of the function. Neighboring a critical point of f , there is a simple form for f in terms of any local coordinates. For convenience the development is restricted to coordinates in which the critical point is the origin.

THEOREM 9.1. *If P is a critical point of f and (x) is a local coordinate system for which $P = (0)$ then there are functions $a_{ij}(x)$ of class C' such that*

$$f(x) - f(0) = a_{ij}(x)x^i x^j,$$

$$a_{ij}(x) = a_{ji}(x), \quad a_{ij}(0) = \frac{1}{2} f_{x^i x^j}(0).$$

The functions a_{ij} are determined by expanding f by Taylor's Theorem with integral form of the remainder in the terms of the second order. One finds

$$\begin{aligned} f(x) - f(0) &= \int_0^1 \int_0^s \frac{d^2}{dt^2} f(tx) dt ds \\ &= a_{ij}(x) x^i x^j \end{aligned}$$

where

$$\begin{aligned} a_{ij}(x) &= \int_0^1 \int_0^s f_{x^i x^j}(tx) dt ds \\ &= \int_0^1 (1-t) f_{x^i x^j}(tx) dt. \end{aligned}$$

The required properties of the $a_{ij}(x)$ follow.

10. The nondegenerate critical point. Suppose that P is a critical point with local coordinates (x_0) . Then P is called a *degenerate* critical point if the determinant $|f_{x^i x^j}(x_0)| = 0$ and a *nondegenerate* critical point otherwise. This is an invariant condition as may be verified by direct computation or from the following discussion which relates the condition to the general background.

Introduce temporarily the notation $f_i = f_{x^i}$. For any covariant vector (λ_i) , recall that the covariant derivative $(\lambda_{i,j})$, given by $\lambda_{i,j} = \lambda_{ij} - \lambda_h \Gamma_{ij}^h$, is a tensor which is covariant of order 2. Thus $(f_{i,j})$ is a covariant tensor of order 2. But $f_{i,j} = f_{x^i x^j}$ at a critical point, so that the covariance of $(f_{x^i x^j}(x_0))$ is established. The invariance of the vanishing of the determinant follows since under a change of variable the matrix $(f_{x^i x^j}(x_0))$ is transformed by congruence.

It is readily seen from use of the implicit function theorem that if (x_0) is a nondegenerate critical point, it is isolated.

The representation of f presented in Theorem 9.1 permits a simple representation of f neighboring a nondegenerate critical point.

THEOREM 10.1. *If P is a nondegenerate critical point there are local coordinates (y) , related to admissible local coordinates (x) on X by a transformation of class C' with nonvanishing Jacobian, such that*

$$f(y) - f(0) = \sum_1^n \epsilon_i y_i^2 \quad \epsilon = \pm 1, \epsilon_i \leq \epsilon_{i+1}.$$

One does not expect the coordinates (y) to belong to the admissible coordinates of the space X . However they facilitate a computation.

For proof, one uses the Lagrange reduction of a quadratic form with constant coefficients to diagonal form as a model. The method is sketched briefly. Two kinds of transformations are used. The first kind is used if at least one of the numbers $a_{ii}(0) \neq 0$. One may suppose it is $a_{11}(0)$, for this may be achieved by renaming variables. Then one sets

$$\begin{aligned} z^1 &= a_{1j}(x)x^j / |a_{11}(x)|^{1/2}, \\ z^i &= x^i, \quad i \geq 2. \end{aligned}$$

Thus

$$f(x) - f(0) = \eta(z^1)^2 + Q(z^2, \dots, z^n), \quad \eta = \text{sgn } a_{11}(0)$$

where Q is a quadratic function in z^2, \dots, z^n with coefficients which are functions of z^1, \dots, z^n . The second kind of transformation is used if all $a_{ii}(0) = 0$. There is a number $a_{ij}(0) \neq 0$ since the determinant $|a_{ij}(0)| \neq 0$. Suppose the fact which may be obtained by renaming variables that $a_{12}(0) \neq 0$. One then makes the preparatory change of variable

$$\begin{aligned} x^1 &= u^1 - u^2, \\ x^2 &= u^1 + u^2, \\ x^i &= u^i \quad i \geq 2, \end{aligned}$$

to write $f(x) - f(0)$ as a quadratic function in u^1, \dots, u^n , with coefficients which are functions of (u) , such that the coefficient of $(u^1)^2$ is not 0 at $(u) = (0)$. Then the first kind of transformation is again available.

By a succession of n transformations of the first kind, with renaming of variables and the intervention of the transformation of the second kind when necessary, $f(x) - f(0)$ is written as a signed sum of squares. A final renaming of variables brings all minus signs together in an initial block so that the form required in the theorem is obtained.

One verifies that the renaming of variables and the transformations of the first and second kind are all of class C' with nonvanishing Jacobian.

The number of $\epsilon_i < 0$ in the representation of f in Theorem 10.1

is called the *index* of the critical point. The linear approximation to the transformation from variables (x) to variables (y) in Theorem 10.1 changes $a_{ij}(0)x^i x^j$ into $\sum_1^n \epsilon_i y_i^2$, so that the index defined in this way is seen to be the same as the classical index of the quadratic form $a_{ij}(0)x^i x^j$ which approximates $f(x) - f(0)$. In particular it is seen to be independent of the choice of transformation.

Theorem 8.4 and Theorem 10.1 permit the computation of the critical groups of a nondegenerate critical point (or of the direct summand contributed by a nondegenerate critical point).

THEOREM 10.2. *If P is a critical set consisting of a single nondegenerate critical point, the groups $H_k(f_c^0 \cup P, f_c^0)$ are trivial except in the dimension q which is the index of P ; $H_q(f_c^0 \cup P, f_c^0) = G$.*

One computes $H_q((f_c^0 \cap U) \cup P, f_c^0 \cap U)$ where U is a neighborhood of P on which the representation of f of Theorem 10.1 holds. One further restricts U for convenience to consist of the points (y) for which $\sum_1^n (y^i)^2 < h$ for suitably small positive h . Two subsets of U are distinguished, the set D where $\sum_1^q (y^i)^2 < h$ and $y_{q+1} = \dots = y_n = 0$, and the subset D_0 where $0 < \sum_1^q (y^i)^2$. Then the collection $((f_c^0 \cap U) \cup P, f_c^0 \cap U, D, D_0)$ is deformed into the collection (D, D_0, D, D_0) (with points of (D, D_0) actually fixed during the deformation) as follows. The image of y^1, \dots, y^n at time $t, 0 \leq t \leq 1$, is $y^1, \dots, y^q, (1-t)y^{q+1}, \dots, (1-t)y^n$. Thus the inclusion map on (D, D_0) to $((f_c^0 \cap U) \cup P, f_c^0 \cap U)$ induces isomorphism of the corresponding homology groups, whence the conclusion of the theorem follows.

11. The inequalities of Morse. The inequalities of Morse, and similar inequalities which come from the use of coefficient groups which are not fields, can be obtained from the following observations about a module over a principal ideal domain G .

Associated with a free module F over G (i.e. a direct sum of copies of G) is the cardinal number of generators in an independent basis for F , called the rank of F and denoted by $R[F]$. A module H over G is a group with operators from G (a useful example being an abelian group H with the usual operation by integers). The rank $R[H]$ is defined as the maximum of numbers $R[F]$ over free modules $F \subset H$.

The usefulness of rank depends on the fact that if θ is a homomorphism of H onto L with kernel K , as expressed in the exact sequence

$$0 \rightarrow K \rightarrow H \xrightarrow{\theta} L \rightarrow 0,$$

then

$$R[H] = R[L] + R[K].$$

It follows that if

$$\cdots \rightarrow H_i \xrightarrow{\theta_i} H_{i-1} \rightarrow \cdots$$

is an exact sequence of finitely generated modules H_i over G , with image-kernel K_i in H_i , then

$$R[H_i] = R[K_i] + R[K_{i-1}].$$

Thus

$$R[H_{i+1}] - R[H_i] + R[H_{i-1}] = R[K_{i+1}] + R[K_{i-2}].$$

Now suppose the critical levels c_i of f are isolated and let a_i be convenient constants which separate them, so that

$$a_0 < c_1 < a_1 < c_2 < \cdots < a_{i-1} < c_i < a_i < \cdots < a_{N-1} < c_N < a_N.$$

To avoid the repeated use of double subscripts, let $f_{a_i} = A_i$. In particular $\emptyset = A_0 \subset A_1 \subset \cdots \subset A_N = X$. The homomorphism sequence of the pair (A_i, A_{i-1}) is then

$$\begin{aligned} \cdots \rightarrow H_k(A_{i-1}) \xrightarrow{\gamma_*} H_k(A_i) \xrightarrow{\alpha_*} H_k(A_i, A_{i-1}) \xrightarrow{\beta_*} H_{k-1}(A_{i-1}) \rightarrow \cdots \\ \cdots \rightarrow H_0(A_{i-1}) \rightarrow H_0(A_i) \rightarrow H_0(A_i, A_{i-1}) \rightarrow 0 \end{aligned}$$

where γ_* and α_* are induced by inclusion maps and β_* is the boundary homomorphism. The notation for the coefficient group G has been suppressed. The sequence is an exact sequence of finitely generated modules⁵ over G . The groups $H_k(A_i, A_{i-1})$ are the critical groups of the level c_i .

The following notations are introduced.

$$\begin{aligned} R_k^i &= R[H_k(A_i)], \\ M_k^i &= R[H_k(A_i, A_{i-1})], \\ b_k^i &= R[\beta_* H_{k+1}(A_i, A_{i-1})], \\ R_k &= R_k^N = R[H_k(X)], \\ M_k &= \sum_{i=1}^N b_k^i, \\ b_k &= \sum_{i=1}^N b_k^i. \end{aligned}$$

⁵ In fact the pair (A_i, A_{i-1}) can be triangulated in this particular problem. The technique introduced in Lemma 13.1 does not require that the groups be finitely generated.

The inequalities of Morse are expressed in the following theorem.

THEOREM 11.1. *The inequalities*

$$M_k \geq R_k \qquad k = 0, 1, \dots, n$$

and the inequalities

$$\begin{aligned} M_0 &\geq R_0, \\ M_1 - M_0 &\geq R_1 - R_0, \\ &\dots, \\ M_k - M_{k-1} + \dots + (-1)^k M_0 &\geq R_k - R_{k-1} + \dots + (-1)^k R_0, \\ &\dots, \\ M_n - M_{n-1} + \dots + (-1)^n M_0 &= R_n - R_{n-1} + \dots + (-1)^n R_0 \end{aligned}$$

are valid.

The proof consists of observing that the exact homology sequence on the pair (A_i, A_{i-1}) implies that

$$R_h^{i-1} - R_h^i + M_h^i = b_h^i + b_h^{i-1}$$

whence, summing on i ,

$$M_h - R_h = b_h + b_{h-1}.$$

This yields the first set of inequalities. Further, the signed sum on h between 0 and k is

$$M_k - M_{k-1} + \dots + (-1)^k M_0 = R_k - R_{k-1} + \dots + (-1)^k R_0 + b_k$$

which yields the second set of inequalities. The equality for the case $k=n$ can be obtained from the inequalities for $k=n, n+1$ and the fact that $M_{n+1}=R_{n+1}=0$.

Observe that the right hand member of the equality is the Euler-Poincaré characteristic of X , multiplied by $(-1)^n$. Observe also that the second set of inequalities implies the first, but not conversely.

The force of the inequalities of Morse lies in the fact that, with a fixed group G , the right hand members depend only on the space X , being combinations of Betti numbers with reference to the coefficient group G , while the left hand members depend only on the function f in arbitrary neighborhoods of the critical sets of f . If all critical points of f are nondegenerate, it follows from Theorems 8.4 and 10.2 that M_k is the cardinal number of critical points of f of index k . If f admits degenerate critical points, Morse has shown, with G a field (see [MM1, Chapter VI] for the case of analytic functions and integers modulo 2 as the coefficients) that any function g with nondegenerate

critical points which approximates f sufficiently closely has at least M_k^i critical points of index k near the critical set σ_i at level c_i , and no other critical points. He thus counts a critical set σ_i as the ideal equivalent of M_k^i nondegenerate critical points of index k . A more precise statement of a stronger theorem will be made in §15.

12. Examples. This section is devoted to examples illustrating what the inequalities of Morse do *not* do.

EXAMPLE 1. The number $\sum_0^N R_k$ does not provide a lower bound to the cardinal number of critical points of f . For let X be the torus, represented by a rectangle with matched opposite edges in which one diagonal has been drawn. Then f can be defined by class C^3 so that it is 0 on the edges and the diagonal and nowhere else, is positive and has an absolute nondegenerate maximum interior to one triangle, is negative and has a proper nondegenerate minimum interior to the other triangle, has a monkey saddle at the point represented by the four vertices, and no other critical points. Thus f has 3 critical points although $R_0 + R_1 + R_2 = 4$. The Morse inequalities do imply that every function on the torus whose critical points are nondegenerate has at least 4 critical points.

This example was taken from Lusternik and Schnirelmann [L-S], where an introduction to their use of *category* to supply a lower bound to the cardinal number of critical points can be found.

EXAMPLE 2. There exist manifolds X which admit no function f with nondegenerate critical points for which

$$M_k = R_k, \quad k = 0, 1, \dots, n.$$

The existence of such examples shows the possibility of stronger inequalities for functions with nondegenerate critical points.⁶ The requirement that the critical points of f be nondegenerate is essential, for the equality is satisfied for the constant functions on any manifold, where there is one critical level and the entire space X is the critical set.

Examples are of two kinds, depending on whether the difficulty is approached through the fundamental group or through torsion.

First, let X be a homology 3-sphere with d -fold universal covering by the 3-sphere S^3 , where $1 < d < \infty$. The Poincaré "sphere," whose description is readily available in [S-T1, Chapter IX], is such a

⁶ This is a different statement from that of Morse (see [M1, 145]) that his inequalities are the only ones which always hold between the numbers M_k and R_k alone. His statement means that any set of numbers M_k and R_k satisfying his inequalities can be realized by a region Σ and a function f on it satisfying appropriate boundary conditions. In our statement above, the space X is specified in advance.

space. Let $f: X \rightarrow R$ be a function on X with nondegenerate critical points. With any admissible coefficient domain $R_0 = 1 = R_3$ and $R_1 = 0 = R_2$; it follows from Theorem 11.1 that $M_0 \geq 1$, $M_1 \geq 0$, $M_2 \geq 0$, $M_3 \geq 1$. The purpose of this example is to show that $M_1 = 0$ (or $M_2 = 0$) is impossible. To that end, let $\phi: S^3 \rightarrow X$ denote the covering map and set $\tilde{f} = f\phi$. Let \tilde{R}_k and \tilde{M}_k refer to Betti numbers and numbers of critical points of S^3, \tilde{f} . Since ϕ is locally a homeomorphism, each inverse of a critical point of f is a critical point of \tilde{f} of the same index, and $\tilde{M}_k = dM_k$. Further $\tilde{R}_0 = 1 = \tilde{R}_3$ and $\tilde{R}_1 = 0 = \tilde{R}_2$. Thus

$$dM_0 \geq 1, \quad dM_1 - dM_0 \geq -1.$$

But $dM_0 \geq d$ so that $dM_1 \geq d - 1 > 0$. Thus $M_1 \geq 1$. Similar argument shows $M_2 \geq 1$ also.

The critical point theory of a function on a space is in fact the critical point theory of the special class of functions on the universal covering space which are invariant under the action of the fundamental group on the universal covering. The writer has developed such a theory, as yet unpublished.

The second kind of difficulty which prevents the realization of equalities $M_k = R_k$ with a function whose critical points are nondegenerate will be described in more general terms. If X has torsion and Z_p denotes the integers modulo p , then there are dimensions k and primes p and coefficient groups G for which $R_k(Z_p) > R_k(G)$. The group notation refers to the coefficient group used in the computation. But M_k is independent of the coefficient group, so that $M_k > R_k(G)$. One might suppose one could deal systematically with this problem through the use of integer coefficients followed by the universal coefficient theorem with appropriate primes p . In fact, the difficulty which is raised by torsion associated with different primes in different dimensions can be overcome for nondegenerate critical points by that approach (see §14). But the similar issue relative to degenerate critical sets is more difficult, because there is also torsion associated with different primes for different critical levels in the same dimension. A direct attack on stronger inequalities in the theory based on a principal ideal domain of coefficients, with integers as the primary example, is desired and will be considered in the sections which follow.

13. Skeleta. The difficulty with counting critical points and, in an idealized fashion, counting critical sets, lies in the fact that the chain groups are too large. This is overcome by the introduction of skeleta, which in application will be taken in the spirit of algebraic

minimal subcomplexes sufficient to describe the homology. Their definition and fundamental properties are undertaken here.

The context of the definition and construction is a Mayer chain complex with associated subcomplexes and quotient complexes. See [K-P, §§1-3 and references to the papers of Mayer] or [E-S, Chapter V]. A *Mayer chain complex* $\{C, \partial\}$ is a system of abelian groups (the chain groups) $C_r, r = -1, 0, 1, \dots$, with $C_{-1} = 0$, and a sequence of homomorphisms (the boundary operation) $\partial_r: C_r \rightarrow C_{r-1}$ of order 2, that is, such that $\partial_{r-1}\partial_r = 0$. A *subcomplex* $\{C', \partial'\} \subset \{C, \partial\}$ is a chain complex $\{C', \partial'\}$ for which $C'_r \subset C_r$ and $\partial'_r = \partial_r|_{C'_r} |_{C'_r}$. The system $\{C, C', \partial\}$ is then a *chain pair*. The *quotient complex* $\{C/C', \partial/C'\}$ has the chain groups C_r/C'_r and the boundary operator ∂/C' defined by $(\partial/C')_r(x) = \partial_r x + C_{r-1}$. Homology groups $H_r(C)$, chain maps (which commute with boundary), induced maps, and the exact homology sequence of a pair are defined as usual.

The groups C_r in $\{C, \partial\}$ will be assumed to be modules over a coefficient group G which is a principal ideal domain. Operation by elements of G commutes with the boundary operator. A principal case is the case that G is the integers. Another case of interest is the case that G is a field. The complex $\{C, \partial\}$ is *free* if the groups C_r are free modules over G .

A *skeleton* of a chain complex $\{C, \partial\}$ is defined as a subcomplex $\{C', \partial'\}$ such that the inclusion maps $i_r: C'_r \rightarrow C_r$ induce homomorphisms $i_{r*}: H_r(C') \rightarrow H_r(C)$ which are isomorphisms onto $H_r(C)$ for all r . It is free if the subcomplex is free. Where more convenient, the skeleton will be regarded as mapped isomorphically into the chain complex rather than as a subcomplex of the chain complex.

The basic lemma states precisely how to construct a skeleton of a complex from a skeleton of a subcomplex and a free skeleton of the quotient complex; the chain groups are direct sums in the respective dimensions and the boundary operation is consistent with those of the subcomplex and quotient complex. The form chosen for the lemma is just strong enough for the use to be made of it here, since the proof is thereby made substantially easier. The requirements about free groups can be removed at the expense of introducing appropriate group extensions, and the chain groups overlying the homology groups can be removed at the expense of constructing appropriate substitutes. See [P1, 2].

LEMMA 13.1. *Suppose*

¹ If $\phi: A \rightarrow C, B \subset A, \phi B \subset D \subset C$, the function $\psi: B \rightarrow D$ defined by $\psi(x) = \phi(x)$ is denoted by $\phi|B||D$. The shortened notation $\phi|B = \phi|B||C$ will be used as required.

$$\{K, \partial^K\} \subset \underset{k}{\{C', \partial'\}} \subset \{C, \partial\}.$$

Suppose $\{K, \partial^K\}$ is a skeleton for $\{C', \partial'\}$, that is, that the homomorphisms k_{r*} are isomorphisms. Suppose $\{Q, \partial^Q\}$ is a free skeleton of $\{C/C', \partial/C'\}$, that is, that chain maps $q_r: Q_r \rightarrow C_r/C_r'$ on free modules Q_r induce isomorphisms q_{r*} . Let $L_r = K_r + Q_r$, and let $n_r: C_r \rightarrow C_r/C_r'$ and $p_r: L_r \rightarrow Q_r$ denote the natural homomorphisms. Then there are homomorphisms $h_r: L_r \rightarrow C_r$ for which

- (i) $h_r|K_r = k_r$,
- (ii) $n_r h_r = q_r p_r$,
- (iii) h_r is an isomorphism into C_r ,
- (iv) $\partial_r^L = h_{r-1}^{-1} \partial_r h_r$ is defined and is of order 2,
- (v) h_{r*} is an isomorphism onto $H_r(C)$.

The complex $\{L, \partial^L\}$ under h is a skeleton of $\{C, \partial\}$.

The proof is in several steps.

(a) If $u_r: Q_r \rightarrow C_r$, defined for each r , is any set of homomorphisms for which $n_r u_r = q_r$, then $\partial_r u_r x - u_{r-1} \partial_r^Q x \in C_{r-1}'$.

Suppose $x \in Q_r$. Then $q_r x = n_r u_r x = u_r x + C_r'$ and $q_{r-1} \partial_r^Q x = n_{r-1} u_{r-1} \partial_r^Q x + C_{r-1}'$. Further, $q_{r-1} \partial_r^Q x = (\partial/C')_r q_r x = (\partial/C')_r n_r u_r x = \partial_r u_r x + C_{r-1}'$. Thus $\partial_r u_r x + C_{r-1}' = u_{r-1} \partial_r^Q x + C_{r-1}'$ in C_{r-1}/C_{r-1}' , whence (a) follows.

(b) There is a homomorphism $v_r: Q_r \rightarrow C_r$ for each r such that $n_r v_r = q_r$ and $\partial_r v_r x - v_{r-1} \partial_r^Q x \in K_{r-1}$.

First, there is a set of homomorphisms $u_r: Q_r \rightarrow C_r$ satisfying the relation $n_r u_r = q_r$ because Q_r is free and n_r is a homomorphism onto C_r/C_r' for each r . The set v is constructed inductively on r by modification of such a set u . The homomorphism v_r is trivial if $r = -1$ (if one makes all chain groups trivial in dimension -2 so that relations involving boundary are trivially satisfied, then no other special considerations are necessary in the first step of the induction). To proceed with the induction, suppose that $v_r, r < s$, and $u_r, r \geq s$, form a set of homomorphisms on Q_r to C_r such that $n_r w_r = q_r$ when $w_r = v_r$ or u_r and such that $\partial_r v_r x - v_{r-1} \partial_r^Q x \in K_{r-1}$ if $r < s$. Of course (a) applies to w_r for all r .

The homomorphism v_s is constructed for the elements of an independent basis of Q_s . Let x be such an element and let $t_1 = \partial_s u_s x - v_{s-1} \partial_s^Q x$. Then $\partial_{s-1} t_1 = -\partial_{s-1} v_{s-1} \partial_s^Q x = -v_{s-2} \partial_{s-1}^Q \partial_s^Q x + t_2 = t_2$ where t_2 by virtue of the inductive hypothesis is a chain (necessarily a cycle) of K_{s-2} and a boundary of C_{s-2}' . Then $t_2 = \partial_{s-1} y_1$ where y_1 is a chain of K_{s-1} because k_{s-2*} is an isomorphism. Thus $t_1 - y_1$ is a cycle of C_{s-1}' and, because k_{s-1*} is an isomorphism, there are a cycle y_2 of K_{s-1}

and a chain z of C_s' such that $t_1 - y_1 = y_2 + \partial_s z$. Finally, $\partial_s(u_s x - z) - v_{s-1} \partial_s^Q x = y_2 + y_1$. On setting $v_s x = u_s x - z$ one finds $n_s v_s x = n_s u_s x = q_s x$ and $\partial_s v_s x - v_{s-1} \partial_s^Q x \in K_{s-1}$. Now v_s is extended linearly over Q_s . This completes the inductive step in the proof of (b).

(c) *The homomorphisms v_r are isomorphisms into C_r .*

Since $n_r v_r = q_r$ is an isomorphism it follows that v_r is.

(d) *The homomorphism h_r defined by $h_r(y, x) = y + v_r x$ has the properties (i)–(iv).*

First, $h_r(y, 0) = y = k_r y$ so that (i) holds. Second, $n_r h_r(y, x) = n_r(y + v_r x) = n_r v_r x = q_r x = q_r p_r(y, x)$ so that (ii) holds. Third, if $h_r(y, x) = 0$ then $n_r h_r x = 0$ so that $x = 0$, and $h_r(y, 0) = k_r y = 0$ so that $y = 0$. Thus (iii) is verified. Fourth, $\partial_r h_r(y, x) = \partial_r y + \partial_r v_r x = \partial_r y + v_{r-1} \partial_r^Q x + z = h_{r-1}(\partial_{r-1} y + z, \partial_r x)$, where $z \in K_{r-1}$. Since h_r is an isomorphism, $h_{r-1}^{-1} \partial_r h_r = \partial_r^L$ is now well defined on L_r to L_{r-1} . Further $\partial_{r-1}^L \partial_r^L(y, x) = h_{r-2}^{-1} \partial_{r-1} \partial_r(y, x) = h_{r-2}^{-1} 0 = 0$, so that ∂_r^L is of order 2 and $h_{r-1} \partial_r^L = \partial_r^L h_r$. Thus (iv) holds and (d) is proved.

(e) *The homomorphism has the property (v).*

Statements (iii) and (iv) show that the homomorphisms h_r form a set of chain maps of the pair $\{L, K, \partial^L\}$ to the pair $\{C, C', \partial\}$ which are isomorphisms into the groups C_r . Accordingly the diagram

$$\begin{array}{ccccccccc}
 \cdots & \rightarrow & H_{r+1}(Q) & \xrightarrow{\partial_{r+1}^L} & H_r(K) & \longrightarrow & H_r(L) & \xrightarrow{p_r^*} & H_r(Q) & \longrightarrow & H_{r-1}(K) & \longrightarrow & \cdots \\
 & & \downarrow q_{r^*} & & \downarrow k_{r^*} & & \downarrow h_{r^*} & & \downarrow q_{r^*} & & \downarrow k_{r^*} & & \\
 \cdots & \rightarrow & H_{r+1}(C/C') & \xrightarrow{\partial_{r+1}^L} & H_r(C') & \longrightarrow & H_r(C) & \xrightarrow{n_r^*} & H_r(C/C') & \longrightarrow & H_{r-1}(C') & \longrightarrow & \cdots
 \end{array}$$

in which lines are the exact sequences of homology on the chain pair $\{L, K, \partial^L\}$ and the chain pair $\{C, C', \partial\}$, is a commutative diagram. The groups Q_r have been identified with the factor groups L_r/K_r , and the maps previously used in the proof have been labeled. Since q_{r^*} and k_{r^*} are isomorphisms onto their respective images for all r , it follows from the five-lemma (see [E-S, p. 16]) that h_{r^*} is an isomorphism onto $H_r(C)$. Thus (e) is proved.

Since h_r is an isomorphism into C_r and h_{r^*} an isomorphism onto $H_r(C)$, the proof that $\{L, \partial^L\}$ is a skeleton of $\{C, \partial\}$ is complete.

The existence of certain especially useful skeleta is shown as follows.

LEMMA 13.2. *Suppose the free chain complex $\{C, \partial\}$ has homology groups $H_k(C)$, $k=0, 1, \dots$, such that $H_k(C)$ has rank ρ_k and η_k torsion coefficients. Then $\{C, \partial\}$ has a skeleton $\{L, \partial^L\}$ for which the chain group L_k has rank $\rho_k + \eta_k + \eta_{k-1}$.*

The group $H_k(C)$ is the internal direct sum of a free module S_k

with rank ρ_k and the submodule T_k of elements of finite order. (See [J1, II, Chapter III, §9].) The latter can be described by an exact sequence

$$0 \rightarrow F_k \xrightarrow{\lambda_k} F_k \xrightarrow{\mu_k} T_k \rightarrow 0$$

in which F_k is a free module on η_k generators. The number η_k is the smallest number of cyclic submodules of which T_k is a direct sum. It is easily shown that the module F_k is the smallest which can be so used to describe T_k in the sense that if $\nu: A \rightarrow T_k$ is a homomorphism of a free module onto T_k and μ_k is factored through ν according to the diagram

$$\begin{array}{ccc} & A & \\ \pi \nearrow & & \searrow \nu \\ F_k & \xrightarrow{\mu_k} & T_k \end{array}$$

so that $\mu_k = \nu\pi$, then π is an isomorphism into A . This will be cited as the minimal property of (F_k, λ_k, μ_k) . Let

$$L_k = S_k + F_k + F_{k-1} \quad (\text{external direct sum})$$

and define

$$\partial_k^L: L_k \rightarrow L_{k-1}$$

by

$$\partial_k^L(x, y, z) = (0, \lambda_{k-1}z, 0).$$

Then $\partial_{k-1}^L \partial_k^L = 0$ so that $\{L, \partial^L\}$ is a free chain complex for which L_k has rank $\rho_k + \eta_k + \eta_{k-1}$. The group $S_k + F_k$ is the group of k -cycles and the group $\lambda_k F_k$ is the group of k -boundaries.

To make $\{L, \partial^L\}$ a skeleton for $\{C, \partial\}$, the imbedding must be defined. The group S_k is imbedded in the group $Z_k(C)$ of cycles of C by a homomorphism θ_k' factoring the identity homomorphism of S_k through the natural homomorphism ϕ_k on $Z_k(C)$ to $H_k(C)$ as indicated in the diagram

$$S_k \xrightarrow{\theta_k'} Z_k(C) \xrightarrow{\phi_k} S_k + T_k = H_k(C).$$

This is possible since ϕ_k is a homomorphism onto $H_k(C)$, and θ_k' is an isomorphism into $Z_k(C)$ since $\phi_k \theta_k'$ is the identity on S_k . The group F_k is imbedded in $Z_k(C)$ by a homomorphism θ_k'' factoring μ_k through ϕ_k so that $\phi_k \theta_k'' = \mu_k$, as indicated in the diagram

$$F_k \xrightarrow{\theta_k''} Z_k(C) \xrightarrow{\phi_k} S_k + T_k = H_k(C).$$

Again, this is possible because ϕ_k is a map onto $H_k(C)$, and θ_k'' is an isomorphism into $Z_k(C)$ because of the minimal property of (F_k, λ_k, μ_k) . Further, the direct sum map

$$\theta_k' + \theta_k'': S_k + F_k \rightarrow Z_k(C)$$

is an isomorphism into $Z_k(C)$ because $\phi_k \theta_k' S_k$ and $\phi_k \theta_k'' F_k$ lie in complementary direct summands of $H_k(C)$.

The group F_{k-1} is imbedded in C_k by a homomorphism θ_k''' which factors $\theta_k'' \partial_k^L | F_{k-1}$ through the homomorphism ∂_k . In order to do this, one must know that $\theta_k'' \partial_k^L F_{k-1} \subset \partial C_k = \phi_{k-1}^{-1}(0)$. But $\theta_k'' \partial_k^L F_{k-1} = \theta_k'' \lambda_{k-1} F_{k-1}$ and $\phi_{k-1} \theta_k'' \lambda_{k-1} F_{k-1} = \mu_{k-1} \lambda_{k-1} F_{k-1} = 0$ as required. Thus θ_k''' is well defined and $\partial_k \theta_k''' = \theta_k'' \partial_k^L | F_{k-1} = \theta_k'' \lambda_{k-1}$. Since λ_{k-1} and θ_k'' are isomorphisms into the respective ranges, so is θ_k''' . Further, the direct sum $\theta_k = \theta_k' + \theta_k'' + \theta_k'''$ is an isomorphism into C_k , since $\theta_k' + \theta_k''$ is an isomorphism into C_k and $\partial_k \theta_k | (S_k + F_k) = 0$ while $\partial_k \theta_k | F_{k-1} = \partial_k \theta_k'''$ is an isomorphism into C_{k-1} .

The imbedding of L_k in C_k with the property $\partial_k \theta_k = \theta_{k-1} \partial_k^L$ has been completed. It remains to show θ_{k*} is an isomorphism onto $H_k(C)$. To that end consider the commutative diagram

$$\begin{array}{ccc} Z_k(C) & \xrightarrow{\phi_k} & H_k(C) = S_k + T_k \\ \uparrow \theta_k' + \theta_k'' & & \uparrow \theta_{k*} \\ S_k + F_k = Z_k(L) & \xrightarrow{\psi_k} & H_k(L) \end{array}$$

in which ψ_k is the natural homomorphism with kernel $\partial_{k+1}^L L_{k+1} = \lambda_k F_k$. Since $\phi_k(\theta_k' + \theta_k'')$ is onto $H_k(C)$, so is θ_{k*} . Since the kernel of $\phi_k(\theta_k' + \theta_k'')$ is $\lambda_k F_k$ and $\psi_k \lambda_k F_k = 0$, it follows that θ_{k*} is an isomorphism onto $H_k(C)$, as required to complete the proof of the lemma.

14. Strengthened inequalities of critical point theory. The situation of §11, with the same notation, will be considered again.

The singular chain complex $\{C(A_i), \partial\}$ with groups $C_k(A_i)$ and boundary operator $\partial_k | C_k(A_i) | \partial C_{k-1}(A_i)$, shortened to ∂ , is free and $C_k(A_{i-1})$ is a direct summand of $C_k(A_i)$. Consequently, the relative homology groups $H_k(A_i, A_{i-1})$ are the homology groups of a free complex. Let R_k^i and η_k^i denote the rank and number of torsion coefficients of $H_k(A_i, A_{i-1})$. Then the quotient complex, with chain groups $C_k(A_i)/C_k(A_{i-1})$, admits a skeleton with a basis of $\overline{M}_k^i = R_k^i + \eta_k^i + \eta_k^{i-1}$ generators according to Lemma 13.2. Set

$$\overline{M}_k = \sum_{i=1}^N \overline{M}_k^i.$$

Then Lemma 13.1 can be applied for $i=1, 2, \dots, N$ to yield the following theorem.

THEOREM 14.1. *The singular chain complex $\{C(X), \partial\}$ of the space X admits a skeleton $\{L, \partial^L\}$ in which the groups L_k have rank \overline{M}_k .*

A sharper statement could have been developed. The idea of a skeleton of a chain complex could be enlarged to the idea of a skeleton of a filtered chain complex, and in those terms, the statement made that $\{L, \partial^L\}$ is a skeleton of $\{C(X), \partial\}$ filtered by the function f . (See [P2].)

The strengthened inequalities relate the numbers \overline{M}_k , the rank R_k of $H_k(X)$ and the number η_k of torsion coefficients of $H_k(X)$ (recall that $\eta_0=0=\eta_n$) as follows.

THEOREM 14.2. *The inequalities*

$$\begin{aligned} \overline{M}_0 &\geq R_0, \\ \overline{M}_1 &\geq R_1 + \eta_1, \\ \overline{M}_2 &\geq R_2 + \eta_2 + \eta_1, \\ &\dots, \\ \overline{M}_k &\geq R_k + \eta_k + \eta_{k-1}, \\ &\dots, \\ \overline{M}_n &\geq R_n + \eta_{n-1} \end{aligned}$$

and the inequalities

$$\begin{aligned} \overline{M}_0 &\geq R_0, \\ \overline{M}_1 - \overline{M}_0 &\geq R_1 - R_0 + \eta_1, \\ &\dots, \\ \overline{M}_k - \overline{M}_{k-1} + \dots + (-1)^k \overline{M}_0 &\geq R_k - R_{k-1} + \dots + (-1)^k R_0 + \eta_k, \\ &\dots, \\ \overline{M}_n - \overline{M}_{n-1} + \dots + (-1)^n \overline{M}_0 &= R_n - R_{n-1} + \dots + (-1)^n R_0, \end{aligned}$$

are valid.

The boundary operator on L can be described by an incidence matrix between basis elements in dimensions $k+1$ and k . (See [S-T1, Chapter III], and [J1, II, Ch. III, §§9, 10].) Let p_k denote its rank. Then

$$\overline{M}_h = R_h + p_h + p_{h-1}.$$

Since $p_h \geq \eta_h$, the first set of inequalities follows. Taking signed sums on h for $0 \leq h \leq k$ and using $p_k \geq \eta_k$, the second set of inequalities follows. The equality in the last case follows from the inequalities for $h = n$ and $n + 1$.

COROLLARY 14.3. *If all critical points of f are nondegenerate, the inequalities of Theorem 14.2 hold with \overline{M}_k computed as the cardinal number of critical points of f of index k .*

Forrester and the writer, working together, produced a proof of the strengthened inequalities based on the universal coefficient theorem, when the added assumption is made that all the critical points are nondegenerate. In outline it is as follows. With p a prime as yet unspecified, and with Z and Z_p denoting integers and integers modulo p , the universal coefficient theorem states that

$$H_h(X; Z_p) = H_h(X; Z) \otimes Z_p + H_{h-1}(X; Z) * Z_p,$$

where the notation specifies coefficient groups and \otimes and $*$ are tensor and torsion products. (See [E-S, Chapter V, particularly p. 161].) Suppose p divides η_k' of the η_k torsion coefficients of $H_k(X; Z)$. Then it follows that

$$R_h(Z_p) = R_h(Z) + \eta_h' + \eta_{h-1}'$$

where the group notation again refers to the coefficient group used. Then from Theorem 11.1

$$M_k - M_{k-1} + \cdots + (-1)^k M_0 \geq R_k - R_{k-1} + \cdots + (-1)^k R_0 + \eta_k',$$

$$R_h = R_h(Z).$$

Letting p be a prime dividing the smallest torsion coefficient in dimension k makes $\eta_k' = \eta_k$, leading to the second set of inequalities in Theorem 14.2 for integer coefficients under the additional assumption. This in turn, implies the first set.

This proof, it must be emphasized, does not prove Theorem 14.2 as it applies to functions with degenerate critical points. In using properties of a specific function to learn about the underlying space, one may be faced inescapably with degenerate critical points.

15. Idealized critical sets and lower semi-continuity of \overline{M}_k . The strengthened inequalities suggest that a critical set for which the critical groups have rank r_k and ζ_k torsion coefficients be regarded as equivalent to $m_k = r_k + \zeta_k + \zeta_{k-1}$ nondegenerate critical points of index k . This point of view will be justified as follows.

With a fixed manifold X of class C^3 associate the admissible class of functions of class C^3 , each with only a finite number of critical levels. Introduce two metrics into the admissible class, defined as

$$d_0(f, g) = \max |f(P) - g(P)|, \quad P \in X$$

$$d_1(f, g) = d_0(f, g) + \max |\text{grad}(f(P) - g(P))|.$$

Suppose σ is the critical set of an admissible function f at level c , with r_k and ζ_k as the rank and number of torsion coefficients of the critical group and $m_k = r_k + \zeta_k + \zeta_{k-1}$. If U is any neighborhood of σ and g is any admissible function which is sufficiently near f in the metric d_1 and has only nondegenerate critical points, then it will be shown that g has at least m_k critical points of index k in U . This is the required justification. Whether the statement remains true with d_1 replaced by d_0 is not clear to the writer.

It will further be shown that in the metric d_0 (and hence also in the metric d_1), the numbers \overline{M}_k of §14 are lower semi-continuous functions on the class of admissible functions.

Suppose first that

$$\alpha = a_0 < c_0 < a_1 < \dots < a_{i-1} < c_i < a_i < \dots < a_{N-1} < c_N < a_N = \beta$$

where α and β are ordinary levels of f and c_i is the only critical level of f on $[a_{i-1}, a_i]$. The notation $f_{a_i} = A_i$ will be continued. The exact homology sequences on the triples (A_i, A_{i-1}, A_0) admit analysis similar to that applied in §14. Let $\eta_k^i(f)$ and $\eta_k(f; \alpha, \beta)$ denote the number of torsion coefficients of $H_k(A_i, A_{i-1})$ and $H_k(f_\beta, f_\alpha)$ respectively and let

$$R_k(f; \alpha, \beta) = R[H_k(f_\beta, f_\alpha)],$$

$$\overline{M}_k^i(f) = R[H_k(A_i, A_{i-1})] + \eta_k^i(f) + \eta_k^{i-1}(f),$$

$$\overline{M}_k(f; \alpha, \beta) = \sum_{i=0}^N M_k^i(f).$$

Then the following theorem holds.

THEOREM 15.1. *The inequalities of Theorem 14.2 are valid when $\overline{M}_h, \overline{R}_h$ and η_h are replaced by $\overline{M}_h(f; \alpha, \beta), R_h(f; \alpha, \beta), \eta_h(f; \alpha, \beta)$.*

Suppose now that $a < c < b$ with c the only critical level of f on $[a, b]$ and let σ denote the critical set at level c .

LEMMA 15.2. *Corresponding to an open set U with $\sigma \subset U \subset f_b$, there is a number $\epsilon > 0$ such that if $d_1(f, g) < \epsilon$, every critical point of g on $f_b \cap f_a^{0'}$ is on U .*

This follows from the fact that $|\text{grad } f|$ is bounded from 0 on $f_b \cap f_a^{\alpha'} \cap U'$.

LEMMA 15.3. *If $d_0(f, g) < h = \min(b-c, c-a)/4$ and $a+h < \alpha' < c-2h$, $c+2h < \beta' < b-h$ then $f_a^{\alpha'} \subset g_{\alpha'}^{\alpha'} \subset g_{\beta'} \subset f_b$. The numbers $\overline{M}_k(g; \alpha', \beta')$ satisfy the inequalities $\overline{M}_k(g; \alpha', \beta') \geq \overline{M}_k(f; a, b)$. If all critical points of g are nondegenerate, and α' and β' are ordinary levels of g , then the cardinal number of critical points of g of index k at levels between α' and β' is at least $\overline{M}_k(f; a, b)$.*

In fact, with $d=c-h$ and $e=c+h$ the inclusions $f_a \subset g_{\alpha'} \subset f_a$ and $f_e \subset g_{\beta'} \subset f_b$ are clear. Then the sequence of inclusion maps

$$(f_e, f_a) \rightarrow (g_{\beta'}, g_{\alpha'}) \rightarrow (f_b, f_a)$$

induces homomorphisms

$$H_k(f_e, f_a) \rightarrow H_k(g_{\beta'}, g_{\alpha'}) \rightarrow H_k(f_b, f_a)$$

for which the composite homomorphism is the identity according to Theorem 4.1. Thus the critical group of σ in dimension k is a direct summand of $H_k(g_{\beta'}, g_{\alpha'})$. The weaker inequalities of Theorem 15.1 apply to the function g on the interval $[\alpha', \beta']$, whence the truth of the lemma.

The first statement in Lemma 15.3 implies the following theorem, in which the notation \overline{M}_k is that of section 14.

THEOREM 15.4. *The numbers \overline{M}_k are lower semi-continuous functions on the class of admissible functions in the metric d_0 .*

Since $d_0(f, g) \leq d_1(f, g)$, Lemmas 15.2 and 15.3 can be combined to yield the following theorem. It is supposed that $\alpha < f(x) < \beta$ on X .

THEOREM 15.5. *Corresponding to open sets U_i of the critical sets σ_i of f , there is a number $\epsilon > 0$ such that if $d_1(f, g) < \epsilon$ and all the critical points of g are nondegenerate, then g has at least $\overline{M}_k^i(f)$ critical points of index k in U_i ; and no critical points outside the neighborhoods U_i .*

REFERENCES

- RB1 R. Bott, *Nondegenerate critical manifolds*, Ann. of Math. vol. 60 (1954) pp. 248-261.
 DGB1 D. G. Bourgin, *Arrays of compact pairs*, Annali di Matematica Pura ed Applicata, Ser. IV vol. XL (1955) pp. 211-221.
 D1 R. Deheuvels, *Topologie d'une fonctionnelle*, Ann. of Math. vol. 61 (1955) pp. 13-72.
 E-S1 S. Eilenberg and N. Steenrod, *Foundations of algebraic topology*, Princeton Mathematical Series, vol. 15, Princeton, 1952.

- J1 N. Jacobson, *Lectures in abstract algebra*, I and II, New York, van Nostrand, 1951 and 1953.
- K-P1 J. L. Kelley and E. Pitcher, *Exact homomorphism sequences in homology theory*, Ann. of Math. vol. 48 (1947) pp. 682–709.
- L-S1 L. Lusternik and L. Schnirelmann, *Methodes topologiques dans les problemes variationnels*, Actualités Scientifiques et Industrielles, no. 188, Paris, Hermann, 1934.
- APM1 A. P. Morse, *The behavior of a function on its critical set*, Ann. of Math. vol. 40 (1939) pp. 62–70.
- MM1 M. Morse, *Calculus of variations in the large*, Amer. Math. Soc. Colloquium Publications, vol. 18, 1934
- MM2 ———, *Functional topology and abstract variational theory*, Ann. of Math. vol. 38 (1937) pp. 386–449.
- P1 E. Pitcher, *A class of group extensions*, Bull. Amer. Math. Soc. Abstract 59-6-585.
- P2 ———, *Chain groups and critical point inequalities*, Abstract, 59-6-714.
- S-T H. Seifert and W. Threlfall, *Variationsrechnung im Grossen*, Leipzig and Berlin, Teubner, 1938.
- W1 H. Whitney, *A function not constant on a connected critical set*, Duke Math. J. vol. 1 (1935) pp. 514–517.

LEHIGH UNIVERSITY