Topology Vol. 2, pp. 299–340. Pergamon Press, 1963. Printed in Great Britain

MORSE THEORY ON HILBERT MANIFOLDS

RICHARD S. PALAIS

(*Received* 22 *July* 1963)

THE TERM 'MORSE THEORY' is usually understood to apply to two analagous but quite distinct bodies of mathematical theorems. On the one hand one considers a smooth, real valued function f on a compact manifold M, defines $M_a = f^{-1}[-\infty, a]$, and given a closed interval $[a, b]$ describes the homology, homotopy, homeomorphism or diffeomorphism type of the pair (M_b, M_a) in terms of the critical point structure of f in $f^{-1}[a, b]$. On the other hand one takes a compact Riemannian manifold V, defines M to be the 'loop space' of piecewise smooth curves joining two points (with a natural topology) and $f: M \to \mathbf{R}$ the length function and again describes the homology and homotopy type of (M_b, M_a) in terms of the critical point structure of f in $f^{-1}[a, b]$ (i.e. in terms of the geodesics joining the two points whose lengths lie between a and b). The classical approach to this second Morse theory is to reduce it to the first Morse theory by approximating M_a (up to homotopy type) by certain compact submanifolds of piecewise broken geodesics. A particularly elegant and lucid exposition of this classical approach can be found in John Milnor's recent *Annals Study* [5].

Our goal in the present paper is to present a Morse theory for differentiable real valued functions on Hilbert manifolds. This encompasses both forms of Morse theory mentioned above in a unified way. In addition the generalization of the Morse theory of geodesics to higher loop spaces (i.e. maps of an n-disk into a manifold with fixed boundary conditions) and even more general situations works smoothly in this framework, whereas previous attempts at such generalizations were thwarted by the lack of a good analogue of the approximating compact manifolds of piecewise broken geodesics.

We have endeavored to make the exposition relatively self contained. Thus the first two sections give a brief resumé of the classical theory of Fréchet on the differential calculus of maps between Banach spaces (details and proofs will be found in [1]) and in sections 3 to 9 we give a brief treatment of the theory of Banach manifolds with particular emphasis on Hilbert manifolds (details and proofs will be found in [4]).

In sections 10 through 12 we prove the

MAIN THEOREM OF MORSE THEORY

Let M be a complete Riemannian manifold of class C^{k+2} $(k \geq 1)$ and $f: M \to \mathbf{R}$ a C^{k+2}-function. Assume that all the critical points of f are non-degenerate and in addition

(C) If S is any subset of M on which f is bounded but on which $\| \nabla f \|$ is not bounded away from zero then there is a critical point of f adherent to S. Then

(a) The critical values of f are isolated and there are only a finite number of critical points of f on any critical level;

(b) If there are no critical values of f in $[a, b]$ then M_b is diffeomorphic to M_a;

(c) If $a < c < b$ and c is the only critical value of f in $[a, b]$ and $p_1, \ldots, p_r$ are the critical points of f on the level c, then M_b is diffeomorphic to M_a with r handles of type $(k_1, l_1), \ldots, (k_r, l_r)$ disjointly C^k-attached, where k_i and l_i are respectively the index and co-index of p_i.

It should be noted that if f is proper, i.e. if $f^{-1}[a, b]$ is compact for every closed interval $[a, b]$, then condition (C) is automatically satisfied, hence the Morse theory for compact manifolds is included in the above theorem. On the other hand if M is infinite dimensional, hence not locally compact, then it is impossible for a real-valued function $f: M \to \mathbf{R}$ to be proper whereas we shall see condition (C) is still satisfied in cases of significant interest.

A theorem similar to the above was obtained independently and essentially simultaneously by S. Smale.

In §13 we show how to interpret the loop space of a complete finite dimensional Riemannian manifold V as a complete infinite dimensional Riemannian manifold M. This is due to Eells [2], however we have followed an approach suggested by Smale. In §14 we show that if we take $f: M \to \mathbf{R}$ to be the 'action integral' then the hypotheses of the Main Theorem are satisfied, thereby deriving the Morse theory of geodesics. In §15 we return to the abstract Morse theory of functions satisfying condition (C) on a Riemannian manifold and in particular derive the Morse inequalities. Finally in §16 we comment briefly on generalizing the Morse Theory of geodesics to higher loop spaces, a subject we hope to treat in detail in a later paper.

§1. DIFFERENTIABILITY

Let V and W be Banach spaces, $\mathcal{O}$ an open set in V and $f: \mathcal{O} \to W$ a function. If $p \in \mathcal{O}$ we say that f is differentiable at p if there exists a bounded linear transformation $T: V \to W$ such that $\| f(p + x) - f(p) - Tx \| / \|x\| \to 0$ as $x \to 0$. It is easily seen that T is uniquely determined and it is called the differential of f at p, denoted by df_p. The following facts are elementary [1, Ch. VIII]:

If f is differentiable at p then f is continuous at p;

If f is differentiable at p and $U \subseteq \mathcal{O}$ is a neighborhood of p then $g = f | U$ is differentiable at p and $dg_p = df_p$;

If f is constant then it is differentiable at p and $df_p = 0$;

If $S: V \to W$ is a bounded linear transformation and $f = S|\mathcal{O}$ then f is differentiable at p and $df_p = S$;

If f is differentiable at p, $g: \mathcal{O} \to W$ is differentiable at p and α, β are real numbers then $(\alpha f + \beta g)$ is differentiable at p and $d(\alpha f + \beta g)_p = \alpha df_p + \beta dg_p$;

If U is a neighborhood of $f(p)$ and $g : U \to Z$ is differentiable at $f(p)$ then if f is differentiable at p, $g \circ f$ is differentiable at p and $d(g \circ f)_p = dg_{f(p)} \circ df_p$.

Now suppose f is differentiable at each point of $\mathcal{O}$. Then $df : p \to df_p$ is a function from $\mathcal{O}$ into the Banach space $L(V, W)$ of bounded linear transformations of V into W (sup. norm). If df is continuous then we say that f is of class C^1 in $\mathcal{O}$. If df is differentiable at a point $p \in \mathcal{O}$ then $d(df)_p = d^2f_p \in L(V, L(V, W))$. We make the usual canonical identification of $L(V, L(V, W))$ with $L^2(V, W)$, the space of continuous bilinear maps of $V \times V$ into W. Thus d^2f_p is interpreted as a bilinear map of $V \times V$ into W and it can easily be shown to be symmetric [1, p. 175]. In case d^2f_p exists at each $p \in \mathcal{O}$ and the map $d^2f : p \to d^2f_p$ is continuous we say that f is of class C^2 in $\mathcal{O}$. Inductively suppose $d^kf : \mathcal{O} \to L^k(V, W)$ exists and is differentiable at p. Then $(d^{k+1}f)_p = d(d^kf)_p \in L(V, L^k(V, W)) \simeq L^{k+1}(V, W)$ and $d^{k+1}f_p$, the $(k + 1)$st differential of f at p, is a bounded, symmetric [1, p. 176] $(k + 1)$-linear map of $V \times \ldots \times V$ into W. If $d^{k+1}f_p$ exists at each point $p \in \mathcal{O}$ and $d^{k+1}f : \mathcal{O} \to L^{k+1}(V, W)$ is continuous then we say that f is of class C^{k+1} in $\mathcal{O}$. If f is of class C^k in $\mathcal{O}$ for every positive integer k we say that f is of class C^∞ in $\mathcal{O}$.

A linear map of $\mathbf{R}$ into a Banach space W is completely determined by its value on the basis element 1. We use this fact to define the derivative of a differentiable function $f : \mathcal{O} \to W$ where $\mathcal{O} \subseteq \mathbf{R}$; namely the derivative of f at p, denoted by $f'(p)$, is defined by $f'(p) = df_p(1)$, so $df_p(\alpha) = \alpha f'(p)$. If f is differentiable at each point of $\mathcal{O}$ we have $f' : \mathcal{O} \to W$, and if f is of class C^2 in $\mathcal{O}$ we can define $f'' = (f')'$ and in general if f is of class C^k we can define $f^{(k)} : \mathcal{O} \to W$. Clearly the relation of $f^{(k)}$ and d^kf is $d^kf_p(x_1, \ldots, x_k) = x_1x_2 \ldots x_k f^{(k)}(p)$. If $g : W \to Z$ is differentiable then $(g \circ f)'_{(p)} = d(g \circ f)_p(1) = dg_{f(p)}(df_p(1)) = dg_{(fp)}(f'(p))$, i.e. $(g \circ f)' = dg_f \circ f'$.

For future reference we note the following. If B is a continuous symmetric bilinear map of $V \times V$ into W then $f : V \to W$ defined by $f(v) = B(v, v)$ is of class C^∞. In fact $df_p(v) = 2B(p, v)$, $d^2f = 2B$ and $d^3f = 0$.

§2. THREE BASIC THEOREMS

Mean Value Theorem

Let V and W be Banach spaces, $p \in V$, $\mathcal{O}$ a convex neighborhood of p in V and $f : \mathcal{O} \to W$ a C^k-map, $k \geq 1$. Then there is a C^{k-1}-map $R_1 : \mathcal{O} \to L(V, W)$ such that if $x = p + v \in \mathcal{O}$

$$f(x) = f(p) + R_1(x)v.$$

COROLLARY (TAYLOR'S THEOREM). *If $m \leq k$ there is a C^{k-m}-map $R_m : \mathcal{O} \to L^m(V, W)$ such that if $x = p + v \in \mathcal{O}$ then*

$$f(x) = f(p) + df_p(v) + \tfrac{1}{2}\, d^2f_p(v, v) + \ldots + \frac{1}{(m-1)!}\, d^{m-1}f_p(v, \ldots v) + R_m(x)(v, \ldots v).$$

Inverse Function Theorem

Let V and W be Banach spaces, $\mathcal{O}$ open in V, and $f : \mathcal{O} \to W$ a C^k-map, $k \geq 1$. Let $p \in \mathcal{O}$ and suppose that df_p maps V one-to-one onto W. Then there is a neighborhood U of

p included in $\mathcal{O}$ such that $f|U$ is a one-to-one map of U onto a neighborhood of $f(p)$ and moreover $(f|U)^{-1}$ is a C^k-map of $f(U)$ onto U.

DEFINITION. *If $\mathcal{O}$ is an open set in a Banach space V then a C^k-vector field in $\mathcal{O}$ is a C^k-map $X : \mathcal{O} \to V$. A solution curve X is a C^1-map σ of an open interval $(a, b) \subseteq \mathbf{R}$ into $\mathcal{O}$ such that $\sigma' = X \circ \sigma$. If $0 \in (a, b)$ we call $\sigma(0)$ the initial condition of the solution σ.*

The following theorem is usually referred to as the local existence and uniqueness theorem for ordinary differential equations (or vector fields).

THEOREM. *Let X be a C^k-vector field on an open set $\mathcal{O}$ in a Banach space V, $k \geq 1$. Given $p_0 \in \mathcal{O}$ there is a neighborhood U of p_0 included in $\mathcal{O}$, an $\varepsilon > 0$, and a C^k-map $\varphi : U \times (-\varepsilon, \varepsilon) \to V$ such that:*

(1) *If $p \in U$ then the map $\sigma_p : (-\varepsilon, \varepsilon) \to V$ defined by $\sigma_p(t) = \varphi(p, t)$ is a solution of X with initial condition p;*

(2) *If $\sigma : (a, b) \to V$ is a solution curve of X with initial condition $p \in U$ then $\sigma(t) = \sigma_p(t)$ for $|t| < \varepsilon$.*

The proofs of these three basic theorems can all be found in [1] or in [4].

§3. DIFFERENTIABLE MANIFOLDS WITH BOUNDARY

If k is a bounded linear functional on a Banach space V, $k \neq 0$ we call $H = \{v \in V | k(v) \geq 0\}$ the (positive) half space determined by k, and $\partial H = \{v \in V | k(v) = 0\}$ is called the boundary of H. A function f mapping an open set $\mathcal{O}$ of H into a Banach space W is said to be of class C^k at a point $p \in \mathcal{O} \cap \partial H$ if there exists a C^k-map $g : U \to W$, where U is a neighborhood of p in V, such that $f|\mathcal{O} \cap U = g|\mathcal{O} \cap U$. It is easily seen that $d^m f_p = d^m g_p$ is then well defined for $m \leq k$ and that if f is of class C^k at each point of $\mathcal{O} \cap \partial H$ and also in $\mathcal{O} - \partial H$ then $d^m f : \mathcal{O} \to L^m(V, W)$ is continuous for $m \leq k$; in this case we say that f is of class C^k in $\mathcal{O}$. Next suppose that f is a one-to-one C^k-map of $\mathcal{O}$ into an open half space K in W. We say that f is a *C^k-isomorphism of $\mathcal{O}$ into K* if $f(\mathcal{O})$ is open in K and if f^{-1} is of class C^k (if $k \geq 1$ then it follows from the inverse function theorem that this will be so if and only if df_p maps V one-to-one onto W for each $p \in \mathcal{O}$).

Invariance of Boundary Theorem

Let H_i be a half space in a Banach space V_i and $\mathcal{O}_i$ an open set in H_i $(i = 1, 2)$. Let $f : \mathcal{O}_1 \to \mathcal{O}_2$ be a C^k-isomorphism. Then if either $k \geq 1$ or $\dim V_i < \infty$ f maps $\mathcal{O}_1 \cap \partial H_1$ C^k-isomorphically onto $\mathcal{O}_2 \cap \partial H_2$.

Proof. In case $k \geq 1$ the result is an immediate consequence of the inverse function theorem. In case $\dim V_i < \infty$ the theorem follows from invariance of domain.

Caution. In case $k = 0$ and $\dim V_i = \infty$ the conclusion of the theorem may well fail. Thus if V is an infinite dimensional Hilbert space and H a half space in V it is a (non-trivial) theorem that H and $H - \partial H$ are homeomorphic.

In the following, to avoid logical difficulties, we shall fix some set S of Banach spaces and whenever we say Banach space we will mean one which belongs to the set S.

A chart in a topological space X is a homeomorphism of an open set $D(\varphi)$ of X onto either an open set in a Banach space or else onto an open set in a half space of a Banach space. Two charts φ and ψ in X, with $U = D(\varphi) \cap D(\psi)$, are called C^k related if $\psi \circ \varphi^{-1}$ is a C^k-isomorphism of $\varphi(U)$ onto $\psi(U)$. A C^k-atlas for X is a set A of pairwise C^k-related charts for X whose domains cover X, and A is called complete if it is not included in any properly larger C^k-atlas for X. It is an easy lemma that if each of two charts φ, ψ in X is C^k-related to every chart in A then φ and ψ are C^k-related. It follows that there is a unique complete C^k-atlas $\tilde{A}$ including A, namely the set of all charts φ in X such that φ is C^k-related to every chart in A. $\tilde{A}$ is called the completion of A.

A C^k-manifold with boundary is a pair (X, A) where X is a paracompact Hausdorff space and A is a complete C^k-atlas for X. In general we will use a single symbol, such as M, to denote both a C^k-manifold (X, A) and its underlying topological space X, and elements of A will be referred to as charts for M. If $p \in M$ a chart at p is a chart for M having p in its domain. If A is a (not necessarily complete) C^k-atlas for X then by the C^k-manifold determined by A we mean the pair $M = (X, \tilde{A})$. If $m < k$ then A is a C^m-atlas for X and so determines a C^m-manifold which we also denote by M (an abuse of notation), so that a C^k-manifold is regarded as a C^m manifold if $m \leq k$.

If M is a C^k-manifold, $k \geq 1$, we define ∂M to be the set of $p \in M$ such that there exists a chart φ at p mapping $D(\varphi)$ onto an open set in a half space H so that $\varphi(p) \in \partial H$. It follows from the invariance of boundary theorem that every chart at p has this property and also that $\{\varphi|\partial M\}$, where φ runs over the charts for M, is a C^k-atlas for ∂M, so ∂M is a C^k-manifold. Moreover we have the obvious, but satisfying relation $\partial(\partial M) = \phi$.

If M and N are C^k-manifolds a function $f: M \to N$ is said to be of class C^k near p if there exists a chart φ at p and a chart ψ at $f(p)$ such that $\psi \circ f \circ \varphi^{-1}$ is of class C^k, and f is said to be of class C^k if the latter holds for each $p \in M$. It is easily seen that $f: M \to N$ is of class C^k if and only if $\psi \circ f \circ \varphi^{-1}$ is C^k for every chart φ for M and ψ for N.

If we define objects to be C^k-manifolds and morphisms to be C^k-maps then the axioms for a category are satisfied.

§4. TANGENT SPACES AND DIFFERENTIALS OF MAPS

Let $\{V_i\}_{i \in I}$ be an indexed collection of Banach spaces and for each $(i, j) \in I^2$ let φ_{ij} be an isomorphism of V_j with V_i (as topological vector spaces) such that $\varphi_{ii} =$ identity and $\varphi_{ij}\varphi_{jk} = \varphi_{ik}$. From the data $\{V_i, \varphi_{ij}\}$ we construct a new Banach space V (by a process we shall call amalgamation) and a canonical isomorphism $\pi_i : V \to V_i$ such that $\pi_i = \varphi_{ij}\pi_j$. Namely V is the set of $\{v_i\}$ in the Cartesian product of the V_i such that $v_i = \varphi_{ij}v_j$. Clearly V is a subspace of the full Cartesian product, hence a topological vector space. We define π_i to be the restriction of the natural projection of the Cartesian product onto V_i. To prove that V is a Banach space and π_i an isomorphism it suffices to note that there is an obvious continuous, linear, two sided inverse λ_j to π_j, namely $\lambda_j(v)_i = \varphi_{ij}(v)$.

Given a second set of data $\{W_k, \psi_{kl}\}$ satisfying the same conditions (with indexing set K) suppose that for each $(i, k) \in I \times K$ we have a bounded linear transformation $T_{ki} : V_i \to W_k$ such that $\psi_{kl} T_{li} \varphi_{ij} = T_{kj}$. Then if W is the amalgamation of $\{W_k, \psi_{kl}\}$ there is a uniquely determined bounded linear map $T : V \to W$ such that $\pi_k T = T_{ki} \pi_i$, namely if $\{v_i\} \in V$ then $T\{v_i\} = \{w_k\}$ where $w_k = T_{ki} v_i$. T is called the amalgamation of the T_{ki}.

Now suppose M is a C^k-manifold $k \geq 1$, $p \in M$ and let I be the set of charts at p. Given $\varphi \in I$ let V_φ be the target of φ (i.e. the Banach space into which φ maps $D(\varphi)$). Then for each $(\varphi, \psi) \in I^2$ $d(\psi \circ \varphi^{-1})_{\varphi(p)}$ is an isomorphism of V_φ onto V_ψ. Clearly $d(\varphi \circ \varphi^{-1})_{\varphi(p)}$ $=$ identity and by the chain rule $d(\lambda \circ \varphi^{-1})_{\varphi(p)} \circ d(\varphi \circ \psi^{-1})_{\psi(p)} = d(\lambda \circ \psi^{-1})_{\psi(p)}$. Hence the conditions for an amalgamation are satisfied. The resulting amalgamation is called the tangent space to M at p and denoted by M_p.

Let N be a second C^k-manifold $f : M \to N$ a C^k-map and K the set of charts at $f(p)$. For each $(\varphi, \psi) \in I \times K$ we have a linear map $d(\psi \circ f \circ \varphi^{-1})_{\varphi(p)}$ of V_φ into W_ψ. Moreover the abstract condition for amalgamating is clearly satisfied, hence we have a well determined amalgamated map $df_p : M_p \to N_{f(p)}$ called the differential of f at p.

§5. THE TANGENT BUNDLE

Let $\pi : E \to B$ be a C^k map of C^k-manifolds and suppose for each $b \in B$ $\pi^{-1}(b) = F_b$ has the structure of a Banach space. We call the triple (E, B, π) a *C^k-Banach space bundle* if for each $b_0 \in B$ there is an open neighborhood U of b_0 in B and a C^k-isomorphism $f : U \times F_{b_0} \approx \pi^{-1}(U)$ such that $v \to f(b, v)$ is a linear isomorphism of F_{b_0} onto F_b for each $b \in U$. If (E', B', π') is a second C^k-Banach space bundle then a C^k-map $f : E \to E'$ is called a C^k-bundle map if for each $b \in B$ f maps F_b linearly into a fiber $F'_{\tilde{f}(b)}$. The map $\tilde{f} : B \to B'$ is then C^k and is called the map induced by f.

Let M be a C^{k+1}-manifold with boundary. Let $T(M) = \bigcup_{p \in M} M_p$ and define $\pi : T(M) \to M$ by $\pi(M_p) = p$. Given a chart φ for M with domain U and target V_φ define $\varphi : U \times V_\varphi \to \pi^{-1}(U)$ by letting $v \to \varphi(p, v)$ be the natural isomorphism of V_φ with M_p. Then it is a straightforward exercise to show that the set of such φ is a C^k-atlas for a C^k-manifold with underlying set $T(M)$ and moreover that $T(M)$ is a C^k-Banach space bundle over M with projection π. If f is a C^{k+1}-map of M into a second C^{k+1}-manifold N we define $df : T(M) \to T(N)$ by $df | M_p = df_p$. Then one shows that df is of class C^k and is a bundle map which clearly has f as its induced map.

The category whose objects are C^k-Banach space bundles and whose morphisms are C^k-bundle maps is called the category of C^k-Banach space bundles. The function $M \to T(M)$, $f \to df$ is then a functor from the category of C^{k+1}-manifolds with boundary to the category of C^k-Banach space bundles. Since each author has his own definition of the tangent bundle functor it is useful to have a general theorem which proves they are all naturally equivalent, i.e. a characterization of T up to natural equivalence in purely functorial terms. To this end we first note two facts. If $\mathcal{O}$ is an open subset of a C^k-manifold M then $\mathcal{O}$ is in a natural way a C^k-manifold called an open submanifold of M: namely a chart for $\mathcal{O}$ is a chart for M whose domain is included in $\mathcal{O}$. If M is a Banach space V or else a half space in a Banach space V then the identity map of M is a chart in M and its unit class is a C^{k+1}-atlas for M. The C^{k+1}-manifold defined by this atlas will also be denoted by M. The corresponding full subcategory of the category of C^{k+1}-manifolds with boundary which we get in this way will be referred to as the subcategory of Banach spaces and half spaces. On this subcategory we have an obvious functor τ into the category of C^k-Banach space bundles; namely with each such C^{k+1}-manifold M we associate the product bundle $\tau(M) = M \times V$ considered as a C^k-Banach space bundle, and if $f: M \to N$ is a C^{k+1}-map, where N is either a Banach space W or else a half space in W, then the induced map $\tau(f): M \times V \to N \times W$ is given by $\tau(f)(m, v) = (f(m), df_m(v))$. We now characterize the notion of a tangent bundle functor.

DEFINITION. *A functor t from the category of C^{k+1}-manifolds to the category of C^k-Banach space bundles is called a tangent bundle functor if:*

(1) *$t(M)$ is a bundle over M and if $f: M \to N$ then f is the induced map of $t(f)$;*

(2) *Restricted to the subcategory of Banach spaces and half spaces t is naturally equivalent to τ;*

(3) *If M is a C^{k+1}-manifold and $\mathcal{O}$ is an open-submanifold and $1 : \mathcal{O} \to M$ the inclusion map then $t(\mathcal{O}) = t(M)|\mathcal{O}$ and $t(1)$ is the inclusion of $t(\mathcal{O})$ in $t(M)$.*

THEOREM. *The functor T defined above is a tangent bundle functor. Moreover any two tangent bundle functors are naturally equivalent.*

§6. INTEGRATION OF VECTOR FIELDS

Let $\sigma : (a, b) \to M$ be a C^{k+1}-map of an open interval into a C^{k+1}-manifold M. We define a C^k-map $\sigma' : (a, b) \to T(M)$, called the canonical lifting of σ, by $\sigma'(t) = d\sigma_t(1)$. We note that $\pi\sigma' = \sigma$ i.e. that σ' is in fact a lifting of σ.

DEFINITION. *A C^k-vector field on a C^{k+1}-manifold M is a C^k-cross section of $T(M)$, i.e. a C^k map $X : M \to T(M)$ such that $\pi \circ X =$ identity. A solution curve of X is a C^1-map σ of an open interval into M such that $\sigma' = X \circ \sigma$. If 0 is in the domain of the solution σ we call $\sigma(0)$ the initial condition of the solution σ.*

The facts stated below are straightforward consequences of the local existence and uniqueness theorem for vector fields and proofs will be found in [4, Chapter IV].

Let M be a C^{k+1}-manifold ($k \geq 1$) with $\partial M = \phi$ and let X be a C^k-vector field on M.

THEOREM (1). *For each $p \in M$ there is a solution curve σ_p of X with initial condition p such that every solution curve of X with initial condition p is a restriction of σ_p.*

The solution curve σ_p in the above theorem is called the maximum solution curve of X with initial condition p. We define $t^+ : M \to (0, \infty]$ and $t^- : M \to [-\infty, 0)$ by the requirement that the domain of σ_p is $(t^-(p), t^+(p))$. They are called respectively the positive and negative escape time functions for X.

THEOREM (2). *If $t^-(p) < s < t^+(p)$ and $q = \sigma_p(s)$ then $\sigma_q = \sigma_p \circ \tau_s$ where $\tau_s : \mathbf{R} \to \mathbf{R}$ is defined by $\tau_s(t) = s + t$. In particular $t^+(q) = t^+(p) - s$ and $t^-(q) = t^-(p) - s$.*

THEOREM (3). *t^+ is upper semi-continuous and t^- is lower semi-continuous. Also if $t^+(p) < \infty$ then $\sigma_p(t)$ has no limit point in M as $t \to t^+(p)$ and if $t^-(p) > -\infty$ then $\sigma_p(t)$ has no limit point in M as $t \to t^-(p)$.*

COROLLARY. *If M is compact then $t^+ \equiv \infty$ and $t^- \equiv -\infty$.*

To state the final and principal result we need the notion of the product of two C^k-manifolds. This is defined if at least one of the two manifolds has no boundary. If $\varphi : D(\varphi) \to V$ is a chart for M and $\psi : D(\psi) \to W$ is a chart for N then $\varphi \times \psi : D(\varphi) \times D(\psi) \to V \times W$ is a chart in $M \times N$ (note that the product of a half-space in V with W is a half-space in $V \times W$). The set of such charts is a C^k-atlas for $M \times N$ and we denote by $M \times N$ the resulting C^k-manifold. If N has no boundary then $\partial(M \times N) = (\partial M) \times N$.

Now we go back to our C^k-vector field X on a C^{k+1}-manifold M with $\partial M = \phi$.

DEFINITION. *Let $D = D(X) = \{(p, t) \in M \times \mathbf{R} | t^-(p) < t < t^+(p)\}$ and for each $t \in \mathbf{R}$ let $D_t = D_t(X) = \{p \in M | (p, t) \in D\}$. Define $\varphi : D \to M$ by $\varphi(p, t) = \sigma_p(t)$ and $\varphi_t : D_t \to M$ by $\varphi_t(p) = \sigma_p(t)$. The indexed set φ_t is called the maximum local one parameter group generated by X.*

THEOREM (4). *D is open in $M \times \mathbf{R}$ and $\varphi : D \to M$ is of class C^k. For each $t \in \mathbf{R}$ D_t is open in M and φ_t is a C^k-isomorphism of D_t onto D_{-t} having φ_{-t} as its inverse. If $p \in D_t$ and $\varphi_t(p) \in D_s$ then $p \in D_{t+s}$ and $\varphi_{t+s}(p) = \varphi_s(\varphi_t(p))$.*

§7. REGULAR AND CRITICAL POINTS OF FUNCTIONS

Let M be a C^1-manifold, $f: M \to \mathbf{R}$ a C^1-function. If $p \in M$ then $\mathrm{d}f_p$ is a bounded linear functional on M_p. If $\mathrm{d}f_p \neq 0$ then p is called a *regular point of* f and if $\mathrm{d}f_p = 0$ then p is called a *critical point* of f. If $c \in \mathbf{R}$ then $f^{-1}(c)$ is called a level of f (more explicitly the c-level of f) and it is called a regular level of f if it contains only regular points of f and a critical level of f if it contains at least one critical point of f. Also we call c a regular value of f if $f^{-1}(c)$ is regular and we call c a critical value of f if $f^{-1}(c)$ is critical.

If f and M are C^2 then there is a further dichotomy of the critical points of f into degenerate and non-degenerate critical points. We consider this next.

LEMMA. *Let φ be a C^k-isomorphism of an open set $\mathcal{O}$ in a Banach space V onto an open set $\mathcal{O}'$ in a Banach space V' ($k \geq 2$). Let $f: \mathcal{O}' \to \mathbf{R}$ be of class C^2 and let $g = f \circ \varphi : \mathcal{O} \to \mathbf{R}$. Then if $\mathrm{d}g_p = 0$, $\mathrm{d}^2 g_p(v_1, v_2) = \mathrm{d}^2 f_{\varphi(p)}(\mathrm{d}\varphi_p(v_1), \mathrm{d}\varphi_p(v_2))$.*

Proof. From the chain rule we get

$$\mathrm{d}g_x = \mathrm{d}f_{\varphi(x)}\, \mathrm{d}\varphi_x \quad \text{and}$$

$$\mathrm{d}^2 g_x(v_1, v_2) = \mathrm{d}^2 f_{\varphi(x)}(\mathrm{d}\varphi_x(v_1), \mathrm{d}\varphi_x(v_2)) + \mathrm{d}f_{\varphi(x)}(\mathrm{d}^2\varphi_x(v_1, v_2)).$$

Putting $x = p$ in the first equation gives $\mathrm{d}f_{\varphi(p)} = 0$ (because $\mathrm{d}\varphi_p$ is a linear isomorphism) and then putting $x = p$ in the second equation gives the desired result.

PROPOSITION. *If f is a C^2-real valued function on a C^2-manifold M and if p is a critical point of f then there is a uniquely determined continuous, symmetric, bilinear form $H(f)_p$ on M_p, called the Hessian of f at p, with the following property: if φ is any chart at p*

$$H(f)_p(v, w) = d^2(f \circ \varphi^{-1})_{\varphi(p)}(v_\varphi, w_\varphi).$$

Proof. Immediate from the lemma.

Given a Banach space V and a bounded, symmetric bilinear form B on V we say that B is *non-degenerate* if the linear map $T: V \to V^*$ defined by $T(v)(w) = B(v, w)$ is a linear isomorphism of V onto V^*, otherwise B is called *degenerate*. Also we define the *index of* B to be the supremum of the dimensions of subspaces W of V on which B is negative definite. The *co-index of* B is defined to be the index of $-B$.

DEFINITION. *If f is a C^2-real valued function on a C^2-manifold M and p is a critical point of f we define p to be degenerate or non-degenerate accordingly as the Hessian of f at p is degenerate or non-degenerate. The index and co-index of f at p are defined respectively as the index and co-index of the Hessian of f at p..*

The finite dimensional version of the following canonical form theorem is due to Marston Morse:

MORSE LEMMA. *Let f be a C^{k+2}- real valued function ($k \geq 1$) defined in a convex neighborhood $\mathcal{O}$ of the origin in a Hilbert space H. Suppose that the origin is a non-degenerate critical point of f and that f vanishes there. Then there is an origin preserving C^k-isomorphism φ of a neighborhood of the origin into H such that $f(\varphi(v)) = \|Pv\|^2 - \|(1 - P)v\|^2$ where P is an orthogonal projection in H.*

Proof. We shall show that there is a C^k-isomorphism ψ of a neighborhood of the origin in H such that $\psi(0) = 0$ and $f(v) = \langle A\psi(v), \psi(v)\rangle$ where $\langle , \rangle$ denotes the inner product in H and A is an invertible self-adjoint operator on H. The remainder of the proof uses the operator calculus as follows. Let h be the characteristic function of $[0, \infty)$. Then $h(A) = P$ is an orthogonal projection. Let $g(\lambda) = |\lambda|^{-1/2}$. Since zero is not in the spectrum of A, g is continuous and non-vanishing on the spectrum of A so $g(A) = T$ is a non-singular self-adjoint operator which commutes with A. Now $\lambda g(\lambda)^2 = sgn(\lambda) = h(\lambda) - (1 - h(\lambda))$ so $AT^2 = P - (1 - P)$. Then

$$\begin{aligned} f(\psi^{-1}Tv) = \langle ATv, Tv\rangle = \langle AT^2v, v\rangle &= \langle Pv, v\rangle - \langle (1 - P)v, v\rangle \\ &= \|Pv\|^2 - \|(1 - P)v\|^2. \end{aligned}$$

It remains to find ψ. By Taylor's theorem with $m = 2$ $f(v) = B(v)(v, v)$ where B is a C^k-map of $\mathcal{O}$ into bounded symmetric bilinear forms on H. Using the canonical identification of the latter space with self-adjoint operators on H we have $f(v) = \langle A(v)v, v\rangle$ where A is a C^k-map of $\mathcal{O}$ into self-adjoint operators on H. Now $d^2f_0(v, w) = 2\langle A(0)v, w\rangle$ and since the origin is a non-degenerate critical point of f, $A(0)$ is invertible, so $A(v)$ is invertible in a neighborhood of the origin which we can assume is $\mathcal{O}$. Define $B(v) = A(v)^{-1}A(0)$. Since inversion is easily seen to be a C^∞-map of the open set of invertible operators onto itself (it is given locally by a convergent power series) B is a C^k-map of $\mathcal{O}$ into $L(H, H)$, and each $B(v)$ is invertible. Now $B(0) =$ identity and since a square root function is defined in a neighborhood of the identity operator by a convergent power series with real coefficients we can define a C^k-map $C : \mathcal{O} \to L(H, H)$ with each $C(v)$ invertible, if $\mathcal{O}$ is taken sufficiently small, by $C(v) = B(v)^{1/2}$. Since $A(0)$ and $A(v)$ are self-adjoint we see easily from the definition of $B(v)$ that $B(v)^*A(v) = A(v)B(v)$ (both sides equaling $A(0)$) and clearly the same relation then holds for any polynomial in $B(v)$ hence for $C(v)$ which is a limit of such polynomials. Thus $C(v)^*A(v)C(v) = A(v)C(v)^2 = A(v)B(v) = A(0)$, or $A(v) = C_1(v)^*A(0)C_1(v)$ where we have put $C_1(v) = C(v)^{-1}$. If we write $\psi(v) = C_1(v)v$ then ψ is of class C^k in a neighborhood of the origin and $f(v) = \langle C_1(v)^*A(0)C_1(v)v, v\rangle = \langle A(0)\psi(v), \psi(v)\rangle$ so it remains only to show that $d\psi_0$ maps H isomorphically, and hence, by the inverse function theorem that ψ is a C^k-isomorphism on a neighborhood of the origin. An easy calculation gives $d\psi_v = C_1(v) + d(C_1)_v(v)$ so in particular $d\psi_0 = C_1(0)$ which in fact is the identity map of H.

COROLLARY. *The index of f at the origin is the dimension of the range of $(1 - P)$ and the co-index of f at the origin is the dimension of the range of P.*

Proof. Let W be a subspace on which d^2f_0 is negative definite. If $w \in W$ and $(1 - P)w = 0$ then $d^2f_0(w, w) = 2\|Pw\|^2 - 2\|(1 - P)w\|^2 = 2\|Pw\|^2 \geq 0$ so $w = 0$. Thus $(1 - P)$ is non-singular on W, hence $\dim W \leq \dim \text{range } (1 - P)$.

q.e.d.

Canonical Form Theorem for a Regular Point

Let f be a C^k-real valued function defined in a neighborhood U of the origin of a Banach space $V (k \geq 1)$. Suppose that the origin is a regular point of f and that f vanishes there. Then there is a non-zero linear functional l on V and an origin preserving C^k-isomorphism φ of a neighborhood of the origin in V into V such that $f(\varphi(v)) = l(v)$.

Proof. Let $l = df_0 \neq 0$. Choose $x \in V$ such that $l(x) = 1$. Let $W = \{v \in V | l(v) = 0\}$. Define $T : V \to W \times \mathbf{R}$ by $T(v) = (v - l(v)x, l(v))$. Then T is a linear isomorphism of V onto $W \times \mathbf{R}$. Define $\psi : U \to W \times \mathbf{R}$ by $\psi(v) = (v - l(v)x, f(v))$. Then ψ is of class C^k and $d\psi_u(v) = (v - l(v)x, df_u(v))$. In particular $d\psi_0 = T$ so by the inverse function theorem $\psi^{-1}T$ is a C^k-isomorphism of a neighborhood of the origin in V into V which clearly preserves the origin. If $v' = \psi^{-1}Tv$ then $(v' - l(v')x, f(v')) = \psi(v') = T(v) = (v - l(v)x, l(v))$, i.e. $f(\psi^{-1}Tv) = l(v)$.

q.e.d.

DEFINITION. *Let M be a C^k-manifold and let N be a closed subspace of M. We call N a* closed C^k-submanifold *of M if the set of charts in N which are restrictions of charts for M form an atlas for N. This atlas is automatically C^k and we denote the C^k-manifold determined by this atlas by N also.*

Smoothness Theorem for Regular Levels

Let f be a C^k-real valued function on a C^k-manifold M $(k \geq 1)$. Let $a \in \mathbf{R}$ be a regular value of f and assume that $f^{-1}(a)$ does not meet the boundary of M. Then

$M_a = \{x \in M | f(x) \leq a\}$ and $f^{-1}(a)$ are closed C^k-submanifolds of M and ∂M_a is the disjoint union of $M_a \cap \partial M$ and $f^{-1}(a)$.

Proof. An immediate consequence of the canonical form theorem for a regular point.

§8. THE STRONG TRANSVERSALITY THEOREM

Let M be a C^{k+1}-manifold without boundary ($k \geq 1$), X a C^k-vector field on M and φ_t the maximum local one parameter group generated by X (see §6). If f is a C^k real valued function on M we define a real valued function Xf on M by $Xf(p) = df_p(X_p)$. In general Xf will be of class C^{k-1} but of course in special circumstances it may be of class C^k or C^{k+1}. If we define $h(t) = f(\varphi_t(p)) = f(\sigma_p(t))$ then $h'(t) = df_{\sigma_p(t)}(\sigma_p'(t)) = df_{\sigma_p(t)}(X_{\sigma_p(t)}) = Xf(\varphi_t(p))$ so that if $Xf \equiv 1$ then $f(\varphi_t(p)) = f(p) + t$.

PROPOSITION. *Assume that $Xf \equiv 1$, $f(M) = (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$, and that $\varphi_t(x)$ is defined for $|t + f(x)| < \varepsilon$. Then $W = f^{-1}(0)$ is a closed C^k-submanifold of M and the map $F: W \times (-\varepsilon, \varepsilon) \to M$ defined by $F(w, t) = \varphi_t(w)$ is a C^k-isomorphism of $W \times (-\varepsilon, \varepsilon)$ onto M which for each $c \in (-\varepsilon, \varepsilon)$ maps $W \times \{c\}$ C^k-isomorphically onto $f^{-1}(c)$.*

Proof. Since at a critical point p of f $Xf(p) = df_p(X_p) = 0$ the condition $Xf \equiv 1$ implies that every real number is a regular value of f, hence that every level $f^{-1}(c)$ and in particular W is a closed C^k-submanifold of M. If $F(w, t) = F(w', t')$ then

$$t = f(w) + t = f(\varphi_t(w)) = f(\varphi_{t'}(w')) = f(w') + t' = t'$$

hence $\varphi_t(w) = \varphi_t(w')$ and since φ_t is one-to-one $w = w'$. We have proved that F is one-to-one. If $m \in M$ then $|-f(m) + f(m)| < \varepsilon$ so $w = \varphi_{-f(m)}(m)$ is well-defined and $f(w) = f(m) - f(m) = 0$ so $w \in W$. Moreover $F(w, f(m)) = \varphi_{f(m)}(\varphi_{-f(m)}(m)) = m$. Hence F is onto and moreover we see that $F^{-1}(m) = (\varphi_{-f(m)}(m), f(m))$ which by Theorem (4) of §6 is a C^k-map of M into $W \times (-\varepsilon, \varepsilon)$. Thus F is a C^k-isomorphism and since $f(F(w, c)) = f(\varphi_c(w)) = f(w) + c = c$ the final statement of the theorem also follows.

q.e.d.

DEFINITION. *A C^k-vector field X on a C^{k+1}-manifold without boundary M ($k \geq 1$) will be said to be C^k-strongly transverse to a C^k-function $f: M \to \mathbf{R}$ on a closed interval $[a, b]$ if for some $\delta > 0$ the following two conditions hold for $V = f^{-1}(a - \delta, b + \delta)$:*

(1) *Xf if of class C^k and non vanishing on V;*

(2) *If $p \in V$ and σ_p is the maximum solution curve of X with initial condition p then $\sigma_p(t)$ is defined and not in V for some positive t and also for some negative t.*

Now given the above, V is clearly an open submanifold of M and by (1) $Y = X/Xf$ is a well-defined C^k-vector field on V. Moreover Yf is identically one on V so the integral curves of Y are just the integral curves of X reparametrized so that $f(\sigma(t)) = f(\sigma(0)) + t$, hence condition (2) is equivalent to the statement that if ψ_t is the maximum local one parameter group generated by Y on V then $\psi_t(p)$ is defined for $a - \delta < f(p) + t < b + \delta$. If we put

$$g = f|V - \frac{a+b}{2}, \quad \varepsilon = \frac{b-a}{2} + \delta$$

we see that the triple (V, g, Y) satisfies the hypotheses made on the triple (M, f, X) in the above proposition. This proves

Strong Transversality Theorem

Let f be a C^k real valued function on a C^{k+1}-manifold without boundary M $(k \geq 1)$. If there exists a C^k-vector field X on M which is C^k-strongly transverse to f on a closed interval $[a, b]$ then $W = f^{-1}(a)$ is a closed C^k-submanifold of M and for some $\delta > 0$ there is a C^k-isomorphism F of $W \times (a - \delta, b + \delta)$ onto an open submanifold of M such that F maps $W \times \{c\}$ C^k-isomorphically onto $f^{-1}(c)$ for all $c \in (a - \delta, b + \delta)$. In particular $f^{-1}([a, b])$ is C^k-isomorphic to $W \times [a, b]$.

COROLLARY. *There is a C^k-map $H : M \times I \to M$ such that if we put $H_s(p) = H(p, s)$ then:*

(1) *H_s is a C^k-isomorphism of M onto itself for all $s \in I$;*

(2) *$H_s(m) = m$ if $m \notin f^{-1}(a - \delta/2, b + \delta/2)$;*

(3) $H_0 =$ identity;

(4) $H_1(f^{-1}(-\infty, a]) = f^{-1}(-\infty, b]$.

Proof. Let $h : \mathbf{R} \to \mathbf{R}$ be a C^∞-function with strictly positive first derivative such that $h(t) = t$ if $t \notin (a - \delta/2, b + \delta/2)$ and $h(a) = b$. Define $H_s =$ identity in the complement of $f^{-1}(a - \delta/2, b + \delta/2)$ and define H_s in $f^{-1}(a - \delta, b + \delta)$ by $H_s(F(w, t)) = F(w, (1 - s)t + sh(t))$.

§9. HILBERT AND RIEMANNIAN MANIFOLDS

Let M be a C^{k+1}-Hilbert manifold $(k \geq 0)$, i.e. M is a C^{k+1}-manifold and for each $p \in M$ M_p is a separable Hilbert space. For each $p \in M$ let $\langle , \rangle_p$ be an admissible inner product in M_p, i.e. a positive definite, symmetric, bilinear form on M_p such that the norm $\|v\|_p = \langle v, v \rangle^{1/2}$ defines the topology of M_p. Let φ be a chart in M having as target a Hilbert space H with inner product $\langle , \rangle$. We define a map G^φ of $D(\varphi)$ into the space of positive definite symmetric operators on H as follows: if $x \in D(\varphi)$ then $d\varphi^{-1}$ is an isomorphism of H onto M_x, hence there is a uniquely determined positive operator $G^\varphi(x)$ on H such that $\langle G^\varphi(x)u, v \rangle = \langle d\varphi_x^{-1}(u), d\varphi_x^{-1}(v) \rangle_x$. Suppose ψ is another chart in M with target H. Let $U = D(\varphi) \cap D(\psi)$ and let $f = \varphi \circ \psi^{-1} : \psi(U) \to \varphi(U)$, so $d\psi_x^{-1} = d\varphi_x^{-1} \circ df_{\psi(x)}$ for $x \in U$. Then

$$\begin{aligned} \langle G^\psi(x)u, v \rangle &= \langle d\varphi_x^{-1}\, df_{\psi(x)}(u), d\varphi_x^{-1}\, df_{\psi(x)}(v) \rangle_x \\ &= \langle G^\varphi(x)\, df_{\psi(x)}(u), df_{\psi(x)}(v) \rangle_x \end{aligned}$$

hence $G^\psi(x) = df^*_{\psi(x)} G^\varphi(x)\, df_{\psi(x)}$, $x \in U$. Since f is of class C^{k+1} it follows that if G^φ is of class C^k in U then so also is G^ψ. Hence it is consistent to demand for each chart φ that G^φ is of class C^k. If this is so we will call $x \to \langle , \rangle_x$ a C^k Riemannian structure for M, and M equipped with this extra structure will be called a C^{k+1}-Riemannian manifold. We will maintain the notation used above. That is given a chart φ in a C^k-Riemannian manifold M we will denote by G^φ the function defined above, and in addition we define Ω^φ on $D(\varphi)$ by $\Omega^\varphi(x) = (G^\varphi(x))^{-1}$. Also we define a function $\| \ \|$ in $T(M)$ by $\|v\| = \langle v, v\rangle^{1/2}$ for $v \in M_p$. Clearly $\| \ \|^2$ is of class C^k on $T(M)$, hence $\| \ \|$ is continuous on $T(M)$ and of class C^k on the complement of the zero section. If $\sigma : [a, b] \to M$ is a C^1-map then $t \to \|\sigma'(t)\|$ is continuous on $[a, b]$ hence

$$L(\sigma) = \int_a^b \|\sigma'(t)\| \, dt$$

is well defined and is called the length of σ. It is easily seen that if x and y are two points in the same component of M then there exists a C^1-curve joining them, hence we can define a metric ρ in each component of M by defining $\rho(x, y)$ to be the infimum of the lengths of all C^1-paths joining x and y. It is clear that ρ is symmetric, satisfies the triangle inequality, and is non-negative. That $\rho(x, y) > 0$ if $x \neq y$ and is hence a metric, and that the topology given by this metric is the given topology of M follows easily from the following two lemmas:

LEMMA (1). *Let H be a Hilbert space, f: $[a, b] \to H$ a C^1-map. Then*

$$\int_a^b \|f'(t)\| \, dt \geq \|f(b) - f(a)\|.$$

Proof. We can suppose $f(b) \neq f(a)$. Let $g(t)(f(b) - f(a))$ be the orthogonal projection of $f(t) - f(a)$ on the one-dimensional space spanned by $f(b) - f(a)$. Then $g : [a, b] \to \mathbf{R}$ is C^1, $g(a) = 0$, $g(b) = 1$ and $f(t) - f(a) = g(t)(f(b) - f(a)) + h(t)$ where $h : [a, b] \to (f(b) - f(a))^\perp$ is C^1. Then $f'(t) = g'(t)(f(b) - f(a)) + h'(t)$ where $h'(t) \perp (f(b) - f(a))$, hence

$$\|f'(t)\|^2 = \|f(b) - f(a)\|^2 \cdot |g'(t)|^2 + \|h'(t)\|^2 \geq \|f(b) - f(a)\|^2 \cdot |g'(t)|^2$$

so

$$\int_a^b \|f'(t)\| \, dt \geq \|f(b) - f(a)\| \cdot \int_a^b |g'(t)| \, dt.$$

But

$$\int_a^b |g'(t)| \, dt \geq \int_a^b g'(t) \, dt = g(b) - g(a) = 1.$$

q.e.d.

LEMMA (2). *Let H be a Hilbert space, $p \in H$, and G a continuous map of a neighborhood of p into the space of positive operators on H. Then there exists $r > 0$ such that G is defined on $B_r(p)$ and positive constants K and L such that:*

(1) *if f: $[a, b] \to B_r(p)$ is a C^1-map with $f(a) = p$*

$$L \int_a^b \langle f'(t), f'(t)\rangle^{1/2} \, dt \leq \int_a^b \langle G(f(t))f'(t), f'(t)\rangle^{1/2} \, dt$$

$$\leq K \int_a^b \langle f'(t), f'(t)\rangle^{1/2} \, dt;$$

(2) *if* $f:[a,b]\to H$ *is a* C^1*-map with* $f(a)=p$ *and* $c=\operatorname{Sup}\{t\in[a,b]\,|\,f([a,t])\subseteq B_r(p)\}$ *then*

$$\int_a^c \langle G(f(t))f'(t), f'(t)\rangle^{1/2}\,dt \geq Lr.$$

Proof. Let $\Omega(x)=G(x)^{-1}$. By continuity of G and Ω we can choose K and $L>0$ such that $\|G(x)\|\leq K^2$ and $\|\Omega(x)\|\leq L^{-2}$ for x in some open ball $B_r(p)$. Then

$$\langle G(f(t))f'(t), f'(t)\rangle \leq K^2\langle f'(t), f'(t)\rangle$$

and

$$\begin{aligned}\langle f'(t), f'(t)\rangle &= \langle \Omega(f(t))G(f(t))f'(t), f'(t)\rangle\\ &\leq L^{-2}\langle G(f(t))f'(t), f'(t)\rangle\end{aligned}$$

if $f(t)\in B_r(p)$. Then (1) follows immediately while (2) follows from (1) and Lemma (1).

q.e.d.

DEFINITION. *If* M *is a* C^{k+1}*-Riemannian manifold then the metric* ρ *defined above on each component of* M *is called the Riemannian metric of* M. *If each component of* M *is a complete metric space in this metric then* M *is called a complete* C^{k+1}*-Riemannian manifold.*

If σ *is a* C^1*-map of an open interval* (a,b) *into a Riemannian manifold* M *we define the length of* σ, $L(\sigma)$, *to be*

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

Of course $L(\sigma)$ may be infinite. However suppose $L(\sigma)<\infty$. Given $\varepsilon>0$ choose $t_0=a<t_1<\ldots<t_n<b=t_{n+1}$ so that

$$\int_{t_i}^{t_{i+1}}\|\sigma'(t)\|\,dt<\varepsilon.$$

Then clearly $\sigma((a,b))$ is included in the union of the ε-balls about the $\sigma(t_i)$ $i=1,2,\ldots,n$. Thus

PROPOSITION (1). *If* M *is a* C^{k+1}*-Riemannian manifold and* $\sigma:(a,b)\to M$ *is a* C^1*-curve of finite length, then the range of* σ *is a totally bounded subset of* M, *hence has compact closure if* M *is complete.*

PROPOSITION (2). *Let* X *be a* C^k*-vector field on a complete* C^{k+1}*-Riemannian manifold* $M(k\geq 1)$ *and let* $\sigma:(a,b)\to M$ *be a maximum solution curve of* X. *If* $b<\infty$ *then*

$$\int_0^b \|X(\sigma(t))\|\,dt=\infty$$

(*in particular* $\|X(\sigma(t))\|$ *is unbounded on* $[0,b)$) *and similarly if* $a>-\infty$ *then*

$$\int_a^0 \|X(\sigma(t))\|\,dt=\infty$$

(*in particular* $\|X(\sigma(t))\|$ *is unbounded on* $(a,0]$).

Proof. Since $\sigma'(t) = X(\sigma(t))$ it follows from Proposition (1) that if

$$\int_0^b \|X(\sigma(t))\| \, dt < \infty$$

then $\sigma(t)$ would have a limit point as $t \to b$, contradicting Theorem (3) of §6.

q.e.d.

Now let $f: M \to \mathbf{R}$ be a C^{k+1} real valued function on a C^{k+1}-Riemannian manifold M. Given $p \in M$ df_p is a continuous linear functional on M_p, hence there is a unique vector $\nabla f_p \in M_p$ such that $df_p(v) = \langle v, \nabla f_p \rangle_p$ for all $v \in M_p$. ∇f_p is called the gradient of f at p and $\nabla f: p \to \nabla f_p$ is called the gradient of f. We claim that ∇f is a C^k-vector field in M. To prove this we compute it explicitly with respect to a chart $\varphi : D(\varphi) \to H$ where H is a Hilbert space with inner product, $\langle , \rangle$. Let T be the canonical identification of H^* with H, so if $l \in H^*$ then $l(v) = \langle v, Tl \rangle$. Since T is a linear isomorphism it is C^∞. Define g on the range U of φ by $g = f \circ \varphi^{-1}$. Then g is of class C^{k+1} hence $dg : U \to H^*$ is class C^k so $T \circ dg = \lambda$ is C^k. Now by definition of G^φ

$$\begin{aligned}\langle G^\varphi(x)\, d\varphi_x(\nabla f_x), v \rangle &= \langle \nabla f_x, d\varphi_x^{-1}(v) \rangle_x = df_x \circ d\varphi_x^{-1}(v) \\ &= dg_{\varphi(x)}(v) = \langle T\, dg_{\varphi(x)}, v \rangle\end{aligned}$$

so $d\varphi_x(\nabla f_x) = \Omega^\varphi(x)\lambda(\varphi(x))$. Since $x \to G^\varphi(x)$ and hence $x \to G^\varphi(x)^{-1} = \Omega^\varphi(x)$ are C^k it follows that $x \to d\varphi_x(\nabla f_x)$ is a C^k-map of $D(\varphi)$ into H. By definition of the C^k-structure on $T(M)$ this means that ∇f is a C^k-vector field on M. We note the following obvious properties of ∇f. First ∇f_p is zero if and only if p is a critical point of f, so the critical locus of f is just the set of zeros of the real valued function $\|\nabla f\|$. Moreover

$$(\nabla f) f(p) = df_p(\nabla f_p) = \langle \nabla f_p, \nabla f_p \rangle_p = \|\nabla f_p\|_p^2$$

so $(\nabla f)f$ is positive off the critical locus of f.

§10. TWO-THIRDS OF THE MAIN THEOREM

In this section we assume that M is a C^{k+2}-Riemannian manifold ($k \geq 1$) without boundary and that $f: M \to \mathbf{R}$ is a C^{k+2}-function on M having only non-degenerate critical points and satisfying condition

(C) If S is any subset of M on which f is bounded but on which $\|\nabla f\|$ is not bounded away from zero then there is a critical point of f adherent to S.

We note that it is an immediate consequence of the Morse Lemma of §7 that a non-degenerate critical point of a C^3 function on a Hilbert manifold is isolated. In particular the critical locus of f is isolated. We will now prove that much more is true. Let $a < b$ be two real numbers and suppose that $\{p_n\}$ was a sequence of distinct critical points of f satisfying $a < f(p_n) < b$. Choose for each n a regular point q_n such that

$$\rho(q_n, p_n) < \frac{1}{n}, \qquad \|\nabla f_{q_n}\| < \frac{1}{n} \quad \text{and} \quad a < f(q_n) < b.$$

Then by condition (C) a subsequence of the $\{q_n\}$ will converge to a critical point p of f.

But clearly the corresponding subsequence of $\{p_n\}$ will also converge to p, contradicting the fact that critical points of f are isolated. Hence

PROPOSITION (1). *If a and b are two real numbers then there are at most a finite number of critical points p of f satisfying $a < f(p) < b$. Hence the critical values of f are isolated and there are at most a finite number of critical points of f on any critical level.*

LEMMA. *Assume M is complete and let $\sigma : (\alpha, \beta) \to M$ be a maximum solution curve of ∇f. Then either $\lim_{t\to\beta} f(\sigma(t)) = \infty$ or else $\beta = \infty$ and $\sigma(t)$ has a critical point of f as a limit point as $t \to \beta$. Similarly either $\lim_{t\to\alpha} f(\sigma(t)) = -\infty$ or else $\alpha = -\infty$ and $\sigma(t)$ has a critical point of f as a limit point as $t \to \alpha$.*

Proof. Let $g(t) = f(\sigma(t))$. Then

$$g'(t) = \mathrm{d}f_{\sigma(t)}(\sigma'(t)) = \mathrm{d}f_{\sigma(t)}(\nabla f_{\sigma(t)}) = \|\nabla f_{\sigma(t)}\|^2.$$

Thus g is monotone, hence has a limit B as $t \to \beta$. Suppose $B < \infty$. Then since

$$g(t) = g(0) + \int_0^t g'(s)\,\mathrm{d}s = g(0) + \int_0^t \|\nabla f_{\sigma(s)}\|^2\,\mathrm{d}s$$

it follows that

$$\int_0^\beta \|\nabla f_{\sigma(s)}\|^2\,\mathrm{d}s < \infty.$$

By Schwartz's inequality we have

$$\int_0^\beta \|\nabla f_{\sigma(s)}\|\,\mathrm{d}s \le \beta^{1/2}\left(\int_0^\beta \|\nabla f_{\sigma(s)}\|^2\,\mathrm{d}s\right)^{1/2}$$

so $\beta < \infty$ would contradict Proposition (2) of §9. Hence $\beta = \infty$. But then clearly $\|\nabla f_{\sigma(s)}\|$ cannot be bounded away from zero for $0 \le s < \infty$ since then the above integral could not converge. Since $f(\sigma(s))$ is bounded for $0 \le s < \infty$ (and in fact lies in the interval $[f(\sigma(0)), B]$ it follows from condition (C) that $\sigma(t)$ has a critical point of f as limit point as $t \to \beta$.

PROPOSITION (2). *If M is complete and f has no critical values in the closed interval $[a, b]$ then ∇f is C^{k+1}-strongly transverse to f on $[a, b]$ hence by the Strong Transversality Theorem (§8) $M_a = \{x \in M \mid f(x) \le a\}$ and $M_b = \{x \in M \mid f(x) \le b\}$ are C^{k+1}-isomorphic.*

Proof. By Proposition (1) of this section there is a $\delta > 0$ such that f has no critical values in $[a - \delta, b + \delta]$. Let $V = f^{-1}(a - \delta, b + \delta)$. Then $(\nabla f)f = \|\nabla f\|^2$ is strictly positive and C^{k+1} in V. Let $p \in M$ and let $\sigma : (\alpha, \beta) \to M$ be the maximal integral curve of ∇f with initial condition p. We must show for some $\alpha < t_2 < 0 < t_1 < \beta$ that $\sigma(t_1)$ and $\sigma(t_2)$ are not in V, i.e. that $f(\sigma(t_1)) \le a - \delta$ and $f(\sigma(t_2)) \ge b + \delta$. Suppose for example that $f(\sigma(t)) < b + \delta$ for $0 < t < \beta$. Then by the lemma $\sigma(t)$ would have a critical point p as limit point as $t \to \beta$. Since f is continuous and $f(\sigma(t))$ monotone it follows that $a - \delta \le f(0) \le f(p) \le b + \delta$ so $f(p)$ would be a critical value of f in $[a - \delta, b + \delta]$, a contradiction.

q.e.d.

Before completing the Main Theorem we must discuss the process of adding a handle to a Hilbert manifold.

§11. HANDLES

Let D^k denote the closed unit ball in a separable Hilbert space H of dimension k ($0 \le k \le \infty$). Note that since $f: H \to \mathbf{R}$ defined by $f(x) = \|x\|^2$ is a C^∞ real valued function in H and zero is the only critical value of f, it follows from the Smoothness Theorem for Regular Levels (§7) that $D^k = f^{-1}(-\infty, 1]$ is a closed C^∞ submanifold of H. Moreover the boundary ∂D^k of D^k is S^{k-1}, the unit sphere in H. We call $D^k \times D^l$ a handle of index k and co-index l. Note that $D^k \times D^l$ is not a differentiable manifold since both D^k and D^l have non-empty boundaries (unless k or $l = 0$). However if we put $\mathring{D}^k = D^k - \partial D^k$ then both $S^{k-1} \times D^l$ and $\mathring{D}^k \times D^l$ are C^∞-Hilbert manifolds.

DEFINITION. *Let M be a C^r-Hilbert manifold and N a closed submanifold of M. Let f be a homeomorphism of $D^k \times D^l$ onto a closed subset h of M. We shall write $M = N \underset{f}{\cup} h$ and say that M arises from N by a C^r-attachment f of a handle of type (k, l) if:*

(1) *$M = N \cup h$;*

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

(3) *$f|\mathring{D}^k \times D^l$ is a C^r-isomorphism onto $M - N$.*

Suppose we have a sequence of C^r-manifolds $N = N_0, N_1, \ldots, N_s = M$ such that N_{i+1} arises from N_i by a C^r-attachment f_i of a handle of type (k_i, l_i). If the images of the f_i are disjoint then we shall say that M arises from N by disjoint C^r-attachments $(f_1, \ldots, f_s)$ of handles of type $((k_1, l_1), \ldots, (k_s, l_s))$.

With the next lemma and theorem we come to one of the crucial steps in seeing what happens when we "pass a non-degenerate critical point".

LEMMA. *Let $\lambda : \mathbf{R} \to \mathbf{R}$ be a C^∞ function which is monotone non-increasing and satisfies $\lambda(x) = 1$ if $x \le 1/2$, $\lambda(x) > 0$ if $x < 1$ and $\lambda(x) = 0$ if $x \ge 1$. For $0 \le s \le 1$ let $\sigma(s)$ be the unique solution of $\lambda(\sigma)/1 + \sigma = \frac{3}{4}(1 - s)$ in the interval $[0, 1]$. Then σ is strictly monotone increasing, continuous, C^∞ in $[0, 1)$ and $\sigma(0) = 1/2$, $\sigma(1) = 1$. Moreover if $\varepsilon > 0$ and $u^2 - v^2 \ge -\varepsilon$ and $u^2 - v^2 - (3\varepsilon/2)\lambda(u^2/\varepsilon) \le -\varepsilon$ then*

$$u^2 \le \varepsilon\sigma\left(\frac{v^2}{\varepsilon + u^2}\right).$$

Proof. Clearly $\lambda(\sigma)/1 + \sigma$ is strictly monotonically decreasing if $0 \le \sigma \le 1$. Since it is one for $\sigma = 0$ and zero for $\sigma = 1$ the theorem that a continuous monotone map of an interval into $\mathbf{R}$ has a continuous monotone inverse gives easily that σ exists, is continuous and monotone. That $\sigma(0) = 1/2$ and $\sigma(1) = 1$ is clear and since $\lambda(\sigma)/1 + \sigma$ has a non-vanishing derivative in $[0, 1)$ it follows from the inverse function theorem that σ is C^∞ in $[0, 1)$. Now consider $f(u, v) = u^2 - \varepsilon\sigma(v^2/(\varepsilon + u^2))$ in the region

$$u^2 - v^2 \ge -\varepsilon, \qquad u^2 - v^2 - \frac{3\varepsilon}{2}\lambda\left(\frac{u^2}{\varepsilon}\right) \le -\varepsilon.$$

For v fixed f is clearly only critical for $u = 0$ where it has a minimum, hence f must assume its maximum on the boundary. On the boundary curve $u^2 - v^2 = -\varepsilon$ we have $v^2/(\varepsilon + u^2) = 1$ so $f(u, v) = u^2 - \varepsilon$. If (u, v) is not also on the other boundary curve then

$$-\frac{3\varepsilon}{2}\lambda\left(\frac{u^2}{\varepsilon}\right) < 0 \quad \text{so} \quad \lambda\left(\frac{u^2}{\varepsilon}\right) > 0$$

so $u^2 < \varepsilon$ so $f(u, v) < 0$. On the other hand if (u, v) is on the boundary

$$u^2 - v^2 - \frac{3\varepsilon}{2}\lambda(u^2/\varepsilon) = -\varepsilon$$

we have

$$\frac{v^2}{\varepsilon + u^2} = 1 - \frac{3}{2(1 + u^2/\varepsilon)}\lambda\left(\frac{u^2}{\varepsilon}\right).$$

Now on this boundary

$$u^2/\varepsilon \geq 1/2 \quad \left(\text{otherwise} \quad \frac{v^2}{\varepsilon + u^2} < 0\right)$$

and clearly

$$\frac{u^2}{\varepsilon} \leq 1 \quad \text{so} \quad \frac{u^2}{\varepsilon} = \sigma(\rho).$$

By definition of $\sigma(\rho)$

$$\frac{v^2}{\varepsilon + u^2} = 1 - \frac{3}{2}\frac{\lambda(\sigma(\rho))}{1 + \sigma(\rho)} = 1 - (1 - \rho) = \rho$$

hence

$$f(u, v) = u^2 - \varepsilon\sigma\left(\frac{v^2}{\varepsilon + u^2}\right) = \varepsilon\sigma(\rho) - \varepsilon\sigma(\rho),$$

i.e. f vanishes on this boundary. Thus $f \leq 0$ everywhere on the boundary of the region and hence also in the interior.

q.e.d.

THEOREM. *Let B be the ball of radius 2ε about the origin in a Hilbert space H. Define $f: B \to \mathbf{R}$ by $f(v) = \|Pv\|^2 - \|Qv\|^2$ where P is an orthogonal projection on a subspace H^l of dimension l and $Q = (1 - P)$ is a projection on a subspace H^k of dimension k. Let*

$$g(v) = f(v) - \frac{3\varepsilon}{2}\lambda(\|Pv\|^2/\varepsilon)$$

where $\lambda: \mathbf{R} \to \mathbf{R}$ is as in the lemma. Then $M = \{x \in B | g(x) \leq -\varepsilon\}$ arises from $N = \{x \in B | f(x) \leq -\varepsilon\}$ by a C^∞ attachment F of a handle h of type (k, l).

Proof. Before commencing on the proof we give a diagram of the case $k = l = 1$ (Fig. 1).

Let D^k and D^l be the unit discs in H^k and H^l respectively. Let h be the set in B where $f \geq -\varepsilon$ and $g \leq -\varepsilon$ so $M = N \cup h$ and $N \cap h \subseteq \partial N$. Define $F: D^k \times D^l \to H$ by

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

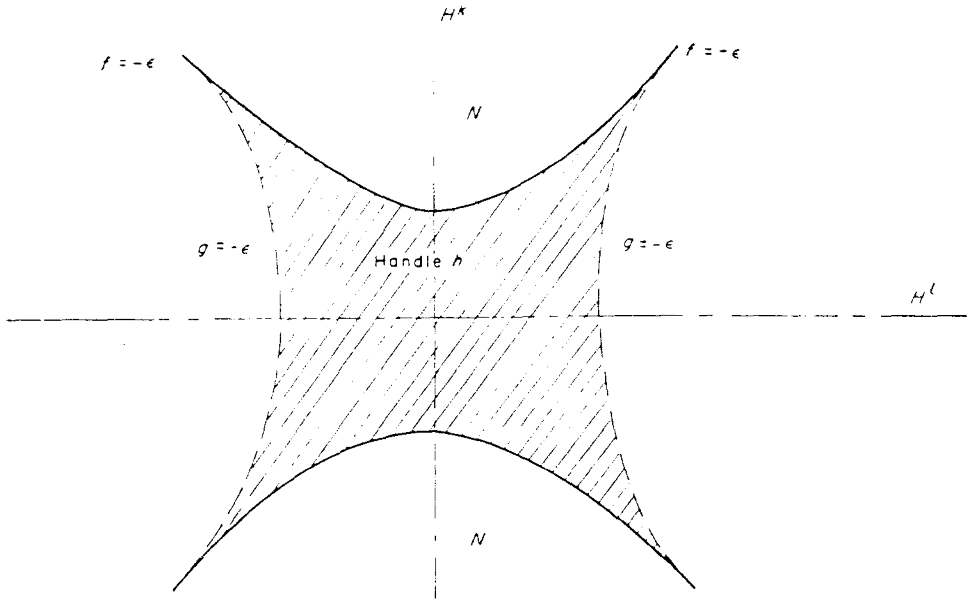

FIG. 1

Where σ is as in the lemma. Then

$$\begin{aligned} f(F(x, y)) &= \varepsilon[\sigma(\|x\|^2)\|y\|^2 - (1 + \sigma(\|x\|^2)\|y\|^2)\|x\|^2] \\ &= \varepsilon[\sigma(\|x\|^2)\|y\|^2(1 - \|x\|^2) - \|x\|^2] \geq -\varepsilon \\ g(F(x, y)) &= \varepsilon[\sigma(\|x\|^2)\|y\|^2(1 - \|x\|^2) - \|x\|^2 \\ &\quad - \tfrac{3}{2}\lambda(\sigma(\|x\|^2)\|y\|^2)] \end{aligned}$$

Since λ is monotone decreasing

$$g(F(x, y)) \leq \varepsilon[\sigma(\|x\|^2)(1 - \|x\|^2) - \|x\|^2 - \tfrac{3}{2}\lambda(\sigma(\|x\|^2))]$$

but $\lambda(\sigma(\|x\|^2)) = \frac{2}{3}(1 + \sigma(\|x\|^2))(1 - \|x\|^2)$ by definition of σ; substituting we see that $g(F(x, y)) \leq -\varepsilon$, hence F maps $D^k \times D^l$ into h. Conversely suppose $w \in h$ and let $u = Pw$, $v = Qw$ so $\|u\|^2 - \|v\|^2 \geq -\varepsilon$ and

$$\|u\|^2 - \|v\|^2 - \frac{3\varepsilon}{2}\lambda(\|u\|^2/\varepsilon) \leq -\varepsilon.$$

Then $\|v\|^2/(\varepsilon + \|u\|^2) \leq 1$ so $x = (\varepsilon + \|u\|^2)^{-1/2}v \in D^k$. Also $\sigma(\|v\|^2/(\varepsilon + \|u\|^2))$ is well defined and by the lemma $\|u\|^2/\varepsilon\sigma(\|v\|^2/(\varepsilon + \|u\|^2)) \leq 1$ so $y = (\varepsilon\sigma(\|v\|^2/(\varepsilon + \|u\|^2)))^{-1/2}u \in D^l$. Thus

$$G(w) = ((\varepsilon + \|Pw\|^2)^{-1/2}Qw, (\varepsilon\sigma(\|Qw\|^2/(\varepsilon + \|Pw\|^2)))^{-1/2}Pw)$$

defines a map of h into $D^k \times D^l$. It is an easy check that F and G are mutually inverse maps, hence F is a homeomorphism of $D^k \times D^l$ onto h. From the fact that σ is C^∞ with non-vanishing derivative in $[0, 1)$ it follows that F is a C^∞-isomorphism on $\mathring{D}^k \times D^l$. On $S^{k-1} \times D^l$ F reduces to

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

which is clearly a C^∞-isomorphism onto $N \cap h$, the set where $f = -\varepsilon$ and $\|Pw\|^2 \leq \varepsilon$.

q.e.d.

§12. PASSING A CRITICAL LEVEL

In this section we will complete the proof of the main theorem by analyzing what happens as we pass a critical level. We will need:

LEMMA. *Let Q and f be bounded, symmetric, non-degenerate bilinear forms on a Hilbert space H, Q positive definite. Then there exists an admissible inner product $\langle,\rangle$ in H such that $Q(v, v) = \langle Gv, v\rangle$ and $f(v, v) = \|Pv\|^2 - \|(1 - P)v\|^2$ where P is an orthogonal projection which commutes with the positive operator G.*

Proof. We essentially proved this in the course of proving the Morse lemma. Note that $Q(u, v)$ is an admissible inner product in H, hence $f(v, v) = Q(Av, v)$ where A is an invertible operator self adjoint with respect to this inner product. Let $G = |A|^{-1}$ and $P = h(A)$, where h is the characteristic function of $[0, \infty)$, and define $\langle u, v\rangle = Q(|A|u, v)$ so $Q(u, v) = \langle Gu, v\rangle$. It is clear that any function of A is self adjoint relative to $\langle,\rangle$ so P is an orthogonal projection and G a positive operator in this inner product, and both being functions of A they commute. Now $|\lambda|^{-1}\lambda = h(\lambda) - (1 - h(\lambda))$ so $GA = P - (1 - P)$ so $f(v, v) = Q(Av, v) = \langle GAv, v\rangle = \|Pv\|^2 - \|(1 - P)v\|^2$.

q.e.d.

We now return to the situation of §10; M is a complete C^{k+2}-Riemannian manifold $(k \geq 1)$ and f is a C^{k+2}-function on M all of whose critical points are non-degenerate, and condition (C) is satisfied. Let c be a critical value of f. We reduce to the case $c = 0$ by replacing f by $f - c$. By §10 there are a finite number of critical points $p_1, \ldots, p_r$ of f with $f(p_i) = 0$. Let k_i and l_i be respectively the index and coindex of f at p_i. By the Morse Lemma (§7) we can find for some $\delta < 1$ a C^k-chart φ_i at p_i whose image is the ball of radius 2δ in a Hilbert space H_i such that $\varphi_i(p_i) = 0$ and $f\varphi_i^{-1}(v) = \|P_iv\|^2 - \|(1 - P_i)v\|^2$ where P_i is an orthogonal projection in H_i of rank l_i and $(1 - P_i)$ has rank k_i. Moreover by the above lemma if G^i is the positive operator in H_i defined by $\langle d\varphi_{p_i}^{-1}(u), d\varphi_{p_i}^{-1}(v)\rangle_{p_i} = \langle G^iu, v\rangle$ then *we can assume that G^i commutes with P_i*. This will be crucial at a later point in the argument.

By Proposition (1) of §10 we can choose $\varepsilon < \delta^2$ so small that 0 is the only critical value of f in $(-3\varepsilon, 3\varepsilon)$. Let $W = f^{-1}(-2\varepsilon, \infty)$. We define a C^k-real valued function g in W by

$$g(\varphi_i^{-1}(v)) = f(\varphi_i^{-1}(v)) - \frac{3\varepsilon}{2}\lambda(\|P_iv\|^2/\varepsilon),$$

where λ is as in the lemma of §11, and $g(w) = f(w)$ if $w \notin \bigcup_{i=1}^{r} D(\varphi_i)$. Note that if $w = \varphi_i^{-1}(v) \in W$ and $f(w) \neq g(w)$ then $\lambda(\|P_iv\|^2/\varepsilon) \neq 0$ so $\|P_iv\|^2 < \varepsilon$ (hence $f(w) < \varepsilon$) and $\|P_iv\|^2 - \|(I - P_i)v\|^2 = f(w) > -2\varepsilon$ so $\|v\|^2 = \|P_iv\|^2 + \|(I - P_i)v\|^2 < 4\varepsilon < 4\delta^2$, so the closure of $\{w \in W \cap D(\varphi_i) \mid f(w) \neq g(w)\}$ is included in the interior of $D(\varphi_i)$, which proves that g is C^k. The above also shows that $\{w \notin W \mid f(w) \leq \varepsilon\} = \{w \in W \mid g(w) \leq \varepsilon\}$. Now it follows immediately from the theorem of §11 that $\{w \in W \mid g(w) \leq -\varepsilon\}$ arises from $\{w \in W \mid f(w) \leq -\varepsilon\}$ by the disjoint C^k-attachment of r handles of type $(k_1, l_1) \ldots, (k_r, l_r)$. We will prove:

LEMMA. *If ε is sufficiently small then ∇f is C^k-strongly transverse to g on $[-\varepsilon, \varepsilon]$.*

It then follows from the strong transversality theorem that there exists a C^k-isomorphism h of W onto itself such that $h(w) = w$ if $|g(w)| \geq -3\varepsilon/2$ and h maps $\{w \in W | g(w) \leq -\varepsilon\}$ C^k-isomorphically onto $\{w \in W | g(w) \leq \varepsilon\} = \{w \in W | f(w) \leq \varepsilon\}$. We can extend h to a C^k-isomorphism of M by defining $h(x) = x$ if $x \notin W$. It follows that $\{x \in M | f(x) \leq \varepsilon\}$ is C^k-isomorphic to $\{x \in M | f(x) \leq -\varepsilon\}$ with r-handles of type $(k_1, l_1), \ldots, (k_r, l_r)$ disjointly C^k-attached. More generally by applying Proposition (2) of §10 to the intervals $[a, -\varepsilon]$ and $[\varepsilon, b]$ we get the third part of the main theorem.

THEOREM. *Let f be a C^{k+2} real valued function on a complete C^{k+2}-Riemannian manifold $M(k \geq 1)$. Assume that all the critical points of f are non-degenerate and that f satisfies condition* (C). *Let $p_1, \ldots, p_r$ be the distinct critical points of f on $f^{-1}(c)$ and let k_i and l_i be the index and coindex of f at p_i. If $a < c < b$ and c is the only critical value of f in $[a, b]$ then $\{x \in M | f(x) \leq b\}$ is C^k-isomorphic to $\{x \in M | f(x) \leq a\}$ with r-handles of type $(k_1, l_1), \ldots, (k_r, l_r)$ disjointly C^k-attached.*

It remains to prove that ∇f is C^k-strongly transverse to g on $[-\varepsilon, \varepsilon]$, if ε is sufficiently small. Let

$$V = \left\{x \in W \,\middle|\, -\frac{5\varepsilon}{4} < g(x) < \frac{5\varepsilon}{4}\right\}.$$

We note that since

$$f - \frac{3\varepsilon}{2} \leq g \leq f$$

and f has no critical value in $(-3\varepsilon, 3\varepsilon)$ except zero, the only possible critical points of f in $\bar{V}$ could be $p_1, \ldots, p_r$. But

$$g(p_i) = -\frac{3\varepsilon}{2}$$

so *f has no critical points in $\bar{V}$.* Now let $p \in V$ and let $\sigma : (\alpha, \beta) \to M$ be the maximal integral curve of ∇f with initial condition p. Then by the lemma of §10 either $f(\sigma(t)) \to \infty$ as $t \to \beta$, so $\sigma(t)$ gets outside V as $t \to \beta$ or else $\sigma(t)$ has a critical point of f as limit point as $t \to \beta$ so again $\sigma(t)$ must get outside V as $t \to \beta$. Similarly $\sigma(t)$ must get outside V as $t \to \alpha$. Thus it remains only to show that $(\nabla f)g$ is C^k and positive in V. Outside $\bigcup_{i=1}^{r} D(\varphi_i), f = g$ so $(\nabla f)g = (\nabla f)f = \|\nabla f\|^2$ which is C^{k+1} and is positive since f has no critical points in V. What is left then is to show that $(\nabla f)g$ is C^k and does not vanish on $D(\varphi_i)$ except at p_i. The following proposition settles this local question.

PROPOSITION (1). *Let $\mathcal{O}$ be a neighborhood of zero in a Hilbert space H with inner product $\langle\ \rangle$, made into a C^{k+1}-Riemannian manifold $(k \geq 0)$ by defining $\langle u, v\rangle_w = \langle G(w)u, v\rangle$ where G is a C^k- map of $\mathcal{O}$ into the invertible positive operators on H. Let P be an orthogonal projection in H which commutes with $G(0)$ and define $f(v) = \|Pv\|^2 - \|(1-P)v\|^2$ and*

$$g(v) = f(v) - \frac{3\varepsilon}{2}\lambda(\|Pv\|^2/\varepsilon)$$

where λ is as in the lemma of §11. *Then $(\nabla f)g$ is C^k and for ε sufficiently small does not vanish on the 2ε ball about the origin except at the origin.*

Proof. Let $\Omega(x) = G(x)^{-1}$ so that $\Omega(0)$ also commutes with P. Let $T(x) = P\Omega(x) - \Omega(x)P$. Note that $\|Px\| \leq \|x\|$ and $\|(2P - 1)x\| = \|x\|$ so

$$\begin{aligned}\langle Px, \Omega(x)(2P - I)x\rangle &= \langle Px, P\Omega(x)(2P - I)x\rangle \\ &= \langle Px, T(x)(2P - I)x\rangle + \langle Px, \Omega(x)Px\rangle \\ &\geq \langle Px, T(x)(2P - I)x\rangle \\ &\geq -\|T(x)\| \cdot \|x\|^2.\end{aligned}$$

Now $\|u\|^2 = \langle u, G(x)\Omega(x)u\rangle \leq \|G(x)\| < u, \Omega(x)u\rangle$ hence

$$\langle (2P - I)x, \Omega(x)(2P - I)x\rangle \geq \|G(x)\|^{-1} \cdot \|x\|^2.$$

Since $\|T(0)\| = 0$ while $\|G(0)\|^{-1} > 0$ we can find a neighborhood U of the origin, independent of ε, such that for $x \in U$

$$\|G(x)\|^{-1} > \tfrac{3}{2}\|T(x)\| \sup|\lambda'|.$$

Since $\lambda' \leq 0$ it follows that for x in U

$$\begin{aligned}4(\langle (2P - I)x, \Omega(x)(2P - I)x\rangle &- \tfrac{3}{2}\lambda'(\|Px\|^2/\varepsilon)\langle Px, \Omega(x)(2P - I)x\rangle) \\ &\geq 4(\|G(x)\|^{-1} - \tfrac{3}{2}|\lambda'(\|Px\|^2/\varepsilon)| \cdot \|Tx\|)\|x\|^2.\end{aligned}$$

which is positive unless $x = 0$. Since the left-hand side is clearly C^k it will suffice to prove that it equals $(\nabla f)g$. From the definition of f and g $df_x(y) = 2\langle (2P - I)x, y\rangle = 2\langle \Omega(x)(2P - I)x, y\rangle_x$ so $\nabla f_x = 2\Omega(x)(2P - I)x$ while

$$\begin{aligned}dg_x(y) &= df_x(y) - 3\lambda'(\|Px\|^2/\varepsilon)\langle Px, y\rangle \\ &= 2(\langle (2P - I)x, y\rangle - \tfrac{3}{2}\lambda'(\|Px\|^2/\varepsilon)\langle Px, y\rangle).\end{aligned}$$

Since $\nabla f_x(g) = dg_x(\nabla f_x)$ the desired expression for $(\nabla f)g$ is immediate.

q.e.d.

This completes the proof of the Theorem. We now consider an interesting corollary of the proof of Proposition (1). Maintaining the notation of the proof let us define $\rho(x) = \|x\|^2 = \|Px\|^2 + \|(1 - P)x\|^2$ so that $(f - \rho)(x) = -2\|(1 - P)x\|^2$ and

$$\begin{aligned}(\nabla f)(f - \rho)(x) &= 8\langle (P - I)x, \Omega(x)(2P - I)x\rangle \\ &= 8\langle (P - I)x, \Omega(x)(P - I)x\rangle + 8\langle (P - I)x, \Omega(x)Px\rangle.\end{aligned}$$

Since $\Omega(x)Px = P\Omega(x)x - T(x)x$ and since $P\Omega(x)x$ is orthogonal to $(P - I)x$ we get

$$(\nabla f)(f - \rho)(x) = 8\langle (P - I)x, \Omega(x)(P - I)x\rangle - \langle (P - I)x, T(x)x\rangle.$$

Recalling the inequality $\langle u, \Omega(x)u\rangle \geq \|G(x)\|^{-1} . \|u\|^2$

$$(\nabla f)(f - \rho)(x) \geq 8\|(P - I)x\|(\|(P - I)x\| \cdot \|G(x)\|^{-1} - \|x\| \cdot \|T(x)\|).$$

Since $\|T(0)\| = 0$, in a sufficiently small neighborhood of the origin we have $\|T(x)\| \leq \frac{1}{2}\|G(x)\|^{-1}$ so in this neighborhood

$$(\nabla f)(f - \rho)(x) \geq 8\|(P - I)x\| \cdot \|G(x)\|^{-1}\left(\|(P - I)x\| - \frac{\|x\|}{2}\right).$$

If $f(x) \leq 0$ then $\|(P - I)x\|^2 \geq \|Px\|^2$ so $2\|(P - I)x\|^2 \geq \|x\|^2$ which implies that

$$\|(P - I)x\| \geq \frac{\|x\|}{2},$$

hence near the origin $f(x) \leq 0$ implies $(\nabla f)(f - \rho)(x) > 0$. It follows that $f - \rho$ is monotonically increasing along any solution curve of ∇f which is close enough to the origin and on which f is negative. Since clearly $f(x) > -\rho(x)$ we see that if ε is sufficiently small and $\sigma(t)$ is the maximum solution curve of ∇f with initial condition p, where $\rho(p) < \varepsilon/2$, then $\rho(\sigma(t)) > \varepsilon$ implies $f(\sigma(t)) > 0$. This proves

PROPOSITION (2). *Let f be a C^3-real valued function on a C^3-Riemannian manifold M and let p be a non-degenerate critical point of f. Then if U is any neighborhood of p there is a neighborhood $\mathcal{O}$ of p such that if σ is a maximum solution curve of ∇f having initial condition in $\mathcal{O}$ then for $t > 0$ either $\sigma(t) \in U$ or $f(\sigma(t)) > f(p)$.*

COROLLARY. *If σ is a maximum solution curve of ∇f and if p is a limit point of $\sigma(t)$ as $t \to \infty$ $(t \to -\infty)$ then $\lim_{t \to \infty} \sigma(t) = p$ $(\lim_{t \to -\infty} \sigma(t) = p)$.*

PROPOSITION (3). *Let M be a complete C^3-Riemannian, f a C^3-real valued function on M which is bounded above (below), has only non-degenerate critical points, and satisfies condition (C). If σ is any maximum solution curve of ∇f then $\lim_{t \to \infty} \sigma(t)$ $(\lim_{t \to -\infty} \sigma(t))$ exists and is a critical point of f.*

Proof. An immediate consequence of the above corollary and the lemma to Proposition (2) of §10.

§13. THE MANIFOLDS $H_1(I, V)$ AND $\Omega(V; P, Q)$

In this section we will develop some of the concepts that are involved in applying the results of the preceding sections to Calculus of Variations problems.

A map σ of the unit interval I into $\mathbf{R}^n$ is called absolutely continuous if either and hence both of the following two conditions are satisfied:

(1) Given $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 \leq t_0 < \ldots < t_{2k+1} \leq 1 \quad \text{and} \quad \sum_{i=0}^{k} |t_{2i+1} - t_{2i}| < \delta$$

then

$$\sum_{i=0}^{k} \|\sigma(t_{2i+1}) - \sigma(t_{2i})\| < \varepsilon.$$

(2) There is a $g \in L^1(I, \mathbf{R}^n)$

$$\left(\text{i.e. } g \text{ is a measurable function from } I \text{ into } \mathbf{R}^n \text{ and } \int_0^1 \|g(t)\| \, dt < \infty \right)$$

such that

$$\sigma(t) = \sigma(0) + \int_0^t g(s) \, ds.$$

The equivalence of these two conditions is a classical theorem of Lebesgue. From the second condition it follows that $\sigma'(t)$ exists for almost all $t \in I$, that σ' ($=g$) is summable and

$$\sigma(t) = \sigma(0) + \int_0^t \sigma'(s)\,ds.$$

From the first condition it follows that if φ is a C^1-map of $\mathbf{R}^n$ into $\mathbf{R}^m$, or more generally if $\varphi : \mathbf{R}^n \to \mathbf{R}^m$ satisfies a Lipshitz condition on every compact set, then $\varphi \circ \sigma$ is absolutely continuous.

For reasons of consistency that will become clear later we will denote the set of measureable functions σ of I into $\mathbf{R}^n$ such that

$$\int_0^1 \|\sigma(t)\|^2\,dt < \infty \quad \text{by} \quad H_0(I, \mathbf{R}^n),$$

rather than the more customary $L^2(I, \mathbf{R}^n)$. Then $H_0(I, \mathbf{R}^n)$ is a Hilbert space under pointwise operations and the inner product $\langle , \rangle_0$ defined by

$$\langle \sigma, \rho \rangle_0 = \int_0^1 \langle \sigma(t), \rho(t) \rangle\,dt$$

where of course $\langle , \rangle$ is the standard inner product in $\mathbf{R}^n$.

We will denote by $H_1(I, \mathbf{R}^n)$ the set of absolutely continuous maps $\sigma : I \to \mathbf{R}^n$ such that $\sigma' \in H_0(I, \mathbf{R}^n)$. Then $H_1(I, \mathbf{R}^n)$ is a Hilbert space under the inner product $\langle , \rangle_1$ defined by $\langle \sigma, \rho \rangle_1 = \langle \sigma(0), \rho(0) \rangle + \langle \sigma', \rho' \rangle_0$ and in fact the map $\mathbf{R}^n \oplus H_0(I, \mathbf{R}^n) \to H_1(I, \mathbf{R}^n)$ defined by $(p, g) \to \sigma$, where

$$\sigma(t) = p + \int_0^t g(s)\,ds,$$

is an isometry onto.

Definition. *We define* $L : H_1(I, \mathbf{R}^n) \to H_0(I, \mathbf{R}^n)$ *by* $L\sigma = \sigma'$ *and we define*

$$H_1^*(I, \mathbf{R}^n) = \{\sigma \in H_1(I, \mathbf{R}^n) | \sigma(0) = \sigma(1) = 0\}.$$

Then the following is immediate:

Theorem (1). *L is a bounded linear transformation of norm one. $H_1^*(I, \mathbf{R}^n)$ is a closed linear subspace of codimension $2n$ in $H_1(I, \mathbf{R}^n)$ and L maps $H_1^*(I, \mathbf{R}^n)$ isometrically onto the set of $g \in H_0(I, \mathbf{R}^n)$ such that*

$$\int_0^1 g(t)\,dt = 0,$$

i.e. onto the orthogonal complement in $H_0(I, \mathbf{R}^n)$ of the set of constant maps of I into $\mathbf{R}^n$.

Theorem (2). *If $\rho \in H_1^*(I, \mathbf{R}^n)$ and λ is an absolutely continuous map of I into $\mathbf{R}^n$ then*

$$\int_0^1 \langle \lambda'(t), \rho(t) \rangle\,dt = \langle \lambda, -L\rho \rangle_0.$$

Proof. Clearly $t \to \langle \lambda(t), \rho(t) \rangle$ is an absolutely continuous real valued function with derivative $\langle \lambda'(t), \rho(t) \rangle + \langle \lambda(t), \rho'(t) \rangle$. Since an absolutely continuous function is the integral of its derivative and since $\langle \lambda(t), \rho(t) \rangle$ vanishes at zero and one, the theorem follows.

q.e.d.

We shall denote the set of continuous maps of I into $\mathbf{R}^n$ by $C^0(I, \mathbf{R}^n)$, considered as a Banach space with norm $\| \ \|_\infty$ defined by $\|\sigma\|_\infty = \sup\{\|\sigma(t)\| \mid t \in I\}$. We recall that by the Ascoli–Arzela theorem a subset S of $C^0(I, \mathbf{R}^n)$ is totally bounded if and only if it is bounded and equicontinuous (the latter means given $\varepsilon > 0$ there exists $\delta > 0$ such that if $|s - t| < \delta$ then $|g(s) - g(t)| < \varepsilon$ for all $g \in S$). Since the inclusion of $C^0(I, \mathbf{R}^n)$ in $H_0(I, \mathbf{R}^n)$ is clearly uniformly continuous it follows that such an S is also totally bounded in $H_0(I, \mathbf{R}^n)$.

The following is a rather trivial special case of the Sobolev inequalities:

THEOREM (3). *If $\sigma \in H_1(I, \mathbf{R}^n)$ then*

$$\|\sigma(t) - \sigma(s)\| \leq |t - s|^{1/2} \|L\sigma\|_0.$$

Proof. Let h be the characteristic function of $[s, t]$. Then

$$\|\sigma(t) - \sigma(s)\| = \left\| \int_s^t \sigma'(x)\,\mathrm{d}x \right\| \leq \int_s^t \|\sigma'(x)\|\,\mathrm{d}x$$
$$= \int_0^1 h(x)\|\sigma'(x)\|\,\mathrm{d}x$$

and Schwartz's inequality completes the proof.

q.e.d.

COROLLARY (1). *If $\sigma \in H_1(I, \mathbf{R}^n)$ then $\|\sigma\|_\infty \leq 2\|\sigma\|_1$.*

Proof. By definition of $\| \ \|_1$ we have $\|\sigma(0)\| \leq \|\sigma\|_1$ and $\|L\sigma\|_0 \leq \|\sigma\|_1$. Now $\|\sigma(t)\| \leq \|\sigma(0)\| + \|\sigma(t) - \sigma(0)\|$ and by the theorem $\|\sigma(t) - \sigma(0)\| \leq \|L\sigma\|_0$.

q.e.d.

COROLLARY (2). *The inclusion maps of $H_1(I, \mathbf{R}^n)$ into $C^0(I, \mathbf{R}^n)$ and $H_0(I, \mathbf{R}^n)$ are completely continuous.*

Proof. Let S be a bounded set in $H_1(I, \mathbf{R}^n)$. Then by Corollary (1) S is bounded in $C^0(I, \mathbf{R}^n)$ and by the theorem S satisfies a uniform Hölder condition of order 1/2, hence is equicontinuous. q.e.d.

THEOREM (4). *If $\varphi : \mathbf{R}^n \to \mathbf{R}^p$ is a C^{k+2}-map then $\sigma \to \varphi \circ \sigma$ is a C^k-map $\bar{\varphi} : H_1(I, \mathbf{R}^n) \to H_1(I, \mathbf{R}^p)$. Moreover if $1 \leq m \leq k$ then*

$$\mathrm{d}^m\bar{\varphi}_\sigma(\lambda_1, \dots, \lambda_m)(t) = \mathrm{d}^m\varphi_{\sigma(t)}(\lambda_1(t), \dots, \lambda_m(t)).$$

Proof. This is a consequence of the following lemma if we take $F = \mathrm{d}^s\varphi$ $0 \leq s \leq k - 1$. [Note that in the lemma if $s = 0$ then we interpret $L^s(\mathbf{R}^n, \mathbf{R}^p)$ to be $\mathbf{R}^p$.]

LEMMA. *Let F be a C^1-map of $\mathbf{R}^n$ into $L^s(\mathbf{R}^n, \mathbf{R}^p)$. Then the map $\bar{F}$ of $H_1(I, \mathbf{R}^n)$ into $L^s(H_1(I, \mathbf{R}^n), H_1(I, \mathbf{R}^p))$ defined by*

$$\bar{F}(\sigma)(\lambda_1, \dots, \lambda_s)(t) = F(\sigma(t))(\lambda_1(t), \dots, \lambda_s(t))$$

is continuous. Moreover if F is C^3 then $\bar{F}$ is C^1 and $\mathrm{d}\bar{F} = \overline{\mathrm{d}F}$.

Proof. We note that

$$(\bar{F}(\sigma)(\lambda_1, \ldots, \lambda_s))'(t) = \mathrm{d}F_{\sigma(t)}(\sigma'(t))(\lambda_1(t), \ldots, \lambda_s(t)) + \sum_{i=1}^{s} F(\sigma(t))(\lambda_1(t), \ldots, \lambda_i'(t)), \ldots, \lambda_s(t))$$

hence

$$\|(\bar{F}(\sigma)(\lambda_1, \ldots, \lambda_s))'(t)\| \leq \|\mathrm{d}F_{\sigma(t)}\| \cdot \|\sigma'(t)\| \cdot \|\lambda_1(t)\| \ldots \|\lambda_s(t)\|$$
$$+ \sum_{i=1}^{s} \|F(\sigma(t))\| \cdot \|\lambda_1(t)\| \ldots \|\lambda_i'(t)\| \ldots \|\lambda_s(t)\|.$$

Since $\|\lambda_i\|_\infty \leq 2\|\lambda_i\|_1$ (Corollary (1) of Theorem (3)) we see that $\|(\bar{F}(\sigma)(\lambda_1, \ldots, \lambda_s))'\|_0 \leq 2^s L(\sigma)\|\lambda_1\|_1 \ldots \|\lambda_s\|_1$ where $L(\sigma) = \mathrm{Sup}\|\mathrm{d}F_{\sigma(t)}\| \cdot \|\sigma'\|_0 + s\,\mathrm{Sup}\|F(\sigma(t))\|$. Also $\|\bar{F}(\sigma)(\lambda_1, \ldots, \lambda_s)\|_\infty \leq 2^s\,\mathrm{Sup}\|F(\sigma(t))\| \cdot \|\lambda_1\|_1 \ldots \|\lambda_s\|_1$. Recalling that $\|\rho\|_1^2 = \|\rho(0)\|^2 + \|\rho'\|_0^2$ we see $\|\bar{F}(\sigma)(\lambda_1, \ldots, \lambda_s)\|_1 \leq K(\sigma)\|\lambda_1\|_1 \ldots \|\lambda_s\|_1$. Since $\bar{F}(\sigma)$ is clearly multilinear it follows that $\bar{F}(\sigma) \in L^s(H_1(I, \mathbf{R}^n), H_1(I, \mathbf{R}^p))$. If $\rho \in H_1(I, \mathbf{R}^n)$ then

$$\|(\bar{F}(\sigma) - \bar{F}(\rho))(\lambda_1, \ldots, \lambda_s)\|_\infty \leq 2^s\,\mathrm{Sup}\|F(\sigma(t)) - F(\rho(t))\| \cdot \|\lambda_1\|_1 \ldots \|\lambda_s\|_1$$

and a straightforward calculation gives

$$\|((\bar{F}(\sigma) - \bar{F}(\rho))(\lambda_1, \ldots, \lambda_s))'\|_0 \leq 2^s M(\sigma, \rho)\|\lambda_1\|_1 \ldots \|\lambda_s\|_1$$

where

$$M(\sigma, \rho) = \mathrm{Sup}\|\mathrm{d}F_{\sigma(t)}\| \cdot \|\sigma' - \rho'\|_0 + \mathrm{Sup}\|\mathrm{d}F_{\sigma(t)} - \mathrm{d}F_{\rho(t)}\| \cdot \|\rho'\|_0 + s\,\mathrm{Sup}\|F(\sigma(t)) - F(\rho(t))\|$$

so

$$|||\bar{F}(\sigma) - \bar{F}(\rho)||| \leq K(\sigma, \rho)$$

where $|||\ |||$ is the norm in $L^s(H_1(I, \mathbf{R}^n), H_1(I, \mathbf{R}^p))$, and $K(\sigma, \rho) \to 0$ if $\mathrm{Sup}\|F(\sigma(t)) - F(\rho(t))\|$, $\mathrm{Sup}\|\mathrm{d}F_{\sigma(t)} - \mathrm{d}F_{\rho(t)}\|$, and $\|\sigma' - \rho'\|_0$ all approach zero. But if $\rho \to \sigma$ in $H_1(I, \mathbf{R}^n)$ then $\|\sigma' - \rho'\|_0 \leq \|\sigma - \rho\|_1$ goes to zero and by Corollary (1) of Theorem (3) $\rho \to \sigma$ uniformly, hence since F and $\mathrm{d}F$ are continuous $F(\rho(t)) \to F\sigma((t))$ uniformly and $\mathrm{d}F_{\rho(t)} \to \mathrm{d}F_{\sigma(t)}$ uniformly, so $K(\sigma, \rho) \to 0$. Thus $|||\bar{F}(\sigma) - \bar{F}(\rho)||| \to 0$ so $\bar{F}$ is continuous. This proves the first part of the lemma. Now suppose F is C^3 so $\mathrm{d}F$ is C^2. By the mean value theorem there is a C^1-map $R : \mathbf{R}^n \to L^2(\mathbf{R}^n, L^s(\mathbf{R}^n, \mathbf{R}^p))$ such that if $x = p + v$ then $F(x) - F(p) - \mathrm{d}F_p(v) = R(x)(v, v)$. Then $\bar{R} : H_1(I, \mathbf{R}^n) \to L^2(H_1(I, \mathbf{R}^n), H_1(I, L^s(\mathbf{R}^n, \mathbf{R}^p)))$ is continuous by the first part of the theorem and if σ and $x = \rho + \sigma$ are in $H_1(I, \mathbf{R}^n)$ $\bar{F}(x) - \bar{F}(\sigma) - \mathrm{d}\bar{F}_\sigma(\rho) = \bar{R}(x)(\rho, \rho)$. It follows that $\bar{F}$ is differentiable at σ and $\mathrm{d}\bar{F}_\sigma = \overline{dF}_\sigma$. Since $\overline{dF}_\sigma$ is a continuous function of σ by the first part of the lemma $\bar{F}$ is C^1.

q.e.d.

The following is trivial:

THEOREM (5). *Consider $\mathbf{R}^m$ and $\mathbf{R}^n$ as complementary subspaces of $\mathbf{R}^{m+n}$. Then the map $(\lambda, \sigma) \to \lambda + \sigma$ is an isometry of $H_1(I, \mathbf{R}^m) \oplus H_1(I, \mathbf{R}^n)$ onto $H_1(I, \mathbf{R}^{m+n})$.*

DEFINITION. *If V is a finite dimensional C^1-manifold we define $H_1(I, V)$ to be the set of continuous maps σ of I into V such that $\varphi \circ \sigma$ is absolutely continuous and $\|(\varphi \circ \sigma)'\|$ locally square summable for each chart φ for V. If V is C^2 and $\sigma \in H_1(I, V)$ we define $H_1(I, V)_\sigma = \{\lambda \in H_1(I, T(V)) | \lambda(t) \in V_{\sigma(t)}$ for all $t \in I\}$. If $P, Q \in V$ we define $\Omega(V; P, Q) = \{\sigma \in H_1(I, V) | \sigma(0) = P, \sigma(1) = Q\}$ and if $\sigma \in \Omega(V; P, Q)$ we define $\Omega(V; P, Q)_\sigma =$*

$\{\lambda \in H_1(I, V)_\sigma | \lambda(0) = 0_P$ and $\lambda(1) = 0_Q\}$. We note that $H_1(I, V)_\sigma$ is a vector space under pointwise operations and that $\Omega(V; P, Q)_\sigma$ is a subspace of $H_1(I, V)_\sigma$.

THEOREM (6). *If V is a closed C^{k+4}-submanifold of $\mathbf{R}^n$ ($k \geq 1$) then $H_1(I, V)$ consists of all $\sigma \in H_1(I, \mathbf{R}^n)$ such that $\sigma(I) \subseteq V$ and is a closed C^k-submanifold of the Hilbert space $H_1(I, \mathbf{R}^n)$. If $P, Q \in V$ then $\Omega(V; P, Q)$ is a closed C^k-submanifold of $H_1(I, V)$. If $\sigma \in H_1(I, V)$ then the tangent space to $H_1(I, V)$ at σ (as a submanifold of $H_1(I, \mathbf{R}^n)$) is just $H_1(I, V)_\sigma$ which is equal to $\{\lambda \in H_1(I, \mathbf{R}^n) | \lambda(t) \in V_{\sigma(t)} t \in I\}$ and if $\sigma \in \Omega(V; P, Q)$ then the tangent space to $\Omega(V; P, Q)$ at σ is just $\Omega(V; P, Q)_\sigma$ which equals $\{\lambda \in H_1(I, V)_\sigma | \lambda(0) = \lambda(1) = 0\}$.*

Proof. That $H_1(I, V)$ equals the set of $\sigma \in H_1(I, \mathbf{R}^n)$ such that $\sigma(I) \subseteq V$ is clear, and so is the fact that $H_1(I, V)_\sigma$ and $\Omega(V; P, Q)_\sigma$ are what they are claimed to be. Since V is closed in $\mathbf{R}^n$ it follows that $H_1(I, V)$ is closed in $C^0(I, \mathbf{R}^n)$, hence in $H_1(I, \mathbf{R}^n)$ by Corollary (2) of Theorem (3). In the same way we see that $\Omega(V; P, Q)$ is closed in $H_1(I, \mathbf{R}^n)$ and that $H_1(I, V)_\sigma$ and $\Omega(V; P, Q)_\sigma$ are closed subspaces of $H_1(I, \mathbf{R}^n)$. Since V is a C^{k+4}-submanifold of $\mathbf{R}^n$ we can find a C^{k+3}-Riemannian metric for $\mathbf{R}^n$ such that V is a totally geodesic submanifold. Then if $E : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ is the corresponding exponential map (i.e. $t \to E(p, tv)$ is the geodesic starting from p with tangent vector v), E is a C^{k+2}-map. Let $\sigma \in H_1(I, V)$ and define $\varphi : H_1(I, \mathbf{R}^n) \to H_1(I, \mathbf{R}^n)$ by $\varphi(\lambda)(t) = E(\sigma(t), \lambda(t))$. Then by Theorems (4) and (5) φ is C^k and clearly $\varphi(0) = \sigma$. Moreover by Theorem (4) $d\varphi_0(\lambda)(t) = dE_0^{\sigma(t)}(\lambda(t))$ where $E^{\sigma(t)}(v) = E(\sigma(t), v)$. By a basic property of exponential maps $dE_0^{\sigma(t)}$ is the identity map of $\mathbf{R}^n$, hence $d\varphi_0$ is the identity map of $H_1(I, \mathbf{R}^n)$ so by the inverse function theorem φ maps a neighborhood of zero in $H_1(I, \mathbf{R}^n)$ C^k-isomorphically onto a neighborhood of σ in $H_1(I, \mathbf{R}^n)$. Since V is totally geodesic it follows that for λ near zero in $H_1(I, \mathbf{R}^n)$, $\varphi(\lambda) \in H_1(I, V)$ if and only if $\lambda \in H_1(I, V)_\sigma$ and similarly if $\sigma \in \Omega(V; P, Q)$ then $\varphi(\lambda) \in \Omega(V; P, Q)$ if and only if $\lambda \in \Omega(V; P, Q)_\sigma$. Consequently φ^{-1} restricted to a neighborhood of σ in $H_1(I, V)$ (respectively $\Omega(V; P, Q)$) is a chart in $H_1(I, V)$ (respectively $\Omega(V; P, Q)$) which is the restriction of a C^k-chart for $H_1(I, \mathbf{R}^n)$, so by definition $H_1(I, V)$ and $\Omega(V; P, Q)$ are closed C^k-submanifolds of $H_1(I, \mathbf{R}^n)$ and their tangent spaces at σ are respectively $H_1(I, V)_\sigma$ and $\Omega(V; P, Q)_\sigma$.

q.e.d.

THEOREM (7). *Let V and W be closed C^{k+4}-submanifolds of $\mathbf{R}^n$ and $\mathbf{R}^m$ respectively ($k \geq 1$) and let $\varphi : V \to W$ be a C^{k+4}-map. Then $\bar{\varphi} : H_1(I, V) \to H_1(I, W)$ defined by $\bar{\varphi}(\sigma) = \varphi \circ \sigma$ is a C^k-map of $H_1(I, V)$ into $H_1(I, W)$. Moreover $d\bar{\varphi}_\sigma : H_1(I, V)_\sigma \to H_1(I, W)_{\bar{\varphi}(\sigma)}$ is given by $d\bar{\varphi}_\sigma(\lambda)(t) = d\varphi_{\sigma(t)}(\lambda(t))$.*

Proof. By a well-known theorem of elementary differential topology φ can be extended to a C^{k+4}-map of $\mathbf{R}^n$ into $\mathbf{R}^m$ and Theorem (7) then follows from Theorems (4) and (6).

DEFINITION. *Let V be a C^{k+4}-manifold of finite dimension ($k \geq 1$) and let $j : V \to \mathbf{R}^n$ be a C^{k+4}-imbedding of V as a closed submanifold of a Euclidean space (such always exists by a theorem of Whitney). Then by Theorem (7) the C^k-structures induced on $H_1(I, V)$ and $\Omega(V; P, Q)$ as closed C^k-submanifolds of $H_1(I, \mathbf{R}^n)$ are independent of j. Henceforth we shall regard $H_1(I, V)$ and $\Omega(V; P, Q)$ as C^k-Hilbert manifolds. If $\varphi : V \to W$ is a C^{k+4}-map then by Theorem (7) $\bar{\varphi} : H_1(I, V) \to H_1(I, W)$ defined by $\bar{\varphi}(\sigma) = \varphi \circ \sigma$ is a C^k-map and $d\bar{\varphi}_\sigma(\lambda)(t) = d\varphi_{\sigma(t)}(\lambda(t))$. We note that $\bar{\varphi}$ maps $\Omega(V; P, Q)$ C^k into $\Omega(W; \varphi(P), \varphi(Q))$.*

THEOREM (8). *The function* $V \to H_1(I, V)$, $\varphi \to \bar{\varphi}$ *is a functor from the category of finite dimensional* C^{k+4}*-manifolds to the category of* C^k*-Hilbert manifolds* ($k \geq 1$).

DEFINITION. *Let* V *be a* C^{k+4}*-finite dimensional Riemannian manifold* ($k \geq 1$). *We define a real valued function* J^V *on* $H_1(I, V)$ *called the* action integral *by*

$$J^V(\sigma) = \frac{1}{2}\int_0^1 \|\sigma'(t)\|^2 \, dt.$$

THEOREM (9). *Let* V *and* W *be* C^{k+4}*-Riemannian manifolds of finite dimension and let* $\varphi : V \to W$ *be a* C^{k+4}*-local isometry. Then* $J^V = J^W \circ \bar{\varphi}$.

Proof. $\bar{\varphi}(\sigma)'(t) = (\varphi \circ \sigma)'(t) = d\varphi_{\sigma(t)}(\sigma'(t))$. Since $d\varphi_{\sigma(t)}$ maps $V_{\sigma(t)}$ isometrically into $W_{\varphi(\sigma(t))}$, $\|\bar{\varphi}(\sigma)'(t)\| = \|\sigma'(t)\|$ and the theorem follows.

COROLLARY (1). *If* V *is a* C^{k+4}*-Riemannian submanifold of the* C^{k+4}*-Riemannian manifold* W *then* $J^V = J^W | H_1(I, V)$.

COROLLARY (2). *If* V *is a closed* C^{k+4}*-submanifold of* $\mathbf{R}^n$ *then* $J^V(\sigma) = \frac{1}{2}\|L\sigma\|_0^2$. *Consequently* $J^V : H_1(I, V) \to \mathbf{R}$ *is a* C^k*-map.*

Proof. By definition $J^{\mathbf{R}^n}(\sigma) = \frac{1}{2}\|L\sigma\|_0^2$, so the first statement follows. Since $J^{\mathbf{R}^n}$ is a continuous quadratic form on $H_1(I, \mathbf{R}^n)$ (Theorem (1)), $J^{\mathbf{R}^n}$ is a C^∞-map of $H_1(I, \mathbf{R}^n)$ into $\mathbf{R}$, hence its restriction to the closed C^k-submanifold $H_1(I, V)$ is C^k.

q.e.d.

COROLLARY (3). *If* V *is a complete finite dimensional* C^{k+4}*-Riemannian manifold then* J^V *is a* C^k*-real valued function on* $H_1(I, V)$.

Proof. By a theorem of Nash [7] V can be C^{k+4}-imbedded isometrically in some $\mathbf{R}^n$, so Corollary (3) follows from Corollary (2).

Remark. Let W be a complete Riemannian manifold, V a closed submanifold of W and give V the Riemannian structure induced from W. Let ρ_V and ρ_W denote the Riemannian metrics on V and W. Then clearly if $p, q \in V$ $\rho_V(p, q) \geq \rho_W(p, q)$ since the right hand side is by definition an Inf over a larger set than the left. Hence if $\{p_n\}$ is a Cauchy sequence in V it is Cauchy in W and hence convergent in W and therefore in V because V is closed in W. Hence V is complete. From this we see that

THEOREM (10). *If* V *is a closed* C^{k+4}*-submanifold of* $\mathbf{R}^n$ *then* $H_1(I, V)$ *is a complete* C^k*-Riemannian manifold in the Riemannian structure induced on it as a closed* C^k*-submanifold of* $H_1(I, \mathbf{R}^n)$.

Caution. The Riemannian structure on $H_1(I, V)$ induced on it by an imbedding onto a closed submanifold of some $\mathbf{R}^n$ depends on the imbedding. To be more precise if V and W are closed submanifolds of Euclidean spaces and $\varphi : V \to W$ is an isometry it does *not* follow that $\bar{\varphi} : H_1(I, V) \to H_1(I, W)$ is an isometry. It seems reasonable to conjecture that $\bar{\varphi}$ is uniformly continuous but I do not know if this is true.

THEOREM (11). *If* V *is a closed* C^{k+4}*-submanifold of* $\mathbf{R}^n$ *and* $P, Q \in V$ *then* $\Omega(V; P, Q)$ *is included in a translate of* $H_1^*(I, \mathbf{R}^n)$, *and* $\Omega(V; P, Q)_\sigma \subseteq H_1^*(I, \mathbf{R}^n)$.

Proof. If σ and ρ are in $\Omega(V; P, Q)$ then $(\sigma - \rho)(0) = P - P = 0$ and $(\sigma - \rho)(1) = Q - Q = 0$, and the first statement follows. The second statement is of course a consequence of the first, but it is also a direct consequence of the definition of $\Omega(V; P, Q)_\sigma$.

COROLLARY (1). *If we regard $\Omega(V; P, Q)$ as a Riemannian submanifold of $H_1(I, \mathbf{R}^n)$ then the inner product $\langle , \rangle_\sigma$ in $\Omega(V; P, Q)_\sigma$ is given by $\langle \rho, \lambda \rangle_\sigma = \langle L\rho, L\lambda \rangle_0$.*

Proof. Immediate from Theorem (1).

COROLLARY (2). *If $S \subseteq \Omega(V; P, Q)$ and if J^V is bounded on S then S is totally bounded in $C^0(I, \mathbf{R}^n)$ and $H_0(I, \mathbf{R}^n)$.*

Proof. Since $J^V(\sigma) = \frac{1}{2}\|L\sigma\|_0^2$ (Corollary (2) of Theorem (9)) $\|L\sigma\|_0$ is bounded on S. Since S is included in a translate of $H_1^*(I, \mathbf{R}^n)$ it follows from Theorem (1) that S is bounded in $H_1(I, \mathbf{R}^n)$, hence by Corollary (2) of Theorem (3) S is totally bounded in $C^0(I, \mathbf{R}^n)$ and $H_0(I, \mathbf{R}^n)$.

COROLLARY (3). *If $\{\sigma_n\}$ is a sequence in $\Omega(V; P, Q)$ and $\|L(\sigma_n - \sigma_m)\|_0 \to 0$ as $n, m \to \infty$ then σ_n converges in $\Omega(V; P, Q)$.*

Proof. By Theorem (11) $\sigma_n - \sigma_m \in H_1^*(I, \mathbf{R}^n)$ hence by Theorem (1) $\{\sigma_n\}$ is Cauchy in $H_1(I, \mathbf{R}^n)$, hence convergent in $H_1(I, \mathbf{R}^n)$. Since $\Omega(V; P, Q)$ is closed in $H_1(I, \mathbf{R}^n)$ the corollary follows.

DEFINITION. *Let V be a closed C^{k+4}-submanifold of $\mathbf{R}^n$ $(k \geq 1)$ and let $P, Q \in V$. If $\sigma \in \Omega(V; P, Q)$ we define $h(\sigma)$ to be the orthogonal projection of $L\sigma$ on the orthogonal complement of $L(\Omega(V; P, Q)_\sigma)$ in $H_0(I, \mathbf{R}^n)$.*

THEOREM (12). *Let V be a closed C^{k+4}-submanifold of $\mathbf{R}^n$ $(k \geq 1)$, $P, Q \in V$ and let J be the restriction of J^V to $\Omega(V; P, Q)$. If we consider $\Omega(V; P, Q)$ as a Riemannian manifold in the structure induced on it as a closed submanifold of $H_1(I, \mathbf{R}^n)$, then for each $\sigma \in \Omega(V; P, Q)$ ∇J_σ can be characterized as the unique element of $\Omega(V; P, Q)_\sigma$ mapped by L onto $L\sigma - h(\sigma)$. Moreover $\|\nabla J_\sigma\|_\sigma = \|L\sigma - h(\sigma)\|_0$.*

Proof. Since $\Omega(V; P, Q)_\sigma$ is a closed subspace of $H_1(I, \mathbf{R}^n)$ (Theorem (6)) and is included in $H_1^*(I, \mathbf{R}^n)$ (Theorem (11)) it follows from Theorem (1) that L maps $\Omega(V; P, Q)_\sigma$ isometrically onto a closed subspace of $H_0(I, \mathbf{R}^n)$ which therefore is the orthogonal complement of its orthogonal complement. Since $L\sigma - h(\sigma)$ is orthogonal to the orthogonal complement of $L(\Omega(V; P, Q)_\sigma)$ it is therefore of the form $L\lambda$ for some $\lambda \in \Omega(V; P, Q)_\sigma$ and since L is an isometry on $\Omega(V; P, Q)_\sigma$ λ is unique and $\|\lambda\|_\sigma = \|L\lambda\|_0 = \|L\sigma - h(\sigma)\|_0$ so it will suffice, by the definition of ∇J_σ, to prove that $dJ_\sigma(\rho) = \langle \lambda, \rho \rangle_\sigma$ for $\rho \in \Omega(V; P, Q)_\sigma$, or by Corollary (1) of Theorem (11), that $dJ_\sigma(\rho) = \langle L\lambda, L\rho \rangle_0 = \langle L_\sigma - h(\sigma), L\rho \rangle_0$ for $\rho \in \Omega(V; P, Q)_\sigma$. Since by definition of $h(\sigma)$ we have $\langle h(\sigma), L\rho \rangle_0 = 0$ for $\rho \in \Omega(V; P, Q)_\sigma$ we must prove that $dJ_\sigma(\rho) = \langle L\sigma, L\rho \rangle_0$ for $\rho \in \Omega(V; P, Q)_\sigma$. Now $J^{\mathbf{R}^n}(\alpha) = \frac{1}{2}\|L\alpha\|_0^2$ (Corollary (2) of Theorem (9)) so $dJ_\sigma^{\mathbf{R}^n}(\rho) = \langle L\sigma, L\rho \rangle_0$ for $\rho \in H_1(I, \mathbf{R}^n)$. Since $J = J^{\mathbf{R}^n}|\Omega(V; P, Q)$ by Corollary (1) of Theorem (9) it follows that $dJ_\sigma = dJ_\sigma^{\mathbf{R}^n}|\Omega(V; P, Q)_\sigma$.

q.e.d

§14. VERIFICATION OF CONDITION (C) FOR THE ACTION INTEGRAL

In this section we assume that V is a closed C^{k+4}-submanifold of $\mathbf{R}^n$ $(k \geq 3)$, $P, Q \in V$ and $J = J^V|\Omega(V; P, Q)$. We recall from the preceding section that $\Omega(V; P, Q)$ is a complete C^k-Riemannian manifold in the Riemannian structure induced on it as a closed submanifold of $H_1(I, \mathbf{R}^n)$ and J is a C^k-real valued function. Our goal in this section is to identify the critical points of J as those elements of $\Omega(V; P, Q)$ which are geodesics of V parameterized proportionally to arc length, and secondly to prove that J satisfies condition (C).

DEFINITION. *We define a C^{k+3}-map $\Omega : V \to L(\mathbf{R}^n, \mathbf{R}^n)$ by $\Omega(p)$ = orthogonal projection of $\mathbf{R}^n$ on V_p. If $\sigma \in \Omega(V; P, Q)$ we define $\bar{\Omega}(V; P, Q)_\sigma$ to be the closure of $\Omega(V; P, Q)_\sigma$ in $H_0(I, \mathbf{R}^n)$ and we define P_σ to be the orthogonal projection of $H_0(I, \mathbf{R}^n)$ on $\bar{\Omega}(V; P, Q)_\sigma$.*

THEOREM (1). *If $\sigma \in \Omega(V; P, Q)$ then $\bar{\Omega}(V; P, Q)_\sigma = \{\lambda \in H_0(I, \mathbf{R}^n) | \lambda(t) \in V_{\sigma(t)}$ for almost all $t \in I\}$ and if $\lambda \in H_0(I, \mathbf{R}^n)$ then $(P_\sigma \lambda)(t) = \Omega(\sigma(t))\lambda(t)$.*

Proof. Define a linear transformation π_σ on $H_0(I, \mathbf{R}^n)$ by $(\pi_\sigma \lambda)(t) = \Omega(\sigma(t))\lambda(t)$. Since $\Omega(\sigma(t))$ is an orthogonal projection in $L(\mathbf{R}^n, \mathbf{R}^n)$ for each $t \in I$ it follows from the definition of the inner product in $H_0(I, \mathbf{R}^n)$ that π_σ is an orthogonal projection. From the characterization of $\Omega(V; P, Q)_\sigma$ in Theorem (6) of §13 it is clear that π_σ maps $H_1^*(I, \mathbf{R}^n)$ onto $\Omega(V; P, Q)_\sigma$. Since $H_1^*(I, \mathbf{R}^n)$ is dense in $H_0(I, \mathbf{R}^n)$ it follows that the range of π_σ is $\bar{\Omega}(V; P, Q)_\sigma$, hence $\pi_\sigma = P_\sigma$. On the other hand it is clear that $\lambda \in H_0(I, \mathbf{R}^n)$ is left fixed by π_σ if and only if $\lambda(t) \in V_{\sigma(t)}$ for almost all $t \in I$. Since the range of a projection is its set of fixed points this proves the theorem.

q.e.d.

COROLLARY (1). *If $\sigma \in \Omega(V; P, Q)$ then*

$$P_\sigma(H_1(I, \mathbf{R}^n)) = H_1(I, V)_\sigma$$

and

$$P_\sigma(H_1^*(I, R^n)) = \Omega(V; P, Q)_\sigma.$$

COROLLARY (2). *If $\sigma \in \Omega(V; P, Q)$ then $P_\sigma L\sigma = L\sigma$.*

Proof. Clearly $(L\sigma)(t) = \sigma'(t) \in V_{\sigma(t)}$ whenever $\sigma'(t)$ is defined, so $L\sigma \in \bar{\Omega}(V; P, Q)_\sigma$.

THEOREM (2). *Let $T \in H_0(I, L(\mathbf{R}^n, \mathbf{R}^p))$ and define for each $\lambda \in H_0(I, \mathbf{R}^n)$ a measureable function $\bar{T}(\lambda) : I \to \mathbf{R}^p$ by $\bar{T}(\lambda)(t) = T(t)\lambda(t)$. Then:*

(1) *$\bar{T}$ is a bounded linear transformation of $H_0(I, \mathbf{R}^n)$ into $L^1(I, \mathbf{R}^p)$;*

(2) *If T and λ are absolutely continuous then so is $\bar{T}\lambda$ and $(\bar{T}\lambda)'(t) = T'(t)\lambda(t) + T(t)\lambda'(t)$;*

(3) *If $T \in H_1(I, L(\mathbf{R}^n, \mathbf{R}^p))$, $\lambda \in H_1(I, \mathbf{R}^n)$ then $\bar{T}\lambda \in H_1(I, \mathbf{R}^p)$.*

Proof. If $n = p = 1$ then (1) follows from Schwartz's inequality, (2) is just the product rule for differentiation and (3) is an immediate consequence of (2). The general case follows from this case by choosing bases for $\mathbf{R}^n$ and $\mathbf{R}^p$ and looking at components.

DEFINITION. *Given $\sigma \in \Omega(V; P, Q)$ we define $G_\sigma \in H_1(I, L(\mathbf{R}^n, \mathbf{R}^n))$ by $G_\sigma = \Omega \circ \sigma$ and we define $Q_\sigma \in H_0(I, L(\mathbf{R}^n, \mathbf{R}^n))$ by $Q_\sigma = G_\sigma'$.*

Remark. That $G_\sigma \in H_1(I, L(\mathbf{R}^n, \mathbf{R}^n))$ follows from Theorem (4) of Section (13).

THEOREM (3). *Let $\sigma \in \Omega(V; P, Q)$. If $\rho \in H_1(I, \mathbf{R}^n)$ then $(LP_\sigma - P_\sigma L)\rho(t) = Q_\sigma(t)\rho(t)$ Given $f \in H_0(I, \mathbf{R}^n)$ define an absolutely continuous map $g : I \to \mathbf{R}^n$ by*

$$g(t) = \int_0^t Q_\sigma(s)f(s)\,\mathrm{d}s.$$

Then if $\rho \in H_1^*(I, \mathbf{R}^n)$

$$\langle f, (LP_\sigma - P_\sigma L)\rho\rangle_0 = \langle g, -L\rho\rangle_0.$$

Proof. Since $P_\sigma\rho(t) = G_\sigma(t)\rho(t)$ and $P_\sigma(L\rho)(t) = G_\sigma(t)\rho'(t)$ by Theorem (1). the fact that $(LP_\sigma - P_\sigma L)\rho(t) = Q_\sigma(t)\rho(t)$ is an immediate consequence of (2) of Theorem (2). By (1) of Theorem (2) $s \to Q_\sigma(s)f(s)$ is summable so g is absolutely continuous. Next note that since $G_\sigma(t) = \Omega(\sigma(t))$ is self-adjoint for all t, $Q_\sigma(t) = G_\sigma'(t)$ is self-adjoint wherever it is defined, hence

$$\langle f, (LP_\sigma - P_\sigma L)\rho\rangle_0 = \int_0^1 \langle f(t), Q_\sigma(t)\rho(t)\rangle\,\mathrm{d}t = \int_0^1 \langle Q_\sigma(t)f(t), \rho(t)\rangle\,\mathrm{d}t$$

$$= \int_0^1 \langle g'(t), \rho(t)\rangle\,\mathrm{d}t.$$

Then if $\rho \in H_1^*(I, \mathbf{R}^n)$ Theorem (2) of §13 gives

$$\langle f, (LP_\sigma - P_\sigma L)\rho\rangle_0 = \langle g, -L\rho\rangle_0 \qquad \text{q.e.d.}$$

We now recall that if $\sigma \in \Omega(V; P, Q)$ then in §13 we defined $h(\sigma)$ to be the orthogonal projection of $L\sigma$ on the orthogonal complement of $L(\Omega(V; P, Q)_\sigma)$ in $H_0(I, \mathbf{R}^n)$. By Corollary (1) of Theorem (1) above it follows that $\langle h(\sigma), LP_\sigma\rho\rangle = 0$ if $\rho \in H_1^*(I, \mathbf{R}^n)$.

THEOREM (4). *If $\sigma \in \Omega(V; P, Q)$ then $P_\sigma h(\sigma)$ is absolutely continuous and $(P_\sigma h(\sigma))'(t) = Q_\sigma(t)h(\sigma)(t)$.*

Proof. If $\rho \in H^*(I, \mathbf{R}^n)$ then

$$\langle P_\sigma h(\sigma), L\rho\rangle_0 = \langle h(\sigma), P_\sigma L\rho\rangle_0 = \langle h(\sigma), (P_\sigma L - LP_\sigma)\rho\rangle_0$$

since $\langle h(\sigma), LP_\sigma\rho\rangle = 0$. Hence by Theorem (3) $\langle P_\sigma h(\sigma), L\rho\rangle_0 = \langle g, L\rho\rangle_0$ if we define g to be the absolutely continuous map of $I \to \mathbf{R}^n$

$$g(t) = \int_0^t Q_\sigma(s)h(\sigma)(s)\,\mathrm{d}s.$$

Then $P_\sigma h(\sigma) - g$ is orthogonal to $L(H_1^*(I, \mathbf{R}^n))$ so by Theorem (1) of §13 $P_\sigma h(\sigma) - g$ is constant. Since g is absolutely continuous so is $P_\sigma h(\sigma)$ and they have the same derivative. But $g'(t) = Q_\sigma(t)h(\sigma)(t)$.

q.e.d.

THEOREM (5). *If σ is a critical point of J then $\sigma \in C^{k+4}(I, V)$ and moreover σ'' is everywhere orthogonal to V. Conversely given $\sigma \in \Omega(V; P, Q)$ such that σ' is absolutely continuous and $(\sigma')'$ is almost everywhere orthogonal to V, σ is a critical point of J.*

Proof. By Theorem (12) of §13 if σ is a critical point of J then $L\sigma = h(\sigma)$. Since

$P_\sigma L\sigma = L\sigma$ (Corollary (2) of Theorem (1) above) it follows that $P_\sigma h(\sigma) = h(\sigma)$ so by Theorem (4) σ' is absolutely continuous (so σ is C^1) and

$$(*) \qquad \sigma''(t) = Q_\sigma(t)\sigma'(t).$$

Now since $\Omega : V \to L(\mathbf{R}^n, \mathbf{R}^n)$ is C^{k+3} and

$$Q_\sigma(t) = \frac{\mathrm{d}}{\mathrm{d}t}\,\Omega(\sigma(t))$$

it follows that if σ is C^m $(1 \le m \le k + 3)$ then $Q_\sigma(t)$ is C^{m-1}, hence by $(*)$ σ'' is C^{m-1} so σ is C^{m+1}. Since we already know σ is C^1 we have a start for an induction that gives $\sigma \in C^{k+4}$. If $\rho \in \Omega(V; P, Q)_\sigma$ then $L\sigma = h(\sigma)$ is orthogonal to $L\rho$, so by Theorem (2) of §13 (and the fact that $\Omega(V; P, Q)_\sigma \subseteq H_1^*(I, \mathbf{R}^n)$—Theorem (11) of §13) σ'' is orthogonal to ρ. Since σ'' and ρ are continuous it follows that $\langle\sigma''(t), \rho(t)\rangle = 0$ for all $t \in I$. Now it is clear that if $t \in I$ is not an endpoint of I and $v_0 \in V_{\sigma(t)}$ then there exists $\rho \in \Omega(V; P, Q)_\sigma$ such that $\rho(t) = v_0$, hence $\sigma''(t)$ is orthogonal to $V_{\sigma(t)}$, and by continuity this holds at the endpoints of I also. Conversely suppose $\sigma \in \Omega(V; P, Q)$ is such that σ' is absolutely continuous and $\sigma''(t)$ is orthogonal to $V_{\sigma(t)}$ for almost all $t \in I$. Then by Theorem (2) of §13 $L\sigma$ is orthogonal to $L(\Omega(V; P, Q)_\sigma)$ so $L\sigma = h(\sigma)$ and by Theorem (12) of §13 σ is a critical point of J.

q.e.d.

COROLLARY. *If $\sigma \in \Omega(V; P, Q)$ then σ is a critical point of J if and only if σ is a geodesic of V parameterized proportionally to arc length.*

Proof. It is a well-known result of elementary differential geometry that $\sigma \in C^2(I, V)$ is a geodesic of V parameterized proportionally to arc length if and only if σ'' is everywhere orthogonal to V†

LEMMA. *Given a compact subset A of V there is a constant K such that*

$$\int_0^1 \|Q_\sigma(t)\rho(t)\|\,\mathrm{d}t \le K\|L\sigma\|_0\|\rho\|_0$$

for all $\rho \in H_0(I, \mathbf{R}^n)$ and all $\sigma \in H_1(I, \mathbf{R}^n)$ such that $\sigma(I) \subseteq A$.

Proof. Let A^* be the compact subset of $\mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}^n$ consisting of triples (p, v, x) such that $p \in A$, v is a unit vector in V_p and x is a unit vector in $\mathbf{R}^n$. Since Ω is C^{k+3}, $(p, v, x) \to \|\mathrm{d}\Omega_p(v)x\|$ is continuous on A^* and hence bounded by some constant K. Since

$$Q_\sigma(t) = \frac{\mathrm{d}}{\mathrm{d}t}\,G_\sigma(t) = \frac{\mathrm{d}}{\mathrm{d}t}\Omega(\sigma(t)) = \mathrm{d}\Omega_{\sigma(t)}(\sigma'(t))$$

it follows that

$$\|Q_\sigma(t)\rho(t)\| \le K\|\sigma'(t)\|\cdot\|\rho(t)\|.$$

Integrating and applying Schwartz's inequality gives the desired inequality.

q.e.d.

We now come to the proof of condition (C).

---

† See EISENHART: *An Introduction to Differential Geometry*, p. 246.

THEOREM (6). *Let $S \subseteq \Omega(V; P, Q)$ and suppose J is bounded on S but that $\|\nabla J\|$ is not bounded away from zero on S. Then there is a critical point of J adherent to S.*

Proof. By Theorem (12) of §13 we can choose a sequence $\{\sigma_n\}$ in S such that $\|\nabla J_{\sigma_n}\| = \|L\sigma_n - h(\sigma_n)\|_0 \to 0$. Since each P_{σ_n} is a projection, hence norm decreasing, it follows from Corollary (2) of Theorem (1) of this section that $\|L\sigma_n - P_{\sigma_n}h(\sigma_n)\|_0 \to 0$, and by Corollary (2) of Theorem (11), §13, we can assume that $\|\sigma_n - \sigma_m\|_\infty \to 0$ as $m, n \to \infty$. It will suffice to prove that $\|L(\sigma_n - \sigma_m)\|_0 \to 0$ as $m, n \to \infty$ for then by Corollary (3) of Theorem (11), §13, σ_n will converge in $(\Omega V; P, Q)$ to a point σ in the closure of S, and since $\|\nabla J\|$ is continuous it will follow that $\|\nabla J_\sigma\| = 0$, i.e. σ is a critical point of J. But

$$\|L(\sigma_n - \sigma_m)\|_0^2 = \langle L\sigma_n, L(\sigma_n - \sigma_m)\rangle_0 - \langle L\sigma_m, L(\sigma_n - \sigma_m)\rangle_0$$

hence it will in turn suffice to prove that $\langle L\sigma_n, L(\sigma_n - \sigma_m)\rangle_0 \to 0$ as $m, n \to \infty$. Now $\|L\sigma_n\|^2 = 2J(\sigma_n)$ is bounded, hence $\|L(\sigma_n - \sigma_m)\|_0$ is bounded, and since $L_{\sigma_n} - P_{\sigma_n}h(\sigma_n) \to 0$ in $H_0(I, \mathbf{R}^n)$ it will suffice to prove that $\langle P_{\sigma_n}h(\sigma_n), L(\sigma_n - \sigma_m)\rangle_0 \to 0$ as $m, n \to \infty$. Recalling that $\sigma_n - \sigma_m \in H_1^*(I, \mathbf{R}^n)$ (Theorem (11) of §13) it follows from Theorem (4) above and Theorem (2) of §13 that

$$\begin{aligned}|\langle P_{\sigma_n}h(\sigma_n), L(\sigma_n - \sigma_m)\rangle_0| &= \left|\int_0^1 \langle Q_{\sigma_n}(t)h(\sigma_n)(t), (\sigma_n - \sigma_m)(t)\rangle\, dt\right| \\ &\le \|\sigma_n - \sigma_m\|_\infty \int_0^1 \|Q_{\sigma_n}(t)h(\sigma_n)(t)\|\, dt\end{aligned}$$

and since $\|\sigma_n - \sigma_m\|_\infty \to 0$ it will suffice to prove that

$$\int_0^1 \|Q_{\sigma_n}(t)h(\sigma_n)(t)\|\, dt$$

is bounded. Let A be a compact set such that $\sigma_n(I) \subset A$ (the existence of A follows from the fact that $\{\sigma_n\}$ being uniformly Cauchy is uniformly bounded). By the lemma there exists K such that

$$\int_0^1 \|Q_{\sigma_n}(t)h(\sigma_n)(t)\|\, dt \le K\|L\sigma_n\|_0\|h(\sigma_n)\|_0.$$

Now it has already been noted that $\|L\sigma_n\|_0$ is bounded, and since $\|L\sigma_n - h(\sigma_n)\|_0 \to 0$ so is $\|h(\sigma_n)\|_0$.

q.e.d.

For the sake of completeness we give here a brief description of the classical conditions that the critical points of J be non-degenerate and of a geometrical form of the Morse Index Theorem†.

Let E denote the exponential map of V_p into V; i.e. if $v \in V_p$ then $E(v) = \sigma(\|v\|)$ where σ is the geodesic starting from P with tangent vector $v/\|v\|$. Then E is a C^{k+2}-map. Given $v \in V_p$ define $\lambda(v) =$ dimension of null-space of dE_v. If $\lambda(v) > 0$ we call v a conjugate vector at P. A point of V is called a conjugate point of P if it is in the image under E of

† For a detailed exposition the reader is referred to I. M. Singer's *Notes on Differential Geometry* (Mimeographed, Massachusetts Institute of Technology, 1962).

the set of conjugate vectors at P. By an easy special case of Sard's Theorem† the set of conjugate points of P has measure zero and in particular is nowhere dense in V.

Given $v \in E^{-1}(Q)$ define $\tilde{v} \in \Omega(V; P, Q)$ by $\tilde{v}(t) = E(t(v))$. Then $\tilde{v}$ is a geodesic parameterized proportionally to arc length (the proportionality factor being $\|v\|$), hence a critical point of J by corollary of Theorem (5), and conversely by the same corollary any critical point of J is of the form $\tilde{v}$ for a unique $v \in E^{-1}(Q)$.

Non-degeneracy Theorem

If $v \in E^{-1}(Q)$ then $\tilde{v}$ is a degenerate critical point of J if and only if v is a conjugate vector at P, hence J has only non-degenerate critical points if and only if Q is not a conjugate point of P. It follows that if Q is chosen outside a set of measure zero in V then $J : \Omega(V; P, Q) \to \mathbf{R}$ has only non-degenerate critical points.

Morse Index Theorem

Let $v \in E^{-1}(Q)$. Then there are only a finite number of t satisfying $0 < t < 1$ such that tv is a conjugate vector at P and the index of $\tilde{v} = \sum_{0<t<1} \lambda(tv)$. In particular each critical point of $J : \Omega(V; P, Q) \to \mathbf{R}$ has finite index.

§15. TOPOLOGICAL IMPLICATIONS

Until now we have given no indication of why one would like to prove theorems such as the Main Theorem. Roughly speaking the answer is that as a consequence of the Main Theorem one is able to derive inequalities relating the number of critical points of a given index with certain topological invariants of the manifold on which the function is defined. These are the famous Morse Inequalities and are useful read in either direction. That is, if we know certain facts about the topology of the manifold they imply existential statements about critical points, and conversely if we know certain facts about the critical point structure we can deduce that the topology of the manifold can be only so complicated.

As a start in this direction we will show that if M is a complete C^2-Riemannian manifold and $f : M \to \mathbf{R}$ is a C^2-function bounded below and satisfying condition (C) then on each component of M f assumes its lower bound. Note that we do *not* assume that the critical points of f are non-degenerate, however since it is clear that a point where f assumes a local minimum is a critical point, and is of index zero if non-degenerate, it follows that if the critical points of f are all non-degenerate then there are at least as many critical points of index zero as there are components of M. This is the first Morse inequality.

In what follows we denote the frontier of a set K by $\dot{K}$.

THEOREM (1). *Let M be a connected C^1-manifold $f : M \to \mathbf{R}$ a non-constant C^1-function and K the set of critical points of f. Then $f(K) = f(\dot{K})$.*

† DE RHAM: *Variete's Differentiable*, p. 10.

Proof. Let $p \in K$. We will find $x \in \dot{K}$ such that $f(x) = f(p)$. Choose $q \in M$ with $f(q) \neq f(p)$ and $\sigma : I \to M$ a C^1-path such that $\sigma(0) = p$ and $\sigma(1) = q$ and let $g(t) = f(\sigma(t))$. Then $g'(t) = df_{\sigma(t)}(\sigma'(t))$ and since g is not constant, g' is not identically zero, so $\sigma(I)$ is not included in K. Let $t_0 = \text{Inf}\{t \in I | \sigma(t) \notin K\}$. Then $x = \sigma(t_0) \in \dot{K}$ and since $g'(t) = 0$ for $0 \leq t \leq t_0$, $f(x) = g(t_0) = g(0) = f(p)$.

q.e.d.

THEOREM (2). *Let M be a C^1-Riemannian manifold, $f: M \to \mathbf{R}$ a C^1-function satisfying condition* (C) *and K the set of critical points of f. Then $f|\dot{K}$ is proper; i.e. given $-\infty < a < b < \infty$, $\dot{K} \cap f^{-1}([a, b])$ is compact (note we do not assume that M is complete).*

Proof. Let $\{p_n\}$ be a sequence in $\dot{K}$ with $a \leq f(p_n) \leq b$. Since K is closed it will suffice to prove that $\{p_n\}$ has a convergent subsequence. Since $p_n \in \dot{K}$ we can choose $q_n \notin K$ arbitrarily close to p_n. In particular since $\|\nabla f\|$ is continuous and $\|\nabla f_{p_n}\| = 0$ we can choose q_n so close to p_n that

$$\|\nabla f_{q_n}\| < \frac{1}{n}, \qquad a - 1 < f(q_n) < b + 1$$

and also

$$\rho(q_n, p_n) < \frac{1}{n}$$

where ρ is the Riemannian metric for M. Then by condition (C) a subsequence of $\{q_n\}$ will converge to a critical point p of f. Since

$$\rho(q_n, p_n) < \frac{1}{n}$$

the corresponding subsequence of $\{p_n\}$ will also converge to p. q.e.d.

Remark. $f|K$ need not be proper—for example if M is not compact and f is constant then f trivially satisfies condition (C) and $K = M$.

THEOREM (3)†. *Let M be a complete C^2-Riemannian manifold without boundary, $f: M \to \mathbf{R}$ a C^2-function satisfying condition* (C) *and $\sigma : (\alpha, \beta) \to M$ a maximum integral curve of ∇f. Then either* $\lim f(\sigma(t)) = \infty$ *or else $\beta = \infty$ and $\sigma(t)$ has a critical point of f as a limit point as $t \to \infty$. Similarly either* $\lim f(\sigma(t)) = -\infty$ *or else $\alpha = -\infty$ and $\sigma(t)$ has a critical point of f as limit point as $t \to -\infty$.*

Proof. This is just the lemma to Proposition (2) of §10 restated verbatim. We simply note that in the proof of that lemma we did not use the standing assumptions of §10 that f was C^3 or that the critical points of f were non-degenerate.

THEOREM (4). *Let M be a complete C^2-Riemannian manifold and $f: M \to \mathbf{R}$ a C^2-function satisfying condition* (C). *If f is bounded below on a component M_0 of M then $f|M_0$ assumes its greatest lower bound.*

† In this regard see also Proposition (3) of §12.

Proof. We can assume that M is connected. Let $B = \operatorname{Inf}\{f(x) | x \in M\}$. Given $\varepsilon > 0$ choose $p \in M$ such that $f(p) < B + \varepsilon$. If $\sigma : (\alpha, \beta) \to M$ is the maximum integral curve of ∇f with initial condition p then by Theorem (3) $\alpha = -\infty$ and $\sigma(t)$ has a critical point q as limit point as $t \to -\infty$. Since $f(\sigma(t))$ is monotonic non-decreasing $f(q) < B + \varepsilon$. Since the theorem is trivial if f is constant we can assume f is not constant and it follows from Theorem (1) that if K is the set of critical points of f we can find x in $\dot{K}$, the frontier of K, such that $f(x) < B + \epsilon$. Choose $x_n \in \dot{K}$ such that

$$B \leq f(x_n) < B + \frac{1}{n}.$$

Then by Theorem (2) a subsequence of $\{x_n\}$ will converge to a point x and clearly $f(x) = B$.
q.e.d.

COROLLARY (1). *If the set of critical points of f has no interior and if f is bounded below on M then f assumes its greatest lower bound.*

Proof. If B is the greatest lower bound of f then for every positive integer n we can choose $x_n \in \dot{K}$ (a minimum of f on some component of M) such that

$$B \leq f(x_n) \leq B + \frac{1}{n}$$

Since K has no interior and is closed $K = \dot{K}$, so by Theorem (2) a subsequence of $\{x_n\}$ will converge to a point x where $f(x) = B$.

q.e.d.

COROLLARY (2). *If V is a C^6 complete Riemannian manifold and $P, Q \in V$ then the action integral J^V assumes its greatest lower bound on each component of $\Omega(V; P, Q)$ and also on $\Omega(V; P, Q)$.*

Proof. We saw in Theorem (6) of §14 that condition (C) is satisfied. If K is the set of critical points of $J^V | \Omega(V; P, Q)$ then by the corollary of Theorem (5) of §14 the elements σ of K are (geodesics) parameterized proportionally to arc length. By making a small parameter change we can get element of $\Omega(V; P, Q)$ arbitrarily close to σ which are not parameterized proportionally to arc length, hence K has no interior.

Remark. If V is a complete Riemannian manifold and $P, Q \in V$ then given an absolutely continuous path $\sigma : I \to V$ with $\sigma(0) = P$, $\sigma(1) = Q$ define the length of σ, $L(\sigma)$, by

$$L(\sigma) = \int_0^1 \|\sigma'(t)\| \, dt.$$

Then by Schwartz's inequality if $\sigma \in \Omega(V; P, Q)$ $L(\sigma) \leq (2J(\sigma))^{1/2}$ and moreover equality occurs if and only if $\|\sigma'\|$ is constant, i.e. if and only if σ is parameterized proportionally to arc length. Now if $\sigma : I \to V$ is absolutely continuous and $\sigma(0) = P$, $\sigma(1) = Q$ we can reparameterize σ proportionally to arc length, getting $\gamma : I \to V$. Then $\gamma \in \Omega(V; P, Q)$ and since arc length is independent of parameterization $L(\gamma) = L(\sigma)$. Since reparameterizing also does not affect the homotopy class of σ we see that if J^V assumes its greatest lower bound on a component of $\Omega(V; P, Q)$ at a point γ (so that γ is a geodesic parameterized proportionally to arc length) then among all absolutely continuous paths joining P to Q and homotopic to γ, γ has the smallest length. Together with the preceding corollary this gives:

THEOREM (5). *If V is a complete C^6-Riemannian manifold, $P, Q \in V$ then given any homotopy class of paths joining $P \in Q$ there is a geodesic in this class whose length is less than or equal to that of any other absolutely continuous path in the class. Moreover there is a geodesic joining P to Q whose length is $\rho(p, q)$.*

Let H_i be a Hilbert space of dimension d_i, $i = 1, \ldots, n$, D_i the closed unit disc in H_i and S_i the unit sphere in H_i. Let $g_i : S_i \to X$ be continuous maps with disjoint images in a topological space X. We form a new space $X \cup_{g_1} D_1 \cup \ldots \cup_{g_n} D_n$ (called the result of attaching cells of dimension $d_1, \ldots, d_n$ to X by attaching maps $g_1, \ldots, g_n$) by taking the topological sum of X and the D_i and identifying $y \in D_i$ with $g_i(y) \in X$. Suppose now that $d_i < \infty$ $i = 1, \ldots, m$ and $d_i = \infty$ $i > m$. Then $X \cup_{g_1} D_1 \cup \ldots \cup_{g_m} D_m$ is a strong deformation retract of $X \cup_{g_1} D_1 \cup \ldots \cup_{g_n} D_n$. It will suffice to prove that if D is the unit disc and S the unit sphere in a Hilbert space H of infinite dimension then S is a strong deformation retract of D, or since D is convex it will suffice to find a retraction $\rho : D \to S$. By a theorem of Klee [2.2 of 3] there is a fixed point free map $h : D \to D$ (to see this note that if $\{x_n\}_{n \in Z}$ is a complete orthonormal basis for H then

$$f(t) = \left(\cos \frac{\pi(t-n)}{2}\right)_{x_n} + \left(\sin \frac{\pi(t-n)}{2}\right) x_{n+1} \qquad n \le t \le n+1$$

defines a topological embedding of $\mathbf{R}$ onto a closed subset F of D. Since F is an absolute retract the fixed point free map $f(t) \to f(t+1)$ of F into F can be extended to a map $h : D \to F$ which is clearly fixed point free). We define $\rho : D \to S$ by $\rho(x) =$ point where the directed line segment from $h(x)$ to x meets S.

It now follows (by excision) that if H_* denotes the singular homology functor with any coefficient group G then

$$H_*(X \cup_{g_1} D_1 \cup \ldots \cup_{g_n} D_n, X) \approx \sum_{i=1}^{m} H_*(D^{d_i}, S^{d_i - 1})$$

hence for any positive integer r

$$H_r(X \cup_{g_1} D_1 \cup \ldots \cup_{g_n} D_n, X) \approx G^{p(r)}$$

where $p(r)$ is the number of indices $i = 1, \ldots, n$ such that $d_i = r$.

Next let N be a Hilbert manifold with boundary and suppose M arises from N by disjoint C^r-attachments $(f_1, \ldots, f_n)$ of handles of type $(d_1, e_1), \ldots, (d_n, e_n)$ (§11). Define attaching maps $g_i : S^{di-1} \to \partial N$ by $g_i(y) = f_i(y, 0)$. (Note that since $f_i : D^{di} \times D^{ei} \to M$ is a homeomorphism each g_i is a homeomorphism.) Then clearly $N \cup f_1(D^{d_1} \times 0) \cup \ldots \cup f_n(D^{d_n} \times 0)$ can be identified with $N \cup_{g_1} D^{d_1} \cup \ldots \cup_{g_n} D^{d_n}$. We shall now prove that

$$N \cup \bigcup_{i=1}^{n} f_i(D^{d_i} \times 0)$$

is a strong deformation retract of M, hence by what we have just proved above that if $d_i < \infty$ $i = 1, \ldots, m$, $d_i = \infty$ $i > m$ then $N \cup_{g_1} D^{d_1} \cup \ldots \cup_{g} D^{d_m}$ is a strong deformation retract of M. It will suffice to prove that $(D^d \times 0) \cup (S^{d-1} \times D^e)$ is a strong deformation retract of $D^d \times D^e$, and since $D^d \times D^e$ is convex it will suffice to define a retraction r of $D^d \times D^e$ onto $(D^d \times 0) \cup (S^{d-1} \times D^e)$. Define $r(x, 0) = (x, 0)$ and if $y \neq 0$ define

$$r(x, y) = \left(\frac{2x}{2 - \|y\|}, 0\right) \quad \text{if} \quad \|x\| \leq 1 - \frac{\|y\|}{2}$$

$$r(x, y) = \left(\frac{x}{\|x\|}, (2\|x\| + \|y\| - 2)\frac{y}{\|y\|}\right) \quad \text{if} \quad \|x\| \geq 1 - \frac{\|y\|}{2}.$$

From the above remarks together with the theorem of §12 we deduce:

THEOREM (6). *Let M be a complete C^3-Riemannian manifold, $f: M \to \mathbf{R}$ a C^3-function satisfying condition* (C) *all of whose critical points are non-degenerate, c a critical value of f, $p_1, \ldots, p_n$ the critical points of finite index on the level c, and let d_i be the index of p_i. If c is the only critical value of f in a closed interval $[a, b]$ then M_b has as a deformation retract M_a with cells of dimension $d_1, \ldots, d_n$ disjointly attached to ∂M_a by homeomorphisms of the boundary spheres. Hence if H_k denotes the singular homology functor in dimension k with coefficient group G then $H_k(M_b, M_a) \approx G^{C(k)}$ where $C(k)$ is the number of critical points of index k on the level c.*

Remark. The surprising fact about Theorem (6) is that the homotopy type of (M_b, M_a) depends only on the critical points of finite index on the level c, those of infinite index being homotopically invisible. This is of course just a reflexion of the theorem of Klee that the unit disc modulo its boundary in an infinite dimensional Hilbert space is homotopically trivial. If it were not for this unexpected phenomenon we would have to make the rather unaesthetic assumption that all critical points were of finite index in order to derive Morse Inequalities.

In deriving the Morse Inequalities we shall follow Milnor closely. Let F denote a fixed field and H_* the singular homology functor with coefficients F. We call a pair of spaces (X, Y) admissible if $H_*(X, Y)$ is of finite type, i.e. each $H_k(X, Y)$ is finite dimensional and $H_k(X, Y) = 0$ except for finitely many k. From the exact homology sequence of a triple (X, Y, Z) it follows that if (X, Y) and (Y, Z) are admissible then so is (X, Z). We call an integer valued function S on admissible pairs subadditive if $S(X, Z) \leq S(X, Y) + S(Y, Z)$ for all triples (X, Y, Z) such that (X, Y) and (Y, Z) are admissible, and S is called additive if equality always holds in the above inequality. Then by an easy induction if $X_n \supseteq X_{n-1} \supseteq \ldots \supseteq X_0$ and each (X_{i+1}, X_i) is admissible it follows that (X_n, X_0) is admissible and

$$S(X_n, X_0) \leq \sum_{i=0}^{n-1} S(X_{i+1}, X_i)$$

if S is subadditive, equality holding if S is additive.

DEFINITION. *For each non-negative integer k we define integer valued functions R_k and S_k on admissible pairs by $R_k(X, Y) = \dim H_k(X, Y)$ and*

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

We define the Euler characteristic χ for admissible pairs by

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

LEMMA. *R_k and S_k are subadditive and χ is additive.*

Proof. Let (X, Y, Z) be a triple of spaces such that (X, Y) and (Y, Z) are admissible. From the long exact homology sequence of 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 derive the usual three short exact sequences

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

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

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

from which follow

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

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

$$R_m(X, Y) = \dim \mathrm{im}(j_m) + \dim \mathrm{im}(\partial_m)$$

hence

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

If we multiply $(*)$ by $(-1)^{k-m}$ and sum over m from $m = 0$ to $m = k$ we get

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

which is negative since in fact $\partial_0 = 0$. Similarly if we multiply $(*)$ by $(-1)^m$ and sum over all non-negative m we get $\chi(X, Z) - \chi(X, Y) - \chi(Y, Z) = 0$ since $\partial_{k+1} = 0$ for k sufficiently large.

q.e.d.

Now let f and M be as in Theorem (6). Let $-\infty < a < b < \infty$ and suppose a and b are regular values of f. Let $c_1, \ldots, c_n$ be the distinct critical values of f in $[a, b]$ in increasing order. Choose a_i, $i = 0, \ldots, n$ so that $a = a_0 < c_1 < a_1 < c_2 < \ldots < a_{n-1} < c_n < a_n = b$ and put $X_i = M_{a_i} = \{x \in M \mid f(x) \leq a_i\}$. Then by Theorem (6) (X_{i+1}, X_i) is admissible and $R_k(X_{i+1}, X_i)$ = number of critical points of index k on the level c_i. Hence

$$S_k(X_{i+1}, X_i) = \sum_{m=0}^{k} (-1)^{k-m} \quad \text{(number of critical points of index } m \text{ on level } c_i)$$

and

$$\chi(X_{i+1}, X_i) = \sum_{m=0}^{\infty} (-1)^m \quad \text{(number of critical points of index } m \text{ on the level } c_i).$$

Hence

$$\sum_{i=0}^{n-1} S_k(X_{i+1}, X_i) = \sum_{m=0}^{k} (-1)^{k-m} \quad \text{(number of critical points of index } m \text{ in } f^{-1}([a, b])$$

while

$$\sum_{i=0}^{n-1} \chi(X_{i+1}, X_i) = \sum_{m=0}^{\infty} (-1)^m \text{ (number of critical points of index } m \text{ in } f^{-1}([a, b])\text{)}.$$

Since S_k and χ are subadditive and additive respectively we deduce

THEOREM (7). (MORSE INEQUALITIES.) *Let M be a complete C^3-Riemannian manifold, $f: M \to \mathbf{R}$ a C^3-function satisfying condition (C) all of whose critical points are non-degenerate. Let a and b be regular values of f, $a < b$. For each non-negative integer m let R_m denote the* mth *betti-number of (M_b, M_a) relative to some fixed field F and let C_m denote the number of critical points of f of index m in $f^{-1}([a, b])$. Then*

$$R_0 \leq C_0$$

$$R_1 - R_0 \leq C_1 - C_0$$

$$\sum_{m=0}^{k} (-1)^{k-m} R_m \leq \sum_{m=0}^{k} (-1)^{k-m} C_m$$

and

$$\chi(M_a, M_b) = \sum_{m=0}^{\infty} (-1)^m R_m = \sum_{m=0}^{\infty} (-1)^m C_m.$$

COROLLARY (1). *$R_m \leq C_m$ for all m.*

COROLLARY (2). *If f is bounded below then the conclusions of the theorem and of Corollary* (1) *remain valid if we interpret $R_m =$* mth *betti-number of M_b and $C_m =$ number of critical points of f having index m in M_b respectively.*

Proof. Choose $a < glb\, f$.

COROLLARY (3). *If f is bounded below then for each non-negative integer m $R^*_m \leq C^*_m$, where R^*_m is the* mth *betti-number of M and C^*_m is the total number of critical points of f having index m.* (Of course either or both of R^*_m and C^*_m may be infinite.)

Proof. By Corollary (2) we have $C^*_m \geq R_m(M_b)$ for any regular value b of f. Hence it will suffice to show that if $R^*_m = \dim H_m(M; F) \geq k$ for some non-negative integer k then $R_m(M_b) \geq k$ for some regular value b of f. Let $h_1, \ldots, h_k$ be linearly independent elements of $H_m(M; F)$, $z_1, \ldots, z_k$ singular cycles of M which represent them, and C a compact set containing the supports of $z_1, \ldots, z_m$. Then as $b \to \infty$ through regular values of f the interiors of the M_b form an increasing family of open sets which exhaust M, hence $C \subseteq M_b$ for some regular value b of f. Then $z_1, \ldots, z_k$ are singular cycles of M_b, moreover no non-trivial linear combination of them could be homologous to zero in M_b since that same combination would *a fortiori* be homologous to zero in M. Hence $R_m(M_b) \geq k$.

q.e.d.

Caution. The assumption that f is bounded below is necessary in Corollary (3) as can be seen by considering the identity map of $\mathbf{R}$ which has no critical points even though $R^*_0(\mathbf{R}) = 1$.

We refer the reader to [8] for more delicate forms of the Morse Inequalities.

Remark. If V is a complete C^6-Riemannian manifold, $P, Q \in V$ define $\Omega_{P,Q}(V)$ to be the set of continuous maps $\sigma : I \to V$ such that $\sigma(0) = P$ and $\sigma(1) = Q$, in the compact open topology. The standard techniques of homotopy theory relate the topology of V and that of $\Omega_{P,Q}(V)$, while Theorem (7) and the results of §14 together give results concerning the topology of $\Omega(V; P, Q)$. Clearly some sort of bridge theorem relating $\Omega_{P,Q}(V)$ and $\Omega(V; P, Q)$ is desirable. Now if V is imbedded as a closed submanifold of $\mathbf{R}^n$ then $\Omega(V; P, Q)$ is a closed submanifold of $H_1(I, \mathbf{R}^n)$. While $\Omega_{P,Q}(V)$ is a subspace of $C^0(I, \mathbf{R}^n)$, hence it follows from Corollary (2) of Theorem (3) (§14) that the inclusion map $i : \Omega(V; P, Q) \to \Omega_{P,Q}(V)$ is continuous. The desired bridge theorem is the statement that i is in fact a homotopy equivalence. A homotopy inverse can be constructed by using smoothing operators of convolution type.

§16. MORE GENERAL CALCULUS OF VARIATIONS PROBLEMS

Given an n-tuple $\alpha = (\alpha_1, \ldots, \alpha_n)$ of non-negative integers let

$$|\alpha| = \sum_{i=1}^{n} \alpha_i$$

and let $D^\alpha = \partial^{|\alpha|}/\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}$. We define norms $\| \; \|_k$ on the space $C^\infty(D^n, \mathbf{R}^m)$ of C^∞-maps of the n-disk into $\mathbf{R}^m$ by $\|f\|_0^2 = \int \|f(x)\|^2 \, du(x)$ (where u is Lebesgue measure on D^n) and

$$\|f\|_k^2 = \sum_{|\alpha| \le k} \|D^\alpha f\|_0^2.$$

Then the completion of $C^\infty(D^n, \mathbf{R}^m)$ relative to the norm $\| \; \|_k$ is a Hilbert space which we denote by $H_k(D^n, \mathbf{R}^m)$. We denote by $H_k^*(D^n, \mathbf{R}^m)$ the closure in $H_k(D^n, \mathbf{R}^m)$ of the set of f in $C^\infty(D^n, \mathbf{R}^m)$ such that $(D^\alpha f)(x) = 0$ if $x \in S^{n-1}$ and $|\alpha| \le k - 1$. Let V be closed C^∞-submanifold of $\mathbf{R}^m$ and let $H_k(D^n, V) = \{f \in H_k(D^n, \mathbf{R}^m) | \mathbf{F}(D^n) \subseteq V\}$. If $g \in H_k(D^n, V)$ we define $\Omega^n(V; g) = \{f \in H_k(D^n, V) | f - g \in H_k^*(D^n, \mathbf{R}^m)\}$. It follows from the Sobolev Inequalities that if $2k > n$ $H_k(D^n, V)$ and $\Omega^n(V; g)$ are closed submanifolds of the Hilbert space $H_K(D^n, \mathbf{R}^m)$. More generally analogous Hilbert manifolds of H_K maps of M into V can be constructed for any compact C^∞ n-manifold with boundary M replacing D^n. Note that for $k = n = 1$, $H_1(D^1, V) = H_1(I, V)$ and $\Omega^1(V, g) = \Omega(V; g(0), g(1))$. A question that immediately presents itself is to find functions $J : \Omega^n(V; g) \to \mathbf{R}$ which are analogues of the action integral and satisfy condition (C). If A is a strongly elliptic differential operator of order $2k$ then $J(f) = \frac{1}{2} \langle Af, f \rangle_0$ is such a good analogue of the action integral provided $Af = 0$ has no solutions f in $H_k^*(D^n, \mathbf{R}^m)$. In particular if L is an elliptic kth order elliptic operator such that $Lf = 0$ has no solutions f in $H_k^*(D^n, \mathbf{R}^m)$ then $J(f) = \frac{1}{2}\|Lf\|_0^2 = \frac{1}{2} \langle L^*Lf, f \rangle_0$ is such a function (taking $k = n = 1$ and $L = \mathrm{d}/\mathrm{d}t$ gives the ordinary action integral). Smale has found an even wider class of functions which also satisfy condition (C).

Now let $n < m$ and regard $\mathbf{0}(n) \le \mathbf{0}(m)$ in the standard way. Define an orthogonal representation of $\mathbf{0}(n)$ on $H_k(D^n, \mathbf{R}^m)$ by $(Tf)(x) = T(f(T^{-1}x))$. If we take $V = S^{m-1}$ then since V is invariant under $\mathbf{0}(n)$ it follows that $H_k(D^n, V)$ is a invariant submanifold of $H_k(D^n, \mathbf{R}^m)$. Moreover if we define $g \in H_k(D^n, V)$ by $g(x) = (x_1, \ldots, x_n, \sqrt{1 - \|x\|^2}, 0 \ldots 0)$ then $Tg = g$ for any $T \in \mathbf{0}(n)$ and it follows that $\Omega^n(V, g)$ is also an invariant submanifold of $H_k(D^n, V)$, hence $\mathbf{0}(n)$ is a group of isometries of the complete Riemannian manifold $\Omega^n(V, g)$. Now suppose A is a strongly elliptic differential operator of order $2k$, $A : C^\infty(D^n, \mathbf{R}^m) \to C^\infty(D^n, \mathbf{R}^m)$, such that $A(Tf) = T(Af)$ for all $T \in \mathbf{0}(n)$, for example $A = \Delta^k$ where Δ is the Laplacian

$$\sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}.$$

Then $J : \Omega(V, g) \to \mathbf{R}$ defined by $J(f) = \frac{1}{2}\langle Af, f\rangle_0$ satisfies $J(Tf) = f$ for any $T \in \mathbf{0}(n)$, hence if f is a critical point of J so is Tf for any $T \in \mathbf{0}(n)$, and since non-degenerate critical points of J are isolated, $Tf = f$ if f is a non-degenerate critical point of J. But $Tf = f$ is equivalent to $R(f(x))$ being a function F of $\|x\|$ where R is the distance measured along the sphere $S^{m-1} = V$ of a point on V to the north pole. Moreover F will satisfy an ordinary differential equation of order $2k$. With a little computation one should be able to compute all the critical points and their indices and hence, via the Morse inequalities, get information about the homology groups of $\Omega^n(V, g)$ (which has the homotopy type of the nth loop space of S^{m-1}). Clearly the same sort of process will work whenever we can force a large degree of symmetry into the situation.

REFERENCFS

1. J. DIEUDONNE: *Foundations of Mondern Analysis*, Academic Press, New York, 1960.
2. J. EELLS: On the geometry of function spaces, *Symposium Internacional de Topologia Algebraica* (*Mexico*, 1956), pp. 303–308, 1958.
3. V. KLEE: Some topological properties of convex sets, *Trans. Amer. Math. Soc.* **78** (1955), 30–45.
4. S. LANG: *Introduction to Differentiable Manifolds*, Interscience, New York, 1962.
5. J. MILNOR: Morse theory, *Ann. Math. Stud.* No. 51, (1963).
6. M. MORSE: The calculus of variations in the large, *Colloq. Lect. Amer. Math. Soc.* **18** (1934).
7. J. NASH: The imbedding problem for Riemannian manifolds, *Ann. Math., Princeton* **63** (1956), 20–63.
8. E. PITCHER: Inequalities of critical point theory, *Bull. Amer. Math. Soc.* **64** (1958), 1–30.

Brandeis University.