ON THE APPROXIMATION OF FUNCTION SPACES IN THE CALCULUS OF VARIATIONS

DAVID A. POPE

Introduction. A basic feature of most of the methods used for the numerical calculation of a variational problem is the reduction of the infinite dimensional problem to a finite dimensional problem by some kind of approximation. One of the most natural approximations is that of replacing a curve or a surface by a finite number of points lying on or near the curve or surface. The points are then connected by simple arcs or surfaces, and the resulting approximation will, if the number of points is sufficiently large, presumably be close to the original curve or surface. The difficulties inherent in this approach to surface problems are well illustrated in the works of Rado [8], [9] on surface area.

The replacement of a curve by an approximating polygon, however, does lead to a usable finite dimensional approximation scheme. Lewy [3] (Chapter IV) gives a proof of the existence of an absolute minimum to the positive regular nonparametric problem by using such an approximation scheme, and his proof could be used to design a numerical process for approximating this minimum.

The methods of algebraic topology which M. Morse ([4] to [7]) applied to the calculus of variations have led to a greater understanding of the relationships between all the extremals to a variational problem. The extremals are classified according to their index types, in analogy with quadratic forms of a finite number of variables. While the extremals with nonzero index are not minimizing, they are of importance in many physical applications.

In this paper we shall treat the problem of computing the nonminimizing extremals as well as those of minimizing type, using the theory developed by Morse, together with a general theory of approximation. In part 1, a brief restatement of some of the principal definitions and theorems of Morse [6] will be given, in the current language of algebraic topology. In part 2, a general theory of approximation to an abstract metric space will be developed, and the convergence of the approximations to the critical levels of the problem defined on this space will be demonstrated. Part 3 will show that the polygonal approximations to curves leads, in the parametric problem, to approximations satisfying the theory of part 2.

The structure of part 2 is given with sufficient abstraction so that

Received July 24, 1961. The preparation of this paper was sponsored by the Office of Ordnance Research, U. S. Army. Reproduction in whole or in part is peimitted for any purpose of the United States Government.

it may be applied to any reasonable method of approximation to a variational problem; thus we are not necessarily restricted to the polygonal approximations described in part 3.

1. Outline of the Morse theory. Given a metric space M, and a real valued function F defined on M, we define the sets F_a as the set of all points x of M such that $F(x) \leqq a$. We make the basic assumption of *bounded compactness*; that is, *the sets F_a are compact for all a.*

We now assign a homology theory to M and its compact subsets. The most generally useful homology for the calculus of variations has been the Čech theory, in the form given by Vietoris (see Vietoris [11] and Morse [6]). In this paper we shall use Čech homology, as defined in Eilenberg and Steenrod [1], Chapter IX, with coefficient group the field of integers modulo 2. We shall largely follow the notation of Eilenberg and Steenrod [1].

For $\alpha \geqq \beta > 0$, we define the inclusion map

$$i_\alpha^\beta\colon (F_a, F_{a-\alpha}) \longrightarrow (F_a, F_{a-\beta})$$

of the compact pairs. Then for $\alpha \geqq \beta > 0$ and for each q we have the homomorphisms

$$\pi_\alpha^\beta\colon H_q(F_a, F_{a-\alpha}) \longrightarrow H_q(F_a, F_{a-\beta})$$

induced by i_α^β, with π_α^α=identity, and $\pi_\beta^\gamma \pi_\alpha^\beta = \pi_\alpha^\gamma$ for $\alpha \geqq \beta \geqq \gamma$.

This set of groups and homomophisms defines a *direct system* (Eilenberg and Steenrod [1] p. 212). We may then take the direct limit

$$\operatorname*{dir\,lim}_{\alpha \to 0} H_q(F_a, F_{a-\alpha}) = H_q(F_a, F_{a-})$$

and define this as the *cap group* of index q at the level $F = a$. Morse [6] defines an equivalent group of cap classes in a different way, using the Vietoris homology theory.

If $H_q(F_a, F_{a-}) \neq 0$, we will say that $F = a$ is a *critical level of F on M, with index q.* Now we derive three lemmas about the cap groups which will prove useful.

LEMMA 1.1. *Given any nonzero element V of $H_q(F_a, F_{a-})$, for every sufficiently small $\alpha > 0$, there exists a nonzero element V_α in $H_q(F_a, F_{a-\alpha})$ such that the projection*

$$\pi_\alpha\colon H_q(F_a, F_{a-\alpha}) \longrightarrow H_q(F_a, F_{a-})$$

maps V_α into V.

Proof: By Lemma 4.3, p. 221 of Eilenberg and Steenrod [1], there exists a positive number γ and an element V_γ in $H_q(F_a, F_{a-\gamma})$ such that

$\pi_\gamma V_\gamma = V$. Now choose α arbitrarily between 0 and γ, and set $V_\alpha = \pi_\alpha^\gamma V_\gamma$. Clearly V_α satisfies the lemma, since $\pi_\alpha V_\alpha = \pi_\alpha \pi_\alpha^\gamma V_\gamma = \pi_\gamma V_\gamma = V$.

LEMMA 1.2. *Suppose there is no critical level $F = c$ of index q in the half-open interval $b < c \leqq a$. Then $H_q(F_a, F_b) = 0$.*

Proof. We shall prove this lemma by contradiction. Suppose $H_q(F_a, F_b) \neq 0$, and let U be a nonzero element of this group. We denote the homomorphisms of $H_q(F_a, F_\alpha)$ into $H_q(F_a, F_\beta)$ induced by inclusion, for $\alpha < \beta < a$, by j_α^β. These are the projections of the direct system defining the cap group at the level $F = a$. This cap group is zero, since $F = a$ is not a critical level of index q.

Next we define s as the supremum of all numbers β in the interval $[b, a]$ such that U is *not* in the kernel of j_b^β. By Lemma 4.4 of Eilenberg and Steenrod [1], p. 221, there is a $\gamma < a$ such that $j_b^\gamma U = 0$, since the direct limit is zero. Hence for all β with $\gamma \leqq \beta \leqq a$, $j_b^\beta U = j_\gamma^\beta j_b^\gamma U = 0$. Therefore by definition, $s \leqq \gamma < a$.

Now suppose $\beta < s$ and $j_b^\beta U = 0$. Then for every γ with $\beta < \gamma$, $j_b^\gamma U = j_\beta^\gamma j_b^\beta U = 0$, and β is an upper bound, contrary to the definition of s. Hence for all $\beta < s$, $j_b^\beta U \neq 0$.

Since, by definition of s, $j_b^\beta U = 0$ for all $\beta > s$, the map j_b^s satisfies the equation

$$j_b^s U = \text{inverse}\ \lim_{\beta > s}\ j_b^\beta U = 0$$

by Lemma 3.11, p. 218 of Eilenberg and Steenrod [1]. Now consider the following portion of the exact sequence of the triple (F_b, F_s, F_a)

$$H_q(F_s, F_b) \xrightarrow{i} H_q(F_a, F_b) \xrightarrow{j_b^s} H_q(F_a, F_s)\ .$$

Since U is in the kernel of j_b^s, it is in the image of i. Hence there is a nonzero element V in $H_q(F_s, F_b)$ such that $iV = U$.

But $F = s$ is not a critical level of index q; hence the direct limit as $\beta \overrightarrow{<} s$ of $H_q(F_s, F_\beta) = 0$. Then by Lemma 4.4 of Eilenberg and Steenrod [1], p. 221, there is a $\gamma < s$ such that V is in the kernel of

$$j \colon H_q(F_s, F_b) \longrightarrow H_q(F_s, F_\gamma)\ .$$

Consider now the following portion of the homomorphism of the exact sequences of the triples (F_a, F_s, F_b) and (F_a, F_s, F_γ) induced by inclusion:

$$\begin{array}{ccc} H_q(F_s, F_b) & \xrightarrow{i} & H_q(F_a, F_b) \\ {\scriptstyle j}\downarrow & & \downarrow{\scriptstyle j_b^\gamma} \\ H_q(F_s, F_\gamma) & \xrightarrow{i_\gamma} & H_q(F_a, F_\gamma) \end{array}$$

The element V in $H_q(F_s, F_b)$ satisfies $iV = U$, and so $j_b iV = i_\gamma jV = i_\gamma 0 = 0$. Hence U is in the kernel of j_b^γ and $\gamma < s$. This contradiction proves there is no nonzero element U in $H_q(F_a, F_b)$, thus proving the lemma.

LEMMA 1.3. *Suppose there are no critical levels $F = c$ of index q or $q+1$ on the half-open interval $b < c \leqq a$. Then the inclusion map of F_b into F_a induces an isomorphism of $H_q(F_b)$ onto $H_q(F_a)$.*

Proof. By Lemma 1.2, $H_q(F_a, F_b) = H_{q+1}(F_a, F_b) = 0$. Hence in the exact sequence of the pair (F_a, F_b) we have

$$0 \xrightarrow{\delta} H_q(F_b) \xrightarrow{i} H_q(F_a) \xrightarrow{j} 0 .$$

Therefore i is an isomorphism, and the lemma follows.

2. Approximations to a metric space. The following problem is now defined: we are given a metric space M of points x and a real valued function F defined on M; we wish to find the critical values of F on M. To do this we define a sequence of approximations to the space M. Using the methods of algebraic topology which M. Morse ([4] to [7]) applied to the calculus of variations, we are able to measure how close to the critical levels of the original problem those of the approximated problem will lie.

Following Morse [6], pp. 29–36, we repeat here some definitions for the convenience of the reader. An *admissible deformation* of a subset E of M is defined as a homotopy $q(p, t)$: $E \times I \to M$, where I is the interval $0 \leqq t \leqq T$, and $q(p, 0) = p$ for all p in E. The curve $q(p, t)$ obtained by holding p constant is called the *trajectory* defined by p. If the points $r_1 = q(p, t_1)$ and $r_2 = q(p, t_2)$ are on the trajectory defined by p with $0 \leqq t_1 \leqq t_2 \leqq T$, then r_1 is said to be an *antecedent* of r_2.

The admissible deformation is said to admit a *displacement function* $\delta(\varepsilon)$ on E if, whenever r_1 is an antecedent of r_2 with the distance from r_1 to r_2 greater than $\varepsilon > 0$, then $F(r_1) - F(r_2) > \delta(\varepsilon)$, where $\delta(\varepsilon)$ is a positive single valued function of ε. If an admissible deformation of E admits a displacement function on each compact subset of E, the deformation is called an *F-deformation.*

The function F is called *upper-reducible* at p if for each constant $c > F(p)$, there exists a neighborhood of p relative to F_c which posesses an F-deformation carrying the neighborhood into a set lying in F_{c-e} for some positive e.

Following Morse, we make the assumptions that the sets F_a are compact for all a, and that F is upper-reducible at all points of M. Under these assumptions Morse ([6], p. 38) proves that each critical level ("cap limit") contains at least one homotopic critical point.

Next we give a set of formal requirements defining a set of approximations to the space M which are admissible with respect to F. To do this we define a sequence $\{p_n\}$ of functions, called *approximations*, with the following properties:

(1) For each n, p_n is a continuous function of M into M. The image of M under p_n will be called M^n.

(2) M^n is a closed subset of M for each n.

(3) F is a continuous function on M^n for each n.

(4) For any real number a, and any $e > 0$, there is an integer N such that

$$F(p_n x) \leqq F(x) + e \tag{2.1}$$

for all $n > N$, and for all x in F_a.

(5) For any $a, e > 0$, $n > N$ of property 4, the composite map

$$ip_n\colon F_a \longrightarrow F_{a+e} \cap M^n \longrightarrow F_{a+e} \tag{2.2}$$

is homotopic to the inclusion map

$$j\colon F_a \longrightarrow F_{a+e}\ . \tag{2.3}$$

The map p_n in (2.2) is regarded as taking points of M into points of M^n, where M^n is regarded as a separate space, with topology induced by the topology of M. The inclusion map i of (2.2) is the map taking points of M^n into themselves, considering M^n as a subset of M in the image. We shall abbreviate by setting

$$F_b^n = F_b \cap M^n\ . \tag{2.4}$$

For the applications of this theory to computing variational problems, we shall demand one further restriction on the subspaces M_n; they are to be *finite dimensional of dimension* r_n, in the sense that they are locally homeomorphic to euclidean space of r_n dimensions. In this manner the problem of finding the critical points and levels of the infinite dimensional space M is reduced to the simpler problem of finding the corresponding objects in the finite dimensional subspace M^n. However, as this restriction is not used in the topological arguments to follow, it is not placed in the set of requirements above.

First, the properties of bounded compactness and upper reducibility must be verified for the subspace M^n.

LEMMA 2.1. *Under the conditions* (2) *and* (3) *of* 2.1, *the subsets* F_b^n *are compact, and* F *is upper reducible on* M^n.

Proof. M^n is closed by requirement (2), hence F_b^n is compact. F

is upper reducible on M^n because any continuous function on a space is upper reducible (Morse [6] p. 37).

Lemma 2.1 allows us now to use the Čech homology theory with the compact subsets F_b^n and the theory developed by Morse to indicate the relationship between the critical value of F on M and those of F on M^n. The theorem below gives this relationship in one case of interest, and show convergence of the critical levels on M^n to those on M, as $n \to \infty$.

THEOREM 2.1. *Suppose $F = c$ is a critical level of index q of F on M, and there are no other critical levels of F on M of index q or $q+1$ on the interval $[c, c+h]$ for some $h > 0$. Suppose futher that there are only a finite number of critical levels of F on M^n, for each n.*

Then for every sufficiently small $e > 0$, there is an integer N such that for $n > N$, the approximating space M^n possesses a critical level c_n of index q, with $c \leqq c_n \leqq c + e$.

Proof. Since $F = c$ is a critical level, the cap group $H_q(F_c, F_{c-})$ is nontrivial. Let V be a nonzero element of this group. Then by Lemma 1.1, for every sufficiently small $\alpha > 0$, there is a nonzero element V_α in $H_q(F_c, F_{c-\alpha})$ such that the projection of this group into the cap group maps V_α into V. Now choose such a number $\alpha < h$, and call it α_0. Next pick arbitrarily a positive number $e < \alpha_0$. Finally choose a positive number $\alpha < e$. Then by property (4) of the approximations we may choose N so that for $n > N$ we have

$$p_n(F_c) \subset F^n_{c+e} \subset F_{c+e}$$

and

$$p_n(F_{c-\alpha}) \subset F^n_{c-\alpha+e} \subset F_c \ ,$$

The composite map ip_n as defined in equation (2.2) then defines a map of the pairs $(F_c, F_{c-\alpha})$ into $(F_{c-e}, F_{c-\alpha+e})$, which is homotopic to the inclusion map of these pairs, by property (5) of the approximations. Therefore the homomorphism

$$i^*p_n^*\colon H_q(F_c, F_{c-\alpha}) \longrightarrow H_q(F_{c+e}, F_{c-\alpha+e})$$

induced by this pair map is the same as the inclusion homomorphism for the pairs, by the homotopy axiom of homology theory.

Now consider the following portion of the map of the exact sequences of the pairs $(F_c, F_{c-\alpha})$ and $(F_{c+e}, F_{c-\alpha+e})$ induced by inclusion:

$$\begin{array}{ccccc} H_q(F_{c-\alpha}) & \longrightarrow & H_q(F_c) & \longrightarrow & H_q(F_c, F_{c-\alpha}) \\ \downarrow i_1^* & & \downarrow i_2^* & & \downarrow i_3^* \\ H_q(F_{c-\alpha+e}) & \longrightarrow & H_q(F_{c+e}) & \longrightarrow & H_q(F_{c+e}, F_{c-\alpha+e}) \ . \end{array}$$

Since there are no critical levels on M of index q or $q+1$ in the intervals $(c-\alpha,\ c-\alpha+e]$ and $(c, c+e]$, Lemma 1.3 allows us to conclude that the inclusion homomorphisms i_1^* and i_2^* are isomorphisms. Then the "five" lemma shows that i_3^* is also an isomorphism. But $i_3^* = i^* p_n^*$. *Therefore the kernel of* p_n^* *must be zero.*

This means that $H_q(F_{c+e}^n, F_{c-\alpha+e}^n)$ contains the nonzero element $p_n^* V_\alpha$. Therefore by Lemma 1.2, there is at least one critical level c_n of F on M^n of index q, in the interval $(c-\alpha+e, c+e]$. Now suppose there are no critical levels of index q of F on M^n in the interval $[c, c+e]$. Then there must be a critical level c_n below c, in the interval $(c-\alpha+e, c)$. But α may be chosen so that $c-\alpha+e$ is arbitrarily close to c. This is clearly contradictory since there are only a finite number of critical levels on M^n, by hypothesis. Therefore there is at least one critical level c_n in the interval $[c, c+e]$, and the theorem is proved.

Next we show convergence of the critical points on M^n to those on M in a simple case.

THEOREM 2.2. *Given the conditions of Theorem* 2.1 *for* $q=0$, *suppose the function* F *attains its absolute minimum on* M *at the point* x, *and this minimum point is unique.*

Then the approximating spaces M^n *contain homotopic critical points of index zero, and a subsequence of these points converges to the point* x.

Proof. Let $F(z) = c$. This level is clearly a critical level of index zero. Then by Theorem 2.1, the approximating spaces M^n contain critical levels c_n of index zero, which approach the level c from above. But Morse [6] p. 38 proves that each critical level c_n contains at least one homotopic critical point of index zero. Choosing one such point for each n, we denote it x_n. Since the infinite set $\{x_n\}$ is contained in the compact set F_{c+e} for some $e > 0$, it has at least one accumulation point, which we may call y, and a subsequence converges to y. But the lower semicontinuity of F implies $F(y) \leqq \lim_n F(x_n) = c$, for any convergent subsequence. Clearly the inequality is impossible, since c is the absolute minimum of F on M. Hence we have $F(y) = c$, and therefore $y = x$, since this minimum is unique.

3. The approximation of the parametric problem with fixed end points. In this part the requirements of § 2 will be applied to approximations to a general class of fixed end point problems in parametric form. The definition of the parametric problem will follow closely that given in Morse [6].

3.1. The curve space. The space M of § 2 will in this application be replaced by the space Ω of all continuous curves (parametrized curve classes) between two fixed points a and b on a compact subset of euclidean n-space, which we shall denote by Σ.

A parametrized curve, or p-curve, on Σ is defined as any continuous function from a real interval into the space Σ.

Suppose two p-curves η_1 and η_2 are given in the form

$$\begin{aligned} \eta_1 &: q = q_1(t) \qquad 0 \leqq t \leqq c \\ \eta_2 &: q = q_2(t) \qquad 0 \leqq t \leqq d \,. \end{aligned} \tag{3.1}$$

Let w be any sense-preserving homeomorphism between $[0, c]$ and $[0, d]$, and let $d(w)$ be the maximum distance between the points $q_1(t)$ and $q_2(wt)$ for t in $[0, c]$. The Frechet distance between η_1 and η_2 is defined as

$$d(\eta_1, \eta_2) = \inf d(w)$$

taken over all possible homeomorphisms w. (Cf. Frechet [2].) The set of all p-curves at zero distance from a given p-curve will be called a curve class, or simply a curve. A p-curve belonging to a curve class will be called a representation or parametrization of that curve.

A p-curve joining a to b on Σ is defined as a p-curve

$$\eta: q = q(t) \qquad 0 \leqq t \leqq c$$

with the property that

$$q(0) = a \text{ and } q(c) = b \,.$$

Clearly every p-curve in the curve class of η has this property. The curves representable by p-curves with this property are the points of the space Ω. The Frechet distance between two curves α and β of Ω is defined as the Frechet distance between any parametrization of α and any parametrization of β. This distance is clearly independent of the choice of the parametrizations.

3.2. μ-length. A special parametrization of curves given the name μ-length by Morse has been used in many connections in the calculus of variations. A summary of some of its important properties will be made here.

To define the μ-length of a curve η, we take any parametrization $q = q(t), 0 \leqq t \leqq c$, of η, and pick a set $\{t_j\}$ of k values of t with $0 \leqq t_1 < t_2 \cdots < t_k \leqq c$. This set defines a partition $P = \{q_j\}$ of the curve η, where $q_j = q(t_j)$. We denote the minimum of the $k-1$ distances (q_j, q_{j+1}) on Σ for $j = 1, 2, \cdots, k-1$, by $m(P)$. Then the sup $m(P)$ taken over all such partitions with k points will be called μ_k. We then set

(3.2) $$\mu_\eta = \sum_{k=2}^{\infty} \frac{\mu_k}{2^{k-1}} .$$

μ_η is the μ-length of η, and is independent of the particular choice of the parametrization $q(t)$ of η. Thus the μ-length $\mu(\tau)$ of the part of $q(t)$ from $t \leqq 0$ to $t \leqq \tau$ may be used as a parametrization for η. This parametrization has the following properties, proved in Morse [5], and listed also in Morse [6] p. 34–35.

(1) If $q = q(t)$ is any parametrization of a curve η, then the μ-parametrization of η has the form

(3.3) $$\eta\colon q = \bar{q}(\mu) = q(t(\mu)) \qquad 0 \leqq u \leqq u_\eta$$

where $t(\mu)$ is a continuous nondecreasing function of μ on the closed interval $[0, \mu_\eta]$.

(2) The value of μ at any point q on the curve η satisfies the inequality

(3.4) $$\frac{d}{2} \leqq \mu \leqq d$$

where d is the diameter of the set of points preceding q on η.

(3) The μ-length μ_η of a curve η is a continous function of η on the curve space Ω.

(4) For a given η, the μ-parametrization of

$$\eta\colon q = q(\mu)$$

is constant with respect to μ on no subinterval of $[0, \mu_\eta]$.

(5) The parametrization $q(\mu; \eta)$ of a curve η of Ω is a continuous function of μ and η for μ in $[0, \mu_\eta]$ and η in Ω.

Suppose η is a straight line of length s joining two points p, q in euclidean space. For any partition P_k with k values $\{q_j\}$ on η, $m(P_k)$ is clearly $\leqq s/k-1$. But the equidistant partition gives $m = s/k-1$. Hence $\mu_k = s/k-1$, and the u-length μ_η of η is

(3.5) $$\mu_\eta = \sum_{k=2}^{\infty} \frac{s}{(k-1)2^{k-1}} = s \log 2 .$$

If η is any rectifiable curve joining p to q with are length s, $m(P_k)$ is still $\leqq s/k-1$ for any partition P_k of k points. Hence $\mu_k \leqq s/k-1$, and therefore

(3.6) $$\mu_\eta \leqq s \log 2 .$$

3.3 The functional F. Having defined the curve space Ω, we now construct the functional F on Ω. We are given a function $f(x_1, \cdots, x_n,$

$r_1 \cdots r_n) = f(x, r)$ of $2n$ variables with the following properties (n is the dimension of Σ):

(A) $f(x, r)$ is of class C^4 in (x, r) for x in Σ, and any set of numbers $(r) \neq (0)$.

(B) f is an invariant under the transformations of local coordinates (cf. Morse [6] p. 64–65).

(C) $f(x, r) > 0$ at every point of Σ and for all r.

(D) f is positive homogeneous of degree 1 in r.

(E) The rank of the determinant

$$|f_{r_i r_j}|$$

is $n - 1$, and all its characteristic roots except the zero root are positive.

Assuming that Σ is arcwise connected, we define a secondary metric $[q_1 q_2]$ on points of Σ as follows. For any two points q_1, q_2 of Σ we consider the class of all rectifiable curves on Σ joining q_1 to q_2; the integral

$$F = \int_{q_1}^{q_2} f(x, x') ds \tag{3.7}$$

is computed over this class, and the inf F over this class of curves is $[q_1 q_2]$.

To define F on an arbitrary curve η of Ω we form the sum

$$s = \sum_{i=1}^{k} [q_i q_{i+1}] \tag{3.8}$$

where the points q_i are partition points on η. The sup of s over all partitions of η is defined as $F(\eta)$. $F(\eta)$ is equal to the integral of f along the curve η if η is rectifiable, and is infinite if η is not. Morse [6] shows that under these conditions, we have bounded compactness of M, and upper reducibility of F.

3.4. Compactness and rectifiability. Before showing properties $1 - 5$ of § 2 are satisfied in the parametric problem, we need the following compactness lemma:

LEMMA 3.1. *If Γ is a compact subset of the curve space Ω, the set A of all the points of Σ which lie on curves of Γ is a compact subset of Σ.*

Proof. (Cf. Morse [6] p. 59.) Let $\{q_j\}$ be an infinite set of points of A. Each point q_j lies on at least one curve γ_j of Γ. Pick one such γ_j corresponding to each q_j and consider the sequence $\{\gamma_j\}$ of curves. We parametrize each curve of Γ with the μ-length defined in 3.2, and therefore we have a unique number μ_j defined as the μ-value of the point

q_j on the curve γ_j. Since Γ is compact, the curves $\{\gamma_j\}$ possess a limit curve γ_0 and the μ-length μ_γ of the curves of Γ is bounded above; hence the sequence $\{\mu_j\}$ possesses a limit point μ_0.

Consider the point $q_0 = \gamma_0(\mu_0)$. q_0 is clearly a limit point of the sequence $\{q_j\}$. Hence the lemma follows.

Now let Σ_a be the set of all points of Σ which lie on curves of F_a; the compactness of F_a then implies the compactness of Σ_a by the previous lemma. Therefore the set T_a of (x, r) space

$$T_a\colon x \text{ in } \Sigma_a\,, \quad \sum_i r_i^2 = 1 \tag{3.9}$$

is also compact.

Therefore the function $f(x, r)$ is bounded above and below by M and $m > 0$ respectively for $x, r \in T_a$. Thus we have for any $p, q \in \Sigma$

$$m(pq) \leqq \int_p^q f ds \leqq M(pq)$$

for any rectifiable curve joining p to q on Σ. Therefore

$$m(pq) \leqq [pq] \leqq M(pq)\ .$$

Therefore on any partition $\{q_j\}$ of a curve η on J_a we have

$$m\Sigma(q_jq_{j+1}) \leqq \Sigma[q_jq_{j+1}] \leqq M\Sigma(q_jq_{j+1})\ .$$

Taking the limit for norm of partitions $\to 0$, we find

$$mL(\eta) \leqq F(\eta) \leqq ML(\eta)\ . \tag{3.10}$$

Thus $L(\eta) \leqq j(\eta)/m \leqq a/m$, where $L(\eta)$ is the arc length of η. Therefore we have shown

LEMMA 3.2. *The curves of F_a are rectifiable and of arc length $\leqq a/m$, where $m > 0$ is the minimum of $f(x, r)$ on T_a.*

3.5. **Definition of the approximations.** Given any curve of Ω, we may parametrize it in terms of the μ-length described in 3.2. If this is done we denote the curve η by

$$\eta\colon q = q(\mu; \eta) \qquad 0 \leqq \mu \leqq \mu_\eta\ . \tag{3.11}$$

Making the linear substitution

$$t = \mu/\mu_\eta \tag{3.12}$$

we have a new parametrization of η:

$$\eta\colon q = \bar{q}(t) = q(t\mu_\eta; \eta)\ . \tag{3.13}$$

This parametrization of the curves of Ω will be called the uniform t-parametrization.

Now to define the approximations p_n, we take a sequence of partitions P_n of the unit interval as follows: P_n will be a set of $n+2$ points $\{t_j\}$ $j = 0, 1, \cdots, n+1$, with $t_0 = 0$ and $t_{n+1} = 1$, and $t_{j+1} > t_j$. The norm of the partition P_n will be denoted δ_n. We require further that $\delta_n \to 0$ (as $n \to \infty$).

We now take a curve η of Ω, parametrized as in equation (3.13) and define the points

$$q_j = \bar{q}(t_j) = q(t_j u_\eta; \eta) . \tag{3.14}$$

The points q_j all lie on the curve η, and $q_0 = a$, $q_{n+1} = b$, the end points of η.

Lemma 3.1 shows that the set of μ-lengths of the curves in F_a is bounded. Suppose $\mu_\eta < M$ for all η in F_a. Then

$$\Delta\mu_j = (t_{j+1} - t_j)\mu_\eta < \delta_n M \tag{3.15}$$

for all curves of F_a. Therefore the diameter d_j of the subarc of η from q_j to q_{j+1} satifies

$$d_j \leqq 2\Delta\mu_j < 2\delta_n M . \tag{3.16}$$

Under the conditions described in 3.3, there is a fundamental distance ρ in the compact set Σ_a (described in § 3.4) with the property that if p, q have distance $(pq) < \rho$, there exists a unique extremal arc joining p to q which gives to F a proper minimum value over the class of all arcs joining p to q, and this extremal arc is a member of a field of extremals covering the ρ-neighborhood of the point p simply (except at p). Hereafter, when dealing with a set F_a, we shall assume that n is large enough so that the diameters d_j of the subarcs of η are all less than ρ.

We then construct the polygon through the points $q_0, q_1, \cdots, q_{n+1}$. The arc of the polygon from q_j to q_{j+1} is defined as the euclidean straight line from q_j to q_{j+1}. This polygon obtained from the curve η is denoted by $p_n(\eta)$, and is a continuous parametrized curve class, which may also be denoted by $p_n(\eta)$, is in Ω. The space $p_n(\Omega)$ of all polygons obtained from the partition P_n will be called Ω_n.

3.6. Verification of the approximation requirements. Having defined the sequence $\{p_n\}$ of approximations to the parametric problem, we now seek to show that they satisfy the topological requirements 1–5 of § 2. The following lemmas dispose of properties 1–3.

LEMMA 3.3. *For each n, the approximation p_n is a continuous* *function of the curve space Ω into itself.*

Proof. The μ-length μ_η of η is a continuous function of η in Ω; and the points $q_j = q(t_j)$, where $q(t)$ is the uniform t parametrization of η are therefore continuous functions of η in Ω. This means, given any $e > 0$, there is a $\delta > 0$ such that $d(\eta, \zeta) < \delta$ in Ω implies the distance $(q_j, r_j) < e$ in Σ for all j, where q_j, r_j are the points on η, ζ respectively with t-value t_j. But if the distance between the corners q_j and r_j of the two polygons is less than e, the Frechet distance in Ω between the polygons is also $< e$. Hence the lemma follows.

LEMMA 3.4. *Ω_k is a closed subset of Ω.*

Proof. The limit of any sequence of polygons with k corners can be only a polygon with k or fewer corners, hence contained in Ω_k. Thus Ω_k is evidently closed.

LEMMA 3.5. *F is a continuous function on Ω_k.*

Proof. By property 3.2, given any curve α in Ω_k and any $\delta > 0$, there is a $\rho > 0$ such that β in Ω_k, $d(\alpha, \beta) < \rho$ implies $|\mu_\alpha - \mu_\beta| < \delta$. Then equation 3.5 of § 2 implies that the arc lengths of α and β differ by less than $\delta \log 2$, since they consists of straight line segments. But Tonelli ([10] vol. 1, p. 304) proves the following theorem:

Given any curve η, and any $\varepsilon > 0$, there exist two numbers $\delta > 0$ and $\rho > 0$ such that if

$$d(\eta, \zeta) > \rho$$

and

$$|L(\eta) - L(\zeta)| < \delta \qquad (L(\eta) = \text{arc length of } \eta)$$

then

$$|F(\eta) - F(\zeta)| < \varepsilon\ .$$

Hence Lemma 3.5 follows immediately.

In order to prove requirements 4 and 5 of § 1.1 we shall use the following lemmas.

LEMMA 3.6. $\lim p_n \eta = \eta$ as $n \to \infty$, uniformly for η in F_a.

Proof. Equation (3.16) of § 3.5 states:

$$d_j \leqq 2\delta_\eta M$$

whenever η is in F_a. Hence the distance from a point on the straight line from q_j to q_{j+1} to any point of η between the same points is closer than $2d_j$, so the Frechet distance between η and $p_n\eta$ is less than $4\delta_n M$. But the δ_n approach zero, hence the lemma follows.

Now we show the uniform convergence of the arc length of $p_n\eta$ to that of the broken extremal associated with $p_n\eta$ for η in F_a.

Consider the set of all extremal arcs for the problem, parametrized by arc length s, joining the points p and q on Σ, and satisfying $(pq) < \rho$, the elementary length defined in 3.5. These extremals satisfy the Euler differential equations

$$\frac{d}{dt}[F_{r_i}(g, g')] = F_{x_i}(g, g') \tag{3.17}$$

which may be written in the form

$$g_i'' = \phi_i(g, g') \tag{3.18}$$

where ϕ is the function obtained by the solution of the implicit equations:

$$\begin{aligned} F_{r_i x_j} g_i' + F_{r_i r_j} g_i'' - F_{x_j} &= 0 \\ g_i' g_i' - 1 &= 0 \end{aligned} \qquad \text{(summed on } i\text{)} . \tag{3.19}$$

Under the assumptions made on F, the equations (3.19) can be solved uniquely for ϕ_i, and ϕ_i is a continuous function of g and g'. Since (g, g') lies in the compact set T_a, $|\phi_i|$ is bounded. Thus for all the geodesics considered,

$$|g_i''(s)| < M . \tag{3.20}$$

A broken extremal consisting of the unique elementary extremal arcs joining the points q_j, q_{j+1} of an approximation $p_n(\eta)$ will be called the broken extremal *associated* with $p_n(\eta)$.

Under these conditions, we can prove the following lemma.

LEMMA 3.7. *The arc length of the polygon $p_n(\eta)$ approaches the arc length of the broken extremal associated with $P_n(\eta)$ as $n \to \infty$ uniformly for η in F_a.*

Proof. Consider the pair of points q_j, q_{j+1} of Σ. We shall compare the arc length of the extremal

$$x_i = g_i(s) \qquad 0 \leqq s \leqq s_1$$

(parametrized by arc length) with the length of the straight line from q_j to q_{j+1}. Consider the family of straight lines drawn form the initial point q_j to the point $q = g(s)$ on the geodesic. We denote the length of the straight line to $g(s)$ by $L(s)$. L is clearly a continuous function of s, with $L(0) = 0$.

In the parametrization by arc length, the extremal g is a function of class C^2. We let $g_i'(0) = r_i$. The law of the mean gives

$$g_i'(s) = r_i + s g_i''(\sigma) \qquad 0 \leqq \sigma \leqq s \tag{3.21}$$

and application of equation (3.20) gives the inequality

$$r_i - Ms \leqq g_i'(s) \leqq r_i + Ms \,. \tag{3.22}$$

Integrating (3.19) from 0 to an arbitrary point s between 0 and s_1, we obtain

$$r_i s - \frac{Ms^2}{2} \leqq g_i(s) - g_i(0) \leqq r_i s + \frac{Ms^2}{2} \,. \tag{3.23}$$

Now suppose $|r_i| > Ms/2$. In this case we have

$$(g_i(s) - g_i(0))^2 \geqq s^2\left(r_i^2 - Ms|r_i| + \frac{M^2 s^2}{2}\right) . \tag{3.24}$$

Therefore the length $L(s)$ of the line from $g(0)$ to $g(s)$ satisfies the inequality

$$\frac{L^2(s)}{s^2} \geqq \Sigma'\{r_i^2 - Ms|r_i|\} \tag{3.25}$$

the sum being taken over all r_i with $|r_i| > Ms/2$.

But

$$\Sigma' r_i^2 = \sum_i r_i^2 - \sum_{|r_i| < Ms/2} r_i^2 \geqq 1 - \frac{nM^2 s^2}{4}$$

and $\sum_i r_i^2 = 1$ implies $\Sigma|r_i| \leqq \sqrt{n}$.

Hence

$$1 \geqq \frac{L^2(s)}{s^2} \geqq 1 - M\sqrt{n}\, s - \frac{M^2 n}{4} s^2 \,. \tag{3.26}$$

Also, since the geodesic and the family of straight lines lies in a point set N on Σ of diameter less than Δ_n, we have

$$L(s) < \Delta_n \qquad 0 \leqq s \leqq s_1 \,.$$

Now if Δ_n is taken to be smaller than the maximum value of the curve $L = s[1 - M\sqrt{n}\, s - (M^2 n/4)s^2]^{1/2}$ the curve $L = L(s)$ giving the length of the line from $g(0)$ to $g(s)$ must lie in a disconnected region of the (L, s) plane. Since L is continuous and $L(0) = 0$, it must lie entirely in the left hand region, which shows

$$\text{(3.27)}\qquad \begin{aligned}&\text{(a)}\quad s \leqq \hat{s} \text{ for all geodesics in the set } N .\\ &\text{(b)}\quad \lim_{\Delta_n \to 0} \frac{L(s)}{s} = 1 .\end{aligned}$$

Thus Lemma 3.7 follows.

Now we can demonstrate that requirement 4 of 1.1 is fulfilled.

LEMMA 3.8. *For any real number a, and any $e > 0$, there is an integer N such that for $n > N$, and for all η in F_a, we have*

$$F(p_n\eta) \leqq F(\eta) + e . \tag{3.28}$$

Proof. Let us denote the broken extremal associated with $p_n\eta$ by $g_n\eta$. Lemmas 3.6 and 3.7 state that given any $\rho > 0$, $\delta > 0$, there is an integer N such that if $n > N$, we have

$$d(p_n\eta, g_n\eta) < \rho$$

and

$$|L(p_n\eta) - L(g_n\eta)| < \delta$$

for all η in F_a.

But Tonelli ([10] vol. 1, p. 304) proves that given any $e > 0$, if δ and ρ are chosen sufficiently small, we will then have

$$|F(p_n\eta) - F(g_n\eta)| < e . \tag{3.29}$$

But from the definition of F and the remark of § 3.5 about the fundamental distance in Σ_a we have

$$F(g_n\eta) \leqq F(\eta) . \tag{3.30}$$

Addition of inequalities (3.29) and (3.30) give the conclusion of the lemma.

Now the homotopy described in property 5 of 1.1 will be set up.

First we describe the standard deformation $\theta(\eta. u)$ of Morse. Let η be any curve in F_a. Let $q = q(\mu)$ be the parametrization of η in terms of μ-length. Taking the points $q_j = q(t_j\mu_\eta)$ of the approximation p_n of η as corner points, we deform η onto a broken extremal $\hat{g}$ consisting of the unique extremals from q_j to q_{j+1}, $j = 0, 1, \cdots, n$.

This deformation is defined as follows: let $\mu_j = t_j\mu_\eta$, the value of μ corresponding to the point q_j. Let $\Delta\mu_j = \mu_{j+1} - \mu_j$. Then at time u, $0 \leqq u \leqq 1/2$, we take $\mu_j(u) = \mu_j + 2u\Delta\mu_j$, and construct the unique extremals from μ_j to $\mu^{j}(u)$. The curve $\theta(\eta, u)$ is then defined as the curve formed by these extremals from μ_j to $\mu_j(u)$ and the original curve $q(\mu)$ from $\mu_j(u)$ to μ_{j+1}.

We now apply this deformation to the polygonal curve $p_n(\eta)$, deforming it onto the same broken extremal $\hat{g}$. If we set $u = 1 - u$ in this latter deformation, and follow the first deformation by the second, we have a deformation $\bar{\theta}(\eta, u)$ which carries η to $\hat{g}$ and then to $p_n(\eta)$ for $0 \leqq u \leqq 1$. During the first half of this deformation F is not increased, and during the second half F is not decreased; thus the deformation takes place in the set F_b, where $b = \max\{F(\eta), F(p_n)\}$. But in Lemma 3.8 it was shown that for n sufficiently large, $b \leqq a + \varepsilon$ for any $\varepsilon > 0$, $\eta \in F_a$.

The deformation $\bar{\theta}(\eta, u)$ is a homotopy, since it is easily seen to be continuous in both η and u. Thus we have shown that the parametric problem with the approximations described above satisfies the properties of § 1.1.

BIBLIOGRAPHY

1. S. Eilenberg and N. Steenrod, *Foundations of Algebraic Topology*, Princeton University Press, 1952.

2. M. Frechet, *Sur une representation parametrique intrinsique de la Courbe continue la plus generale*, Journal de Math., 9th Series, **4** (1925), 281-297.

3. H. Lewy, *Aspects of the Calculus of Valriations* (Notes by J. W. Green), University of California Press, 1939.

4. M. Morse, *The calculus of variations in the large*, Amer. Math. Soc. Colloq. Publ. No. 18, 1934.

5 ———, *A special parametrization of curves*, Bull. Amer. Math. Soc. **42** (1936), 915-922.

6. ———, *Functional Topology and Abstract Variational Theory*, Memorial des Sciences Mathematiques, no. 92, Paris, 1939.

7. ———, *Rank and span in functional topology*, Annals of Math., **41** (1941), 419-454.

8. T. Rado, *On the problem of Plateau*, Ergebnisse der Mathematik, Chelsea, 1951.

9. ———, *Length and area*, Amer. Math. Soc. Colloq. Publ. No. 30, 1948.

10. L. Tonelli, *Fondamenti di Calcolo delle Variazioni* (2 vols.), Bologna, 1925.

11. L. Vietoris, *Über den hoheren Zusammenhang kompakter Raume und eine Klasse von zusammenhangstreuen Abbildungen*, Mathematische Annalen, **97** (1927), 454-472.

UNIVERSITY OF CALIFORNIA, DAVIS AND LOS ANGELES