INTRODUCTION TO ANALYSIS IN THE LARGE

by

Marston Morse

The Institute for Advanced Study

First Edition - February 1947

Second Edition - November 1951

INTRODUCTION TO ANALYSIS IN THE LARGE

By

Marston Morse

February 1947

§1. Preliminary theorems in the calculus of variations. Following the heuristic remarks of the first lecture, I shall define the index of an extremal g_o joining two fixed points in a manner analogous to the defining of the index of a quadratic form*

(1.1) $$Q(x) = a_{ij}x_ix_j \qquad (x) = (x_1, \ldots, x_m)$$

Recall that the index k of Q could be defined as the maximum dimension of an r-plane on which Q is negative definite. The form Q will be replaced by the second variation

(1.2) $$I(u) = 2\int_{a^1}^{a^2} \Omega(u, u')dx$$

based on the extremal g_o, and the index of I will be defined as follows. One is concerned with a cartesian space of coordinates $(x, v_1, \ldots, v_n)$. One admits arcs of the form

$$v_i = u_i(x) \qquad (i = 1, \ldots, n)$$

where u_i vanishes at a^1 and a^2 and is a class D' on the closed interval $[a^1, a^2]$; that is where u_i is continuous together with u_i', except at most at a finite number of points x at which right and left derivatives of u_i shall exist.

For each fixed j (j = 1, ..., m) let u_{ij} be a function of the type of u_i above. If the columns (i variable) of the matrix u_{ij} are linearly

---

* We agree to sum with respect to a repeated index.
All r-planes will be taken through the origin or null element.

independent on $[a^1, a^2]$ then

$$v_i = c_j u_{ij}(x) \qquad \begin{array}{l}(i = 1, \ldots, n)\\ (j = 1, \ldots, m)\end{array}$$

represents an m-plane of admissible arcs.

<u>The index of I is by definition the maximum dimension h of those hyperplanes of admissible arcs on which I is negative definite (i.e. negative unless $(v) = (0)$).</u>

We shall evaluate this index of I as the number of conjugate points of $x = a$ on the interval $a^1 < x < a^2$. For this purpose we review a little of the classical variation theory.

We shall consider a function f with values

$$f(x; y_1, \ldots, y_n; p_1, \ldots, p_n) = f(x, y, p)$$

defined on a region R in a Cartesian (n+1)-space (x, y) and for arbitrary (p). We suppose f of class C''' on the product space of coordinates (x, y, p). We consider integrals of the form

$$J(y) = \int_{a^1}^{a^2} f(x, y, y')dx$$

taken along any arc $y_i = y_i(x)$ of class D' on $[a^1, a^2]$. We suppose that the set of derivatives

$$f_{p_i p_j}(x, y, p) \qquad (i, j = 1, \ldots, n)$$

for each fixed (x, y, p), are the coefficients of a positive definite quadratic form. In this case f is termed <u>positive regular</u>.

The Euler equations take the form

$$(1.2) \qquad \frac{d}{dx} f_{p_i}(x, y, y') - f_{y_i}(x, y, y') = 0 \qquad (i = 1, \ldots, n)$$

Solutions $y_i(x)$ of the Euler equations of class C" are called <u>extremals.</u> Given x^o there is a unique extremal for which the initial point (x^o, y^o) on R and slope p^o is prescribed. The family of such extremals takes the form

$$y_i = h_i(x, x^o, y^o, p) \qquad (i = 1, \ldots, n) \tag{1.3}$$

where h_i is of class C" in its arguments.

If (x^o, y^o) is held fast while (p^o) is replaced by a variable set (b), then (1.3) yields a family of extremals

$$y_i = k_i(x, b)$$

through the point (x^o, y^o). Let b^o determine a particular extremal g_o, and evaluate the Jacobian,

$$\frac{D(k_1, \ldots, k_n)}{D(b_1, \ldots, b_n)} = D(x, x^o)$$

for $(b) = (b^o)$. At $x = x^o$ we have

$$k_i = y_i^o, \quad \frac{\partial k_i}{\partial b_j} = 0, \qquad (i, j = 1, \ldots, n)$$

$$\frac{\partial k_i}{\partial x} = b_i, \quad \frac{\partial^2 k_i}{\partial x \partial b_j} = \delta_{ij}$$

so that D vanishes to exactly the nth order in x at x^o. In particular D does not vanish other than for $x = x^o$ for $|x - x^o| \leqq e$ provided $o < e < e_1$ and e_1 is a suitably chosen positive constant independent of x^o on $[a^1, a^2]$.

The points on g_o determined by values of $x \neq x^o$ at which D vanishes are called the <u>conjugate</u> points of (x^o, y^o) on g_o. The following theorem is fundamental.

Theorem 1: Let f be positive regular and let g_o be an extremal, on the interval $[a^1, a^2]$. If there is no conjugate point on g_o of the initial point of g_o then

$$J_g > J_{g_o} \qquad (g \neq g_o)$$

for every arc g: $y_i = y_i(x)$, of class D' joining the end points of g_o in a sufficiently small neighborhood of g_o.

For a proof of this theorem the reader may refer to the first chapter of my lectures "Calculus of Variations in the Large." Amer. Math. Soc. Coll. Publ. XVIII (1934).

§2. The second variation I. Further information concerning conjugate points will be obtained in terms of the so-called second variation I. This is an integral of the form

$$(2.1) \qquad I(u) = \int_{a_1}^{a^2} 2\Omega(u, u')dx$$

where $2\Omega(u,u') = f^o_{p_i p_j} u'_i u'_j + 2f^o_{p_i y_j} u'_i u_j + f^o_{y_i y_j} u_i u_j$ and the superscript zero indicates evaluation of (x, y, y') at the point x on the base extremal g_o. The Euler equations of (2.1) take the form

$$(2.2) \qquad \frac{d}{dx}\Omega_{u'_i} - \Omega_{u_i} = 0 \qquad (i = 1, \ldots, n)$$

and are called the Jacobi equations. They are linear in (u) (u') and (u'') and, under our assumption that f is regular, can be solved for the u''_i as linear functions of the components of (u') and (u) with coefficients which are continuous in x. A solution of (2.2) of class C" is called a secondary extremal. The set $(u) = (0)$ is such a secondary extremal.

It is important to see that conjugate values x as defined for a secondary extremal of a second variation I based on an extremal g_o, agree with conjugate values x as defined for g_o. A link connecting the Euler and Jacobi equations is the observation of Jacobi that if,

$$y_i = F_i(x, b) \qquad (i = 1, \ldots, n) \tag{2.3}$$

is a 1-parameter family of extremals which reduces to g_o when $b = b_o$, then

$$u_i(x) = \frac{\partial F_i(x, b_o)}{\partial b} \tag{2.4}$$

defines a secondary extremal. It is assumed that $F_i(x, b)$ is of class C" for x on $[a^1, a^2]$ and $|b - b_o|$ sufficiently small. The proof of Jacobi's theorem is immediate on making use of the fact that F satisfies the Euler equations for each b near b_o, differentiating these Euler equations with respect to b, and finally making the substitution (2.4) and the substitution

$$u_i'(x) = \frac{\partial^2 F_i(x, b_o)}{\partial x \partial b}$$

In particular the jth column of the Jacobian of §1

$$\frac{D(h_1, \ldots, h_n)}{D(b_1, \ldots, b_n)} = D(x, x_o) \tag{2.4}$$

evaluated on g_o, defines a secondary extremal

$$u_{ij}(x, x_o) \qquad (i, j = 1, \ldots, n) \tag{2.5}$$

which satisfies the initial conditions

$$u_{ij}(x_o, x_o) = 0 \qquad u_{ij}'(x_o, x_o) = \delta_{ij}$$

and for which the determinant

$$|u_{ij}(x, x_o)| = D(x, x_o) \tag{2.6}$$

Instead of using the determinant (2.6) to define values x conjugate to x_o one could equally well use the determinant $D_1(x, x_o)$ of any n independent secondary extremals which vanish at x_o, since it is clear that for x_o fixed

$$(2.7) \qquad D(x, x_o) = C\, D_1(x, x_o)$$

where C is a non-null constant. A necessary and sufficient condition that x_1 be conjugate to x_o is clearly that there exists a non-null secondary extremal for which (u) vanishes at x_o and x_1. This fact shows that when x_o is conjugate to x_1, x_1 is conjugate to x_o. The number of independent secondary extremals for which (u) vanishes at x_o and x_1 is called the <u>multiplicity</u> of x_1 as a conjugate point of x_o. This multiplicity is at most n.

To determine conjugate pairs x, x_o relative to any secondary extremal E, one observes that, relative to E, I is its own second variation. Hence relative to E the values x conjugate to x_o are the zeros x of $D(x, x_o)$, and so are independent of E and agree with conjugate pairs (a, x_o) defined for the base extremal g_o.

We continue with a classical lemma.

<u>Lemma 2.1. If (u) defines non-null secondary extremal E, and vanishes at x = a and x = b, then on E</u> -

$$(2.7) \qquad \int_a^b 2\Omega\, dx = 0.$$

Because of the homogeneity of Ω in u_i and u_i' we can write the integral (2.7) as

$$\int_a^b (u_i \Omega_{u_i} + u_i' \Omega_{u_i'})\, dx$$

and on integrating by parts as

$$[u_i \Omega_{u_i}]_a^b + \int_a^b u_i[\Omega_{u_i} - \frac{d}{dx}\Omega_{u_i'}]\,dx = 0$$

from which the lemma follows.

If x^1 and x^2 are two values of x which are not mutually conjugate there exist sets of n secondary extremals $v_{ij}(x)$ and $w_{ij}(x)$ for which

$$v_{ij}(x^1) = \delta_{ij} \qquad v_{ij}(x^2) = 0$$

$$w_{ij}(x^1) = 0 \qquad w_{ij}(x^2) = \delta_{ij}$$

There then exists a family of secondary extremals of the form

$$u_i(x) = a_j\, v_{ij}(x) + b_j\, w_{ij}(x) \tag{2.8}$$

on which (u) = (a) at x^1, and (u) = (b) at (x^2). The representation (2.8) will be of general use.

§3. The index form. We need to show that I has a finite index. It is also necessary to show that there is at most a finite set of values x on $[a^1, a^2]$ conjugate to a^1. For this and later purposes we introduce an index form Q(z), defined as follows.

We shall consider broken secondary extremals, that is continuous arcs made up of a finite sequence of secondary extremals. The corners and end points on these broken extremals E shall lie at points x_j for which

$$a^1 = x_1 < x_2 < \dots < x_m < x_{m+1} = a^2 \tag{3.1}$$

It follows from the nature of the determinant $D(x, x_o)$ whose zeros define conjugate points of x_o that there exists a positive constant e, independent of x_o on $[a^1, a^2]$, such that there is no conjugate point x of x_o for which $|x - x_o| \leqq e$. We suppose the points (3.1) so chosen that

$$|x_j - x_{j+1}| < e \qquad (j = 1, \ldots, m)$$

One can prescribe the value of (u) on E at each x_j given by (3.1). We shall take $(u) = (0)$ at a^1 and a^2. On each interval (x_j, x_{j+1}) the secondary extremal arcs of E can be linearly represented in terms of the values of (u) at x_j and x_{j+1} as in (2.8). Let the set of $q = n(m-1)$ values of (u)

$$(3.2) \qquad [u_1(x_2), \ldots, u_n(x_2), \ldots, u_1(x_m) \ldots u_n(x_m)]$$

be respectively denoted by

$$[z_1, \ldots, z_q] = [z]$$

or more specifically by $[z^u]$. When the broken extremal (u) is piece-wise represented as in (2.8) in terms of the parameters $[z^u]$ then

$$I(u) = Q[z] = Q[z^u]$$

where Q is a quadratic form in the variables $[z]$. We term Q an <u>index form</u> based on the extremal g_o.

If (v), more generally, represents an arbitrary arc which is admissible on $[a^1, a^2]$ (i.e. of class D' with $(v) = (0)$ at a^1 and a^2) there exists a unique broken extremal (u) such that $(v) = (u)$ at the points x_j of (3.1). The parameters $[z^u]$ are thus uniquely determined by (v) and will be denoted by $[z^v]$ as well as by $[z^u]$. From the minimizing property of each separate extremal arc of (u) we have the relation*

$$(3.5) \qquad I(v) > I(u) = Q(z^v) \qquad (u) \neq (v)$$

We can prove the following theorem.

---

* The separate extremal arcs are minimizing without limitation as to neighborhood because the Weierstrass fields involved are without such limitation for secondary extremals.

Theorem 3.1. The index of I equals the index of Q. [Cf. §1 for definition].

Let H_s be an s-plane in the space of admissible arcs (v) on which I is negative definite. Let K_r be the r-plane in the space (z) of sets (z') determined by the arcs (v) of H_s. It follows from (3.5) that Q is negative definite on K_r. Clearly $r \leqq s$. But if $r < s$ there would exist a (v) on H_s such that $I(v) < 0$ and $(z^v) = (0)$. This is impossible by virtue of (3.5). Hence $r = s$ and the index of Q is at least that of I.

But the index of I is at least that of Q since a $(z) \neq (0)$ determines a broken extremal on which $(u) \not\equiv (0)$ and

$$I(u) = Q(z)$$

Theorem 3.1 follows.

The augmented index of I is defined as the maximum of the dimensions of hyperplanes on which $I \leqq 0$ in the space of admissible arcs (u).

Theorem 3.2. The augmented index of I equals that of Q.

The proof of this theorem is similar to that of Theorem 3.1 provided it is true that when $(v) \neq (0)$ and $I(v) \leqq 0$, then $(z^v) \neq (0)$. This is clear if (v) defines an admissible broken extremal, because the relation $(v) \neq (0)$ then implies $(z) \neq (0)$; but if (v) is not an admissible broken extremal, (3.5) takes the form

$$I(v) > I(u) = Q(z^v)$$

so that $I(v) \leqq 0$ implies $(z^v) \neq (0)$.

Theorem 3.2 follows.

Lemma 3.1. The augmented index of I is at least the number of conjugate points of a^1 on the interval $a^1 < x \leqq a^2$.

Suppose the conjugate points of a^1 include the points c_j for which

$$a^1 < c_1 < c_2 < \dots < c_r \leqq a^2$$

Corresponding to c_j there exists a secondary extremal (w^j) which vanishes at a^1 and c_j. Let (w^j) be extended in definition so that

$$(w^j) \equiv (0) \qquad\qquad (c_j \leqq x \leqq a^2)$$

The vector functions (w^j) so extended are linearly independent. In fact there could be no linear relation between these r vectors which effectively involved (w^r), since $(w^r) \not\equiv (0)$ for $c_{r-1} \leqq x \leqq c_r$, and each other vector $(w^j) \equiv (0)$ on this interval. Similarly (w^{r-1}) cannot be effectively involved, nor (w^{r-2}) etc., so that no proper linear relation is possible.

It follows however from Lemma 2.1 that $I(w^j) = 0$, and hence $I = 0$ on the r-plane with base (w^1), ..., (w^r). The lemma follows from the definition of the augmented index.

The following is a particular consequence of the lemma.

<u>Theorem 3.3.</u> <u>There is at most a finite number of conjugate points of a^1 on the interval</u> $[a^1, a^2]$.

§4. <u>The nullity of Q and I.</u> The nullity of the quadratic form $Q(z)$ may be defined as the number of linearly independent critical points* (z) of Q. An admissible arc which is an extremal** of I will be called a <u>critical extremal</u> and the nullity of I is defined as the number of linearly independent critical extremals of I. We wish to prove that the nullity of Q equals that of I.

---

* That is sets $(z_1, \dots, z_q)$ for which all the partial derivatives of Q vanish.

** We drop the qualification "secondary" when there is no ambiguity.

To that end let (u) and (v) represent arcs of class D' on $[a^1, a^2]$, and set

$$(4.0) \qquad H(u, v) = \int_{a^1}^{a^2} [f^o_{p_i p_j} u_i' v_j' + f^o_{p_i y_j}(u_i' v_j + u_j v_i') + f^o_{y_i y_j} u_i v_j]dx$$

so that $H(u, v) = H(v, u)$ and $H(u, u) = I(u)$.

Observe that

$$I(u+ev) = H(u, u) + 2eH(u, u) + e^2 H(v, v)$$

and recall that a necessary and sufficient condition in the fixed end-point problem that the first variation based on an arc (u) of class D' vanish, is, that (u) be an extremal.

Hence a necessary and sufficient condition that (u) be an extremal on $[a^1, a^2]$ is that the first variation $2H(u, v) = 0$ for every admissible arc (v).

As in §3 an admissible arc (u) determines a set $(z) = (z^u)$.

Lemma 4.1. A necessary and sufficient condition that an admissible broken extremal (u) be a critical extremal of I is that (z^u) be a critical point of Q.

Let (u) be a critical extremal of I and let (v) be an arbitrary admissible broken extremal. Then for any constant e

$$(4.1) \qquad Q(z^u + ez^v) = I(u + ev)$$

Differentiation with respect to e gives the result

$$(4.2) \qquad z_p^v \, Q_{z_p}(z^u) = 2H(u, v), \qquad (p = 1, \ldots q)$$

Since (u) is an extremal, $H(u, v) = 0$. Since z_p^v is arbitrary we infer that (z^u) is a critical point of Q. The condition of the theorem is accordingly necessary.

Suppose then that (z^u) is a critical point of Q determined by an admissible broken extremal (u). Let (y) represent an arbitrary admissible arc (of class D' with (y) = (0) at a^1 and a^2). Then (y) can be represented as a sum

$$(y) = (v) + (v^*)$$

where (v) is an admissible broken extremal and (v^*) vanishes at each of the points x_j of (3.1). Then

$$H(u, y) = H(u, v) + H(u, v^*). \tag{4.3}$$

Now H(u, v) satisfies a relation (4.2) and is null since (z^u) is a critical point of Q. But $H(u, v^*)$ is the sum of integrals of the type (4.0) with limits x_j and x_{j+1}, evaluated along an extremal arc of I, and for an arc of (v^*) which vanishes at x_j and x_{j+1}. Hence each such contribution to $H(u, v^*)$ is null, making $H(u, v^*)$ null. Thus H(u, y) = 0 in accordance with (4.3), so that (u) is a critical extremal.

The proof of the lemma is complete. The following theorem is a consequence of the lemma.

Theorem 4.1. *The nullity of I equals the nullity of Q.*

The augmented index of Q equals its index plus its nullity as is readily seen on reducing Q to a sum of squared terms with coefficients ±1 using a real affine transformation. We can accordingly prove the following theorem.

Theorem 4.2. *The augmented index of I equals its index plus its nullity.*

This follows from the equality of the nullity, index and augmented index of Q to those, respectively of I. Since the augmented index of Q equals its index plus its nullity the same is true of I.

§5. Evaluation of the index of I. The following lemma will afford the basis for an evaluation of the index of I.

Lemma 5.1. Let $Q(z, c)$ be a quadratic form in the variables (z) with a parameter c on the closed interval $[a, b]$ with the following properties:

(1) The coefficients of the form Q vary continuously with c.

(2) The index $h(c)$ and augmented index $k(c)$ of Q never decrease as c increases.

(3) The values of c at which Q is degenerate are finite in number.

(4) For $c = a$ the form Q is positive definite.

Then $Q(z, b)$ has an index equal to the sum of the nullities $v(c)$ of Q for the values of c on $a < c < b$ at which Q is degenerate.

Let s be a value of c on the interval $a < c < b$ for which $Q(z, s)$ is degenerate. It is only at such values s that $h(c)$ and $k(c)$ can change. In particular

$$h(s^+) = k(s^+), \tag{5.1}$$

where we indicate a limit of h from the right at s by suffixing a '+'.

We shall verify the relations

$$h(s^-) = h(s), \qquad k(s) = k(s^+). \tag{5.2}$$

First $h(s^-) \leqq h(s)$ by virtue of (2). But the continuous dependence of the characteristic roots of Q on c implies that $h(s)$ negative characteristic roots exist and remain negative for c sufficiently near s. Hence $h(s^-) \geqq h(s)$. We conclude that $h(s^-) = h(s)$.

We have $k(s) \leqq k(s^+)$ by virtue of (2). The continuity of characteristic roots as functions of c implies that the $k(s^+)$ roots which are negative as c tends to s from above, remain non-positive at s, so that $k(s) \geqq k(s^+)$.

Hence (5.2) holds. From the relation

$$h(s) + v(s) = k(s)$$

and relations (5.1) and (5.2) one finds that

$$h(s^-) + v(s) = h(s^+)$$

The lemma follows from this relation together with (4), and the relation $h(b^-) = h(b)$.

Theorem 5.1. The index of I equals the number of conjugate points of a^1 on the interval $a^1 < x < a^2$, counting each conjugate point with its multiplicity.

The index of I equals the index of the corresponding index form Q. To apply the preceding lemma we shall replace the upper limit a^2 of I by c where $a^1 < c \leqq a^2$ thereby defining an integral $I^c(u)$. Here (u) represents an arc which is admissible on $[(a^1, c)]$. To define a corresponding index form $Q^c(z)$ let each value of x_j in (3.1) be replaced by a value $x_j' = cx_j/a^2$, supposing for simplicity that $a^1 = 0$. On using broken extremals whose corners and end points occur respectively when $x = x_j'$, $Q^c(z)$ can be defined in the same manner as $Q(z)$.

We shall see that Q^c, so defined, satisfies the conditions of Lemma 5.1, provided in that lemma one takes $b = a^2$ and $a > a^1$ so near a^1 that I^a is positive definite. Conditions (1) and (4) of the lemma are then satisfied.

Condition (3) is satisfied, since Q^c is degenerate only if I^c has a positive nullity. This happens only if c is a conjugate point of a^1, and we have seen in Theorem 3.3 that there is at most a finite number of points on $[a^1, a^2]$ conjugate to a^1.

To establish Condition (2) of Lemma 5.1 for Q^c it is sufficient to show that the index and augmented index of I^c never decrease as c increases, since these indices are the same for I^c and Q^c.

To this end let h(c) be the index of I^c. There then exists an h(c)-plane X in the space of arcs which are admissible on $[a^1, c]$ such that I^c is negative definite on X. If c' is a value on $(a^1, a^2]$ with $c' > c$, one can obtain an h(c)-plane of arcs which are admissible on $[a^1, c']$ by extending every arc of X along the x-axis from $x = c$ to $x = c'$. Hence the index of I^c is non-decreasing as c increases. Similarly k(c) is non-decreasing.

The theorem now follows from the lemma and the fact that the nullity of $Q^c(z)$ equals the multiplicity of c as a conjugate point of a^1.

Since one can interchange the role of a^1 and a^2 in the preceding proof, one has the immediate corollary:

Corollary. The number of conjugate points of a^1 on the open interval (a^1, a^2) equals the number of conjugate points of a^2 on that interval.

We shall generalize Sturm's famous separation theorem to the case where the number n of ordinates exceeds 1. Sturm's theorem concerns a single second-order differential equation, and such an equation is always the Jacobi equation of a suitably chosen form I(u). A lemma is needed.

Lemma 5.2. If in a quadratic form P(z), n of the variables (z) are set equal to 0 to define a form $P_1(z)$, then

$$h(P) = h(P_1) + r \qquad (0 \leqq r \leqq n) \tag{5.3}$$

where h(P) is the index of the form P(z).

It is obvious that $h(P) \geqq h(P_1)$. It remains to show that

$$h(P) \leqq h(P_1) + n \tag{5.4}$$

Let X be a linear space in the space of coordinates (z) of dimension h(P), with P negative definite on X. The subspace Y of X on which the n given z's equal 0 has a dimension at least h(P)- n. On Y, P_1 is negative definite so that $h(P_1)$ is at least h(P)- n. Relations (5.4) and (5.3) follow.

Conjugate points x of x_o have been defined by the zeros $x \neq x_o$ of the determinant

$$D(x, x_o) = | u_{ij}(x, x_o) |$$

We shall now consider the zeros of $D(x, x_o)$ not excepting $x = x_o$. Each zero $x \neq x_o$ will be counted with a multiplicity equal to the multiplicity of x as a conjugate point of x_o. The zero $x = x_o$ will be counted with a multiplicity n, since there are n independent extremals which vanish at x_o. It can be shown that the multiplicity of a zero x of $D(x, x_o)$ equals the ordinary order of vanishing of D. See Coll. Lect.p.47. No use will be made of this fact.

Let (b, c) be an open subinterval of (a^1, a^2). Let N(a: b, c) be the number of zeros of D(x, a) on (b, c) counting these zeros with their multiplicities. In accordance with the preceding corollary,

(5.5) $$N(b\colon b, c) = N(c\colon b, c)$$

Since the conjugate points of a point b are isolated one can decrease c slightly without altering the numbers (5.5). Similarly one can increase b slightly.

<u>Theorem 5.3.</u> <u>The number of zeros of</u> $D(x, x_o)$ <u>on any open subinterval</u> (b, c) <u>of</u> $[a^1, a^2]$ <u>differ for different values of</u> x_o <u>by at most</u> n.

The lemma follows immediately from the relation.

(5.6) $$N(a\colon b, c) = N(b\colon b, c) + r \qquad (0 \leqq r \leqq n)$$

which we shall now prove. In proving (5.6) we can suppose that a is not conjugate to b since a sufficiently small increase in b will not affect the numbers N in (5.6) and provide that a and b are not conjugate.

There are essentially three cases to be considered

(5.7) $\quad a < b < c \qquad$ (or $b < c < a$),

(5.8) $\quad a = b \qquad$ (or $a = c$),

(5.9) $\quad b < a < c$.

We begin with the case $a < b < c$ and consider the index form Q_a^c for which the basic interval is $[a, c]$. In constructing the broken extremals we can suppose that for some j, $x_j = b$. If one sets the ordinates of the vertices on $x_j = b$, equal to 0, Q_a^c reduces to a sum

$$Q_a^b + Q_b^c .$$

In accordance with Lemma 5.2

$$h[Q_a^c] = h[Q_a^b] + h[Q_b^c] + r \qquad (0 \leqq r \leqq n) \tag{5.10}$$

The left member of (5.10) is the number of conjugate points of <u>a</u> on (a, c); that is on (a, b) and (b, c) since b is not conjugate to a. Thus (5.10) takes the form

$$N(a\colon a, b) + N(a\colon b, c) = N(a\colon a, b) + N(b\colon b, c) + r$$

Relation (5.6) follows when $a < b < c$; similarly (5.6) holds when $b < c < a$, and holds trivially with $r = 0$ when $a = b$ or $a = c$.

There remains the case $b < a < c$. To evaluate the zeros of $D(x, a)$ on (b, c) as called for in (5.6), one counts these zeros on (b,a), (a,c) and at $x = a$ respectively. Thus

$$N(a\colon b, c) = N(a\colon b, a) + N(a\colon a, c) + n. \tag{5.11}$$

With $b < a < c$ one can replace (a, b, c) in (5.6) by (b, a, c), so that from (5.6)

(5.12) $\qquad N(b\colon a, c) = N(a\colon a, c) + r_1 \qquad (0 \leq r_1 \leq n).$

From (5.11) and (5.12) one finds that

$$N(a\colon b, c) = N(b\colon b, a) + N(b\colon a, c) - r_1 + n$$
$$= N(b\colon b, c) - r_1 + n.$$

Thus (5.6) holds in all cases; and the lemma is thereby established.

The separation Theorem 5.3 holds if the open interval (b, c) is replaced by an arbitrary subinterval of $[a^1, a^2]$. One has only to note that the zeros of $D(x, x_0)$ and $D(x, x)$ on $(b, c]$, for example, are identical with these on (b, c') if $c' > c$ and $c' - c$ is sufficiently small. The theorem holds for (b, c') and hence for $(b, c]$.

One can use the properties of the indices of Q_a^b as functions of a and b to prove that the r-th conjugate point of $x = a$ (if it exists) is an increasing function of a, and a continuous function of any parameter s appearing in $f(x, y, y', s)$ with f of Class "". Or again the r-th conjugate point of $x = a$ is a decreasing function of the characteristic parameter s introduced in the next section. Details are left to the reader.

§6. Characteristic roots and solutions. The characteristic values s of a quadratic form $P(z) = a_{rp}z_r z_p$ are the roots of the determinant

(6.1) $\qquad |a_{rp} - s\,\delta_{rp}| = 0 \qquad (r, p = 1, \ldots, q)$

If one sets

(6.2) $\qquad P(z, s) = P(z) - s\, z_p z_p$

one sees that (6.1) is the condition that the system of linear equations in (z),

(6.3) $\qquad P_{z_p}(z, s) = 0 \qquad (p = 1, \ldots, q)$

possess a non-null solution (z) for some value s. Such a value s_o is called characteristic, and a corresponding solution (z) of (6.3) a characteristic solution of $I(u)$ with root s_o. The multiplicity of the root s_o is taken as the number of linearly independent solutions (z) of (6.3) when $s = s_o$.

A useful form of statement is that a characteristic root s_o is a value of s for which the form $P(z, s)$ has a non-null critical point.

It is this form of statement which we can immediately extend to the second variation $I(u)$, evaluated on admissible arcs (u).

With (u) and (v) admissible arcs it will be convenient to introduce the inner product

$$(u, v) = \int_{a^1}^{a^2} u_i(x)v_i(x)dx \qquad (i = 1,\ldots,n)$$

The analogue of $P(z, s)$ is then

$$I(u, s) = I(u) - s(u, u) \tag{6.4}$$

or perhaps more suggestively

$$= H(u, u) - s(u, u)$$

If one sets

$$L_i(u) = \frac{d}{dx}\Omega_{u_i'} - \Omega_{u_i} \qquad (i = 1, \ldots, n)$$

then the analogue of (6.3) is the Jacobi differential system

$$L_i(u) - s\, u_i = 0 \qquad (i = 1, \ldots, n) \tag{6.5$'$}$$

$$u_i(a^1) = u_i(a^2) = 0 \tag{6.5$''$}$$

A value s_o and an arc $(u) \not\equiv (0)$ of class C'' satisfying (6.5) are called a characteristic value and solution, respectively, of $I(u)$. We shall then refer to (s_o, u) as a characteristic set. If in the preceding italicized statement concerning Q one replaces "critical point" by "admissible extremal arc" one has the following definition.

A characteristic value s_o *of* $I(u)$ *is a value of* s *for which the integral* $I(u, s)$ *admits a non-null admissible extremal arc.*

As previously, the *multiplicity* of a characteristic value s_o of $I(u)$ is the number of linearly independent admissible extremal arcs of $I(u, s_o)$. This number is at most n, since there are just n linearly independent extremals of $I(u, s_o)$ for which $(u) = (0)$ at a^1.

If (u) and (v) are admissible arcs, and if the inner product $(u, v) = 0$, one says that (u) and (v) are orthogonal. The following lemma is classical.

Lemma 6.1. If (s^0, u^0) *and* (s^1, u^1) *are characteristic sets for which* $s^0 \neq s^1$ *then* (u^0) *is orthogonal to* (u^1).

It will be convenient to introduce the bilinear form

$$B(u, v, s) = H(u, v) - s(u, v) = B(v, u, s). \tag{6.6}$$

For $k = 0, 1$, (u^k) is an admissible extremal arc of $B(u, u, s^k)$ so that

$$B(u^k, v, s^k) = B(v, u^k, s^k) = 0 \tag{6.7}$$

for every admissible arc (v). But by definition of B

$$B(u, v, s^0) = B(u, v, s^1) + (s^1 - s^0)\,(u, v)$$

In particular

$$B(u^0, u^1, s^0) = B(u^0, u^1, s^1) + (s^1 - s^0)\,(u^0, u^1)$$

Since $s^1 \neq s^0$, and since

$$B(u^0, u^1, s^0) = B(u^0, u^1, s^1) = 0$$

we conclude that $(u^0, u^1) = 0$ as required.

The following lemma shows that the number of characteristic roots s less than a constant c is finite. We already know that the index of $I(u, c)$ is finite.

Lemma 6.2. The index of $I(u, c)$ *is at least the number of distinct characteristic roots* $s < c$.

Let

$$(s^1, u^1), \ldots, (s^m, u^m)$$

be any finite ensemble of characteristic sets (s, u) for which $s^1 < \ldots < x^m < c$. Let X be the linear space determined by the solutions

$$(u^1), \ldots, (u^m)$$

There can be no proper linear relation between these solutions because of their orthogonality. Hence the dimension of X is m. If

$$(v) = b^p(v^p) \qquad (p = 1, \ldots, m)$$

is a non-null arc (v) in X. We have for $(p, q = 1, \ldots, m)$

$$\begin{aligned} I(v, c) = B(v, v, c) &= b^p b^q B(u^p, u^q, c) \\ &= b^p b^q [B(u^p, u^q, s^p) + (s^p - c)(u^p, u^q)] \end{aligned}$$

Since (u^p, s^p) is a characteristic set

$$B(u^p, u^q, s^p) = 0 \qquad (p = 1, \ldots, m)$$

On using the orthogonality relations one then finds that

$$I(v, c) = b^p b^p (s^p - c)\ (u^p, u^p) < 0\ ;$$

and therefore the index $h \geqq m$.

The number of characteristic roots less than c is accordingly finite so that these roots are bounded below (if in fact there exist any such roots).

The proof of the next lemma involves the following theorem of Kronecker (Kowalewski, Determinanten (1909) §102): A quadratic form $a_{ij}w_iw_j$ is positive definite if, and only if, all m-rowed principal minors

(6.8) $$A_m = \det(a_{ij}) \geqq 0\ , \qquad (1 \leqq i, j \leqq m)$$

Lemma 6.3. There exists an s^* such that $I(u, s)$ is positive definite if $s < s^*$.

$I(u, s)$ is the integral of a quadratic form

$$2\Omega(u, u') - s\, u_i u_i\,, \qquad (i = 1, 2, \ldots, n) \tag{6.9}$$

with coefficients depending continuously on x, $a^1 \leqq x \leqq a^2$. We arrange the variables in (6.9) in the order $(u_1', \ldots, u_n', u_1, \ldots, u_n)$. For $m \leqq n$ the minors (6.8) for (6.9) are equal to the corresponding minors of the Hessian matrix

$$(f^o_{p_i p_j})$$

whose associated quadratic form is supposed to be positive definite in the positive regular problem, and thus are positive. Denote the determinant of the Hessian by A. Then the leading term of A_m for $m = n+k$ is

$$(-s)^k A\,.$$

Hence each $A_m (m > n)$ approaches + infinity as $-s$ approaches + infinity. It follows that if $-s$ is sufficiently large the quadratic form (6.9) will be positive definite for each x. Lemma 6.3 follows.

To evaluate the index of $I(u, s)$ we make use of the equality of the index of $I(u, s)$ to that of the index form $Q(z, s)$, as defined in §3. In order that the construction of broken extremals of §3 be possible we confine s to a finite interval $[a, b]$. An extremal (u) of $I(u, s)$ depends continuously on x, s, and initial values $(x^o, u^o, u^{o\prime})$. The x-derivative (u') is similarly continuous: In particular the elements of the determinant $D(x, x_o)$ and their x-derivatives are uniformly continuous in (x, x_o, s) for x and x_o on $[a^1, a^2]$ and s on $[a, b]$. There accordingly exists a constant $e > 0$ such that there are no pairs of conjugate points on a subinterval of $[a^1, a^2]$ of length less than e. The construction of §3 is accordingly possible with a single set of points.

$$a^1 = x_1 < x_2 < \dots < x_m < x_{m+1} = a^2$$

serving for all s on [a, b].

We can now prove a fundamental theorem.

Theorem 6.1. The index of I(u, b) equals the number of characteristic roots $s < b$ of I(u, s), each root counted with its multiplicity.

As in the preceding paragraph we construct the form Q(z, s) for $a \leq s \leq b$. In accordance with Lemma 6.3 one can take a so that I(u, a) and hence Q(z, a) is positive definite. We apply Lemma 5.1 to evaluate the index of Q(z, b). The conditions of Lemma 5.1 on Q are verified as follows:

(1) The coefficients of the form Q(z, s) vary continuously with s, because for fixed (u), s enters into I(u, s) through the term s(u, u), and because u_i and u_i' on an admissible broken extremal depend linearly on (z) with coefficients which are continuous in (x, s).

(2) The index and augmented index of Q never decrease as s increases since

$$Q(z, s') > Q(z, s'')$$

whenever $s' < s''$ and $(z) \neq (0)$. This is because s enters into I(u, s) in the form $-s(u, u)$.

(3) The form Q(z, s) is degenerate for at most a finite set of values of $s < b$. In fact each such value s_0 is a characteristic root of I(u) since $Q(z, s_0)$ and $I(u, s_0)$ have the same nullity.

(4) By virtue of the choice of a, Q(z, a) is positive definite.

The theorem follows from Lemma 5.1 and from the fact that the nullity of $Q(z, s_0)$, equals the nullity of $I(u, s_0)$, and hence by definition, equals the multiplicity of s_0 as a characteristic root.

<u>Corollary 6.1.</u> <u>The number of characteristic values s of I(u) with</u> $s < c$, <u>equals the number of conjugate points on</u> $a^1 < x < a^2$ <u>of the point</u> a^1, <u>taking conjugate points relative to</u> I(u, s) <u>and counting conjugate points and characteristic roots with their multiplicities.</u>

<u>Corollary 6.2.</u> <u>A necessary and sufficient condition that</u> $I(u) \geqq 0$ <u>for admissible arcs</u> (u) <u>is that there be no negative characteristic roots.</u>

The condition is necessary, for if there are negative characteristic roots the index of I would be positive. The condition is sufficient because

$$I(u) \geqq Q(z^u)$$

and $Q \geqq 0$; since the index of I and hence of Q is zero.

<u>Corollary 6.3.</u> <u>A necssary and sufficient condition that I be positive definite is that each characteristic root be positive.</u>

It is necessary that no root be negative by the preceding corollary. If a root is null, $I(u) = 0$ on the corresponding characteristic solution, contrary to hypothesis.

The condition is sufficient because the positive values of the characteristic roots imply that the index and the nullity of I are zero. Hence the augmented index k of I is zero, and I is positive definite by virtue of the definition of k.

<u>The case n = 1.</u> In case $n = 1$ the determinant $D(x, x_0)$ whose zeros $x \neq x_0$ determine the points x conjugate to x_0, reduces to a solution $u(x, x_0)$ of the single second order Jacobi equation such that

$$u(x_0, x_0) = 0 \qquad\qquad u'(x_0, x_0) = 1$$

Hence Corollary 6.1 gives a classical result: When $n = 1$ the number of characteristic roots of I with values less than c equals the number of zeros on $a^1 < x < a^2$ of a non-null extremal of $I(u, c)$ vanishing at a^1. The multiplicities of the characteristic roots, and of the zeros of the extremal are all unity in this special case. The generalization to the case $n > 1$, as given in Corollary 6.1, is due to the writer.

We have seen that the characteristic roots are bounded below, and are finite in number on any bounded interval. We conclude with the following.

Theorem 6.2. There are infinitely many characteristic roots.

Let r be any positive integer. We shall show that if c is sufficiently large the index of $I(u, c)$ is at least r. It then follows from Theorem 6.1 that there are at least r roots less than c.

One determines a suitable value of c as follows. Let $E_1, \ldots, E_r$ be r disjoint subintervals of $[a^1, a^2]$. Let (u^j) be an admissible arc which is null except on E_j, but is not identically null on E_j. The arcs E_j are clearly linearly independent. One can take c so large that

$$I(u^j, c) < 0 \qquad (j = 1, \ldots, r)$$

since c enters only through the term $-c(u, u)$. If

$$(v) = a^j(u^j) \qquad (j = 1, \ldots, r)$$

represents any arc on the r-plane determined by the arcs (u^j)

$$I(v, c) = a^j a^j \, I(u^j, c) < 0 \qquad (a) \neq (0)$$

so that the index of $I(u, c)$ is at least r.

The theorem now follows from Theorem 6.1.

The theory which has been developed in the preceding sections can be applied to situations where the construction of the index form Q is not possible. It is sufficient that the general quadratic form in question be at least as great as a form of the type studied here. The general theory so obtained goes much beyond the calculus of variations and includes integro-differential systems of great generality with general self-adjoint boundary conditions.

§7. A coordinate manifold M. Problems in the large in variational theory are geometric in character and this usually means that the spaces in which the problems are defined are of a type given by overlapping coordinate systems. We shall give a set of axioms which define a topological space, then a Hausdorff space, and finally a coordinate n-manifold.

A topological space T is a set of points together with an aggregate of subsets U of T, termed open sets of T such that:

(1) The null set and T itself are open.

(2) The union of any aggregate of open sets is open.

(3) The intersection of any finite aggregate of open sets is open.

Two topological spaces T_1 and T_2 are said to be topologically equivalent or the topological images of each other if their points admit a 1 - 1 correspondence in which open sets of either correspond to open sets of the other. Any open set is termed a neighborhood of each of its points.

A topological space is termed a Hausdorff space if each pair of disjoint points is in at least one pair of disjoint neighborhoods. A topological space is termed connected if not the union of two non-void disjoint open sets.

We suppose that M is a connected Hausdorff space that is a coordinate n-manifold in the following sense. There should exist a subset of the open sets of M, termed coordinated regions, with the following properties.

(a) The coordinate regions cover M.

(b) Each coordinate region N is the topological image of a region R in an Euclidean n-space E_n. The rectangular coordinates (x) of a point of R are termed coordinates of the image point on N.

(c) If coordinate regions N_1 and N_2 with point coordinates (x) and (y) respectively intersect in N, then sets (x) and (y) which represent the same point on N shall stand in a relation

$$y^i = y^i(x) = y^i(x^1, \ldots, x^n) \qquad (i = 1, \ldots, n) \tag{7.1}$$

where $y^i(x)$ is of class C^4 in its arguments and the transformation from (x) to (y) has a non-vanishing jacobian. Any subregion N of M will be <u>admitted</u> as a coordinate region if (b) and (c) are satisfied on adding N to the set of coordinate regions.

On a coordinate manifold, tensors are defined as usual. In particular let P be a point of M represented by admissible coordinates (x), (y), (z), etc. A <u>contravariant</u> vector is defined at P if its n components

$$(r), \ (\rho), \ (\sigma) \text{ etc.}$$

are given at the points (x), (y), (z) in the systems (x), (y), (z), etc. respectively, and are related as are the differentials (dx), (dy), (dz), etc. For example,

$$\sigma^h = \frac{\partial z^h}{\partial x^i} r^i \qquad (h, i = 1, \ldots, n) \tag{7.2}$$

A <u>covariant</u> vector is defined at P if its n components are given at (x), (y), (z), etc. respectively and are related as are the partial derivatives of an invariant function

$$X(x) = Y(y) = Z(z) = \ldots\ldots$$

For example

$$X_{x^i} = \frac{\partial z^h}{\partial x^i} Z_{z^h}$$

The integral to be studied has an integrand F defined in each coordinate system. In typical coordinates (x), F is a function $F(x, r)$ of class C^3 of the point (x) and of a contravariant vector $(r) \neq (0)$. In the space (x) one admits a <u>piece-wise regular arc</u>

$$x^i = x^i(t) \qquad (0 \leqq t \leqq 1) \qquad (i = 1, \ldots, n)$$

in which $x^i(t)$ is of class C^1 on each of a finite number of closed subarcs $[a, b]$ of $[0, 1]$ with

$$\dot{x}^i\dot{x}^i \neq 0 \qquad \text{on } [a\ b]$$

The integral has the form

$$J = \int_0^1 F(x, \dot{x})dt \tag{7.3}$$

In the intersection of two regions of coordinates (x) and (z) one sets

$$G(z, \sigma) = F(x, r)$$

subject to (7.2) so that

$$J = \int_0^1 G(z, \dot{z})dt$$

along any piece-wise regular arc in the space (z).

As is easily shown, a necessary and sufficient condition that the integral J in the coordinate system (x) be independent of admissible changes of the parameter t is that

$$F(x, kr) \equiv k\,F(x, r) \qquad (k > 0) \tag{7.4}$$

for every real positive constant k. One admits changes of parameter $t' = a(t)$ in which $a(t)$ is of class C' and $\dot{a}(t) > 0$ on each of a finite set of closed subintervals covering $[0, 1]$. The interval for t' shall again be $[0, 1]$. On differentiating (7.4) with respect to k, and with respect to r^i one infers that

$$r^i F_{r^i}(x, kr) \equiv F(x, r) \tag{7.5)'}$$

$$(7.5)'' \qquad F_{r^i}(x,\ kr) \equiv F_{r^i}(x,\ r)$$

and from (7.5)", on setting $F_{r^i r^j} = F_{ij}$,

$$(7.6) \qquad r^i F_{ij}(x,\ r) \equiv 0$$

It follows from (7.6) that the determinant

$$(7.7) \qquad \left| F_{ij} \right| \equiv 0$$

On any piece-wise regular arc of class C" the Euler operator is

$$(7.8) \qquad E_i(x) = \frac{d}{dt} F_{r^i} - F_{x^i} = F_{r^i r^j} \ddot{x}^j + F_{r^i x^j} \dot{x}^j - F_{x^i}$$

On account of (7.7) the Euler equations $E_i = 0$ cannot be solved by Cramer's rule for the $\ddot{x}^j$. To meet this difficulty one introduces the bordered determinant

$$(7.9) \qquad D = \begin{vmatrix} F_{11}, & \ldots, & F_{1n} & u_1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ F_{n1}, & \ldots, & F_{nn} & u_n \\ v_1, & \ldots, & v_n & 0 \end{vmatrix} = -A^{ij} u_i v_j$$

Here A^{ij} is the cofactor of F_{ij} in the determinant (7.7).

The determinant D vanishes with $r^i u_i$. To see this suppose for definiteness that $r^1 \neq 0$, recalling that $(r) \neq (0)$. On multiplying the i-th row of D by r^i and adding to the first row for each i, the elements in the first row reduce to zero by virtue of (7.6), possibly excepting the last element $r^i u_i$. Hence the determinant D vanishes with $r^i u_i$. On operating similarly on the columns of D one sees that D vanishes with $r^j v_j$. For fixed (x) and

(r), D is quadratic in u_i, v_j so that

$$(7.10) \qquad A^{ij}u_i v_j \equiv F_1(x, r)\,(r^i u_i)\,(r^j v_j)$$

where F_1 does not depend on u_i and v_j.

In the plane, $F_1(x, r)$ was introduced by Weierstrass with the relations,

$$(7.11) \qquad A^{ij} = F_1 u_i v_j ,$$

which follows at once from (7.10). Weierstrass's derivation was different. If one sets $u_i = r^i$ and $v_j = r^i$ one has a formula

$$F_1 = \frac{A^{ij} r^i r^j}{(r^i r^i)^2}$$

showing the continuity of F_1 when $(r) \neq (0)$. As we shall see, the hypothesis that $F_1 \neq 0$, corresponds in the non-parametric form of the integral to the hypothesis that

$$\left| f_{p_i p_j} \right| \neq 0.$$

The Euler operator is a covariant tensor. That is, when (x) and (z) are admissible coordinates

$$(7.12) \qquad \frac{d}{dt} F_{r^i} - F_{x^j} = \frac{\partial z^h}{\partial x^i}\left[\frac{d}{dt} G_{o^h} - G_{z^g}\right]$$

Relation (7.12) can be verified in detail on using the homogeneity relations (7.5) and (7.6) and the relation

$$(7.13) \qquad r^i F_{r^i x^j} \equiv F_{x^j}$$

obtained on differentiating (7.5)' as to x^j. One can derive (7.12) more simply on equating the first variation of J in the system (x) to the first variation of J in the system (z). Another relation of importance is

(7.14) $$\dot{x}^i \left[\frac{d}{dt} F_{r^i} - F_{x^i} \right] \equiv 0 \ .$$

This relation is an immediate consequence of (7.6) and (7.13) setting $r^i = \dot{x}^i$.

§8. The Euler equations. To introduce a metric for our coordinate manifold, we shall use a function $f(x, r)$ defined and homogeneous in (r) in each coordinate system (x), of the precise character of $F(x, r)$ except that $f(x, r)$ shall be positive. The f-length s along any piece-wise regular arc $[x(t)]$ will be given by the integral

$$s = \int_0^t f(x, \dot{x})dt$$

The Euler equations do not determine the parameter t along an extremal. One can make $s = ct$ by adjoining the condition $f \equiv c$ where c is a positive constant independent of t. We are thus led to the system

(8.1)' $$\frac{d}{dt} F_{r^i} - F_{x^i} = 0$$

(8.1)" $$f(x, \dot{x}) = c \ .$$

To arrive at the general solution of (8.1) we shall immerse the system (8.1) in the system

(8.2)' $$\frac{d}{dt} (F_{r^i} + mf_{r^i}) - (F_{x^i} + mf_{x^i}) = 0$$

(8.2)" $$f(x, \dot{x}) = c \ ,$$

where m is an unknown function of t to be added to the unknown $x^i(t)$. The relation of (8.1) to (8.2) is given by the following lemma.

Lemma 8.1. In any solution $m(t)$ and $[x(t)]$ of (8.2), $m(t)$ is constant. Hence if $m = 0$ initially, then (8.2) $\longleftrightarrow$ (8.1).

We have established the identity in t

$$\dot{x}^i\left(\frac{d}{dt}F_{r^i} - F_{x^i}\right) \equiv 0 ,$$

and a similar identity holds for f. Upon multiplying (8.2)' by $\dot{x}^i$ and summing, one then finds that

$$\dot{x}^i f_{r^i}\frac{dm}{dt} - f\frac{dm}{dt} = 0$$

from which the lemma follows.

We shall proceed to solve (8.2) without limiting m to the value 0. It will clarify matters if we set $\dot{x}^i = r^i$ and write (8.2) as a set of conditions on (x, r) and m in the form

$$\text{(8.3)}' \qquad \frac{d}{dt}(F_{r^i} + mf_{r^i}) - (F_{x^i} + mf_{x^i}) = 0$$

$$\text{(8.3)}'' \qquad \frac{d}{dt}x^i = r^i, \qquad f(x, r) = c.$$

The system (8.3) can be transformed into an equivalent set of conditions on (x) and on a covariant vector (v) by setting

$$\begin{array}{ll} \text{(8.4)}' & \left\{\begin{array}{l} v_i = F_{r^i}(x, r) + mf_{r^i}(x, r) \qquad\qquad (i = 1, \ldots, n) \\ \\ c = f(x, r). \end{array}\right. \\ \text{(8.4)}'' & \end{array}$$

We regard (8.4) as a transformation from (x, r, m) to (x, v), with c a parameter of the transformation. We suppose a set (x_o, r_o, m_o) given for t = 0 with $m_o = 0$, and use (8.4) to define

$$v_{io} = F_{r^i}(x_o, r_o), \qquad c_o = f(x_o, r_o) .$$

With this initial solution (x_o, r_o, v_o, c_o) of (8.4) given, we solve (8.4) for x and m as functions

(8.5)' $$r^i = R^i(x, v, c)$$

(8.5)'' $$m = M(x, v, c)$$

of class C'' for (x, v, c) sufficiently near (x_o, v_o, c_o). The relevant jacobian of the right hand members of (8.4) with respect to the variables (r) and m is

$$J = \begin{vmatrix} F_{ij} & f_{r^i} \\ f_{r^j} & 0 \end{vmatrix} = -F_1 (r^i f_{r^i})(r^j f_{r^j}) = -F_1 f^2$$

and $J \neq 0$, since we are assuming that $F_1 \neq 0$.

Under the transformation (8.4), or its inverse (8.5), (8.3) takes the form

(8.6)' $$\frac{dx^i}{dt} = R^i(x, v, c)$$

(8.6)'' $$\frac{dv_i}{dt} = H_i(x, v, c)$$

where we have set

(8.7) $$H_i(x, v) = F_{x^i}(x, R) + M f_{x^i}(x, R).$$

The parameter c is required to be positive and independent of t.

Lemma 8.2. The conditions (8.6) on (x, v) and the conditions (8.3) on (x, r, m) with c a positive constant, are equivalent under the transformation (8.4) or its inverse (8.5).

That (8.3) implies (8.6) under (8.4) or (8.5) is immediately obvious.

Conversely let (x, v) be a solution of (8.6). Let m and (r) be defined as functions of t by (8.5). Then (x, r, v, r) satisfy (8.4) as functions of t. From (8.5)' and (8.6)' we have $(\dot{x}) = (r)$. Relations (8.4) and (8.3) then imply (8.3).

Equations (8.6) have solutions of the form

(8.8) $$x^i = h^i(t, x_o, v_o, c)$$

$$v_i = k_i(t, x_o, v_o, c)$$

which take on the values (x_o, v_o) when $t = 0$ and for which the functions h^i and k_i are of class C" in their arguments for t on an appropriate interval $[t_1, t_2]$ and (t_o, x_o, v_o, c) sufficiently near a particular initial set of the same character.

We do not desire the general solution (x, v) of (8.6) but only those solutions which, under (8.4), yield solutions of (8.2), that is solutions of (8.3) for which $m_o = 0$. Given (x_o, r_o), for m_o to be zero under (8.4), it is necessary and sufficient that

(8.9) $$V_{io} = F_{r^i}(x_o, r_o) \qquad (i = 1, \ldots, n)$$

for, on multiplying (8.4)' by r_o^i and summing we have

$$m_o r_o^i f_{r^i}(x_o, r_o) = m_o f_o = 0$$

as a consequence of (8.9). The necessity of (8.9) is trivial.

To obtain the general solution of (8.1) one then makes the substitution (8.9) in (8.8). On using the relation $c = f(x_o, r_o)$ of (8.4)", (8.8) yields solutions

(8.10) $$x^i = X^i(t, x_o, r_o)$$

of (8.1), where the functions X^i are of class C" in their arguments for t on $[t_1, t_2]$ and (x_o, r_o) sufficiently near a particular initial set. Moreover

(8.11) $$x_o^i \equiv X^i(0, x_o, r_o)$$

$$r_o^i \equiv X_t^i(0, x_o, r_o).$$

The last relation is a consequence of the given initial relation $R^i(x_o, v_o, c) = r_o^i$ and (8.6)'.

§9. <u>The field of extremals through a point.</u> The solutions X^i ($i = 1, \ldots, n$) of the Euler equations obtained in §8 require interpretation.

In the space of n-coordinates r^i the locus

$$(9.1) \qquad f(x, r) = 1$$

for a fixed (x) is called the <u>indicatrix</u> of f at (x). If (a) is a point on the unit-sphere $a^i a^i = 1$ in the space (r), there is a unique point on the ray $r^i \equiv \rho a^i$ and the indicatrix, namely, a point at the distance

$$\rho = \frac{1}{f(x, a)}$$

from the origin. Thus the indicatrix at (x) is a star-shaped (n-1)-manifold which is compact and contains the origin (r) = (0) in its interior.

In accordance with the homogeneity relation

$$f(x, kr) = k\, f(x, r),$$

f(x, r) can be extended in definition by setting

$$f(x, 0) = 0$$

and so extended is continuous without exception in the space (r). In general however f(x, r) will have no partial derivatives when (r) = (0). With this extension the condition $f(x, r) < e$, where e is a positive constant, defines a region in the space (r) which includes the origin, and is bounded by the (n-1)-manifold obtained by diminishing radial distances to the indicatrix $f(x, r) = 1$ in the ratio of 1 to e.

The condition $f(x, \dot{x}) \equiv c$ implies that

$$(9.2) \qquad \dot{s} \equiv f(x, \dot{x}) \equiv c, \qquad s \equiv t\, f(x_o, r_o),$$

provided one makes s = 0 when t = 0, and lets (x_o, r_o) denote the initial values of $(x, \dot{x})$.

Lemma 9.1. Let $[x(t)]$ be a solution of (8.1) with t on $[0, a]$ and $f(x, \dot{x}) \equiv c$. Set

$$t = kt_1, \qquad x^i(kt_1) = y^i(t_1) \qquad (k > 0)$$

with k a positive constant. Then $y^i(t_1)$ is again a solution of the Euler equations for t_1 on the transformed interval, with $f(y, \dot{y}) = kc$.

This statement is a ready consequence of the relations $t = kt_1$, $dt = kdt_1$, $(\dot{y}) = k(\dot{x})$,

$$f_{r^i}(y, \dot{y}) \equiv f_{r^i}(x, \dot{x}) \qquad f_{x^j}(y, \dot{y}) \equiv kf_{x^j}(x, \dot{x})$$

where $(\dot{y})$ is a t_1-derivative and $(\dot{x})$ a t-derivative.

This leads us to distinguish between an extremal and a solution of the Euler equation as a particular parametrization of an extremal. Two such solutions are regarded as identical if and only if they are identical as functions of t. Thus $[x(t)]$ and $[y(t)]$ given above are distinct solutions of the Euler equations when $k \neq 1$, but are parametrizations of the same curve or extremal.

The basic homogeneity relation (9.4). The solution

$$x^i = X^i(t, x_0, r_0) \qquad (0 \leqq t \leqq a) \qquad (i = 1, \ldots, n) \tag{9.3}$$

of the Euler equations yields a class

$$x^i = X^i(kt, x_0, r_0) \qquad (k > 0, \quad 0 \leqq t \leqq a/k)$$

of such solutions all of which represent the same extremal. Here $\underline{a}$ depends on (x_0, r_0). We now establish the following basic identity.

$$X^i(kt, x_0, r_0) = X^i(t, x_0, kr_0) \qquad (0 \leqq t \leqq a_1). \tag{9.4}$$

For a solution $x^i(t)$ $(i = 1, \ldots, n)$ of (8.1) to be identical with the right member of (9.4) as a function of t it is necessary and sufficient that

$$x^i(0) = x_0^i, \quad \dot{x}^i(0) = kr_0^i \qquad (i = 1, \ldots, n).$$

The left member of (9.4) is a solution of (8.1) with just these properties. Hence (9.4) holds.

The above functions X^i can be continuously extended by a definition of X^i when $(r_0) = (0)$, namely

$$(9.5) \qquad X^i(t, x_0, 0) = x_0^i \qquad (i = 1, \ldots, n)$$

The continuity at $(r_0) = (0)$ of the extended functions X^i follows from (9.4) whose left member is defined and continuous even when $k = 0$, and gives the value x_0^i. It is not assumed that X^i remains of class C" when the point $(r_0) = (0)$ is added to its domain.

<u>The interval for</u> t. Let (u) be a contravariant vector of unit f-length, that is with $f(x_0, u) = 1$. If $(r_0) = (u)$ in (9.3), and if we take $s = 0$ at (x_0), then $s \equiv t$ in (9.3) and

$$(9.6) \qquad x^i = X^i(s, x_0, u) \qquad (i = 1, \ldots, n)$$

gives an intrinsic representation of the extremal considered. There exists an interval

$$(9.7) \qquad 0 \leqq s \leqq e(x_0) \qquad [e(x_0) > 0]$$

on which (9.6) represents an extremal. The interval limit e depends <u>a priori</u> both upon (x_0) and (u). But (u) ranges over a compact set, the indicatrix at (x_0), and it follows that a choice $e(x_0)$ of e can be made independent of (u). This is best seen on recalling the origin of the functions X^i as particular evaluations of the functions h^i of (8.10). The limit t_1 in the interval $(0 \leqq t \leqq t_1)$ for t can be chosen so as to be valid not only for <u>one</u> initial set (x_0, v_0, c) but for a neighborhood N of a particular initial set. A compact ensemble of initial sets (x_o, v_o, c) can be covered by a finite aggregate of neighborhoods N. This applies to the case at hand where (u) ranges over a compact set.

When (u) is not a unit vector, the parameter t in X^i is not s but t where

$$s = tf(x_o, r_o), \tag{9.8}$$

as has been seen in (9.2). Hence for

$$0 \leqq tf(x_o, r_o) < e(x_o) \qquad (r_o) \neq (0), \tag{9.9}$$

X^i is a valid representation of an extremal arc.

The following theorem establishes the existence of a field of extremals issuing from the point (x_o).

Theorem 9.1. *There exists a topological* transformation* T *of a neighborhood* N_x *of* (x_o) *in the n-space* (x) *onto a neighborhood* N_y *of the origin in an Euclidean n-space* (y) *with the following properties.*

(1) *The transformation is of the form*

$$x^i = A^i(x_o, y) \qquad (i = 1, \ldots, n) \tag{9.10}$$

and is valid for

$$f(x_o, y) < \rho(x_o) \tag{9.11}$$

where $\rho(x_o)$ *is a positive constant dependent on* (x_o).

(2) *The functions* $A^i(x_o, y)$ *are of class C' with respect to* (y), *with a non-vanishing Jacobian.*

(3) *An Euclidean ray on* N_y *from the origin to a point* $(y) \neq (0)$ *corresponds to an extremal arc on* N_x *issuing from* (x_o) *with an initial contravariant direction* (y), *and the point* (y) *on* N_y *corresponds to the point on the extremal at which* $s = f(x_o, y)$.

---

* That is, 1-1 and continuous both ways.

If one considers sets (y) for which

(9.12) $$f(x_o, y) < e(x_o),$$

then the interval (9.9) for t, with $(y) = (r_o)$, includes $t = 1$. We can accordingly introduce the transformation

(9.13) $$X^i(1, x_o, y) = A^i(x_o, y) = x^i$$

defining A^i in this way. Observe that A^i is of class C" in (y) for $(y) \neq (0)$. We shall show that A^i is of class C' also at $(y) = (0)$, and that

(9.14)a $$\frac{\partial A^i}{\partial y^j}(x_o, 0) = \delta_{ij} .$$

We set $(y) = t(u)$ in (9.13), where (u) is a contravariant vector of unit f-length, and take $t < e(x_o)$ in agreement with (9.12). Then from (9.13) and (9.4),

(9.14) $$A^i(x_o, tu) = X^i(1, x_o, tu) = X^i(t, x_o, u)$$

For $t < e(x_o)$ relation (9.14) holds, more generally, in the form

(9.15) $$A^i(x_o, tr) = X^i(t, x_o, r)$$

provided (r) is sufficiently near a unit contravariant vector (u). For, the condition $t < e(x_o)$ implies the condition

$$t\, f(x_o, r) < e(x_o)$$

if (r) is sufficiently near a unit vector.

On differentiating (9.15) with respect to r^j we have

(9.16) $$t \frac{\partial A^i}{\partial y^j}(x_o, tr) = \frac{\partial X^i}{\partial r^j}(t, x_o, r) \qquad ((r) \neq (0)).$$

Recall that

$$X^i(0, x_o, r) = x_o^i \,; \quad \frac{\partial X^i}{\partial r^j}(0, x_o, r) = 0$$

The right member of (9.16) thus vanishes when $t = 0$. On using Taylor's theorem with an integral remainder to represent the right member of (9.16) and evaluating for $r = u$ we have

$$(9.17) \qquad \frac{\partial A^i}{\partial y^j}(x_o, tu) = \int_0^i \frac{\partial^2 X^i}{\partial t \partial r^j}(at, x_o, u)da$$

For (u) on the indicatrix and for $t < e(x_o)$, the right member of (9.17) is continuous in t and (u), not excepting $t = 0$. In particular for $t = 0$ the integrand has a value δ_{ij} independent of (u), and this is also the value of the integral.

In terms of the variables (y) with $(y) = t(u)$

$$t = f(x_o, y) \qquad u^i = \frac{y^i}{f(x_o y)} \qquad [(y) \neq (0)]$$

and the limit of the integrand in (9.17) is again δ_{ij} as (y) tends to (0). As a function of (y) the integral in (9.17) becomes continuous in (y) provided δ_{ij} is taken as the value of the integrand when $(y) = (0)$. It follows that the partial derivative of A^i with respect to y^j exists*, equals δ_{ij} when $(y) = (0)$, and is continuous on a neighborhood of $(y) = (0)$.

In accordance with (9.14)a the Jacobian

$$\frac{D(A^1, \ldots, A^n)}{D(y^1, \ldots, y^n)} = 1 \qquad [(y) = (0)]$$

By virtue of the implicit function theorem the transformation defined by (9.13) is topological and satisfies (1) and (2) in the theorem, provided the neighborhood of $(y) = (0)$ defined by (9.11) is sufficiently small. Property (3) of the

---

* In proving this one makes use of the following fact. When a function $H(y)$ of a single variable variable y has a continuous derivative on an open interval $[0, b]$ tending to a limit as y tends to 0 then $H'(0)$ exists and equals that limit.

theorem results from (9.4) on interpreting $(y) = t(u)$ for variable t as a ray on N_y, and interpreting $A^i(x_o, tu)$, $i = 1, \ldots, n$, as an extremal arc on N_x, with an initial contravariant direction (y) where $(y) \neq (0)$, and with

$$s = f(x_o, tu) = f(x_o, y).$$

In Theorem 9.1 we have the limitation

$$f(x_o, y) \leqq \rho\,(x_o), \qquad (0 \leqq s \leqq \rho(x_o))$$

on the "normal" coordinates (y) and their image points on the extremal arcs through (x_o). The dependence of $\rho(x_o)$ on (x_o) can be removed if we assume (as we do) that our coordinate manifold is compact, in that every aggregate of open sets covering M contains a finite subaggregate covering M.

In fact, the functions $A^i(x_o, y)$ and their y^j-derivatives are seen to be continuous on a product domain (x_o, y) defined by conditions of the form

$$f(x_o, y) \leqq e_1, \quad \left| x_o^i - a^i \right| \leqq e_2, \qquad (i = 1, \ldots, n)$$

where (a) is a point on the coordinate region R_x while e_1 and e_2 are positive constants. A choice ρ of $\rho(x_o)$ such that Theorem 9.1 holds can then be made independent of the value of (x_o), on a neighborhood N of (a) defined by the conditions

$$\left| x_o^i - a^i \right| < e_2\,, \qquad (i = 1, \ldots, n).$$

Since M can be covered by a finite aggregate of such neighborhoods N, a choice of ρ can be made independent of the point P (locally (x_o)) on M. We thus have

Theorem 9.2. *There exists a positive constant* ρ *such that the extremal arcs issuing from an arbitrary point* P *on* M, *with* $0 \leqq s \leqq \rho$, *cover the closure of a neighborhood of* P *in a 1 - 1 manner*, P *excepted, and in some coordinate region* R_x, *with* P *given by* (x_o), *admit a representation in terms*

<u>of normal coordinates</u> (y) <u>with</u>

$$f(x_o, y) \leqq \rho$$

<u>as stated in Theorem 9.1.</u>

§10. <u>The combining of overlapping coordinate systems.</u> Our analysis of an extremal arc will be simpler if we know that the arc lies in a single coordinate system. An extremal arc originating in a coordinate system (x) may be continued through any finite sequence of overlapping coordinate systems making use of the covariant character of the Euler operator and of the invariant character of the function f(x, r) used to define f-length s. According to Theorem 10.1 below, an extremal arc g always lies in a suitably chosen single coordinate system.

Recall that a coordinate system (x) is defined by a topological mapping of a coordinate region R_x in the space (x) onto an open subset of our coordinated manifold M. Points in two different coordinate regions R_x and R_y with the same image point on M are termed <u>equivalent</u>. The relation between equivalent points (x) and (y) is (by definition) a one-to-one <u>coordinate transformation</u> of a given class, here class C''', with a non-vanishing jacobian. The equivalence relation is transitive; that is, if (x) is equivalent to (y), and (y) equivalent to (z), then (x) is equivalent to (z).

Let R_x and R_y be two coordinate regions in which a subregion R_x^y of R_x is equivalent to a subregion R_y^x of R_y. Let MR_x be the image on M of a coordinate region R_x. We seek a coordinate region R_z <u>combining</u> R_x and R_y in the sense that

$$(10.1) \qquad MR_z = MR_x + MR_y \qquad [\ + = \text{"union"}]$$

Necessary and sufficient conditions that a coordinate region R_z exist such that (10.1) holds are as follows.

(a) *The set* R_z *admits a 1-1 mapping* T *onto the union of* R_x *and* R_y, *regarding points of* R_x *and* R_y *as identical if and only if they have the same images on* M.

(b) *The mappings of* R_x *and of* R_y *into* R_z *under* T *are representable as coordinate transformations (of Class* C''' *with non-vanishing jacobian and 1-1.)*

If (a) and (b) hold, a point (z) of R_z is made to correspond to the same point of M as does its image (x) or (y) under T.

With the aid of the above conditions for the existence of R_z we can establish the following lemma.

Lemma 10.1. Let R_x *and* R_y *be coordinate regions in which* R_x^y *in* R_x *is equivalent to* R_y^x *under a coordinate transformation*§ Y. *Necessary and sufficient conditions that* R_y *be a subset of some coordinate region* R_y^* *such that*

$$MR_y^* = MR_x + MR_y \tag{10.2}$$

are that Y *be extensible*§§ *from a transformation of* R_x^y *to a coordinate transformation* Y^* *of* R_x *in such a manner that*

$$Y^*(R_x - R_x^y) \cdot (R_y - R_y^x) = 0 \tag{10.4}$$

One can then take R_y^* *as*

$$R_y^* = R_y + Y^*R_x \tag{10.5}$$

and require that (x) *in* R_x *be equivalent to* $Y^*(x)$. *Relation (10.2) then holds.*

---

§ Of the form

$$y^i = Y^i(x) \qquad (i = 1, \ldots, n;\ (x) \text{ in } R_x^y). \tag{10.3}$$

§§ That is, capable of definition over a domain which includes the original domain of definition R_x^y without altering the transformation over R_x^y as originally defined.

The conditions are clearly necessary. To prove them sufficient let $R_x + R_y$ denote the union of R_x and R_y, subject to the equivalence relation Y. The region R_y is mapped onto a subset of R_y^* by the identity, and R_x into R_y^* by Y^*. These mappings combine into a <u>single-valued</u> mapping T of $R_x + R_y$ into R_y^* if and only if the coordinate transformation Y^* of R_x is an extension of Y. The mapping T of $R_x + R_y$ into R_y^* is <u>one-to-one</u> if and only if (10.4) holds. When Y^* is both an extension of Y and satisfies (10.4), the transformation T satisfies (a) and (b) and the lemma holds.

<u>A coordinate system containing a given extremal arc</u> E. We can suppose E covered by a finite set of coordinate regions R_y in each of which E is represented by functions

(10.6) $$y^i(s) \qquad (i = 1, \ldots, n)$$

of class C''' of the s-length measured from the initial point of E. We shall now prove the following theorem.

<u>Theorem 10.1.</u> <u>Let</u> g <u>be a simple regular arc, of class C''' in terms of</u> f-<u>length</u> s. <u>There exists a coordinate region</u> R_x <u>in which</u> g <u>is represented by the</u> x^n<u>-axis with</u> $s \equiv x^n$ <u>on this axis.</u>

We shall first show that the theorem is true for any subarc g_1 of g on which $c - e < s < c + e$, provided e is a sufficiently small positive constant.

Let P be the point of g for which $s = c$. Let $z^i = a^i(s)$, $i = 1, \ldots, n$, be a representation of g neighboring P in a coordinate region R_z. For simplicity suppose that $c = 0$. For at least one value of i, say n, $\dot{a}^i(0) \neq 0$. Suppose that $(z) = (0)$ at P. With this understood we make the transformation of coordinates

$$z^i = x^i + a^i(x^n) \qquad (i = 1, \ldots, n-1)$$
(10.7)
$$z^n = a^n(x^n)$$

from a neighborhood of the point $(x) = (0)$ onto a neighborhood of the point $(z) = (0)$. This transformation has a jacobian equal to $\dot{a}^n(0)$ at $(x) = (0)$, and so defines a locally admissible change of coordinates. The image in the space (x) of the arc $z^i = a^i(s)$ consists of points (x) which satisfy the conditions

$$a^i(s) = x^i + a^i(x^n) \qquad (i = 1, \ldots, n-1)$$
(10.8)
$$a^n(s) = a^n(x^n).$$

Since the n-th condition in (10.8) has no local solution other than $x^n = s$, the local solution of (10.8) has the form

$$x^1 = \ldots = x^{n-1} = 0, \quad x^n = s .$$

The theorem is accordingly true for the subarc g_1, if e is sufficiently small.

By virtue of the preceding result g can be covered by a finite set of overlapping arcs for each of which the theorem is true. The following lemma then implies the truth of the theorem in general.

Lemma 10.2. *Let g_x and g_y be two subarcs of g with common inner point P: $s = s_o$; with g_x and g_y admissibly represented by the x^n-axis and y^n-axis respectively in coordinate regions S_x and S_y; with $x^n \equiv s$ on g_x and $y^n \equiv s$ on g_y. The theorem is then true for the arc $g_x + g_y$.*

For simplicity suppose that $s_o = 0$, that $(x) = (y) = (0)$ at P, that in the sense of increasing s the initial point of g_y precedes that of g_x, and that the final point of g_x follows that of g_y. Equivalence between points on S_x and S_y neighboring P is given by a transformation of the form

(10.9) $$y^i = a^i_j x^j + m^i(x) \qquad (i, j = 1, \ldots, n)$$

where a^i_j is a constant $\neq 0$, and $m^i(x)$ is a function of class C''' with a null differential at $(x) = (0)$. Since $s = x^n = y^n$ along g near $s = 0$, we see that $a^n_n = 1$, and $m^n(x) \equiv (0)$ on the x^n-axis. Without loss of generality we can suppose that (10.9) has the form

$$y^i = x^i + m^i(x) \qquad (i = 1, \ldots, n) \tag{10.10}$$

since a preliminary transformation of the form

$$\bar{x}^i = a^i_j\, x^i$$

would bring this about. We note the existence of a positive constant e so small that for $|x^i| < 2e$,

$$\left| \frac{\partial m^i}{\partial x^j} \right| < 1 \qquad (i, j = 1, \ldots, n)$$

To continue we shall need a function h(s) which is of class C''' for all real values of s, with

$$h(s) \equiv 1 \quad (s \leqq e), \qquad (|h| \leqq 1)$$
$$h(s) \equiv 0 \quad (s \geqq 2e),$$

Such a function clearly exists.

We replace (10.10) by the transformation

$$y^i = x^i + h(x^n)m^i(x) \qquad (i = 1, \ldots, n) \tag{10.11}$$

for (x) on a region R_x to be determined. To that end observe that on the x^n-axis for $x^n \geqq -e$, one has $y^n = x^n$ and $m^i(x) \equiv 0$ with

$$\frac{\partial y^i}{\partial x^j} = \delta_{ij} + h \frac{\partial m^i}{\partial x^j} \qquad (j = 1,\ldots,n-1;\ i = 1,\ldots,n)$$

$$\frac{\partial y^i}{\partial x^n} = \delta_{ni} \qquad (i = 1, \ldots, n)$$

The jacobian J of the transformation (10.11) is accordingly not zero on the x^n-axis for $x^n \geqq -e$. Let R_x consist of those points of S_x for which

$$(10.12) \qquad x^n > -e, \quad |x^i| < d < e, \qquad (i = 1, \ldots, n-1)$$

We can suppose d so small that (10.11) is an admissible coordinate transformation of R_x (of class C''' with $J \neq 0$, and one-to-one).

We now prepare for the application of Lemma 10.1.

The subregion R_x^y of R_x for which $-e < x^n < e$ consists of points (x) equivalent to points (y) under (10.10) or (10.11). Let the transformation (10.10) taken over R_x^y be denoted by Y. Let R_y be the union of the set

$$R_y^x = Y R_x^y$$

and of the subset of S_y for which

$$(10.13) \qquad y^n < 0, \quad |y^i| < d \qquad (i = 1, \ldots, n-1)$$

Let Y^* denote the transformation (10.11) taken over R_x. Then Y^* is an extension of Y. Under Y^* the x^n-axis in R_x is invariant; it follows from the continuity of Y^* that, if the constant d limiting R_x and R_y in (10.12) and (10.13) respectively is sufficiently small, then

$$Y^*(R_x - R_x^y) \cdot (R_y - R_y^x) = 0$$

We conclude from Lemma 9.1 that the set

$$R_y^* = R_y + Y^* R_x$$

is a coordinate region. On this region the arc $g_x + g_y$ is represented by the y^n-axis as required.

The lemma follows, and the proof of the theorem is immediate.

§11. <u>Conjugate points</u>. Let g be an extremal arc. As has been seen in §10, we can suppose that g is interior to a coordinate region R_x.

The extremal g is said to afford a <u>proper</u> relative minimum to the integral J in the fixed end point problems if

$$J(g) < J(g') \qquad (g \neq g') \tag{11.1}$$

for any admissible curve g' joining g's end-points in some neighborhood N of g. As previously, we admit piece-wise regular curves g' of class D'. There are two principal conditions together sufficient that g afford a proper relative minimum to J, namely the condition that there be no point on g <u>conjugate</u> to its initial point, and that F be <u>positive regular</u>. These conditions will be defined.

In order to define a conjugate point we need a local parametric representation of the indicatrix $f(x, r) = 1$. Observe that the relation

$$r^i f_{r^i} = f = 1$$

on the indicatrix implies that at least one of the partial derivative f_{r^i} fails to vanish at a point (u) on the indicatrix. Hence the points on the locus $f(x, r) = 1$ neighboring (u) admit a representation of the form

$$r^i = r^i(v), \qquad r^i(0) = u^i \qquad (i = 1, \ldots, n) \tag{11.2}$$

where (v) is a set of n-1 parameters in terms of which the functions $r^i(v)$ are of class C''' with a matrix of partial derivatives of rank n-1 at the set (v) = (0) corresponding to (u). We term this representation <u>regular of class C'''</u>.

Let $x^i = g^i(s)$ be a representation of g in terms of f-length s. Let $(u) = [\dot{g}(s_0)]$ be the unit contravariant vector tangent to g at a point $s = s_0$.

Let (11.2) be a regular representation of unit vectors neighboring (u), tangent to extremals issuing from $[g(s_0)] = (x_0)$. In terms of the solutions $X^i(t, x_0, r_0)$ of the Euler equations obtained in §8 the extremals issuing from the point (x_0) on g with directions neighboring (u) can be represented in the form

$$x^i = X^i(s-s_0, g(s_0), r(v)) = H^i(s, s_0, v) \tag{11.3}$$

Consider the jacobian

$$M(s, s_0) = \frac{D(H^1, \ldots, H^n)}{D(s, v_1, \ldots, v_{n-1})} \qquad [(v) = (0)] \tag{11.3'}$$

Observe that for $(i = 1,\ldots,n)$ $(j = 1,\ldots,n-1)$

$$H^i(s_0, s_0, v) = x_0^i, \quad \dot{H}^i(s_0, s_0, v) = r^i(v) \tag{11.4'}$$

$$H^i_{v^j}(s_0, s_0, v) = 0, \quad \dot{H}^i_{v^j}(s_0, s_0, 0) = r^i_{v^j}(0) \tag{11.4''}$$

In accordance with (11.4) the first column of $M(s_0, s_0)$ is $[r(0)] = [u]$, while the last n-1 columns consist of null elements. Using the integral form of the law of the mean we can factor $s-s_0$ out of each of the elements of $M(s, s_0)$ in the last n-1 columns, and write

$$M(s, s_0) = (s - s_0)^{n-1} N(s, s_0), \tag{11.5}$$

where $N(s, s_0)$ is continuous in s and s_0 on the interval [0, a] of s on g. We shall prove the following

$$N(s_0, s_0) = \left| u^i, r^i_{v^j}(0) \right| \neq 0 \qquad (i = 1,\ldots,n;\ j = 1,\ldots,n-1) \tag{11.6}$$

To establish (11.6) one starts with the identity in (v), $f(x, r(v)) \equiv 1$, valid for (v) sufficiently near (v) = 0. On differentiating with respect to v^j and recalling the homogeneity relation we have for each i

$$r^i_{v^j}\, f_{r^i}(x, u) = 0 \qquad (j = 1, \ldots, n-1) \tag{11.7)'$$

$$u^i f_{r^i}(x, u) = 1 \tag{11.7)''$$

Regarding the equations (11.7) as linear conditions on the n variables f_{r^i} one sees that the "augmented" matrix of the system has the rank n. Since the equations (11.7) are consistent, the determinant (11.6) of the system must also have the rank n, so that (11.6) holds.

The conjugate points of the point $s = s_0$ on g are defined as the points $s \neq s_0$ at which $M(s, s_0) = 0$.

A first conclusion to be drawn from the nature of $N(s, s_0)$ in (11.5) is as follows.

(a) If there is no point on g conjugate to its initial point $s = 0$, then on an extension g^* of g as an extremal there is no point on g conjugate to a point on g^* preceding g for which $|s_0|$ is sufficiently small.

Since there is no point on g conjugate to $s = 0$ it follows that for s on $[0, a]$ (the interval for s on g) $N(s, 0) \neq 0$. But $N(s, s_0)$ is clearly continuous in (s, s_0) for s on $[0, a]$ and $|s_0|$ sufficiently small, so that for s on $[0, a]$ and $|s_0|$ sufficiently small $N(s, s_0) \neq 0$. Statement (a) follows from (11.5).

A second conclusion based on our assumption that the coordinate manifold is compact is as follows.

(b) There exists a constant $e > 0$ independent of the point P on M such that on any extremal g issuing from P with $s = 0$ at P there is no conjugate point of P on g for which $|s| < e$.

To establish (b) we rewrite (11.3) in the form

$$x^i = X^i[s,\ x_o,\ r(v)]$$

and the jacobian (11.3)' in the form

(11.8) $$s^{n-1}Q(s,\ x_o,\ u)$$

proving as previously that

$$Q(0,\ x_o,\ u) \neq 0.$$

We observe that Q is continuous in its arguments $(s,\ x_o,\ u)$ for $|s| < e_1$ and $(x_o,\ u)$ on a neighborhood N of a particular set $(x_o^*,\ u^*)$. Here e_1 is a positive constant dependent on N. The product domain, with factors M and the indicatrix associated with each point of M, is compact and can be covered with a finite aggregate of neighborhoods N. Hence e_1 can be replaced by a positive constant e independent of P and of the initial direction of g at P. Statement (b) follows.

§12. The condition of positive regularity and the Hilbert integral.

Upon setting

$$F_{r^i r^j} = F_{ij}(x,\ r) \qquad (i,\ j = 1,\ldots,n)$$

recall that the determinant $|F_{ij}| = 0$. Hence the quadratic form in (z)

(12.1) $$F_{ij}(x,\ r)z^i z^j \qquad [(r) \neq (0)]$$

cannot be positive definite. More explicitly, it vanishes for (z) proportional to (r), since

$$r^i\, F_{ij}(x,\ r) = 0$$

(a) The condition of positive regularity is that when (z) is not linearly dependent on (r) the form (12.1) is positive.

A first consequence of positive regularity is that the Weierstrass function $F_1(x, r)$ is not zero.

In fact, when F is positive regular, n-1 characteristic roots of the matrix $\|F_{ij}\|$ must be positive and one root null. Hence the rank of $\|F_{ij}\|$ must be n-1; we conclude that some cofactor $A_{ij} \neq 0$. It follows from the relation

$$A_{ij} = F_1 r^i r^j$$

that $F_1 \neq 0$.

If (z) is a contravariant vector, the form (12.1) is invariant so that the condition of positive regularity is independent of the coordinates used.

<u>The Weierstrass function</u> $E(x, r, q)$. This function has the definition

$$E(x, r, q) = F(x, q) - q^i F_{r^i}(x, r) \qquad [(r) \neq (0)] \tag{12.2}$$

where (r) and (q) are contravariant vectors. Using the relation

$$F(x, r) = r^i F_{r^i}(x, r) \qquad (i = 1, \ldots, n)$$

one sees that

$$E(x, r, q) = F(x, q) - F(x, r) - (q^i - r^i)F_{r^i}(x, r) \tag{12.3}$$

The terms following $F(x, q)$ in (12.3) are those of a Taylor's development of $F(x, q)$ about (r) for a fixed (x).

Since F_{r^i} is a covariant vector we see from (12.2) that E is an invariant. Observe that E is positive homogeneous of the first order in (q) and of the zero-th order in (r), and that

$$E(x, r, kr) = 0 \qquad (k \geqq 0) \tag{12.4}$$

where k is a constant.

The Weierstrass condition in its strong form requires that

$$E(x, u, v) > 0 \tag{12.5}$$

for (u) and (v) arbitrary distinct unit vectors.

We shall need the formula

$$E(x, r, -r) = E(x, q, -r) + E(x, q, r), \tag{12.6}$$

valid for any two non-null vectors (r) and (q). This formula is verified at once on using the definition of E.

Lemma 12.1. The condition of positive regularity implies the strong Weierstrass condition.

We shall regard E(x, u, q) as a function of (q) and expand E about (u) by means of Taylor's formula with a remainder of the second order. Suppose first (Case I) that (u) and (v) are linearly independent. Then

$$E(x, u, v) = \tfrac{1}{2}(v^i - u^i)(v^j - u^j)F_{ij}(x, w) \tag{12.7}$$

where

$$w^i = u^i + \theta(v^i - u^i) \qquad (0 < \theta < 1)$$

One observes that $(w) \neq (0)$ since (w) represents a point in euclidean n-space on the line L joining the point (u) to the point (v) and L does not intersect the origin. Thus $F_{ij}(x,w)$ is well defined.

The right member of (12.7) is positive in accordance with the condition of positive regularity unless (v)-(u) is linearly dependent on (w). As a vector in euclidean n-space (v)-(u) has the direction of L while (w) has the direction of a line joining the origin to a point of L; these directions are different since L does not intersect the origin. Hence (v)-(u) is independent of (w) and E(x, u, v) > 0 in Case I.

There remains Case II where (u) and (v) are non-null but with opposite euclidean directions. In this case it will be sufficient to show that $E(x, u, -u) > 0$. This is a consequence of (12.6) on taking $(r) = (u)$ and (q) as any non-null vector not linearly dependent on (u), using the fact already established that

$$E(x, q, u) > 0, \quad E(x, q, -u) > 0. \tag{12.8}$$

It follows that $E(x, u, v) > 0$ for (u) and (v) distinct vectors of f-length 1, and the proof of the theorem is complete.

The case in which (u) and (v) have opposite directions has been erroneously treated in the literature, and the errors laboriously corrected by means of Behaghel's formula. The above treatment by means of (12.6) was first given by the author.

The Hilbert integral. Consider a (n-1)-parameter family of extremals of the form

$$x^i = x^i(x, v) \qquad (v) = (v^1, \ldots, v^{n-1}) \tag{12.9}$$

in which the functions $x^i(s, v)$ are of class C" in the variables (s, v) for s on some interval $[s_1, s_2]$ and (v) on a region R_v of euclidean (n-1)-space, with a jacobian

$$\frac{D(x^1, \ldots, x^n\ )}{D(s, v^1, \ldots, v^{n-1})} \neq 0$$

It will be convenient to set

$$V_i(s, v) = F_{r^i}[x(s, v), \dot{x}(s, v)] \qquad (i = 1, \ldots, n)$$

We shall establish an identity in s and (v)

$$\frac{\partial}{\partial v^h}\left(V_i \frac{\partial x^i}{\partial v^k}\right) - \frac{\partial}{\partial v^k}\left(V_i \frac{\partial x^i}{\partial v^h}\right) \equiv C_{hk} \tag{12.10}$$

analogous to Abel's integral for any two solutions of a linear second order differential equation. Here $i = 1, \ldots, n$; $h, k = 1, \ldots, n-1$; and C_{hk} is independent of s, but may depend upon (v). There is thus one relation (12.10) for each pair (h, k).

To establish (12.10), we evaluate the integral J along an extremal $(v) = \text{const.}$, from $s = s_1$, to an arbitrary s on $[s_1, s_2]$, and thereby obtain a function $J(s, v)$. Upon differentiating J under the integral sign and integrating the terms involving $F_{r^i} = V_i$ by parts in the usual way we find that

$$\frac{\partial J}{\partial v^h} = \left[V_i \frac{\partial x^i}{\partial v^h}\right]_{s_1}^{s}$$

The condition that the second partial derivative of J with respect to v^h and v^k equals that with respect to v^k and v^h, takes the form

$$\frac{\partial}{\partial v^k}\left[V_i \frac{\partial x^i}{\partial v^h}\right]_{s_1}^{s} = \frac{\partial}{\partial v^h}\left[V_i \frac{\partial x^i}{\partial v^k}\right]_{s_1}^{s} . \tag{12.11}$$

On transferring the terms in (12.11) with upper limit s to the left number, and those with lower limit s_1 to the right, the two members appear equal to a function C_{hk} with the same value at s as at s_1, and accordingly independent of s. Relations (12.10) follow.

To define the Hilbert integral we suppose that the region R_v is simply-connected, and that the family (12.9) is a field of extremals covering a domain S_x in the space (x) in a 1-1 manner. That is, we suppose that (12.9) maps the product of the interval $[s_1, s_2]$ and R_v topologically onto S_x. The point (x) on S_x has an inverse image

$$s = s(x) \qquad v^h = v^h(x) \qquad (h = 1, \ldots, n-1)$$

in which $s(x)$ and $v^h(x)$ are of class C". One sets

$$\dot{x}^i[s(x), v(x)] = u^i(x) \qquad (i = 1, \ldots, n)$$

The vector $(u(x))$ is thus the direction at (x) of the extremal of the field through (x), and is termed <u>the direction of the field</u> at (x). The Hilbert integral is a line integral of the form

$$I = \int P_i(x)dx^i \qquad \text{where } [P_i(x) = F_{r^i}(x, u(x))] \tag{12.12}$$

In terms of the parameters (s, v) the Hilbert integral takes the form

$$I = \int C(s, v)ds + D_h(s, v)dv^h \qquad (h = 1, \ldots, n-1) \tag{12.12)'$$

where

$$C(s, v) = F_{r^i}[x(s, v), \dot{x}(s, v)]\, \dot{x}^i(s, v) = F[x(s, v), \dot{x}(s, v)]$$

$$D_h(s, v) = F_{r^i}[x(s, v), \dot{x}(s, v)]\, \frac{\partial x^i}{\partial v^h}(s, v) \qquad (h = 1,\ldots,n-1)$$

Since R_v is simply connected, necessary and sufficient conditions that I be independent of the path in the space (s, v) are that

$$\frac{\partial C}{\partial v^h} - \frac{\partial D_h}{\partial s} = 0 \qquad (h = 1,\ldots,n-1) \tag{12.13)'$$

$$\frac{\partial D_h}{\partial v^k} - \frac{\partial D_k}{\partial v^h} = 0 \qquad (h,k = 1,\ldots,n-1) \tag{12.13)"$$

Upon using the function $V_i(s, v)$ these conditions take the form

$$\frac{\partial F}{\partial v^h} - \frac{\partial}{\partial s}\left(V_i \frac{\partial x^i}{\partial v^h}\right) = 0 \tag{12.14)'$$

$$(12.14)'' \qquad \frac{\partial}{\partial v^h}\left(V_i \frac{\partial x^i}{\partial v^k}\right) - \frac{\partial}{\partial v^k}\left(V_i \frac{\partial x^i}{\partial v^h}\right) = 0$$

Condition (12.14)' takes the form

$$F_{x^i}\frac{\partial x^i}{\partial v^h} + F_{r^i}\frac{\partial^2 x^i}{\partial s \partial v^h} = \frac{\partial x^i}{\partial v^h}\frac{\partial V_i}{\partial s} + V_i \frac{\partial^2 x^i}{\partial s \partial v^h}$$

and reduces the identity since, with (s, v) the independent variables,

$$F_{r^i} = V_i ; \qquad F_{x^i} = \frac{\partial V_i}{\partial s} \text{ (the Euler equations)}$$

Conditions (12.14)" are absent if $n = 2$. For $n > 2$ they are not satisfied in general. In the important case in which the field of extremals issues from a point $s_o < s_1$ on g^*, the constant $C_{hk} = 0$ in (12.10); for the partial derivatives of $x^i(\dot{s}, v)$ as to v^h vanish for $s = s_o$, since $x^i(s_o, v) = x^i_o$. Hence (12.14)" is satisfied. We thus have the theorem

<u>Theorem 12.1.</u> <u>The Hilbert integral, defined for a simply connected field of extremals issuing from a point, is independent of the path of integration in the field.</u>

From the representation (12.12)' of the Hilbert integral and the fact that $C \equiv F$ in this integral, we see that along an extremal g of the field $I = J$. For along such an extremal $dv^h = 0$, $h = 1, \ldots, n-1$.

Corresponding to any point P on M there exists, in accordance with Theorem 9.1, a positive constant R(P) so small that the following is true. The family of extremals issuing from P with $0 \leqq s \leqq R(P)$ covers a neighborhood of P in a 1-1 manner (P excepted), and has a representation of the form

$$(12.15) \qquad x^i = A^i(x_o, y), \qquad [0 \leqq f(x_o, y) \leqq R(P)]$$

in a suitably chosen coordinate space (x) in which (x_o) represents P, where the A^i are of class C' in the variables (y) and have a non-vanishing jacobian. It is clear that there are infinitely many choices of R(P).

We term R(P) a field radius at P.

Lemma 12.2. There is no conjugate point of P at which $0 < s \leqq R(P)$ on the extremals issuing from P.

To establish the lemma one must return to the definition of a conjugate point in §11. Let (u) be any vector of unit f-length at P. Let $r(v)$ be a regular representation of class C" of vectors (r) on the f-indicatrix, $f(x_o, r) = 1$, neighboring (u) with $[r(0)] = (u)$. By definition of A^i

$$A^i(x_o, sr(v)) = X^i(x, s_o, r(v)) \qquad [0 \leqq s \leqq R(P)]$$

The jacobian $M(s, s_o)$ which by definition determines the conjugate points of P is

$$\frac{D(X)}{D(s, v)} = \frac{D(A)}{D(s, v)} = \frac{D(A)}{D(y)} \frac{D(y)}{D(s, v)} , \tag{12.16}$$

where we have set $y^i = sr^i(v)$, $i = 1,\ldots,n$. Here $D(A)/D(y) \neq 0$ as we have stated in connection with (12.15), while for $(v) = (0)$ the jacobian

$$\frac{D(y)}{D(s, u)} = s^{n-1} \left| u^i, \underset{v}{r}{}^i_h(0) \right| , \qquad (i = 1,\ldots,n;\ h = 1,\ldots,n-1)$$

is not zero for $s > 0$, as noted in (11.6). The lemma follows from (12.16).

The reader will have noted that in the original definition of a conjugate point of a point s_o on an extremal g the zeros of $M(s, s_o)$ in (11.3)' are independent of the choice of the regular representation $[r(v)]$ of class C" of the indicatrix $f(x_o, r) = 1$ neighboring the initial direction (u). If another regular representation of the indicatrix in the neighborhood of (u)

is used in terms of parameters $(\bar{v})$, the parameters (v) will stand in a relation $v^i = m^i(\bar{v})$ $(i = 1,\ldots,n)$ to the parameters $(\bar{v})$, where the functions m^i are of class C'' and have a non-vanishing jacobian at $(\bar{v}) = (0)$. The jacobian $M(s, s_o)$ will be replaced by

$$M(s, s_o)\,\frac{D(v)}{D(\bar{v})}$$

which vanishes at the same points s as does $M(s, s_o)$.

The preceding relations between conjugate points and fields would be simpler if a conjugate point of a point P on an extremal g could be defined as a point $Q \neq P$ at which the family of extremals issuing from P with directions arbitrarily near that of g at P fail to cover some neighborhood of P. This definition is valid in the analytic case as Morse and Littauer have shown (Proc. Natl. Acad. Scs. 1932) but no proof has been published in the non-analytic case. In the plane, the definition is always valid. If Q is not a conjugate point, a neighborhood of Q is covered in a 1-1 manner, but if Q is a conjugate point is it always true that there is no 1-1 covering? The above definition is purely topological and would afford an excellent topological basis for conjugate point theory.

§13. Sufficient conditions for a relative minimum of J.

A "field" of extremals on which the Hilbert integral I is independent of the path is called a Mayer field.

Theorem 13.1. *Let S_x be a region covered by a Mayer field of extremals, with the unit*§ *vector $u^i(x)$ $(i = 1,\ldots,n)$ tangent to the extremal through (x). Suppose that*

§ To avoid ambiguity one might call this vector f-unitary since $f(x, u) = 1$.

(13.1) $$E(x, u(x), q) > 0$$

for (x) on S_x and $(q) \neq k[u(x)]$, $k > 0$. Then any extremal subarc g of the field will afford a proper minimum to J relative to admissible curves on S_x joining g's end points.

Let a be an admissible curve joining the end points of g on S_x, of the form

(13.2) $$x^i = a^i(s) \qquad (0 \leqq s \leqq b;\ i = 1,\dots,n)$$

As we have seen in §12, $J_g = I_g$; and $I_g = I_a$ since we have a Mayer field. Hence

$$J_a - J_g = J_a - I_a$$

But

$$I = \int F_{r^i}(x, u(x))dx^i .$$

Hence for $x^i = a^i(s)$ $(i = 1, \dots, n)$

$$J_a - J_g = \int_0^b F(a, \dot{a})ds - \int_0^b F_{r^i}(a, u(a))\dot{a}^i ds,$$

(13.3) $$= \int_0^b E(a, u(a), \dot{a})ds \geqq 0$$

In accordance with hypothesis (13.1) the inequality in (13.3) holds only if the arc $x^i = a^i(s)$ satisfies the differential equations

(13.4) $$\frac{dx^i}{ds} = u^i(x) \qquad (i = 1,\dots,n)$$

These equations (13.4) are satisfied by the extremals of the field and by no other admissible arcs, so that $[a(s)]$ satisfies (13.4) only if the arc $a = g$. Hence the equality in (13.3) is excluded when $g \neq a$. The proof of the theorem is complete.

The condition of positive regularity at a point (x) implies that for this (x),

$$E(x, r, q) > 0$$

for any non-null (r) and (q), with $(q) \neq k(r)$ when $k > 0$.

The <u>strong conjugate point condition</u> on an extremal arc g is that there be no point on g conjugate to its initial point $s = 0$. In accordance with (a) of §11, a point P_o on the extremal extension g^* of g for which $s_o < 0$ and for which $|s_o|$ is sufficiently small will have no conjugate point on g. As we have seen in Theorem 12.1, a family of extremals issuing from P with directions near that of g^* at P_o will form a Mayer field covering a neighborhood of g. Hence we have the following corollary of Theorem 13.1.

<u>Corollary 13.1.</u> <u>A sufficient condition that an extremal arc</u> g <u>afford a proper minimum to</u> J <u>relative to admissible arcs on a sufficiently small neighborhood of</u> g <u>is that</u> g <u>satisfy the strong conjugate point condition, and that</u> F <u>be positive regular at each point</u> (x) <u>on some neighborhood of</u> g.

It is possible to weaken the conditions of the corollary still further by requiring merely that F be positive regular at each point (x) on g. It can then be shown (see Morse, <u>op. cit.</u>, 122-123) that F is positive regular at each point on some neighborhood of g.

<u>Transversality.</u> In order to establish the existence of short arcs furnishing an absolute minimum to J we shall suppose that $F(x, r) > 0$ for $(r) \neq 0$. We say then that F is <u>positive definite</u>. We shall also suppose that

$$f = F, \quad s = J = \int_0^t F(x\ \dot{x})dt.$$

Transversality is a generalization of orthogonality. A contravariant vector $(r) \neq (0)$ is said to be cut transversally by a contravariant vector (dx) at a point (x) if

$$F_{r^i}(x, r)dx^i = 0 \qquad (i = 1, \ldots, n) \tag{13.5}$$

In the special case in which

$$F(x, r) = (r^i r^i)^{\frac{1}{2}} \tag{13.6}$$

the transversality condition (13.5) reduces to the orthogonality condition

$$r^i dx^i = 0.$$

Unlike the orthogonality relation, which is symmetric, the transversality relation is not generally symmetric. When (13.5) holds, the relation

$$F_{r^i}(x, dx)\, r^i = 0$$

does not always hold.

This may be seen on supposing that (r) and (dx) are F-unitary vectors. F_{r^i} is then the euclidean direction of the normal to the indicatrix K of F at (x). We have seen that K is star-shaped, and apart from conditions of differentiality on F it is arbitrarily star-shaped. With this understood, one sees that when (dx) is a F-unitary direction tangent to K at a point (r) on K, then (r) will not in general have the direction of a tangent to K at the point (dx) on K. Thus the transversality relation is in general not symmetric.

We are supposing F positive definite and that $F = f$. Let $R(P)$ be a "field radius". On the extremals issuing from P let Z_c denote the locus of points at which $J = c$ measuring J from P, with $0 < c \leq R(P)$. We prove the following lemma.

<u>Lemma 13.1.</u> <u>The locus</u> Z_c <u>is a regular</u> (n-1)-<u>dimensional manifold topologically equivalent to the</u> n-1-<u>sphere</u>. <u>If</u> $[u(x)]$ <u>is the direction of the field of extremals issuing from</u> P <u>at a point</u> (x) <u>on</u> Z_c <u>then</u>

$$F_{r^i}(x, r(x))dx^i = 0 \tag{13.7}$$

<u>for every direction</u> (dx) <u>tangent to</u> Z_c <u>at</u> (x).

The extremals issuing from P neighboring an extremal g issuing from P can be given a representation

$$x^i = x^i(s, v) \qquad\qquad (v) = (v^1, \ldots, v^{n-1})$$

of class C" in terms of parameters s = J and (v), with a non-vanishing jacobian for $0 < s \leq R(P)$. The locus J = c on these extremals has the representation $x^i = x^i(c, v)$. Since the functional matrix

$$\left\| \frac{\partial x^i(c, v)}{\partial v^h} \right\| \qquad\qquad (i=1,\ldots,n;\ h=1,\ldots,n-1)$$

has the rank n-1, this local representation of Z_c in terms of the parameters (v) is regular. In terms of these parameters, the integral along the extremals issuing from P is a function

$$J(s, v) = \int_0^s F[x(s, v), \dot{x}(s, v)]ds.$$

An arc b on Z_c has the form $v^h = v^h(t)$ $(h = 1,\ldots,n)$. Since J = s = c on Z_c and in particular on b, then for s = c and $v^h = v^h(t)$

$$0 = \dot{J} = F_{r^i}[x, u(x)]\dot{x}^i \ ,$$

Relation (13.7) follows.

To show that Z_c is a topological (n-1)-sphere one uses the functions A^i of Theorem 9.1 to represent Z_c in the form

$$x^i = A^i(x_o, cr) \qquad (i = 1, \ldots, n) \tag{13.8}$$

where (x_o) represents P, and (r) is an arbitrary point on the F-indicatrix K at P. This representation indicates a topological mapping of K onto Z_c so that Z_c, like K, is a topological (n-1)-sphere.

The proof of the lemma is complete.

In accordance with this lemma we term Z_c a manifold transverse to the extremals issuing from P.

Lemma 13.2. Of the manifolds transverse to the extremals issuing from P let Z_c and Z_d, with $0 < c < d \leqq R(P)$, bound a region S. Let a: $x^i = a^i(t)$, $i = 1, \ldots, n$, be an admissible arc on $\overline{S}$ leading from a point of Z_c to a point of Z_d. If F is positive definite and positive regular, then J_a exceeds d-c unless a is a subarc g of an extremal issuing from P.

The extremals from P form a field on $\overline{S}$ when $n > 2$, $\overline{S}$ is simply connected and the Hilbert integral I for this field is independent of the path. When $n = 2$, Z_c is a topological circle for whose tangent directions (dx), (13.7) holds. Hence $I = 0$ along Z_c. It follows that I is independent of the path on $\overline{S}$ even when $n = 2$. More generally $I = 0$, along paths on Z_c or Z_d since (13.7) holds. It follows that I is independent of the path joining Z_c to Z_d on $\overline{S}$.

To establish the lemma let g, with $c \leqq s \leqq d$, be a subarc of an extremal issuing from P. We have

$$J_a - J_g = J_a - I_a = \int_a E(a, u(a), \dot{a})ds$$

where $[u(x)]$ is the field direction. It follows that $J_a - J_g > 0$ except in the case where a is a subarc of an extremal from P.

This completes the proof.

For fixed P, let the least upper bound of field radii $R(P)$ be denoted by $\rho(P)$ and termed the radial field limit at P.

We shall presently show that $\rho(P)$ is finite and bounded independent of P on M.

The following theorem is fundamental.

Theorem 13.2. When F is positive definite and positive regular any extremal arc g issuing from P for which $J < \rho(P)$ affords an absolute proper minimum to J relative to all admissible arcs b which join its end points.

We shall apply Lemma 13.2 taking $d = J_g$ in Lemma 13.2. For $0 < c < d$ there exists a piece-wise regular subarc b^c of b, on $\overline{S}$ of Lemma 13.2, joining a point of Z_c to a point of Z_d. In accordance with Lemma 13.2,

$$J_{b^c} \geqq d - c = J_g - c \tag{13.9}$$

Since c is arbitrary subject to the condition $0 < c < d$ we conclude that

$$J_b \geqq J_g \, . \tag{13.10}$$

Let us show that $J_b = J_g$ only if $b = g$. Let b^o be the subarc of b preceding b^c on b. Since b^o joins the end points of an extremal arc on which J varies from 0 to c we conclude from (13.9) that

$$J_{b^o} \geqq c \, . \tag{13.11}$$

We have

$$J_b \geqq J_{b^o} + J_{b^c} \geqq c + (J_g - c) = J_g \tag{13.12}$$

so that if $J_b = J_g$ the equality must hold in (13.11) and (13.9). In accordance with Lemma 13.2, b^c must then be a subarc of an extremal issuing from P leading from Z_c to Z_d. If $J_b = J_g$, this must hold for each c for which $0 < c < d$.

Hence the closed set B must contain an extremal subarc g', issuing from P, on which s increases from 0 to d. But b can include no subarc other than g' if $J_b = J_g$, since $J_g = J_{g'}$. Thus b = g'. But b joins the end points of g. Hence g = g' = b.

The proof of the theorem is complete.

Lemma 13.3. *The radial field limit* $\rho(P)$ *is bounded independently of* P *on* M.

We have seen in Theorem 9.2 that there exists a field radius ρ independent of P on M. Let the set of points on M whose J-distance from P is less than ρ be termed a J-*disc with center* P *and radius* ρ. Since M is compact, M can be covered by a finite number m of J-discs of radius ρ. We are assuming that M is connected. On using arcs which are subarcs of extremals on J-discs issuing from the center of these discs, it is seen that at most 2m such arcs are required to join any two points on M. Since any extremal arc issuing from P with J-length at most R(P) is minimizing it follows that

$$R(P) \leqq 2m\,\rho$$

The lemma follows at once.

The radial field limit $\rho(P)$ *is never a field radius* R(P). In fact, the locus Z of points on extremals issuing from P on which $J = \rho(P)$ must either contain a point Q conjugate to P on some extremal issuing from P on which $0 < J \leqq \rho(P)$, or at least two of these radial extremal arcs intersect in their end points on Z. In any other case it is easy to see that $\rho(P) + e$ is a field radius R(P) for $e > 0$ and sufficiently small, contrary to the definition of $\rho(P)$.

As an example of a radial field limit consider a 2-sphere of unit radius with J the arc length. Here $\rho(P) = \pi$, and the point diametrically opposite to P is conjugate to P on each extremal issuing from P.

As a second example let P be a point on the inner equator of a torus. Suppose that the circle C generating the torus by revolution has a length 2b, exceeding that of the inner equator. Then $\rho(P) = b$ and the two geodesic semi-circles of C issuing from P intersect again at their ends, at which $J = b$ measured from P.

We have seen that the radial field limit $\rho(P)$ is bounded above, and bounded from zero. One can also prove that $\rho(P)$ is a continuous function of P.

<u>The most important conclusion is that there exists a field radius</u> $\rho > 0$ <u>independent of</u> P, <u>and that when</u> F <u>is positive definite and positive regular any extremal arc on</u> M <u>whose</u> J-<u>length is at most</u> ρ, <u>affords a proper absolute minimum to</u> J.

§14. <u>Extremals joining two points.</u> We need a formula analogous to a well known formula for the straight line joining two distinct points (x_o) and (x_1) in the plane, namely

$$(14.1)\qquad x^1 = (x_1^1 - x_o^1)s/r + x_o^1, \quad x^2 = (x_1^2 - x_o^2)s/r + x_o^2$$

where r is the distance between (x_o) and (x_1). Observe that the right members of (14.1) are functions of (x_o), (x_1) and s, with singularities when $(x_o) = (x_1)$. In the case of extremals there may be many extremals joining two given points (a_o) and (a_1) of M. Speaking generally there are formulas similar to (14.1) corresponding to each extremal g joining (a_o) to (a_1). The precise theorem is as follows.

Theorem 14.1. *Let* g *be an extremal in a coordinate region* R_x, *with J-length* s_1 *with end points* (a_o) *and* (a_1), *and with* (a_1) *not conjugate to* (a_o) *on* g_1. *There exist neighborhoods* N_o *and* N_1 *of* (a_o) *and* (a_1), *respectively, so small that any point* (x_o) *on* N_o *can be joined to any point* (x_1) *on* N_1 *by an extremal which is unique among extremals with the following properties.*

The extremal has the form

$$x^i = Z^i(s, x_o, x_1) \qquad (i = 1,\dots,n) \tag{14.2}$$

where Z^i *is of class* C'' *for* (x_o) *on* N_o, (x_1) *on* N_1, *and* s *on some open interval including* $[0, s_1]$. *The* J-*length of the extremal from* (x_o) *to* (x_1) *is a function* $s(x_o, x_1)$ *of class* C''. *The representation* (14.2) *yields* (x_o) *when* $s = 0$, (x_1) *when* $s = s(x_o, x_1)$, *and reduces to* g *for* $(x_o) = (a_o)$, $(x_1) = (a_1)$ *and for* s *on* $[0, s_1]$.

We seek extremals *neighboring* g in the sense that their initial points (x_o), directions (r), and J-lengths are near those of g. We suppose that $F(a_o, r_o) = 1$. For fixed (x_o) we regularly represent the indicatrix $F(x_o, r) = 1$ for (x_o, r) near (a_o, r_o) in the form $r^i = r^i(v, x_o)$. Here (v) is a set of parameters $(v^1, \dots, v^{n-1})$ near a set (0) such that $r_o^i = r^i(0, a_o)$, and the functions $r^i(v, x_o)$ are of class C'' for (v, x_o) near $(0, a_o)$. We are concerned with the extremals

$$x^i = X^i(s, x_o, r(v, x_o)) \qquad (i = 1,\dots,n) \tag{14.3}$$

and seek to satisfy the conditions

$$X^i(s, x_o, r(v, x_o)) = x_1^i \qquad (i = 1, \dots,n), \tag{14.4}$$

near the initial solution $s = s_1$, $(x_o) = (a_o)$, $(v) = (0)$. Since (a_1) is not conjugate to (a_o) on g the jacobian

$$\frac{D(X^1, \ldots, X^n)}{D(s, v^1, \ldots, v^{n-1})} \neq 0$$

at the initial solution of (14.4). There accordingly exist solutions $s(x_0, x_1)$, $v^h = v^h(x_0, x_1)$, $h = 1, \ldots, n-1$, of (14.4) of class C" for (x_0, x_1) sufficiently near (a_0, a_1). On replacing (v) in (14.3) by $[v(x_0, x_1)]$ a representation (14.2) is obtained with the properties stated in the theorem, provided N_0 and N_1 are sufficiently small.

The preceding theorem holds under the Weierstrass condition $F_1 \neq 0$. In case F is positive definite and positive regular, and the J-length of the given extremal arc g is less than a field radius ρ, then g is the unique absolute minimizing arc joining its end points.

Let the J-distance J(A, B) between any two points A and B on M be defined as the greatest lower bound of the J-lengths of admissible curves joining A to B. We shall presently see that if $A \neq B$ there is at least one extremal joining A to B whose J-length is J(A, B).

An extremal E(A, B) whose J-length is at most a field radius ρ will be termed elementary, and affords an absolute minimum to J relative to admissible arcs which join A to B.

The results of Theorem 14.1 yield the following:

Theorem 14.2. The point on an elementary extremal E(A, B) at which

$$J = t\, J(A, B) \qquad (0 < t \leq 1) \tag{14.5}$$

admits a representation in a suitable coordinate region S_x by coordinates which are functions of class C" of t and of the coordinates of (A) and (B).

The parameter t is called the reduced J-length along E(A, B).

Theorem 14.3. The end-points P, G of an admissible arc k with J-length less than $\rho(P)$ can be joined by an elementary extremal.

Let the variable point on k be represented as a function P(s) of the J-length measured along k from P. As long as P(s) remains on a J-disc with center at P and J-radius less than $\rho(P)$ it is clear that P and P(s) are the end-points of an elementary extremal. But this is true for s sufficiently small, and remains true for each point P(s) on k since the J-length of k from P to P(s) remains less than $\rho(P)$.

The J-distance J(A, B) satisfies the triangle axiom

$$J(A, C) \leqq J(A, B) + J(B, C)$$

as follows from the definition of J(A, B). As a consequence one can prove the following theorem.

Theorem 14.4. The J-distance J(A, B) is a continuous function of the pair (A, B).

We shall prove J(A, B) continuous at a pair (A_o, B_o). By virtue of the triangle axiom one has

$$J(A, B) \leqq J(A, A_o) + J(A_o, B_o) + J(B_o, B)$$

Let e be a prescribed positive constant. If A is sufficiently near A_o and B sufficiently near B_o,

$$J(A, A_o) < \frac{e}{2}, \quad J(B_o, B) < \frac{e}{2}.$$

One can see by using a particular coordinate system S_x containing A_o (or B_o) and joining A to A_o (or B_o to B) by a straight line on which J can be estimated. Hence

$$J(A, B) \leqq J(A_o, B_o) + e$$

One similarly proves that

$$J(A_o, B_o) \leqq J(A, B) + e$$

provided A and B are sufficiently near A_o and B_o respectively. The theorem follows.

The existence of an extremal of minimum type joining A to B for arbitrary A and B.

Lemma 14.1. *If the first variation of J vanishes for a piece-wise regular arc* k: $x^i = x^i(s)$ $(i = 1,\ldots,n)$ *of class* D', *and if* F *is positive regular at each point* (x) *of* k, *then* k *is without corners.*

In the classical treatment of the first variation, k is compared with arcs of a 1-parameter family of arcs of the same character as k, and the Euler equations derived along k in the integral form

$$F_{r^i}(x, \dot{x}) = \int_0^s F_{x^i}(x, \dot{x})ds + C_i \qquad (i = 1,\ldots,n) \tag{14.6}$$

here (14.6) holds on each regular subarc of k with C_i a constant independent of the subarc chosen. If P is a corner point of k, and if (r) and (q) are the two F-unit vectors giving the two directions positively tangent to k at P, then (14.6) implies that

$$F_{r^i}(x, q) = F_{r^i}(x, r) \tag{14.7}$$

On multiplying the members of (14.7) by q^i and summing one finds that

$$q^iF_{r^i}(x, q) - q^iF_{r^i}(x, r) = 0$$

$$F(x, q) - q^iF_{r^i}(x, r) = E(x, r, q) = 0.$$

Since F is regular at (x) the E-function vanishes only if (r) = (q) and the proof of the lemma is complete.

In particular, if k affords a relative or absolute minimum to J then k can have no corners.

Theorem 14.5. If F is positive definite and positive regular, any two distinct points A and B can be joined by an extremal whose J-length equals the J-distance J(A, B).

Let ρ be a universal field radius for M, and let m be any integer so large that

$$m\,\rho > J(A,\ B)$$

We shall admit broken extremals k joining A to B consisting of a sequence of m-elementary extremals each with a J-length at most ρ. In accordance with the definition of J(A, B) as the greatest lower bound of J-lengths of admissible arcs joining A to B there exist arcs h whose J-lengths J_h differ arbitrarily little from J(A, B). Any arc h for which $m\,\rho > J_h$ can be replaced by an admissible broken extremal k such that

$$J_k \leqq J_h \,. \tag{14.8}$$

One has merely to select a sequence

$$P_o,\ P_1,\ \dots,\ P_m \qquad (P_o = A;\ P_m = B) \tag{14.9}$$

of successive points on h such that the J-lengths of the successive subarcs $P_{i-1}P_i$ of h are at most ρ. The elementary arcs

$$E(P_{i-1},\ P_i) \qquad (i = 1,\dots,m) \tag{14.10}$$

exist in accordance with Theorem 14.3, and together define a broken extremal k for which (14.8) holds.

(i) To minimize J among admissible arcs joining A to B it is accordingly sufficient to minimize J among broken extremals composed of m successive elementary extremals, provided always that $m\rho > J(A, B)$.

A set (P) of vertices of the form (14.9) selected arbitrarily subject to the conditions that $P_o = A$, $P_m = B$, and

$$J(P_{i-1}, P_i) \leqq 0 \qquad (i = 1,\ldots,m) \tag{14.11}$$

will define an admissible broken extremal $k(P)$ joining A to B. Such sets (P) form a compact subset H of points on the product manifold M^{m+1}. The J-length of $k(P)$ is a continuous function of the set (P) on H, and so assumes its absolute minimum J_o at some set (Q) of H. It is clear that

$$J_o = J(A, B).$$

The broken extremal $k(Q)$ affords a minimum to J among admissible curves joining A to B in accordance with (i). It follows from Lemma 14.1 that $k(Q)$ is without corners, and is accordingly an extremal.

The proof of the theorem is complete.

It is easy to modify the preceding proof to show that there exists a minimizing extremal joining A to B of a given homotopy type (that is, deformable into a given arc joining A to B) and affording an absolute minimum to J among admissible arcs of this homotopy type. The same methods of proof permit one to establish that there is a closed extremal of any given non-trivial homotopy type[§], affording a minimum to J relative to closed curves of the same homotopy type. For example if M is of the topological type of the torus there are infinitely many of these minimizing closed extremals. These minimizing extremals are not necessarily unique, but if not unique their existence in general implies the existence of extremals of the given homotopy type which are no longer minimizing. The latter extremals are properly called <u>unstable</u>.

§ Among closed curves mutually deformable into each other.

In typical problems they exist with infinitely many kinds of unstability and are illustrative of general physical phenomena. The subsequent theory in the large will solve the problem of the existence and classification of unstable extremals.

§15. <u>Reduction to a non-parametric integral.</u> The transition from an integral in parametric form to one in non-parametric form is necessarily limited to a special class of curves. For our purposes it will be sufficient to consider a coordinate space R_x of m-variables

$$(x^1, \ldots, x^m) = [y_1, \ldots, y_n, x] \qquad (n = m-1) \tag{15.1}$$

written in the alternative form as indicated, and to consider piece-wise regular arcs on which $\dot{x}^m > 0$. On such arcs it is possible to make the transformation from t to x^m as a parameter.

For arcs on which $\dot{x}^m \neq 0$ the last Euler equation is a consequence of the first m-1 Euler equations, since one has the identity

$$\dot{x}^i\left[\frac{d}{dt} F_{r^i} - F_{x^i}\right] = 0 \qquad (i = 1,\ldots,m) \tag{15.2}$$

In this way one accounts for the fact that there are m Euler equations in the parametric form and m-1 Euler equations in the non-parametric form.

The integrand in the non-parametric form is defined in terms of the function

$$F(x^1, \ldots, x^m; r^1, \ldots, r^m)$$

by the equation (with n = m-1)

$$F(y_1, \ldots, y_n, x; p_1, \ldots, p_n, 1) = f(x, y_1, \ldots, y_n; p_1, \ldots, p_n) \tag{15.3}$$

For these arguments one sees that

(15.4) $$F_{x^h} = f_{y_h}, \quad F_{r^n} = f_{p_n}, \quad F_{r^h r^k} = f_{p_h p_k},$$

for h, k = 1, ..., n. The value of J along an arc on which $\dot{x}^m > 0$, and which has been given a second representation $y_h = y_h(x)$ with $x^m = x = t$ is

$$\int_0^t F(x, \dot{x})dt = \int_0^x F(y_1, \ldots, y_n, x; y_1', \ldots, y_n', 1)dx$$

$$= \int_0^x f(x, y, y')\, dx$$

We have seen in §9 that a solution of the Euler equations in which t is a parameter, remains a solution as a function of t_1 if one replaces t by kt_1 and dt by kdt_1, $(k > 0)$. The formal proof of this makes no use of the fact that k is a constant. On replacing t by $h(t_1)$ and dt by $\dot{h}\, dt_1$ (where the derivative $\dot{h}$ is piece-wise of class C" and positive) we infer that a solution of the Euler equations in which t is the parameter, is replaced by a solution of the Euler equations in which t_1 is the parameter. In particular for solutions along which $\dot{x}^m > 0$ one can change to $x^m = x$ as a parameter, and obtain a solution $[y_1(x), \ldots, y_n(x), x]$ of the m Euler equations in parametric form. The first n = m-1 of these equations then become

$$\frac{d}{dx} F_{r^h}(y_1, \ldots, y_n, x; y_1', \ldots, y_n', 1) - F_{x^h}(y_1, \ldots, y_n, x; y_1', \ldots, y_n', 1)$$

$$= \frac{d}{dx} f_{p_h}(x, y, y') - f_{y_h}(x, y, y') = 0 \qquad (h = 1, \ldots, n)$$

As we have seen in connection with (15.2) the m-th Euler condition on F is a consequence of the other conditions when $\dot{x}^m > 0$.

<u>Thus a solution of the Euler equations in parameter form along which</u> $\dot{x}_m > 0$, <u>when represented in terms of</u> $x^m = x$ <u>as the parameter, becomes a solu-</u>

tion of the Euler equat ions in non-parametric form.

We state the lemma.

Lemma 15.1. For $r^m > 0$ the condition that

$$F_{r^i r^j}(x, r)\, z^i z^j > 0 \qquad (i, j = 1, \ldots, m) \tag{15.5}$$

for (z) linearly independent of (r), implies the condition (with $n = m - 1$)

that

$$Q(w) = f_{p_h p_k}(x, y, p) w^h w^k > 0 \qquad (h, k = 1, \ldots, n) \tag{15.6}$$

for $(w) \neq (0)$, provided

$$(x^1, \ldots, x^m) = [y_1, \ldots, y_n, x]$$

and the sets (r) and $(p_1, \ldots, p_n, 1)$ define the same direction.

Note first that (15.5) implies $Q(w) \geqq 0$. If $Q(w) = 0$, the form in (z) given by (15.5) must vanish for

$$(z^1, \ldots, z^m) = (w^1, \ldots, w^n, 0).$$

Since (z) must then be linearly dependent on (r) with $r^m \neq 0$, we conclude that $(z) = (0)$. Hence $(w) = (0)$ and the lemma follows.

We add the following.

The Weierstrass non-singularity condition $F_1(x, r) \neq 0$ is equivalent, when $r^m \neq 0$, to the condition

$$\left| f_{p_h p_k} \right| \neq 0 \qquad (h, k = 1, \ldots, n;\ n = m-1)$$

for variables (x, y, p) and (x, r) related as in Lemma 15.1.

This follows from the relation

$$A_{ij} = F_1 r^i r^j$$

of (7.11), on setting $i = j = m$.

Lemma 15.2. *The conjugate points of a point P on an extremal arc g on which $\dot{x}^m(s) > 0$, as defined in the parametric theory, agree with the conjugate points of P as defined in the non-parametric theory with $x = x^m$.*

Suppose that $s = 0$ at P. The conjugate points of P on g are the zeros on g other than $s = 0$ of the jacobian of a family of extremals§.

$$x^i = x^i(s, v^1, \ldots, v^n) \qquad (i = 1,\ldots,m;\ m = n+1) \tag{15.7}$$

issuing from P when $s = 0$. Here $(v) = (0)$ defines g, and evaluated for $(v) = (0)$ the jacobian

$$\frac{D(x^1, \ldots, x^m)}{D(x, v^1,\ldots,v^n)} = D_1(s) \tag{15.8}$$

characteristically vanishes to the n-th order in s at $s = 0$. Since $\dot{x}^m(s) > 0$ on g, we can take $x^m = x$ as a parameter along these extremals provided (v) is sufficiently near zero. This amounts to a substitution $s = s(x, v)$ in (15.7), where $s_x(x, 0) > 0$ along g. We set

$$x^i[s(x, v), v] = T^i(x, v) \qquad (i = 1,\ldots,m) \tag{15.9}$$

and observe that

$$\frac{D(x^1, \ldots, x^m)}{D(s, v^1,\ldots,v^n)} \cdot \frac{\partial s}{\partial x} = \frac{D(T^1, \ldots, T^m)}{D(x, v^1, \ldots, v^n)} \tag{15.10}$$

Since $T^m = x$, the latter jacobian takes the form along g

$$\frac{D(T^1, \ldots, T^n)}{D(v^1, \ldots, v^n)} = D_2(x) \qquad [(v) = (0)] \tag{15.11}$$

It is clear that $D_1(s)$ and $D_2(x)$ vanish at corresponding points on g. But the columns of the determinant D_2 are solutions of the Jacobi equations in non-parametric form in accordance with the theorem of Jacobi. Moreover,

§. Taken as in (11.3)' where n is replaced by n-1.

each of these columns vanishes at the point $x = 0$ defining P. Finally $D_2(x) \neq 0$. Accordingly the columns of D_2 form a base for solutions of the Jacobi differential equations which vanish at $x = 0$. As we have seen in connection with (2.7) the zeros of $D_2(x)$ other than $x = 0$ define the conjugate points of P on g.

The proof of the lemma is complete.

The <u>index</u> of g as defined in the non-parametric theory is the sum of the multiplicities of the conjugate points of $s = 0$ on the open subarc of g. When the final end point of g is not conjugate to the initial point of g this index will turn out to be a basic semi-topological characteristic of g. It is important to show that this index is independent of the coordinate system in which the conjugate points are defined.

In the parametric theory a conjugate point s_1 of $s = 0$, defined by the vanishing of the jacobian $D_1(s)$ of (15.8) may be said to have a <u>multiplicity</u> equal to the nullity§ of $D_1(s_1)$. The $n = m-1$ angular parameters (v) used in (15.7) may suffer any non-singular transformation near $(v) = (0)$. The effect will be merely to multiply the jacobian matrix of the last n columns of $D_1(s_1)$ by a non-singular n-square matrix.

To change to any other admissible coordinate system (z) within which g lies will in effect multiply the matrix of $D_1(s_1)$ by the non-singular m-square functional matrix of the coordinate transformation from (x) to (z), evaluated at the point s_1 on g. Neither of these changes will alter the multiplicity of s_1 as a conjugate point of $s = 0$.

If finally one has an extremal along which $\dot{x}_m > 0$ one can change to coordinates (x, y) with $x^m = x$ as in the proof of the preceding lemma. The matrix

§. The nullity of a determinant is its order minus its rank.

of $D_2(x)$ is then obtained by multiplying the matrix of $D_1(s)$ by the matrix of the jacobian

$$\frac{D(s, v^1, \ldots, v^n)}{D(x, v^1, \ldots, v^n)}$$

under the transformation $s = s(x, v)$. The multiplicities of $D_1(s)$ and $D_2(x)$ at a conjugate point on g thus are equal. We have proved the following lemma.

Lemma 15.3. The multiplicity of a conjugate point of $s = 0$ on an extremal g is independent of the angular parameters (v) used to define conjugate points, of the coordinate region in which g is represented, and of any admissible change to a non-parametric system of coordinates (x, y).

§16. Rectifiable arcs. It will be necessary to broaden our class of admissible arcs to include arcs which are rectifiable in any coordinate region R_x in which they lie. Let h be an arc represented by continuous functions $x^i(t)$ $(i = 1, \ldots, m;\ 0 \leqq t \leqq t_1)$. A necessary and sufficient condition that h be rectifiable is that each $x^i(t)$ be of bounded variation. If h is rectifiable on R_x it admits a representation $x^i = a^i(\sigma)$ in terms of its euclidean arc length[§] σ with $0 \leqq \sigma \leqq b$, where b is the total length of h. Then $\dot{a}^i(\sigma)$ exists almost everywhere on [0, b] and

$$\dot{a}^i \dot{a}^i = 1 \qquad (i = 1, \ldots, m)$$

It follows in particular that $a^i(\sigma)$ is absolutely continuous (written A.C.). However, it is not necessary that $x^i(t)$ be A.C. in order that h be rectifiable although any rectifiable arc admits an A.C. representation, as just stated.

If however, $x^i(t)$ is A.C. for each i, $x^i(t)$ is in particular of bounded variation, so that h will be rectifiable. We shall admit rectifiable arcs

§. See Saks, Théorie de l'Integrale, Warsaw 1933; pp.57-60 and require that $x^i(t)$ be A.C. We shall establish the following. (A) If $x^i(t)$ $(i = 1, \ldots, m)$ is A.C. on $[0, t_1]$, then $F[x(t), \dot{x}(t)]$ is measurable on $[0, t_1]$.

To that end, for t on $[0, t_1]$, set

$$r_n^i(t) = [x^i(t + 1/n) - x^i(t)]n, \qquad (n = 1,2, \ldots)$$

extending the definition of $x^i(t)$ beyond t_1 in any continuous manner so that $r_n^i(t)$ is well defined and continuous for each n. For almost all t on $[0, t_1]$

$$\lim_{n = \infty} r_n^i(t) = \dot{x}^i(t) .$$

For each n, $F[x(t), r_n(t)]$ is continuous and hence measurable* in t, while for almost all t

$$\lim_{n = \infty} F[x(t), r_n(t)] = F[x(t), \dot{x}(t)]$$

Statement (A) follows.

Lemma 16.1. When h has an A.C. representation $x^i(t)$ $(i = 1, \ldots, m)$ the integral

$$J_h = \int_0^{t_1} F[x(t), \dot{x}(t)]dt \tag{16.1}$$

exists as a Lebesgue integral.

Under the hypothesis h has a representation $a^i(\sigma)$ $(i = 1, \ldots, m)$ in terms of its euclidean arc length σ. Since $|\dot{a}^i| \leqq 1$ almost everywhere the integral

$$\int_0^{\sigma_1} F[a(\sigma)\ \dot{a}(\sigma)]d\sigma \tag{16.2}$$

exists.§ However σ equals an A.C. monotone function $\sigma(t)$ of t, and

* See Titchmarsh, The theory of functions, Oxford, 1932; p.331

§ Cf. Titchmarsh, op.cit., p.333.

according to the Lebesgue theorem[§§] on change of variable of integration the integral (16.2) equals the integral.

$$\int_0^{t_1} F[a(\sigma(t)),\ \dot{a}(\sigma(t))]\,\frac{d\sigma}{dt}\,dt\ . \tag{16.3}$$

But for almost all t on $[0,\ t_1]$,

$$\frac{dx^i}{dt} = \frac{da^i}{d\sigma}\ \frac{d\sigma}{dt} \qquad (i = 1,\ldots,m);$$

so that (16.3) reduces to (16.1) on using the homogeneity of $F(x,\ r)$ in (r).

We shall show that the theorems which affirm that an extremal arc affords a minimum relative to piece-wise regular arcs hold if the arcs admitted include rectifiable arcs on the given domain. To see this we review certain steps in the proofs of these theorems. The first concerns the Hilbert integral, or more generally a theorem on line integrals.

Lemma 16.2. *If* $P_i(x)$ $(i = 1,\ldots,m)$ *is continuous in* (x) *on* R_x, *and if the line integral*

$$I = \int P_i(x)dx^i$$

is independent of the path among piece-wise regular arcs on R_x, *then* I *is also independent of the path among arcs with A.C. representations* $x^i(t)$ $(i = 1,\ldots,m)$.

We choose a point $(x) = (a)$ on R_x and write

$$\int_{(a)}^{(x)} P_i(x)dx^i = H(x),$$

using piece-wise regular arcs from (a) to (x). As is well known, H(x) is of class C' and

$$H_{x^i} = P_i(x)\ . \qquad (i = 1,\ldots,m). \tag{16.4}$$

---

§§ Cf. Titchmarsh, *op. cit.*, p.377

To establish the lemma it will be sufficient to suppose that $x^i(0) = a^i$ and to prove that

$$H[x(t)] = \int_0^t P_i[x(t)]\,\dot{x}^i(t)dt \, . \tag{16.5}$$

To that end we make use of (16.4), and with the aid of the law of the mean infer that

$$\frac{d\,H[x(t)]}{dt} = P_i[x(t)]\,\dot{x}^i(t) \tag{16.6}$$

at each point t at which the derivatives $\dot{x}^i(t)$ all exist. We note also that $H[x(t)]$ is A.C. in t. Relation (16.5) then follows from Lebesgue's theorem§ on the integral of the derivative of an A.C. function.

In using the Hilbert integral we have found it convenient to represent our arcs in terms of a parameter

$$s = \int_0^t f[x(t), \dot{x}(t)]dt. \tag{16.7}$$

If t is the euclidean arc length along the arc $[x(t)]$ in (16.7) with $0 \leqq t \leqq t_1$ then (16.7) defines a change of parameter $s = s(t)$. For almost all values of t on $[0, t_1]$, $c_1 < \dot{s} < c_2$ where c_1 and c_2 are positive constants. The inverse of $s(t)$ has the same character as $s(t)$ and leads to an A.C. representation $x^i = a^i(s)$ of the given arc.

The final step in establishing a <u>proper</u> relative minimum involved the f-unit vector $u^i(x)$ $(i = 1, \ldots, m)$ giving the direction of a Mayer field at the point (x). The functions $u^i(x)$ are of class C' (at least) at each point of the field. We need the following lemma.

<u>Lemma 16.3.</u> <u>If</u> h <u>is an arc with an A.C. representation</u> $x^i(s)$ $(i = 1, \ldots, m;\ 0 \leqq s \leqq b)$ <u>which satisfies the differential equations</u>

---

§ Cf. Titchmarsh, <u>op. cit.</u>, pp.365-366.

$$\frac{dx^i}{ds} = u^i(x) \qquad (i = 1, \ldots, m) \tag{16.8}$$

<u>for almost all values of</u> s <u>on</u> [0, b] <u>then each</u> $x^i(s)$ <u>is of class</u> C" <u>in</u> s <u>on</u> [0, b] <u>and</u> h <u>reduces to a uniquely defined solution of</u> (16.8).

By the Lebesgue theorem on integrating the derivative of an A.C. function

$$x^i(s) = x^i(0) + \int_0^s u^i[x(s)]ds \tag{16.9}$$

It appears from (16.9) that h satisfies (16.8) for <u>all</u> values of s, and the lemma follows.

With the aid of these lemmas the proofs of the theorems which involve the Hilbert integral go through when the curve compared with the extremal g has an A.C. representation. In particular one has the following theorem.

<u>Theorem 16.1.</u> <u>When</u> F <u>is positive definite and positive regular an elementary extremal arc</u> E(A, B) <u>affords a proper, absolute minimum to</u> J <u>relative to rectifiable arcs which join</u> A <u>to</u> B.

§17. <u>A non-degenerate critical point of</u> H(x). The quadratic analysis with which these lectures began was preliminary to the topological characterization of the level manifolds of a function near a critical point P by means of the quadratic form appearing in a representation of H about P. Before turning to functions of arcs such as J it will be illuminating to study a critical point P of a function H(x) of n variables $(x_1, \ldots, x_n)$.

<u>We shall suppose that</u> P <u>is a non-degenerate critical point of</u> H <u>in the sense that the Hessian of</u> H <u>at</u> P <u>is not zero. We shall assume that</u> H(x) <u>is of class</u> C"' <u>near</u> P.

A non-degenerate critical point (x_0) has the property that it is isolated among critical points of H. For the equations

$$H_{x_i} = 0 \qquad (i = 1, \ldots, n) \tag{17.1}$$

have (x_0) as an initial solution, and since the jacobian of the functions H_{x^i} is not zero at (x_0), there is no solution of (17.1) other than (x_0) in a sufficiently small neighborhood of (x_0). In particular, if $H(x)$ is defined on a neighborhood of a compact subset K on which the critical points of H are non-degenerate then the number of critical points of H on K is finite .

Suppose for simplicity that $(x_0) = (0)$. Employing Taylor's formula with the integral form of the remainder [Cf. Jordan, Cours d'Analyse, vol.I; p.249] we find that

$$H(x) - H(0) = a_{ij}(x)\, x_i x_j \qquad (i,j = 1,\ldots,n) \tag{17.2}$$

in a sufficiently small neighborhood of (0), where

$$a_{ij}(x) = \int_0^1 (1 - u) H_{x_i x_j}(ux_1,\ldots,ux_n)\,du$$

It follows that

$$a_{ij}(0) = \tfrac{1}{2} H_{x_i x_j}(0)$$

In particular $|a_{ij}(0)| \neq 0$ since (0) is a non-degenerate critical point.

We shall establish the following theorem.

Theorem 17.1. If $H(x)$, of class C''', has a non-degenerate critical point of index k at the point $(x) = (0)$ there exists a non-singular transformation $y_i = y_i(x)$ of class C' neighboring $(x) = (0)$ under which

$$H(x) - H(0) \equiv -y_1^2 - \ldots - y_k^2 + y_{k+1}^2 + \ldots + y_n^2 \tag{17.3}$$

If the coefficients $a_{ij}(x)$ in (17.2) were constant one could obtain a representation (17.3) of H by making the classical Lagrange reduction of the form $a_{ij}x_ix_j$. (See Bôcher, Higher Algebra, p.131). Proceeding formally, as if the coefficients $a_{ij}(x)$ were constants, we can still effect this reduction in a sufficiently small neighborhood of $(x) = (0)$. In particular if $a_{11}(0) \neq 0$, then the difference§

$$a_{ij}x_ix_j - \frac{1}{a_{11}}[a_{11}x_1 + \dots + a_{1n}x_n]^2 = b_{pq}(x)z_pz_q$$

where p, q has the restricted range $(2, \dots, n)$. The substitution

$$(17.4) \qquad z_1 = \frac{a_{1j}x_j}{a_{11}}, \qquad z_2 = x_2, \dots, z_n = x_n \qquad (j = 1, \dots, n)$$

yields the relation

$$a_{ij}x_ix_j = a_{11}z_1^2 + b_{pq}z_pz_q \qquad (p, q = 2, \dots, n)$$

If $a_{11}(0) = 0$, but $a_{rr}(0) \neq 0$ for some $r > 1$, a substitution of the form (17.4) is applicable after interchanging the variables x_1 and x_r. If each of the coefficients $a_{ii}(0) = 0$, at least one of the coefficients $a_{1r}(0) \neq 0$. After a change of variables of the form

$$x_1 = \bar{x}_1 - \bar{x}_r \qquad x_r = \bar{x}_1 + \bar{x}_r$$

the new coefficient $\bar{a}_{11}(0) \neq 0$ and a substitution of the form (17.4) will be possible. Thus in any case one is led to a quadratic remainder

$$b_{pq}z_pz_q \qquad (p, q = 2, \dots, n)$$

to which the same method of reduction is applicable, leading finally to a representation (17.3).

---

§ The functions $b_{pq}(x)$ are again of class C'.

The transformation (17.4) is of class C' near (x) = (0) since $a_{ij}(x)$ is of class C'. It has a jacobian 1 at the origin. It is accordingly locally non-singular. We point out that this holds under the hypothesis that H(x) is of class C''' near (x) = (0); were H(x) merely of class C'' $a_{ij}(x)$ would be merely continuous in general, and the transformation would not always be 1-1. For example, a transformation

$$z = c(x)x \qquad [c(0) = 1]$$

in which c(x) is merely continuous near x = 0, is not necessarily 1-1.

The preceding reduction holds formally if each coefficient $a_{ij}(x)$ is replaced by $a_{ij}(0)$. According to the "law of inertia" of quadratic forms the integer k appearing in (17.3) is the index of the form

$$a_{ij}(0)\, x_i x_j$$

(See Bôcher, _op.cit._, p.146]. The proof of the theorem is complete.

H-deformations. We shall refer to a deformation T of a subset A of the region R_x on which H(x) is defined. The _time_ t during which T acts shall increase from 0 to a. A deformation T of A on R_x during the time interval [0, a] is defined by a continuous map

$$Q = f(P, t); \qquad \text{with } P = (P, 0)$$

of the product A×[0, a] into R_x. For P fixed and t variable the image f(P, t) of [0, a] is called the trajectory of P. A deformation is called an _H-deformation_ if H is monotonically decreasing on each trajectory as t increases. We say that a point (x) is _below_ c if H(x) < c.

Theorem 17.2. Suppose that H(x) _has a non-degenerate critical point of index_ k > 0 _at a point_ (x_o) _at which_ $H(x_o) = c$. _There then exists a_ k-_cell_

D_k *with* (x_o) *in its interior, and on which* (x) *is below* c *except at* (x_o), *and an* n-*cell* $D_n \supset D_k$ *with* (x_o) *in its interior, such that* D_n *admits an* H-*deformation of itself onto* D_k *during which* D_k *remains point-wise fixed.*

Without loss of generality, we can suppose $(x_o) = (0)$, and refer the neighborhood of the critical point to coordinates $(y_1, \ldots, y_n)$ such that (17.3) holds for $y_i y_i \leqq e^2$ $(e > 0)$. One can take D_k as the subspace on which

$$y_1^2 + \ldots + y_k^2 \leqq e^2, \qquad y_{k+1} = \ldots = y_n = 0 ,$$

for it is clear that $H < c$ on D_k except when $(y) = (0)$. We take D_n as the n-cell defined by $y_i y_i \leqq e^2$. The cell D_n can then be H-deformed on itself onto D_k by letting

$$y_i(t) = (1 - t)y_i \qquad (j = k+1, \ldots, n;\ 0 \leqq t \leqq 1).$$

Note that $D_k = D_n$ when $k = n$, and that the deformation of the theorem is the identity. When $k = 0$, the origin affords a proper relative minimum to H on D_n.

We now have the difficult problem of obtaining an analogue of Theorem 17.2 for our integral J taken along rectifiable curves in the neighborhood of an extremal g. In this connection the following remark will be useful. The terms of second order in the Taylor's development of H(x) about the origin are given by the formula

$$\left[\frac{d^2H}{de^2}\ (ex_1,\ \ldots,\ ex_m)\right]_{e=0} = H_{x^i}(0)\ x^i x^j .$$

This is a consequence of the formula

$$\frac{dH}{de}\,(ex) = H_{x^i}(ex)x^i .$$

§18. The index[§] function $J(z_1,\dots,z_q)$. Let g_o be an extremal represented by an arc $[a^1, a^2]$ of the x-axis in a space $(x, y) = (x, y_1,\dots,y_n)$ in which

$$J = \int_{a^1}^{a^2} f(x, y, y')dx .$$

Here $f(x, y, p)$ is of class C''', for (x, y) in a neighborhood N of g_o and (p) any set of slopes. Instead of the broken secondary extremals of §3, we shall here consider a finite sequence of primary extremals joining the point $x = a^1$ on g_o to the point $x = a^2$ on g_o with vertices on $(n-1)$-planes on which x is respectively one of the constants

$$a^1 = x_o < x_1 < ,\dots, x_{m-1} < x_m = a^2 .$$

As in §3, we suppose that there is no pair of conjugate points on any one of the intervals $[x_j, x_{j+1}]$. Suppose also that f is positive regular for (x, y) on N. Let

$$(y^j) = (y_1^j, \dots, y_n^j) \qquad (j = 1,\dots,m-1) \tag{18.0}$$

be a point on the $(n-1)$-plane $x = x_j$. With $q = n(m-1)$ set

$$(y_1^1, \dots, y_n^1, \dots, y_1^{m-1}, \dots, y_n^{m-1}) = (z_1, \dots, z_q). \tag{18.1}$$

If the set (z) is sufficiently near the set $(z) = (0)$ there exists a broken extremal E_z' joining the end points of g_o with vertices at the successive points (y^j) defining (z). The value of J on E_z' will be denoted by $J(z,\dots,z_q)$ and termed an index function belonging to g_o.

As in §3, the set (z) may be used to define a broken secondary extremal E_z'' with the same end-points and vertices as E_z'. The value of the second variation

§ Introduced by the author.

$$\int_{a^1}^{a^2} 2\Omega(u, u')dx$$

taken along E''_z has been denoted by $Q(z)$ and termed the <u>index form</u> belonging to g_o. It is understood that the second variation is based on g_o. The principal theorem of this section is as follows.

<u>Theorem 18.1. If</u> $J(z)$ <u>is the index function and</u> $Q(z)$ <u>the index form belonging to</u> g_o, <u>then</u>

(18.2) $$J_{z_h z_k}(0)\, z_h z_k = Q(z) \qquad (h, k = 1,\ldots,q)$$

We begin with a lemma.

<u>Lemma 18.1. For</u> (z) <u>sufficiently near</u> $(z) = (0)$, <u>the index function</u> $J(z)$ <u>is of class</u> C''' <u>and has a critical point at</u> $(z) = (0)$.

To establish the lemma let

$$y_i = Y_i(x, z) \qquad (i = 1, \ldots, n)$$

represent the broken extremal E'_z. On each interval $[x_{j-1}, x_j]$ the functions Y_i and Y_{ix} are of class C'' in (x, z) for (z) sufficiently near (0). Let w be one of the variables (z). Suppose w given by a coordinate $w = y^j_h$. As a function of x, Y_{iw} will be null except on the interval $[x_{j-1}, x_{j+1}]$ and will vanish at the end points of this interval. For a fixed h and j, and for $i = 1, \ldots, n$,

$$Y_i(x_j, z) = y^j_i\ ; \qquad Y_{iw}(x_j, z) = \delta_{ih}\ . \qquad [w = y^j_h\ ]$$

It follows that

(18.3) $$J_w = \int_{a^1}^{a^2} [f_{y_i} Y_{iw} + f_{p_i} Y_{ixw}]\, dx\ .$$

On integrating by parts on the intervals (x_{j-1}, x_j) and (x_j, x_{j+1}), (18.3) takes the form

$$J_w = \Big[f_{p_i}(x, Y, Y_x)\, Y_{iw} \Big]_{x_j^-}^{x_j^+}\ ,$$

$$(18.4) \qquad J_w = \Big[f_{p_h}(x, Y, Y_x) \Big]_{x_j-}^{x_j+} \qquad [w = y_h^j]$$

When $(z) = (0)$, E_z' reduces to g_o, and $J_w = 0$. That J_w is of class C", and hence J of class C"', is obvious from (18.4).

Proof of Theorem 18.1. We shall make use of the formula

$$(18.5) \qquad \frac{d^2 J^o}{de^2}(ez_1, \ldots, ez_q) = J^o_{z_r z_s} z_r z_s \qquad (r,s = 1,\ldots,q)$$

where the superscript o indicates evaluation, after differentiation, at $e = 0$. Set

$$Y_i(x, ez) = y_i(x, e) \qquad (i=1,\ldots,n)$$

Then

$$\frac{dJ}{de} = \int_{a^1}^{a^2} [f_{y_i} y_{ie} + f_{p_i} y_{iex}]\, dx \qquad (i=1,\ldots,n)$$

If we set $y_{ie}(x, 0) = u_i(x)$ we find that $(i = 1, \ldots, n)$,

$$(18.6) \qquad \frac{d^2 J^o}{de^2} = \int_{a^1}^{a^2} 2\Omega(u, u')dx + \int_{a^1}^{a^2} [f^o_{y_i} y^o_{iee} + f^o_{p_i} y^o_{ieex}]dx$$

The second integral in (18.6) vanishes since g_o is an extremal and

$$f^o_{y_i} = f_{y_i}(x, 0, 0), \quad f^o_{p_i} = f_{p_i}(x, 0, 0);$$

while y^o_{iee} vanishes at $x = a^1$ and a^2. Hence

$$(18.7) \qquad \frac{d^2 J^o}{de^2} = \int_{a^1}^{a^2} 2\Omega(u, u')dx$$

It remains to show that the integral in (18.7) is the index form $Q(z)$. One first notes that (y^o_e), taken on any one of the intervals $[x_{j-1}, x_j]$, is an extremal of the second variation based on g_o (Jacobi's theorem). Moreover, as a broken secondary extremal, $(u) = (y^o_e)$ is determined by (z) since

$$y_i(x_j, e) = ey_i^j, \qquad y_{ie}(x_j, 0) = y_i^j.$$

Here $i = 1, \ldots, n$ and $j = 1, \ldots, m-1$. Thus (u) represents the secondary extremal E_z'' and the integral in (18.7) equals $Q(z)$. From (18.7)

$$\frac{d^2J^o}{de^2} = Q(z) ;$$

and (18.2) follows from (18.5).

<u>A primary metric on</u> M. To continue with the analysis of rectifiable curves near a given extremal g_o we need a metric on M. The J-distance $J(A, B)$ between two points A and B is not suitable since $J(A, B)$ does not in general equal $J(B, A)$. We shall introduce a distance

$$AB = \max[J(A, B), J(B, A)] .$$

It is clear that $AB = BA$, that $AB = 0$ if and only if $A = B$, and that

$$AB = AC + BC$$

The metric defined by distances AB will be termed <u>primary</u>.

As in the Appendix on "A special parametrization of curves", we distinguish between parameterized curves (written p-curves) and classes of equivalent p-curves, termed curve classes or curves. We depart from the notation of the Appendix in that curves will be denoted here by letters, g, g', etc., while parameterized curves will be denoted by the same letter with the parameter as index, - g^t, g'^t, etc.

The Fréchet distance between two p-curves g^t and r^s will be denoted by g^tr^s, while the Fréchet distance between the curves g and r will be denoted by gr. As shown in the Appendix

$$gr = g^tr^s$$

and a necessary and sufficient condition that $g^tr^s = 0$ is that the curve class $g = r$. These Fréchet distances are defined with the aid of the primary distance AB, so that $qr = rq$, and qr satisfies the triangle axiom. Moreover,

$qr = 0$ if and only if the curve class $q = r$.

The total μ-length μ_g of an arc g is a continuous bounded function of g on the Fréchet space of arcs g. The maximum distance AB on M is a bound for μ_g. In the representation of g in terms of μ-length it will simplify matters if we set $\mu = \mu_g \sigma$ and let σ range on the interval $[0, 1]$. We term σ the <u>reduced μ-length</u>. The set of curves g joining two distinct points A_g and B_g on M admits a representation

$$P = P(\sigma, g) \qquad (0 \leqq \sigma \leqq 1)$$

in which $P(\sigma, g)$ is a parameterization of g in terms of reduced μ-length σ, where

$$P(0, g) = A_g , \qquad P(1, g) = B_g ,$$

and where $P(\sigma, g)$ maps (σ, g) continuously into a point on M. Here (σ, g) is an element on the product $W \times [0, 1]$ of the Fréchet space W of curves joining distinct points on M and of the interval $[0, 1]$ for σ. On any compact subset W_o of W, the mapping function $P(\sigma, g)$ is, of course, uniformly continuous.

<u>A J-deformation</u> D_m. Let W_o be any compact subset of W. Corresponding to W_o there exists an integer m so large that for g on W_o successive arcs of g between points

$$P(0, g), P(\tfrac{1}{m}, g), \ldots, P(\tfrac{m-1}{m}, g) \; P(1, g)$$

of g have J-diameters less than some universal field radius ρ. We shall define a J-deformation of g into a broken extremal $E_m(g)$ which consists of the sequence of elementary extremals which joins the successive points

$$P(\tfrac{r}{m}, g) = P_r , \qquad (r = 0, \ldots, m)$$

Let h_r, $r = 1,\ldots,m$, be the r-th subarc of g with end points P_{r-1} and P_r. As the time t increases from 0 to 1, h_r will be deformed into the elementary extremal $E(P_{n-1}, P_r)$ as follows. Let h_r^t be the subarc of h_r on which the increment of σ is the fraction t of the total increment of σ on h_r. At the time t, h_r^t shall be replaced by the elementary extremal which joins its end points, while the remainder of h_r shall be unaltered. This replacement of subarcs of h_r, performed for $r = 1,\ldots,n$, defines a J-deformation D_m of W_o in which g on W_o is replaced when $t = 1$ by the broken extremal $E_m(g)$. In defining this deformation it is not necessary to explicitly give a parameterization of the curve which replaces g at the time t; for we are concerned with curves and not p-curves. A parameterization in terms of reduced μ-length is always available if desired.

We now suppose that fixed, distinct points A and B are joined by an extremal g_o, and consider all rectifiable curves g joining A to B on a Fréchet neighborhood N of g_o.

If the neighborhood N of g_o is sufficiently small, and the integer m is sufficiently large, the preceding J-deformation will be applicable to N even though N is not in general compact. As previously, let $E_m(g)$ be the broken extremal which replaces g under D_m when $t = 1$.

As in the definition of the index function J(z) belonging to g_o, let g_o be represented by a segment $[a^1, a^2]$ of the x-axis in a coordinate space (x, y) and let the points§

$$A^1 = x_o < x_1 < , \ldots, x_{m-1} < x_m = a^2$$

be the successive vertices of the elementary extremals of $E_m(g_o)$. If N is

§ In §3 we set $a^1 = x_1$. To set $a^1 = x_o$ is more consistent with what follows.

sufficiently small (and we suppose this the case) the broken extremal $E_m(g)$ defined by a curve g on N will meet the successive (n-1)-planes $x = x_j$ in points $Q_j(g)$, ($j = 0, \ldots, m$), which are single-valued continuous functions of g, and these points $Q_j(g)$ will divide g into successive arcs each with a J-diameter less than the above field radius ρ.

The J-deformation D'_m. This is a deformation of broken extremals $E_m(g)$ for g on the preceding neighborhood N of g_o. In this deformation $E_m(g)$ is finally replaced by the broken extremal $E'_m(g)$ whose respective elementary extremals join the successive points $Q_j(g)$ ($j = 0, \ldots, m$). The manner in which the subarc $Q_{j-1}Q_j$ of $E_m(g)$ is deformed into $E(Q_{j-1}, Q_j)$ is similar to the manner in which the arc $P_{r-1}P_r$ of g was deformed into $E(P_{r-1}, P_r)$, and the definition need not be repeated.

The J-deformation D''_m. The sets $Q_j(g)$ ($j = 1,\ldots,m-1$) are determined by the sets $(z_1,\ldots,z_q)$ of their y-coordinates, so that the broken extremals $E'_m(g)$ can be given a J-deformation by subjecting the sets (z) to a deformation in which the index function J(z) never increases along a trajectory of a point (z). In particular, we can subject the sets (z) to the type of J-deformation (an H-deformation of points (x)) affirmed to exist in Theorem 17.1. The deformations D_m, D'_m, D''_m, successively performed in the order written effect the result affirmed in Theorem 18.2.

Theorem 18.2. Let g_o be an extremal whose end points are not conjugate and whose index k exceeds 0. Corresponding to a sufficiently small Fréchet neighborhood N_o of g_o among rectifiable curves joining the end points of g_o, there exists a J-deformation D on N_o of a sub-neighborhood $N \subset N_o$ of g_o, in which g_o remains fixed and the terminal image of N is a topological r-disc of curves g of N_o on which $J_g < J_{g_o}$ except when $g = g_o$.

It is necessary to have some knowledge of the manner in which the deformation D of Theorem 18.2 affects the r-disc K of the theorem. If D left K invariant there would be no difficulty. Recall that a deformation D of K gives a 1-parameter family of mappings D^t of K. Here D^0 is the identity and D^1 maps $N \cap K$ into K. The following theorem is needed.

Theorem 18.3. The terminal mapping

$$D^1: \quad N \cap K = K^0 \longrightarrow K$$

induced by D is deformable into the identity on K with g_0 fixed, and $K^0 - g_0$ deformed on $K - g_0$, provided N is sufficiently small.

Suppose that g in K^0 is replaced by g^t under D at the time t, $0 \leqq t \leqq 1$. If N is sufficiently small the broken extremal g^t will intersect the successive (n-1)-planes $x = x_j$ $(j = 0, 1, \ldots, m)$ used in defining the index function $J(z)$, in points $P_j(g^t)$ which will vary continuously with (g, t) for g in K and $0 \leqq t \leqq 1$. The points $P_j(g^t)$ will determine the set $(z_1, \ldots, z_q)$ of their y-coordinates. The final image of (z) under the preceding deformation D''_m will determine a point $\bar{g}^t$ in K, if N is sufficiently small.

Let D^* be a deformation of K^0 on K in which g in K^0 is replaced by $\bar{g}^t$ at the time t. It is clear that $\bar{g}^t$ varies continuously with (g, t) and that

$$J_g > J_{g^t} > J_{\bar{g}^t} .$$

Hence D^* deforms $K^0 - g_0$ on $K - g_0$. The point g_0 is fixed in D^*. As in the case of D, the initial mapping of D^* is the identity since $g = g^0 = \bar{g}^0$. Since $g^1 = \bar{g}^1$ the terminal mapping of D^* agrees with that of D. Thus D^* deforms the mapping D^1 into the identity in the manner stated.

§19. The lower semi-continuity of the integral J_g. To show that J_g is a lower semi-continuous function of rectifiable arcs g on M it will be sufficient to establish that J_g is a lower semi-continuous function of rectifiable arcs on any single coordinate region. This is seen as follows. An arc g is a finite sequence of arcs $g^{(1)}, \ldots, g^{(h)}$ each of which lies in a single coordinate system. If a countably infinite sequence of arcs g_n converges to g in the sense of Fréchet each arc g_n can be broken up into a finite sequence of arcs $g_n^{(1)}, \ldots, g_n^{(h)}$ such that for fixed j, $g_n^{(j)}$ converges to $g^{(j)}$ as n becomes infinite. The lower semi-continuity of J_g at $g^{(j)}$, $(j = 1,\ldots,h)$ will then imply the lower semi-continuity of J at g. Our proofs will accordingly be carried out in a single coordinate region.

The mode of proof is essentially that of McShane in his thesis Semi-continuity in the calculus of variations and absolute minima for isoperimetric problems, Contributions to the calculus of variations, University of Chicago Press (1930). We begin with a lemma.

Lemma 19.1. Let $[g_n]$ be a sequence of arcs in an m-dimensional coordinate region R_x, whose euclidean lengths are at most a constant K, with g_n converging in the sense of Fréchet to an arc g. If the arcs $[g_n]$ be referred to reduced arc length t as a parameter $(0 \leq t \leq 1)$, there exists a subsequence $[p_n]$ of arcs with representations $[x_n(t)]$ $(n = 1, 2, \ldots)$ such that $[x_n(t)]$ converges uniformly as a function of t to a representation $[x(t)]$ of g. The vectors $[\dot{x}_n(t)]$ and $[\dot{x}(t)]$ exist for almost all t on [0, 1] and have uniformly bounded euclidean lengths.

A representation

$$x^i = X_n^i(t) \qquad (i=1,\ldots,m;\ n=1,2,\ldots)$$

of g_n in terms of reduced arc length is such that the euclidean length of $[\dot{X}_n]$ exists and is at most K for almost all values of t on [0, 1]. It follows that for any two values t' and t" on [0, 1]

$$(19.1) \qquad \left| X_n^i(t') - X_n^i(t'') \right| \leqq K \left| t' - t'' \right|$$

so that the functions X_n^i are equi-continuous. For n sufficiently large these functions are also uniformly bounded since g_n converges to g in the sense of Fréchet. It follows from a theorem§ of Ascoli that a subsequence $[p_n]$ of $[g_n]$ exists with a representation $[x_n(t)]$ of p_n which converges uniformly in the ordinary sense to a representation [x(t)] of an arc g'. We see that g'g = 0 so that g' = g. See Appendix p. 10. The vector [x(t)] satisfies the Lipschitz condition (19.1) so that $[\dot{x}(t)]$ exists for almost all t on [0, 1] and has a bounded euclidean length.

Another lemma§§ is needed of which a proof is here given considerably simpler than the one to which reference is made.

<u>Lemma 19.2.</u> <u>Let</u> $A_n(t)$, (n = 1,2,...) <u>be an absolutely continuous function of</u> t, 0 < t < 1, <u>which converges uniformly to zero with</u> 1/n <u>and for which</u>

$$|\dot{A}_n(t)| < k \quad \text{(a constant)} \qquad (n=1,2,\ldots)$$

<u>If</u> H(t) <u>is any integrable function of</u> t, <u>then</u>

$$\lim_{n=\infty} \int_0^1 |\dot{A}_n(t)|\, H(t)\, dt = 0$$

---

§ Cf. Hobson, The theory of functions of a real variable, Vol. II, p.168 The theorem of Ascoli (or Arsela) requires an obvious extension for the above application.

§§ See Hobson, <u>op. cit.</u>, Vol. II, p.422.

The lemma is immediate if $H(t)$ is of class C'. One has merely to integrate $\dot{A}_n H$ by parts and pass to the limit.

In the general case let e be an arbitrary positive constant and recall that there exists a function $H_e(t)$ of class C' such that

$$\int_0^1 |H(t) - H_e(t)| \quad dt < e$$

Then

$$\int_0^1 \dot{A}_n H dt = \int_0^1 \dot{A}_n (H - H_e) dt + \int_0^1 \dot{A}_n H_e dt \,. \tag{19.2}$$

The second integral on the right tends to zero with $1/n$ and it follows that

$$\limsup_{n=\infty} \left| \int_0^1 \dot{A}_n H dt \right| \leqq k\, e \tag{19.3}$$

The lemma follows from the fact that e is arbitrarily small.

The theorems on lower semi-continuity depend upon the hypothesis that $F(x, r)$ is <u>quasi-regular</u>, that is that

$$E(x, r, q) \geqq 0$$

for (x) in any coordinate region R_x and (r) and (q) non-null vectors. Quasi-regularity is clearly implied by positive regularity, but not conversely. It implies that for fixed (x), $F(x, r)$ is <u>convex</u> in (r).

<u>Lemma 19.3.</u> <u>If</u> F <u>is quasi-regular in a coordinate system</u> R_x <u>then</u> J_g <u>is lower semi-continuous on the class of all curves on</u> R_x <u>whose euclidean lengths are at most a constant</u> K.

If the lemma were false there would exist a sequence $[g_n]$ of curves on R_x with euclidean lengths at most K which converge in the sense of Fréchet to an arc g on R_x for which

$$\limsup_{n=\infty} J_{g_n} < J_g \tag{19.4}$$

We make use of the subsequence $[p_n]$ of $[g_n]$ affirmed to exist in Lemma 19.1 and write

$$J_{p_n} - J_g = \int_0^1 F(x_n, \dot{x}_n)dt - \int_0^1 F(x, \dot{x})dt$$

$$(19.5) \qquad J_{p_n} - J_g = \int_0^1 [F(x_n, \dot{x}_n) - F(x, \dot{x}_n)]\, dt$$
$$+ \int_0^1 [F(x, \dot{x}_n) - \dot{x}_n^i F_{r^i}(x, \dot{x})]dt$$
$$+ \int_0^1 (\dot{x}_n - \dot{x}) F_{r^i}(x, \dot{x})dt$$

One precaution must be observed: $F_{r^i}(x, \dot{x})$ is not defined when $(\dot{x}) = (0)$; and since this may occur on a set w of values of t of positive measure we replace $(\dot{x})$ on the set w by a constant non-null vector. Relation (19.5) remains valid.

From (19.5) it follows that

$$(19.6) \qquad \text{lim sup. } (J_{p_n} - J_g) = \text{lim sup. } \int_0^1 E(x, \dot{x}, \dot{x}_n)dt.$$

For the first integral in (19.5) tends to zero with $1/n$ since $(\dot{x}_n)$ is bounded in length and (x_n) converges uniformly to (K). The last integral in (19.5) converges to zero by virtue of Lemma 19.2. The integrand of the remaining integral is $E(x, \dot{x}, \dot{x}_n)$, so that (19.6) follows. From (19.6) and the quasi-regularity of $F(x, r)$ it appears that

$$\text{lim sup. } J_{p_n} \geqq J_g$$

contrary to (19.4).

The lemma follows.

Theorem 19.1. If $F(x, r)$ is quasi-regular and positive definite on M, then J_g is lower semi-continuous on the class of all rectifiable curves.

As pointed out in the initial paragraph of this section the theorem is true if true for curves confined to any coordinate system R_x. According to the preceding lemma it is true in R_x if one considers sequences g_n on R_x whose euclidean lengths $L(g_n)$ on R_x are uniformly bounded. But if $L(g_n)$ becomes infinite with n it follows from the positive definiteness of F that J_{g_n} becomes infinite with n. In this special case the relation

$$\liminf. J_{g_n} > J_g$$

is trivial. The theorem follows.

It would be awkward to restate Lemma 19.3 in a form applicable to M rather than to the coordinate region R_x without further hypothesis on F, or without using some intrinsically defined metric on M. The following theorem (which will not be used) makes a natural extension without assuming F positive definite.

Theorem 19.2. If $F(x, r)$ is quasi-regular on M while $f(x, \dot{x})$(with integral I) is positive definite and positive regular on M, then J_g is lower semi-continuous on the class of curves on M whose I-length are at most a constant K.

§20. Integral J-length J_g and Peano[§] J-length $J(g)$. We termed a curve g rectifiable if any subarc of g in a coordinate region R_x is rectifiable in R_x. The value of the integral J along such a rectifiable arc has been denoted

§ Peano J-length is termed abstract length by Morse, Functional topology and abstract variational theory, Mémorial des sciences mathématiques, Paris (1939), p.52.

by J_g. In euclidean space Peano (and Archimedes) defined the length $L(g)$ as the least upper bound of lengths of polygons inscribed in g. The Peano J-length $J(g)$ which we shall now define is a generalization of the Peano length $L(g)$. For this purpose we assume that $F(x, r)$ is positive definite and positive regular.

Let g^t be any p-curve joining a point A to a point B. Let $(P) = (P_o, P_1, \ldots, P_n)$ be a set of successive points on g^t. The set (P) will be termed a <u>partition</u> of g^t of <u>norm</u> d if the maximum of the J-diameters of the arcs P_iP_{i+1} is less than d. We term

$$S(P) = \sum J(P_i, P_{i+1}) \qquad (i = 0, 1, \ldots, n-1) \tag{20.1.}$$

a <u>sum approximating</u> $J(g^t)$, and define $J(g^t)$ as the least upper bound of such sums for all partitions (P) of g^t. We admit the possibility that $J(g^t)$ may be infinite. It appears that $J(g^t)$ is the same for all p-curves in a curve class g. We accordingly write

$$J(g) = J(g^t)$$

The J-length so defined will be termed the Peano J-length of g.

If the vertices (P) of a partition of g^t are a subsequence of the vertices (Q) of a second partition (P) of g^t we term (Q) a <u>subdivision</u> of (P) on g^t. It is clear that the sums S(P) never decrease with subdivision of (P) on g^t.

<u>Theorem 20.1.</u> <u>The Peano length</u> $J(g)$ (<u>possibly infinite</u>) <u>is the limit of any sequence</u> S_n <u>of sums approximating</u> $J(g)$, <u>provided the norm of these partitions tends to zero with</u> $1/n$.

Let a be any constant, $a < J(g)$. We shall show that there exists a positive number d so small that

(20.2) $$S(P) > a$$

for any partition (P) of g whose norm is at most d.

Let b be any constant with $a < b < J(g)$. By definition of J(g) there exists a partition $(Q) = (Q_o, \ldots, Q_n)$ of g such that $S(Q) > b$. Let d be chosen so that

(20.3) $$0 < nd < b - a$$

Let (P) be a partition of g^t of norm at most d. Let each point Q_i of (Q) be replaced by the first point of the set (P) which follows or coincides with Q_i on g^t, ordering points by their parameter values t. Each point Q_i will thereby suffer a J-displacement at most d. The partition (Q) will thereby be replaced by a partition (Q') again with n+1 points. It follows from the triangle axiom that

$$J(Q'_i, Q'_{i+1}) \geqq J(Q_i, Q_{i+1}) - 2d \qquad (i = 0,\ldots,n-1),$$

whence

(20.4) $$S(Q') \geqq S(Q) - 2dn.$$

But each point Q'_i is a point of (P) whence

(20.5) $$S(P) \geqq S(Q') .$$

The relations (20.3), (20.4) and (20.5) yield the inequality

$$S(P) \geqq S(Q) - (b - a)$$

and (20.2) follows, since $S(Q) > b$.

The preceding theorem needs completion as follows.

Theorem 20.2. *Any arc with a finite Peano J-length is rectifiable in any coordinate region in which it lies.*

Let g be any subarc of the given arc which lies in a single coordinate system R_x. We must prove that g is rectifiable in R_x.

Without loss of generality we can suppose that $\overline{R}_x$ is compact and itself interior to a coordinate region. It follows that the values of $F(x, r)$ for x on $\overline{R}_x$ and for directions (r) with $r^i r^i = 1$, have a positive minimum K. Let E be an extremal arc in R_x with euclidean length L(E). On using the integral for J one sees that

$$J(E) \geqq K\, L(E)$$

A straight line joining the end points of E would then have a euclidean length d(E) such that $J(E) \geqq K\, d(E)$. Let g be approximated by a broken extremal composed of a sequence of elementary extremals E_i whose successive end points form a partition (P) of g of norm not greater than d. The straight arcs which subtend the arcs E_i will define a polygon with vertices on g. On letting the norm d tend to zero one infers that

$$J(g) \geqq K\, L(g)$$

Thus L(g) is finite and g rectifiable.

<u>Theorem 20.3.</u> <u>For a rectifiable arc</u> g <u>the Peano length</u> J(g) <u>and the integral length</u> J_g <u>are equal.</u>

If g_n is a broken extremal with a finite number of vertices then

(20.6) $$J(g_n) = J_{g_n} .$$

If g_n has vertices given by a partition (Q) of g and is composed of elementary extremals (of minimizing type)

(20.7) $$J_{g_n} \leqq J_g .$$

With the aid of (20.6), (20.7) yields the relation

$$J(g_n) \leqq J_g \tag{20.8}$$

Suppose that the norm of (Q) tends to zero as n becomes infinite. Then g_n tends to g and

$$\lim J(g_n) = J(g) \tag{20.9}$$

so that (20.8) and (20.9) imply that

$$J_g \geqq J(g) . \tag{20.10}$$

From (20.9) and (20.6) we see that $\lim J_{g_n}$ exists, while from the lower semi-continuity of the integral J-length we infer that

$$\lim J_{g_n} \geqq J_g . \tag{20.11}$$

Thus (20.11) and (20.9) yield the result

$$J(g) \geqq J_g$$

which taken with (20.10) implies that $J(g) = J_g$. This completes the proof.

Corollary. The integral J-length J_g exists if and only if the Peano J-length $J(g)$ is finite, and in that case $J(g) = J_g$.

§21. The compactness of $W_c(A, B)$. Let A and B be two points of M and let W(A, B) be the Fréchet space of sensed curves joining A to B. We shall prove that the subset W_c of W(A, B) for which $J(g) \leqq c$ is compact. Here c is any finite constant and we are assuming that F is positive definite and positive regular.

This compactness of W_c is a generalization of a theorem of Hilbert on arcs of bounded length joining two fixed points in euclidean m-space. There are two additional difficulties in the case at hand: (1) We are here concerned not with one but with two metrics, defined by J-distances $J(P, Q)$ and primary distances PQ respectively; the J-distances enter into the definition of J-length and the primary distances into the definition of the Fréchet distance gg'. (2) The application of the classical Ascoli lemma to prove the theorem requires care since the curves of $W(A, B)$ do not in general lie in only one coordinate system; a method of proof independent of coordinate systems is to be preferred.[§]

The first two lemmas concern the relation of J-distances $J(P, Q)$ to primary distances PQ.

Lemma 21.1. The distance PQ is less than a prescribed positive constant e whenever $J(P, Q)$ is less than a suitably chosen positive constant d dependent upon e but not upon P and Q.

If the lemma were false there would exist an infinite sequence of pairs of points P_n, Q_n such that $J(P_n, Q_n)$ tends to zero with $1/n$, while P_nQ_n is bounded from 0 for all n. Let P^o, Q^o be a cluster pair of the pairs P_n, Q_n. We see that $J(P^o, Q^o) = 0$, since $J(P, Q)$ is a continuous function of the pair (P, Q). Hence $P^o = Q^o$ so that $P^oQ^o = 0$. The distance P_nQ_n cannot then be bounded from zero. From this contradiction we infer the truth of the lemma.

Definition of n-sets on g. If g is an arc of finite J-length an ordered set of $n+1$ successive points on g, including the end points of g and dividing g into n-successive arcs of equal J-length will be called an n-set on g.

Lemma 21.2. The arcs of $W(A, B)$ on which J is at most a finite constant c are divided by n-sets into subarcs whose diameters (using the primary metric) tend to zero uniformly with $1/n$.

---

§ See Morse, op. cit., ("Mémorial"), p.54.

Beside the primary diameter referred to in the lemma we introduce the J-diameter of a set as the least upper bound of J-distances between points of the set. Observe that the J-diameter of an arc is at most its J-length. If $J(g) \leqq c$ the J-diameters of the arcs h_i into which g is divided by an n-set are at most c/n, and so tend to zero uniformly with $1/n$. It follows from the preceding lemma that the (primary) diameters of the arcs h_i tend to zero uniformly with $1/n$. This completes the proof.

The principal theorem follows.

Theorem 21.1. The set of curves of $W(A, B)$ whose J-lengths are at most a finite constant c form a compact subset W_c of $W(A, B)$.

Let H be an infinite sequence of curves of W_c. Because of the compactness of M there will exist a subsequence H_1 of H such that the 2-sets on arcs of H_1 converge to a set of three points on M. Proceeding inductively we see that there will exist a sequence $H_1, H_2, \ldots$ of subsequences of H such that H_m is a subsequence of H_{m-1} and the 2^m-sets on the curves of H_m converge to a set of points

$$P_m^0,\ P_m^1,\ \ldots,\ P_m^{2m}$$

on M. We shall define a p-curve $P = P(t)$ on M with $0 \leqq t \leqq 1$. For each $m > 0$ set

$$P[r/2^m] = P_m^r \qquad (r = 0,1,\ldots,2^m) \tag{21.1}$$

observing that the definitions of (21.1) are consistent for successive values of m. If $t' \geqq t''$ are any two values of t for which $P(t)$ is defined, we see that

$$J[P(t'), P(t'')] \leqq c\,|t' - t''| \tag{21.2}$$

<u>Completion of definition of</u> $P(t)$. Let t^* be an arbitrary value of t on the interval $(0, 1)$, and let t_n be an infinite sequence of values of t for which $P(t)$ is defined and which tend to t^* as n tends to ∞. It follows from (21.2) that the points $P(t_n)$ form a Cauchy sequence relative to the J-metric and hence (by Lemma 21.1) also relative to the primary metric. The points $P(t_n)$ converge on M to a point Q independent of the sequence t_n converging to t^*. We set $P(t^*) = Q$, and observe that (21.2) then holds for all values of t' and t'' on $[0, 1]$. Let g be the curve defined by $P = P(t)$.

<u>Convergence of a subsequence of curves of</u> H <u>to</u> g. Let (e_k) be a sequence of positive constants tending to zero as k becomes infinite. Corresponding to e_k Lemma 21.2 implies the existence of an integer $m(k)$ so large that each of the arcs into which a curve p of W_c is divided by its $2^{m(k)}$-set has a diameter at most e_k. We suppose $m(k)$ also so large that the arcs of $P(t)$ for which

$$\frac{r-1}{2^{m(k)}} \leqq t \leqq \frac{r}{2^{m(k)}} \qquad (r = 1,\ldots,2^m) \tag{21.3}$$

have diameters at most e_k. With $m(k)$ so chosen let p_k be a curve of $H_{m(k)}$ such that the respective points of the $2^{m(k)}$-set of p_k are at distances at most e_k from the corresponding points $P^r_{m(k)}$ on g. If p^r_k is the r-th of the arcs into which the $2^{m(k)}$-set divides p_k, and g^r is the arc of g for which (21.3) holds we see that

$$p^r_k g^r \leqq 3e_k, \quad p_r g \leqq 3e_k .$$

The sequence p_r thus converges to g in the sense of Fréchet.

It follows from (21.2) that the J-length of g is at most e; and the proof of the theorem is complete.

Corollary 21.1. The Peano J-length $J(g)$ is lower semi-continuous without any restriction as to the finiteness of $J(g)$ or the rectifiability of g.

The corollary affirms that when g_n converges to g as n becomes infinite then

$$\text{lim inf. } J(g_n) \geqq J(g)$$

This is true when $J(g)$ is finite since $J(g)$ then equals the integral J-length J_g, known to be lower semi-continuous. It is also true when $J(g)$ is infinite; for no subsequence of $J[g_n]$ can then be bounded by a finite constant K unless $J(g) \leqq K$, in accordance with the preceding theorem.

§22. The space $W(A, B)$. The space $W(A, B)$ consists of the set of curves, p, g, etc. joining a fixed point A to a fixed point B on M. We admit the possibility that $A = B$ and put no restrictions on the curves of $W(A, B)$ as to rectifiability. With the aid of a positive definite, positive regular integrand $F(x, \dot{x})$, a J-length $J(g)$ has been attached to each curve of W. $J(g)$ is finite when g is locally rectifiable, otherwise infinite; moreover $J(g) \geqq 0$, with $J(g) = 0$ if and only if g reduces to the point $A = B$.

With the aid of $J(g)$ we have attached a primary metric to W with distances pg which are never negative, null if and only if $p = g$, and symmetric ($pg = gp$). The distances pq satisfy the triangle axiom. In terms of this metric $J(g)$ is lower semi-continuous even when infinite; that is, when $g_n g$ tends to zero with $1/n$,

$$\text{Lim inf. } J(g_n) \geqq J(g).$$

For each finite c let W_c denote the subspace of W on which $J(g) \leq c$, and W_{c-} the subspace on which $J(g) < c$. In the last section we have shown that W_c is compact. We continue with other general properties of the space W.

Lemma 22.1. No element of W has a neighborhood with compact closure.

Suppose the lemma were false. Then there would exist an element g in W and a positive constant e such that the set $\overline{N}$ of points, p of W, for which $pg \leq e$, would be a compact subset of W.

Let h be an arc of diameter e on M, with B an initial point. Let h^{-1} be h reversed in sense. Let g_n be the arc§

$$g_n = g(h\, h^{-1})^n \qquad (n = 1,2,\ldots)$$

That $gg_n \leq e$ is readily seen on parameterizing g so that B on g is represented by a parameter interval, (not a point) and observing that the distance of points on hh^{-1} from B is at most e. Since $\overline{N}$ is compact the arcs p on $\overline{N}$ can be given a parameterization on M

$$P = P(\sigma, p) \qquad (0 \leq \sigma \leq 1, \text{ p on } \overline{N}) \tag{22.1}$$

in terms of reduced μ-length σ in which $P(\sigma, p)$ is uniformly continuous. In particular there exists a positive constant d so small that a subarc of p on which $|\triangle\sigma| < d$ has a diameter less than e. Each arc g_n is represented by (22.1). On g_n there is a sequence of 2n consecutive subarcs of diameter e.

§ If $h_1, \ldots, h_n$ are arcs on M in which the terminal point of h_i coincides with the initial point of h_{i+1}, then $h_1 h_2 \ldots h_n$ shall represent an arc in which h_1 is an initial subarc and h_n a terminal subarc, and on which the arcs h_i appear as successive subarcs. If the arcs h_i are identical we write $h_1 h_2 \ldots h_n = (h_1)^n$.

Since σ is restricted to an interval $[0, 1]$ it follows that $2n < 1/d$, so that not all arcs g_n are represented by (22.1). The lemma follows from this contradiction.

<u>The hypothesis of non-degeneracy</u>. We shall assume that each extremal arc g which joins A to B is non-degenerate, in the sense that B is not conjugate to A on g. It is known (Morse, <u>op. cit.</u>, "Colloquium Lectures", p.233) that this hypothesis is fulfilled by "almost all" pairs of points A, B on M neighboring a given pair A_o, B_o on M. It will be shown that when g is a non-degenerate extremal arc of W, then g is isolated among extremal arcs of W, in the sense that there are no extremal arcs on W, other than g, in a sufficiently small Fréchet neighborhood of g. For this purpose the following lemma is needed.

<u>Lemma 22.2</u>. <u>If a sequence g_n of extremal arcs joining A to B converges in the sense of Fréchet to an extremal arc g, then the initial direction of g_n converges to that of g, and $J(g_n)$ converges to $J(g)$.</u>

(This lemma would be false if stated for a sequence of regular arcs g_n of class C'' joining A to B and converging to g).

Set $J(g) = b$. Let s be J-length measured along the extremals g_n and g from A. Extending g_n as an extremal G_n, if necessary, let p_n be the initial subarc of G_n of J-length b. The arcs g_n are given as converging to g in the sense of Fréchet. Let each arc g, g_n and p_n be represented in terms of μ-length. A point on g_n at which $\mu < \mu(g)$ (where $\mu(g)$ is the total μ-length of g) will converge to the point of g with this same parameter μ as n becomes infinite; from this we see that the initial direction of g_n must converge to that of g as n becomes infinite. With this established it follows from the dependence of extremals issuing from A on their initial direction and parameter s, that

p_n converges to g in the sense of Fréchet. Hence

$$(22.2) \qquad \lim \mu(g_n) = \lim \mu(p_n) = \mu(g) .$$

If g_n extends beyond p_n let h_n be the subarc of g_n in excess of p_n; if p_n extends beyond g_n let h_n be the subarc of p_n in excess of g_n. If $p_n = g_n$, let h_n be the point B. In any case the representation of g_n and p_n on M in terms of μ-length includes a representation of h_n on which the variation of μ is the number

$$(22.3) \qquad | \mu(g_n) - \mu(p_n) |$$

which tends to zero with $1/n$. Since the arcs g, g_n, p_n, $(n = 1,2,\ldots)$ form a compact subset Z of W, the point $Q(\mu, p)$ on any one of these arcs with the parameter μ, is a uniformly continuous function of μ and of the arc p on Z. Hence $Q(\mu, p)$ is an equi-continuous point function of μ for p on Z, and the diameters of the arcs h_n must tend to zero with the difference (22.3). Since the J-length of p_n is constantly b, it follows that the J-length of g_n must converge to b; and the proof of the lemma is complete.

We continue with the following lemma.

<u>Lemma 22.3.</u> <u>If</u> g <u>is a non-degenerate extremal of</u> W(A, B) <u>there is no other extremal in</u> W(A, B) <u>on a sufficiently small Fréchet neighborhood of</u> g.

By virtue of the preceding lemma the theorem would be false only if there exists a sequence of extremal arcs g_n of W with initial directions converging to that of g and with J-lengths converging to $J(g)$. As seen in §14, if B is not conjugate to A on g, the extremals issuing from A with directions sufficiently near that of g form a field near B, or more precisely their subarcs on which

$$J(g) - e < s < J(g) + e$$

will form such a field if e is sufficiently small and positive. It follows that the terminal point of g_n cannot coincide with B if n is sufficiently large, From this contradiction we infer the truth of the lemma.

Lemma 22.4. The set of extremals on W_c form a closed subset of W_c.

If possible, let g be an arc on W_c which is the limit in the sense of Fréchet of a sequence g_n of extremal arcs of W_c. Let η_n be the initial direction§ of g_n and s_n its J-length. The sequence (η_n, s_n) has at least one limit pair (η^o, s^o). Let g^o be the extremal arc with initial direction η^o at A and J-length s^o. It is clear that g^o belongs to W_c and that g_n converges to g^o. Hence $g^o = g$ and W_c is closed.

From the isolated character of non-degenerate extremal arcs, the compactness of the sets W_c, and the closed character of the set of all extremal arcs on W_c we infer the following major result.

Theorem 22.1. The extremal arcs of W_c, if non-degenerate, are finite in number.

If each extremal arc of W is non-degenerate regardless of J-length the total number of such extremals is countable.

These results are illustrated by the sphere, taking J as ordinary lengths. If A and B are not opposite points on the sphere, B is never conjugate to A on any extremal arc joining A to B. If b is the minimum distance from A to B there are extremal arcs joining A to B with each of the lengths

$$b, \quad 2n\pi \pm b \qquad (n = 1,2,\ldots)$$

§ Represented by a contravariant vector on the F-indicatrix at A.

§23. The deformation $\triangle_c$ of W_c. A point w of W(A, B) which represents an extremal arc on M will be called a critical point of J(g) on W(A, B) and the value J(w) will be called a critical value.

In this section we shall establish a basic theorem which is a generalization of a theorem on a function $f(x_1, \ldots, x_n)$ of n-variables of class C″ throughout cartesian n-space. The trajectories orthogonal to the level manifolds f = const. are given by the differential equations

$$\frac{dx_i}{dt} = -f_{x_i}(x_1, \ldots, x_n) \tag{23.1}$$

where the -sign is added in order that when t increases, f is stationary or decreasing along a trajectory in accordance with the relation

$$\frac{df}{dt} = -f_{x_i} f_{x_i} \qquad (i = 1,\ldots,n)$$

At a critical point (x_0) of f the level manifold has no orthogonal trajectory in a strict sense. Nevertheless the solution of (23.1) for which $(x) = (x_0)$ is well defined and reduces to a point. Let

$$x_i = X_i(t, a) \qquad (i = 1,\ldots,n)$$

be the solution of (23.1) which reduces to (a) when t = 0. We can define an f-deformation X by requiring that each point (x) which coincides with (a) when t = 0, shall be replaced by the point [X(t, a)] when t > 0. This deformation has the fundamental property:

(a). Under the f-deformation X, each ordinary (non-critical) point of f is deformed into a point at a lower f-level.

We shall not be able to generalize this result for J over all of W. We shall, however, establish the existence of a deformation $\triangle_c$ of W_c which lowers the J-level of every ordinary point of W_c to which it is applied.

This holds for each finite constant c. The definition of $\triangle_c$ will depend on c. The amount by which the J-level of a point p is lowered by $\triangle_c$ will depend upon p, and, as in the case of the deformation X, will not in general be bounded from zero.

A special convention is needed. In case A = B the point A = B will be regarded as a null-extremal on M and as a critical point of J on W.

In §18 we have defined a deformation D_m of a compact subset of W (such as W_c). In defining D_m use was made of a universal field radius ρ. A given arc g on M was divided into m successive arcs k_i of equal variation of reduced μ-length σ, measuring σ from the initial point of g. The values of σ at the points of division of the interval $0 \leq \sigma \leq 1$ were

$$\sigma_j = \frac{j}{m} \qquad (j = 1,\ldots,m-1)$$

The deformation D_m was defined for g whenever the J-diameter of each arc k_i was less than ρ. The final image E_m of g under D_m was the broken extremal whose successive elementary extremals subtended the arcs k_i. It can happen that some of the elementary extremals of E_m reduce to points, or that $E_m = g$. This leads us to an important statement.

(b) For m sufficiently large the J-deformation D_m is applicable to each arc g of W_c, and lowers the J-level of g except in the case in which g equals its final image E_m under D_m.

In the exceptional case in which $E_m = g$, another deformation is needed which lowers the J-level of E_m, provided E_m is not an extremal.

The deformation§ D_m^*. Such a deformation is defined exactly as was D_m, except that each point σ_j, $(j = 1,\ldots,m-1)$ of division of the interval

§ Morse, Memorial *op. cit.*, p.59. Cf. The deformation $\Theta(r)$.

$0 \leq \sigma \leq 1$ is replaced by a point $\sigma_j^o < \sigma_j$, on the open interval preceding σ_j and so near σ_j, that the resulting arcs k_i^o into which a curve g of W_c is divided still have J-diameters less than ρ. When $E_m = g$, and E_m is not an extremal, there will be some arc k_i^o of E_m which will contain a corner point of E_m, and so not be an extremal. Hence D_m^* will lower the J-length of E_m.

The deformation $\triangle_c$. Set

$$\triangle_c = D_m^* D_m , \tag{23.2}$$

understanding that D_m is applied to the arc g of W_c, and D_m^* to the final images E_m under D_m of the arcs g. The time t can be supposed to run from 0 to 2 but we prefer to introduce a new time parameter $t' = t/2$ which runs from 0 to 1. The principal theorem here follows.

Theorem 23.1. *There is a J-deformation* $\triangle_c$ *of* W_c *with the following properties. If* g *is on* W_c *and* E(g) *is the final image of* g *under* $\triangle_c$, *then* J(E(g)) *is a continuous function of* g *on* W_c, *and*

$$J(g) > J[E(g)] \tag{23.3}$$

whenever g *is a non-critical point of* W_c; *a critical point of* W_c *remains fixed under* $\triangle_c$.

The continuity of the map[§] of $\triangle_c \times I$ into W_c defining the deformation $\triangle_c$ follows from the manner in which the curves of W_c are represented in terms of reduced μ-length (Cf. Theorem 4, Appendix) and the way in which elementary extremals vary with their end points. Cf. Theorems 14.1 and 14.2. The continuity of the function J(E(g)) is a consequence of the fact that the J-length of an elementary extremal is a continuous function of its end points and that

§ I is the unit time interval.

the vertices of E_m and hence of $E(g)$ vary continuously with g.

Proper and weak J-deformations. A continuous deformation of a set A, which replaces a point p in A by a point p^t, $0 \leqq t \leqq 1$, is called a weak J-deformation, if $J(p^t) \leqq J(p)$, and proper if

$$J(p^t) < J(p), \qquad (\text{whenever } p^t \neq p).$$

The terminology is open to the objection that a weak J-deformation is not in general a J-deformation. Understood, this need cause no difficulty.

The deformations explicitly defined up to this point have been proper J-deformations and hence proper and weak.

We shall not make use of the qualification "proper" until we come to the characterization of a homotopic critical point in §27. In the next section we shall prove certain theorems where use is made of the deformation $\triangle_c$. The fact that $\triangle_c$ is a proper weak J-deformation is all that is needed. In extensions of the theory beyond the fixed end point problem it is frequently difficult, if not impossible, to set up the required deformations as J-deformations. Proper, weak J-deformations are however available, and it is these deformations which we shall depend upon from this point on. The word "weak" could be deleted from the next section.

§24. Weak J-deformations into W_b. We are concerned with $J(g)$ for g on $W(A, B)$ where A and B are fixed points of M. We are assuming that every extremal joining A to B is non-degenerate. As has been see in §22, there are then only a finite number of critical points of J at which J is less than a given constant. We shall prove the following theorem.

Theorem 24.1. Let b be a critical value of J and $(w_1,\dots,w_r) = (w)$ the set of critical points of J on W at which $J = b$. If $c > b$ is any constant such that there are no critical values of J on the interval $b < J \leqq c$ there exists a weak J-deformation of W_c into a subset of $(w) + W_{b-}$, holding the set (w) fast.

This theorem will be established with the aid of three lemmas, of which the first follows.

We shall make use of the deformation $\triangle_c$ of W_c, affirmed to exist in §23. If $E(g)$ is the final image of g under $\triangle_c$ it has been seen that $J[E(g)]$ varies continuously with g on W_c while

$$\Theta(g) = J(g) - J[E(g)] \tag{24.1}$$

is lower semi-continuous.

The product deformation $\triangle_c^n$. Here n is a positive integer. The deformation $\triangle_c^n$ begins with an application of $\triangle_c$ to W_c; $\triangle_c$ is next applied to the resultant terminal image thereby defining the deformation $\triangle_c^2$ of W_c. Then $\triangle_c$ is applied to the terminal image of W_c under $\triangle_c^2$ thereby defining $\triangle_c^3$, and so on until $\triangle_c^n$ is defined. Let $E^n(g)$ be the terminal image of g under $\triangle_c^n$. A point p which is given as an image $E^n(g)$ of a point g on W_c will be said to have a deformation index n.

Lemma 24.1. If $[a, c]$ is a closed interval of ordinary values of J, then $\Theta(g)$ is bounded from zero for g on $W_c - W_{a-}$.

If the lemma were false there would exist a sequence of points p_r on $W_c - W_{a-}$ such that $\Theta(p_r)$ converges to 0 with $1/r$. Without loss of generality we can suppose that p_r converges to a point p of W_c. There are two cases to be considered.

<u>Case</u> I. $J(p) < a$: In this case $J[E(p)] < a$ and

$$\lim \theta(p_r) \geqq a - \lim J[E(p_r)] = a - J[E(p)] > 0 \tag{24.2}$$

contrary to the choice of $[p_r]$.

<u>Case</u> II. $J(p) \geqq a$: Since $\theta(g)$ is lower semi-continuous

$$\lim \theta(p_r) \geqq J(p) - J[E(p)] \tag{24.3}$$

In Case II p is ordinary, so that the right member of (24.3) is positive, (cf. Th. 23.1) again contrary to the choice of $[p_r]$.

This completes the proof.

The preceding lemma implies the existence of an integer n such that the terminal image of W_c under $\triangle_c^n$ is below the level a of the lemma. If as in the theorem, [b, c] is an interval on which b is the only critical value, then for g on W_c

$$\lim_{n=\infty} \max_g J[E^n(g)] = b \qquad [g \text{ in } W_c] \tag{24.4}$$

The limit cannot exceed b by virtue of the preceding lemma, and the value b is taken on by the critical points w_i at the level b. The points w_i remain fixed under $\triangle_c^n$.

<u>Lemma 24.2.</u> <u>Let</u> [b, c] <u>be an interval in which</u> b <u>is the only critical value, let</u> (w) <u>be the critical set at the level</u> b <u>and</u> N <u>a neighborhood of</u> (w) <u>relative to</u> W. <u>If</u> n <u>is sufficiently large</u> $\triangle_c^n$ <u>deforms</u> W_c <u>into a subset of</u> $N + W_{b-}$.

If the lemma were false there would exist a sequence $[p_r]$ of points p_r on W_c but not below b, with deformation indices $n(r) > 1$ which become infinite with r, while

$$\lim J(p_r) = b , \tag{24.5}$$

with no p_r on N. This statement requires use of (24.4). Let q_r be a point such that $E(q_r) = p_r$. Without loss of generality we can suppose that p_r and q_r converge respectively to points p and q on W_c. There are three cases to be considered.

<u>Case</u> I. $J(q) < b$. Here $J(E(q)) < b$ and $\lim J(p_r) = \lim J(E(q_r)) = J(E(q)) < b$ contrary to (24.5).

<u>Case</u> II. $J(q) = b$. The point p is not a point w_i. The point q is not a point w_i; otherwise q_r would converge to w_i with $1/r$; then $E(q_r) = p_r$ would likewise converge to w_i with $1/r$, since $E(w_i) = w_i$ and $E(g)$ is continuous. This is impossible if p is not a point w_i. Thus q is ordinary.

On making use of the lower semi-continuity of $\Theta(g)$, and of the fact that $\triangle_c$ lowers the J-level of an ordinary point q we have

$$\text{(24.6)} \qquad \lim\inf. \ \Theta(q_r) \geqq \Theta(q) > 0.$$

On using the relation (24.5) and the relation $\lim J(q_r) = b$ (derived from (24.4) and the equality $J(q) = b$, using the lower semi-continuity of J) we see that

$$\text{(24.7)} \qquad \lim \Theta(q_r) = \lim J(Q_r) - \lim J(p_r) = b - b = 0.$$

The relations (24.6) and (24.7) are in contradiction, so that Case II is impossible.

<u>Case</u> III. $J(q) > b$. On account of (24.4), and the relation $b = \lim\inf. \ J(q_r) \geqq J(q)$, this cannot occur.

The proof of the lemma is complete.

<u>Extending a</u> J-<u>deformation locally defined to all of</u> W. We need to modify and extend the deformation affirmed to exist in Theorem 18.1 in accordance

with the following lemma. In this lemma we shall refer to the e-neighborhood of a point p of W, meaning thereby the set of all points g of W for which $pg < e$.

Lemma 24.3. *Corresponding to a J-deformation U of the closure $\overline{N}(e)$ of an e-neighborhood N(e) of a point p of W, there exists a J-deformation V of W for which points initially on a d-neighborhood N(d) of p, $(0 < d < e)$ are deformed as in U while points initially on W - N(e) are held fast.*

In the deformation U suppose that the time t increases from 0 to 1. Under V points g initially on N(d) or W - N(e) shall be deformed as stated in the lemma. Let g be a point such that

$$d \leqq gp \leqq e .$$

The point g shall have the same image g^t as under U until the time reaches t_g, where $1 - t_g$ divides the interval [0, 1] in the ratio in which gp divides [d, e]. For $t > t_g$, g^t shall remain fixed. In particular when $gp = d$, $t_g = 1$, and g is deformed as under U for $0 \leqq t \leqq 1$. Where $gp = e$, $t_g = 0$ and g is held fast for all t. The resulting deformation V is continuous, taking account of the continuity of U over $\overline{N}(e)$. The deformation V is clearly a J-deformation; and the proof of the lemma is complete.

Proof of Theorem 24.1. Theorem 18.1 was stated for locally rectifiable arcs. In reality, with Peano J-length well defined, it can be stated with no essential change in the proof for arbitrary arcs on W. Applying Theorem 18.1 so extended to the critical points w_i of Theorem 24.1 we can state the following. If e is a sufficiently small positive constant there exists a weak J-deformation U_i of an e-neighborhood $N(e, w_i)$ of w_i which holds w_i fast, and carries

$N(e, w_i)$ into a subset[§] of $w_i + W_{b-}$. We suppose e so small that the neighborhoods $N(e, w_i)$ $(i = 1,\ldots,r)$ are disjoint for different values of i. In accordance with the preceding lemma there exists a weak J-deformation V_i of W which deforms points initially on $N(d, w_i)$, for some positive $d < e$, as does U_i while points initially on $W - N(e, w_i)$ are held fast.

By virtue of Lemma 24.2, the deformation $\triangle_c^n$ for n sufficiently large will carry W_b into a subset of

$$N(d, w_1) + \ldots + N(d, w_r) + W_{b-}$$

For such an n the product deformation

$$V_r \ldots V_1 \triangle_c^n$$

will carry W_c into a subset of $(w) + W_{b-}$.

This completes the proof of Theorem 24.1.

§25. The space $W(A, B)$ and its components. A first theorem is as follows.

Theorem 25.1. If (A, B) is a pair of distinct points on the coordinate manifold M and (A', B') a second such pair, $W(A, B)$ is homeomorphic with $W(A', B')$.

There exists a homeomorphism of M with itself in which A corresponds to A and B corresponds to any point B_1 in a sufficiently small spherical

§ Theorem 18.1 is stated for the case of extremals with a positive index k. In case $k = 0$ the extremal affords an isolated relative minimum to J. In this case the above deformation U_i can be affirmed to carry $N(e, w_i)$ into w_i. The proof of Theorem 18.1 requires a trivial modification at the end to establish this extension.

neighborhood N of B. For $\overline{N}$ can be mapped homeomorphically onto itself in such a manner that the boundary points of N correspond to themselves and B corresponds to B_1. Since M is connected and compact a finite sequence of homeomorphisms of the above sort (including homeomorphisms that move A as well as B) will suffice to map M topologically onto itself in such a fashion that A corresponds to A', and B to B'.

The preceding proof fails in case A = B. The following theorem, although different, is adequate for our purposes, and admits the case A = B.

Theorem 25.2. Let (A_1, B_1) and (A_2, B_2) be arbitrary pairs of points on M including the cases in which $A_1 = B_1$ or $A_2 = B_2$. For the sets

$$W_1 = W(A_1, B_1) \qquad W_2 = W(A_2, B_2)$$

the homology groups

$$H_1 = H^k(W_1, G) \qquad H_2 = H^k(W_2, G)$$

are isomorphic. $(k = 0, 1, \ldots)$.

We shall presently define a mapping S of W_1 into W_2, and a mapping R of W_2 into W_1, such that RS and SR are deformable on W_1 and W_2 respectively into the identity I. To say that RS is deformable into the identity on W_1, is to say that there exists a deformation D of W_1, with a deformation parameter t $(0 \leqq t \leqq 1)$, such that the initial mapping of W_1 into W_1 defined by D is I, and the final mapping is RS. It follows from Theorem 7.1 that any cycle z of W_1 is homologous to its final image $\overline{D}z$ under D. If S^* and R^* are the homomorphisms of H_1 into H_2, and of H_2 into H_1, induced respectively by S and R in accordance with Lemma 9.3, then

$$R^*S^* = I, \qquad S^*R^* = I.$$

It follows that R^* and S^* are isomorphisms.

To define S and R let a and b be arcs on M leading respectively from A_2 to A_1 and from B_1 to B_2. Under S a curve g_1 of W_1 shall be replaced by $a\, g_1 b$, while under R a curve g_2 of W_2 shall be replaced by $a^{-1} g_2 b^{-1}$. Under RS therefore g_1 is replaced by

$$g = a^{-1} a\, g_1 b\, b^{-1} .$$

The curve g is deformable into g_1 holding A_1 and B_1 fast, by shrinking $a^{-1}a$ on a into A_1, and bb^{-1} on b to B_1. Thus RS is deformable into the identity. Similarly SR is deformable into the identity. The theorem follows.

The following theorem is needed.

<u>Theorem 25.3.</u> <u>The deformation classes of curves of</u> W(A, B) <u>are finite in number or countably infinite.</u>

The theorem will be established if one can exhibit a countably infinite subset Z of curves of W = W(A, B), such that any curve of W can be deformed on W into some curve of Z.

To that end let (P_n) be a countably infinite, everywhere dense set of points on M. Such a set is readily obtained as the union of the points with rational coordinates on each of a finite set of coordinate regions covering M. Let ρ be a field radius for M. For each integer m let (E_m) be the set of broken extremals of W whose component elementary arcs have J-lengths less than ρ, and let (E) be the union of the sets (E_m) for m = 1, 2, ... Let (E') be the subset of (E) consisting of broken extremals whose vertices are confined to points of (P_n). Any curve of W is in the deformation class of some curve of (E) and hence in the deformation class of a nearby curve of (E'). The theorem follows.

Each deformation class of W is a component of W consisting of points (curves) of W arc-wise connected on W. If the coordinate manifold M is a sphere, W has but one component. If M is the torus, W has a countably infinite set of components. In each component there is at least one extremal joining A to B which gives an absolute minimum to J in that component.

Recall that a deformation of a subset X of W which replaces a point p of X by a point p^t, $0 \leq t \leq 1$, is termed a weak§ J-deformation if $J(p^t) \leq J(p)$, and a proper J-deformation if $J(p^t) < J(p)$ whenever $p^t \neq p$.

The reducibility of W at infinity. We shall say that W is reducible at infinity if corresponding to any compact subset X of W there exists a weak J-deformation D^X which carries X into some subset W_c of W. Given X, the deformation D_m of §18 will serve as D^X, provided the integer m is so large that the definition of D_m over X is possible.

In general it will be seen that there exists no one J-deformation of W as a whole into a set W_c. In fact, if a is a given finite number, in general cases (the n-sphere for example) there will exist compact subsets X of W which admit no weak J-deformation into W_a. This will become clear in the next sections.

§26. A positive, non-degenerate, lower semi-continuous function F on a metric space S.

The preceding study of J on $W(A, B)$ and similar studies of other boundary value problems in the large make possible an objectively useful choice of axioms suitable for a theory of a positive non-degenerate, lower semi-continuous

§ A weak J-deformation is not necessarily a J-deformation. The term J-deformation alone could be dropped.

function F on a metric space S. These terms will be defined in a new topological setting . Here S will replace a component of W(A, B) and F will replace J. Several defintions are necessary.

An F-neighborhood of a point g of S. Given a metric space S and a single-valued positive function F defined over S, an F-neighborhood of a point g of S at which F(g) is a finite number c, will be defined as any set $N \cap S_{c+e}$ where N is an ordinary neighborhood of g, e an arbitrary positive constant, and S_{c+e} is the set of points p of S at which $F(p) \leqq c+e$.

The essential topological properties of non-degenerate critical points as extremals motivate the next definition.

Points g with Property C. A point g of S at which F(g) is finite will be said to have Property C if there exists a proper, weak F-deformation D_g $(0 \leqq t \leqq 1)$ of some neighborhood N(g) of (g) such that:

(1) D_g leaves g invariant and deforms N(g) into a (topological)r-disc K(g) $(r \geqq 0)$ which contains g as an interior point and is below F(g) except for g when $r > 0$, and which reduces to g when $r = 0$.

(2) The terminal mapping of D_g, restricted to $K(g) \cap N(g)$, is deformable into the identity on K(g), holding g fast and deforming K(g) - g on K(g) - g.

When $r = 0$, a point g with Property C affords a proper relative minimum to F. For in deforming N(g) into g each point of N(g) - g is displaced and so lowered in F-value since D_g is proper. The integer r is termed the index of a point g with Property C.

The following are the axioms from which all properties of F and S will be derived. In accordance with earlier notation S_{c-} shall denote the subset of S below c, that is the set of points p for which $F(p) < c$.

Axiom I. The metric space S shall be arc-wise connected, the function F positive, and the sets S_c compact for each finite c.

Axiom II. F shall be reducible at infinity§ in the sense that each compact subset of S shall admit a weak F-deformation into some subset S_c of S.

Axiom III. There shall exist a set (η) of points η with the following properties:

(1) Each point η shall have Property C.

(2) The number of points η below any given constant c shall be finite.

(3) Corresponding to any finite constant c there exists a proper weak F-deformation $\triangle_c$ of S_c in which the points of (η) in S_c are invariant and all other points of S_c are displaced, while the value of F(E(p)) at the terminal image E(p) of p in S_c varies continuously with p in S_c.

The above axioms have been seen to be satisfied by the integral J defined on a connected component of W(A, B), provided $A \neq B$, and the points (η) are interpreted as the extremals joining A to B, provided also that A and B have been chosen so that each of these extremals is non-degenerate. The integrand of J is supposed of the positive, positive regular type.

A first theorem follows.

Theorem 26.1. Axiom I implies that F is lower semi-continuous.

In particular let p_n be a sequence of points in S which converge to a point p. We must show that

$$\text{lim. inf. } F(p_n) \geqq F(p) \tag{26.1}$$

§ Axiom II conditions all sets on which F is unbounded, even if F is never infinite.

Let $c = \lim. \inf. F(p_n)$. If $c = \infty$ (26.1) holds. If c is finite let e be any positive constant. There will exist a subsequence $[q_r]$ of $[p_n]$ such that $[q_r]$ is in S_{c+e}. The set S_{c+e} is closed, so that the limit point p is in S_{c+e}. That is $F(p) \leq c+e$. Since e is arbitrary $F(p) \leq c$ and the proof of the theorem is complete.

Axiom II implies that every k-cycle in S is homologous to a k-cycle on some set S_c, and if a k-cycle x on some S_c bounds in S it bounds on some S_a.

The proof of the following theorem is formally the same as the proof of Theorem 24.1.

Theorem 26.2. Let b be a value of F at one of the points (η), and $[\eta]_b$ the finite subset of points of (η) at which $F(\eta) \leq b$. If there are no values $F(\eta)$ on the interval $b < F \leq d$ there exists a proper weak F-deformation of S_d into a subset of $[\eta]_b + S_{b-}$ holding the set $[\eta]_b$ fast.

Theorem 26.3. Let $b < F < c$ be an interval on which there are no values $F(\eta)$. The respective homology groups of S_{c-} and of $Y = S_{b-} + [\eta]_b$ are isomorphic, with each homology class U in S_{c-} corresponding to the subclass of U in Y.

This theorem is a consequence of Lemma 10.1 of the Appendix once the following statements have been established.

(a) Each cycle x in S_{c-} is homologous in S_{c-} to a cycle in Y.

(b) A cycle z in Y which is bounding in S_{c-} is bounding in Y.

Proof of (a). Let $\triangle_c$ be the deformation in Axiom III. Then

$$\partial \,\bar{\bar{\triangle}}_c x = \bar{\triangle}_x x - x \qquad \text{(in } S_{c-}\text{)}$$

(Cf. Theorem 7.1, Appendix). The value of F at the terminal image E(p) of p is a continuous function of p, and hence on the compact carrier of x at most a constant $d < c$. We may suppose $b < d < c$. Set $x' = \bar{\triangle}_c x$. Since x' is in S_d we may apply the deformation D of S_d of Theorem 26.2 to x' with

$$\partial \bar{\bar{D}} x' = \bar{D} x' - x' \qquad \text{(in } S_d)$$

Hence x is homologous in S_{c-} to $\bar{D}x'$ in Y. Thus (a) holds.

Proof of (b). Suppose that $z = \partial w$ where z is a k-cycle in Y and w a (k+1)-chain in S_{c-}. As in (a), $\bar{\triangle}_c z$ and $\bar{\triangle}_c w$ are on some set S_d with $b < d < c$. On setting $z' = \bar{\triangle}_c z$ and $w' = \bar{\triangle}_c w$ and using the deformation D of Theorem 26.2,

$$\bar{D}z' = \partial\bar{\bar{D}}w' , \quad \text{or} \quad \bar{D}z' \sim 0 , \text{ (in Y)}$$

But z in Y is deformed in Y under $\triangle_c$ into z', and z' is deformed in Y under D into $\bar{D}z'$, so that z bounds in Y with $\bar{D}z'$. Thus (b) holds.

The theorem follows from Lemma 10.1 of the Appendix.

A function F satisfying the above axioms and in particular Axiom III is termed non-degenerate.

§27. Property C. It will be convenient to order the points (η) of Axiom III in agreement with their increasing F-values, ordering points of (η) with the same F-value arbitrarily. Let g be any point of the set (η). Let $[\eta]_g$ denote the finite subset of points of (η) whose order is at most that of g. Let $F(g) = b$. Set

(27.1) $$X(g) = S_{b-} + [\eta]_g$$

We shall be concerned with relative (written rel.) cycles in X(g) with a modulus X(g) - g. A rel. cycle in X(g) shall always mean a cycle in X(g) mod. X(g) - g; rel. bounding in X(g) shall mean bounding in X(g) mod. X(g) - g, etc.

Lemma 27.1. Given e, any rel. cycle z in $X(g)$ is rel. homologous in $X(g)$ to a rel. cycle in an e-neighborhood of g.

Let $B^n z$ be the n-th barycentric subdivision of z. Use is made of the relation [Cf. Theorem 8.1 Appendix].

$$\partial \rho z = B^1 z - z - \rho \partial z \tag{27.2}$$

Since ∂z is in $X(g) - g$, $\rho\partial z$ is likewise, so that $z \sim B^1 z$ rel in $X(g)$. Hence $B^n z \sim z$ rel. in $X(g)$. Let $B^n z$ be written in the form

$$B^n z = g_i \sigma_i = u + v \qquad (i = 1, \ldots, n) \tag{27.3}$$

where u is the sum of the terms $g_i \sigma_i$ for which σ_i is in $X(g) - g$. Since

$$\partial 0 = (B^n z - v) - u,$$

$v \sim B^n z$ rel. in $X(g)$. If however n is sufficiently large the norm§ of each σ_i in (27.3) will be less than e, and v will be in the e-neighborhood of g (v possibly null). In resume

$$z \sim B^n z \sim v \qquad [\text{rel. in } X(g)]$$

and the lemma follows.

We shall refer to the r-disc $K(g)$ associated with a point g which has property C. For chains in $K(g)$ the modulus will always be $K(g) - g$, and the term relative will refer to this modulus. The following lemma is proved as was Lemma 27.1.

Lemma 27.2. Any rel. cycle in $K(g)$ is rel. homologous in $K(g)$ to a cycle in an e-neighborhood of g.

The case where g has the index 0. Both Lemma 27.1 and 27.2 are trivial in case the index r of g is 0. In this case $K(g) = g$ so that the modulus $K(g) - g$ is the empty set. A rel. cycle in $K(g)$ is then a cycle in $K(g)$.

§ The norm of a singular cell is the diameter of its carrier.

A rel. n-cycle in X(g) in this case is an n-cycle in g plus an arbitrary n-chain in X(g) - g. The k-dimensional rel. homology group of K(g) is the k-th homology group of K(g), and consists of a null element except when k=0.

The following theorem is basic.

Theorem 27.1. The n-th rel.§ homology group of X(g) is isomorphic with the n-th rel. homology group of K(g); if U is a homology class of rel. n-cycles in X(g), the corresponding homology class U' of rel. n-cycles in K(g) is the subclass of those chains of U which are in K(g).

The proof of this theorem will be made to depend upon statements (a) and (b) which follow.

(a) Each rel. homology class U of X(g) contains at least one rel. cycle in K(g).

In accordance with Lemma 27.1 there exists a chain z of U in N(g). (Cf. Property C). On making use of the deformation D_g affirmed to exist when g has Property C, we have the relation of Theorem 7.1 Appendix.

$$\partial\bar{\bar{D}}_g z = \bar{D}_g z - z - \bar{\bar{D}}_g \partial z \tag{27.4}$$

Here $\bar{\bar{D}}_g \partial z$ is in X(g) - g since ∂z is in this modulus, $\bar{D}_g z$ is in K(g), and $\bar{\bar{D}}_g z$ is in X(g). Thus z is rel. homologous in X(g) to $\bar{D}_g(z)$ in K(g).

(b) A rel. n-cycle x in K(g) which is rel. bounding in X(g), is rel. bounding in K(g).

Let U' be the rel. homology class in K(g) which contains x. Let e be so small a positive constant that points of K(g) within a distance e of g are in N(g). By hypothesis in (b)

$$\partial w \equiv x \qquad \text{mod.}[X(g) - g]^{\S\S}$$

§ As stated, the modulus in X(g) is X(g) - g, that in K(g), K(g) - g.
§§ The sign $\equiv$ indicates equality up to some chain in the modulus.

for some $(n+1)$-chain w in $X(g)$. Let w and x be subdivided so many times (say r times) that the norms of the simplices in the resulting chains are less than e. We have

$$\partial B^r w = B^r \partial w \equiv B^r x \qquad \text{mod. } [X(g) - g]$$

From the chain $B^r w = g_i \sigma_i$ let all simplices σ_i in $X(g) - g$ be dropped giving an $(n+1)$-chain w'. Let x' be similarly obtained from $B^r x$. Then

$$\partial w' \equiv x' \qquad \text{mod. } (X(g) - g) \tag{27.5}$$

It is clear as in the proof of (a) that $B^r x$ and hence x' is in U'. Moreover w' and x' are in $N(g)$, so that one can apply $\overline{D}_g$ to both members of (27.5), with the result

$$\partial \overline{D}_g w' \equiv \overline{D}_g x' . \qquad \text{mod. } (X(g) - g) \tag{27.6}$$

The terminal mapping D_g^1 of D_g restricted to

$$K^o = K(g) \cap N(g) ,$$

is deformable into the identity in $K(g)$ (Property C), deforming $K^o - g$ in $K(g) - g$, so that

$$\overline{D}_g x' \sim x' \qquad [\text{rel. in } K(g)] \tag{27.7}$$

From (27.6) we infer that $x' \sim 0$ rel. in $K(g)$. With x' each other chain x in U' is rel. bounding in $K(g)$.

Statement (b) is accordingly true and the theorem follows from Lemma 10.1 of the Appendix.

<u>Corollary 27.1.</u> <u>If</u> g <u>is a point of</u> (η) <u>of index</u> r, <u>the Betti numbers</u> P_k <u>of</u> $X(g)$ <u>mod.</u> $X(g) - g$ <u>are</u> δ_r^k .

---

§ The sign $\equiv$ indicates equality up to some chain in the modulus.

By virtue of Theorem 27.1 the Betti numbers of $X(g)$ mod. $X(g) - g$ are the Betti numbers P_k of the r-disc $K(g)$ mod. $K(g) - g$, and equal δ_r^k in accordance with Lemma 10.2 of the Appendix.

Homotopic critical points. A point p of S at which F is finite will be called homotopically ordinary if some F-neighborhood of p admits a proper weak F-deformation $(0 \leq t \leq 1)$ which ultimately displaces p. A point p at which F is finite and which is not homotopically ordinary will be termed a homotopic critical point.

We begin with the following lemma.

Lemma 27.3. A point p of S which has the Property C is a homotopic critical point.

Suppose that $F(p) = b$ and set

$$X = S_{b-} + p \,.$$

The proof of Theorem 27.1 and its Corollary show that the Betti number P_k of X mod. X - p is δ_r^k, where r is the index of p. There is accordingly an r-cycle z in X mod. X - p which is rel. non-bounding.

If the lemma were false there would exist a proper weak F-deformation D of some F-neighborhood of p, which displaces p. By modifying D as in the proof of Lemma 24.3 one can obtain a proper weak F-deformation D_o, defined over all of X and identical with D in its deformation of points of X sufficiently near p. Referring to the rel. cycle z of the preceding paragraph

$$\partial\bar{\bar{D}}_o z = \bar{D}_o z - z - \bar{\bar{D}}_o \partial z \tag{27.7}$$

Since the carrier of z is below b except at p, and D_o displaces p, $\bar{D}_o z$ is below b. The chain $\bar{\bar{D}}_o \partial z$ is likewise below b. Thus z is rel. bounding in X in accordance with (27.7). From this contradiction we infer the truth of the lemma.

The following theorem characterizes the points (η) of Axiom III.

<u>Theorem 27.2.</u> <u>The set of homotopic critical points of</u> F, <u>the set of points of</u> S <u>with Property C, and the set</u> (η), <u>are identical.</u>

Lemma 27.3 together with statement (c) will suffice to prove the theorem.

(c) <u>Each point</u> p <u>of</u> S <u>at which</u> F <u>has a finite value</u> b <u>and which is not a point</u> (η) <u>is homotopically ordinary.</u>

According to Axiom III there will exist a proper weak F-deformation Δ_c which deforms p into a point below b, and hence displaces p. Thus p is homotopically ordinary and statement (c) is proved.

By virtue of Lemma 27.3 and statement (c), the set of homotopic critical points of S is identical with the set (η) of Axiom III. A point p which fails to have Property C is not a point of (η), and so by (c) is homotopically oordinary if F(p) is finite. Hence the set of homotopic critical points of S is the set of points of S with Property C.

This completes the proof of the theorem.

We shall refer to the points (η) as <u>critical</u> points, dropping the adjective homotopic.

The reader will of course be well aware that for a function $f(x_1,\ldots,x_n)$ of n variables a point (x^0) may be a differential critical point without being a homotopic critical point. For example, the point $x = 0$ is a differential critical point of x^3, but not a homotopic critical point. The preceding theorems imply that a non-degenerate critical point of a function of n variables of Class C''' is a homotopic critical point; for such a non-degenerate critical point enjoys Property C as we have shown in Theorem 17.2. The deformation used in proving Theorem 17.2 was proper.

§28. Critical points of linking or non-linking types. Let a value of F at a critical point be termed a critical value. We shall add a classification of the critical points (η) which depends in part on their ordering at any given critical level.

Suppose g is a critical point with index r. According to Corollary 27.1 there is exactly one non-trivial r-dimensional homology class U in X(g), mod. X(g) - g. If U contains an "absolute"[§] r-cycle λ_r, g will be said to be of linking type, otherwise of non-linking type. In the latter case μ_r shall denote any rel. cycle in U. The cycle λ_r or the rel. cycle μ_r forms a rel. homology base for rel. r-cycles in X(g), according as g is of linking or non-linking type.

The following lemma makes the origin of the term "linking" clear.

Lemma 28.1. A necessary and sufficient condition that a critical point g be of linking type is that the boundary of every rel. k-cycle z in X(g) bound in X(g) - g.

The condition is proved necessary as follows. Let r be the index of g. Suppose that k = r. When g is of linking type, λ_r is a rel. homology base for r-cycles in X(g), so that a relation

$$(28.1) \qquad \partial w = z - a\lambda_r + u \qquad [u \text{ in } (X(g) - g)]$$

holds with a in G, and w an (r+1)-chain in X(g). Relation (28.1) implies that $\partial z = \partial u$, so that ∂z bounds u in X(g) - g. If k $\neq$ r, z is relatively bounding [Cf. Corollary 27.1], and ∂z accordingly bounds in X(g) - g. (Cf. Lemma 9.1, Appendix.)

---

§ An "absolute" r-cycle is an ordinary r-cycle. The term is used to emphasize this.

The condition is proved sufficient as follows. According to Corollary 27.1 there exists an r-cycle z relatively non-bounding in X(g). By hypothesis ∂z bounds a chain u in X(g) - g. Hence z - u is an absolute cycle λ_r in the rel. homology class of z. Hence g is of linking type.

Examples. If S is a torus with its axis in a horizontal position and F the vertical height on S then every critical point of F is of linking type. If S is a hemisphere with a horizontal base, and F the vertical height on S then the point of maximum F is a critical point of non-linking type. If S is the surface of the earth, and F the distance from the surface to the earth's center, the sets S_c can be supposed flooded with water up to the level c. A saddle point g of F will be of linking or non-linking type, according as the flooding of g creates a new island or connects two previous-disconnected oceans. The latter statement is on the assumption that no two critical points are at the same level. If there were two or more critical points at the same level and these were flooded in some order by a tidal wave the classification according to the creation of islands or connection of oceans would depend upon the order of flooding.

The k-th Betti number of a set X, over G, will be denoted by $R_k(X, G)$. When the k-th Betti numbers of X(g) and X(g) - g are finite we shall set

$$\triangle_g R_k = R_k[X(g), G] - R_k[X(g) - g, G]$$

The following lemma is fundamental.

Lemma 28.2. If the Betti numbers of X(g) - g are finite[§] and g is a critical point of index r, then $\triangle_g R_k = 0$ except that

§ This is an inductive hypothesis. That the Betti numbers of X(g) - g are finite will be established by an induction with respect to the ordered set (η).

$$\triangle_g R_k = 1 \qquad \text{or} \qquad \triangle_g R_{r-1} = -1$$

according as g is of linking or non-linking type.

Case I. g of linking type. The proof in this case consists in showing that a minimal k-homology base for X(g) is obtained by adding $\delta^k_r \lambda_r$ to a minimal k-homology base B_k for X(g) - g. To that end let

w = a (k+1)-chain in X(g)

z = a k-cycle in X(g)

u = a k-cycle in X(g) - g

a = an element in G.

(1) Recall that $\delta^k_r \lambda_r$ is a minimal homology base in X(g), mod. X(g) - g. Hence, given z,

$$\partial w = z - a\delta^k_r \lambda_r + u \tag{28.2}$$

for a suitable choice of w, a, and u. Relation (28.2) shows that $\delta^k_r \lambda_r$ and B_k form a k-homology base for X(g). It remains to show that this base is minimal.

(2) When k = r there is no relation

$$\partial w = \lambda_r + u\ , \tag{28.3}$$

since λ_r would then be rel. bounding in X(g). When $k \neq r$ there is no relation

$$\partial w = u \qquad [\text{for } u \not\sim 0 \text{ in } X(g) - g], \tag{28.4}$$

for (28.4) implies that w is a re. cycle in X(g). According to Lemma 28.1 ∂w bounds in X(g) - g whenever g is of linking type, contrary to the nature of u. Thus neither (28.3) nor (28.4) can hold, so that $\delta^k_r \lambda_r$ with B_r, forms a minimal homology base for X(g).

Case II. g is of non-linking type. For $k \neq r-1$ let B_k be a minimal homology base for $X(g) - g$. When $k = r-1$, $\partial\mu_r$ is non-bounding in $X(g) - g$ (Cf. Lemma 28.1) so that there exists a minimal $(r-1)$-homology base for $X(g) - g$ composed of $\partial\mu_r$ and of a set of $(r-1)$-cycles B_{r-1}. We shall show that B_k is a k-homology base for $X(g)$, including the case $k = r-1$.

(1) Since $\delta^k_r \mu_r$ is a minimal k-homology base in $X(g)$ mod. $X(g) - g$, given z

$$\partial w = z - a\delta^k_r \mu_r + u \tag{28.5}$$

for a suitable choice of w, a, and u. In (28.5) every chain except $a\delta^k_r \mu_r$ is a cycle. Since μ_r is not a cycle, $a\delta^k_r = 0$. This shows that a k-homology base for $X(g) - g$ is a base for $X(g)$. Hence B_k is a k-homology base for $X(g)$ (recalling that $\partial \mu_r$ bounds in $X(g)$). It remains to prove that B_k is minimal in $X(g)$.

(2) Suppose that a relation

$$\partial w = u_i g_i \neq 0 \qquad (u_i \text{ in } B_k) \tag{28.6}$$

holds. Then w would be a rel. non-bounding $(k+1)$-cycle in $X(g)$. Cf. Lemma 9.1, Appendix. This is possible at most if $k+1 = r$. But if $k+1 = r$, $w - a\mu_r$ would be rel. bounding in $X(g)$ for some a in G. (Cf. Corollary 27.1.) Hence the absolute cycle

$$\partial w - a\partial\mu_r = g_i u_i - a\partial\mu_r$$

would be bounding in $X(g) - g$ (Cf. Lemma 9.1, Appendix) contrary to the nature of B_{r-1}. Thus no relation (28.6) holds and B_k is minimal in $X(g)$.

The theorem follows.

Theorem 28.3. The Betti numbers of each of the sets $X(g)$, S_c and S_{c-} are finite.

The Betti numbers[§] $R_k[X(g)]$. Let the critical points (η) be written in order $\eta_1, \eta_2, \ldots$. If $g = \eta_1$, $F(g)$ is an absolute minimum of F and $X(g) = g$. In this case $R_k[X(g)] = \delta_0^k$.

Proceeding inductively we assume that each $R_k[X(p)]$ is finite for $p = \eta_{r-1}$, and seek to prove that $R_k[X(g)]$ is finite for $g = \eta_r$. In case $F(p) = F(g)$, $X(g) - g = X(p)$ and the Betti numbers of $X(g) - g$ are finite by inductive hypothesis. If $F(g) > F(p)$, and if one sets $F(g) = c$, then

$$X(g) - g = S_{c-} \tag{28.7}$$

In accordance with Theorem 26.3, with $X(p) = Y$ therein,

$$R_k(S_{c-}) = R_k(X(p)), \qquad [k = 0, 1, \ldots] \tag{28.8}$$

and the Betti numbers of $X(g) - g$ are again finite by inductive hypothesis. It follows from T. . .em 28.2 that the numbers $R_k[X(g)]$ are finite.

The Betti numbers $R_k(S_{c-})$. Let g be the last point in (η) for which $F(g) < c$. It follows from Theorem 26.3, with $X(g) = Y$ therein, that

$$R_k(S_{c-}) = R_k[X(g)] \qquad [k = 0, 1, \ldots]$$

The Betti numbers $R_k(S_c)$. Let g be the last point in (η) for which $F(g) \leqq c$. In accordance with Axiom III there exists a proper weak F-deformation of S_c into $X(g)$ and the proof of Theorem 26.3 shows that

$$R_k(S_c) = R_k(X(g)) \qquad [k = 0, 1, \ldots]$$

This completes the proof of the theorem.

Lemma 28.2 together with relations (28.7) and (28.8) now give the basic theorem.

---

§ We are dropping the G from $R_k[X(g), G]$ for the sake of brevity.

Theorem 28.4. *If g and g' are successive critical points in (η), with g' of index r, then*

$$R_r[X(g')] - R_r[X(g)] = +1 \qquad \text{Case I}$$

or

$$R_{r-1}[X(g')] - R_{r-1}[X(g)] = -1 \qquad \text{Case II}$$

according as g' is of linking or non-linking type, while all other Betti numbers of X(g) and X(g') are equal.

The proof of Lemma 28.2 together with Theorem 26.3 permit us to state a theorem which is stronger than Theorem 28.4.

Theorem 28.5. *If g and g' are successive critical points in (η) with g' of index r, then in Case I a minimal r-homology base for X(g') may be obtained from one for X(g) by the addition of a suitable r-cycle in X(g'); and in Case II a minimal (r-1)-homology base for X(g') may be obtained by removing a suitable (r-1)-cycle from a suitably chosen minimal (r-1)-homology base for X(g); while any minimal k-homology base of X(g) for which $k \neq r$ in Case I, or k = r in Case II, remains a minimal k-homology base for X(g'). Cases I and II are mutually exclusive.*

The following lemma will be used in §29.

Lemma 28.3. *If there is a last critical point (g) in (η) among points of index k and k+1, then $R_k(S) = R_k(X(g))$.*

This lemma will follow from Lemma 10.1 of the Appendix once statements (a) and (b) are proved.

(a) Any k-cycle x in S is homologous in S to a k-cycle in X(g).

(b) Any k-cycle z in X(g) which is bounding in S is bounding in X(g).

Proof of (a). It follows from Axiom II that x is homologous on S to a k-cycle x' in some set S_{b-}. We can suppose that $b > F(g)$. From Theorem 26.3 and 28.5 we make the inference: (c) A minimal k-homology base B_k for $X(g)$ is a minimal k-homology base for S_{b-}. It follows from (c) that x' is homologous in S_{b-} to a k-cycle in $X(g)$ so that (a) is true.

Proof of (b). It follows from Axiom II that any k-cycle z in $X(g)$ which is bounding in S is bounding in some set S_{b-} with $b > F(g)$. If $z_1, \ldots, z_n$ is a minimal k-homology base for $X(g)$ we have $z \sim g_i z_i$ in $X(g)$. As stated in (c) $z_1, \ldots, z_n$ is a minimal k-homology base for S_{b-} as well as for $X(g)$, so that each $g_i = 0$, since z is bounding in S_{b-}. Hence $z \sim 0$ in $X(g)$ and (b) is true. The lemma follows as stated.

§29. Relations between the type numbers M_o, M_1, ... and the Betti numbers R_o, R_1, ... when each M_k and R_k is finite. In the next sections, R_k shall denote the k-dimensional Betti number of S. We shall introduce the following additional notation:

M_k = The number of critical points of index k.

a_k = The number of critical points of index k of linking type.

b_k = The number of critical points of index k of non-linking type.

In this section we shall assume that M_k and R_k are finite for every k. ($k = 0, 1, \ldots$). In the case where S represents a connected component $W(A, B)$ of the space of paths all known examples yield finite R_k, but this finiteness has not been established in general. Among more general spaces S which satisfy our axioms it is easy to construct examples in which some R_k is infinite.

Example 1: Let Y be the closed sector of the (x, y) plane for which

$$\frac{\pi}{4} \leq \theta \leq \frac{3\pi}{4},$$

where Θ is the angle in conventional polar coordinates. Let an open circular disc with radius 0.1 and center at the point (0, n) be removed from Y for each positive n, to form a space S. In this space let $F = y$. Here $R_o = 1$, $R_1 = \infty$ and every other Betti number is 0. The subspaces S_c are compact. The critical points are the origin, at which F has an absolute minimum, and the points $(x, y) = (0, n + 0.1)$ $(n = 1, 2, \ldots)$ of index 1. Our three axioms on S and F are satisfied. On each compact subset F is bounded, so that Axiom II is trivial.

When each M_k and R_k is finite

$$M_k = a_k + b_k \qquad (b_o = 0) \tag{29.1)'}$$

$$R_k = a_k - b_{k+1} \tag{29.1)''}$$

Relation (29.1)" is a consequence of Lemma 28.3.

Points of index k = 0 are isolated relative minima of F. The only relative cycles associated with such points are absolute cycles, non-bounding only when having the form $g\sigma$, where $g \neq 0$ and σ is the image of an euclidean point. Such critical points are of linking type, so that $b_o = 0$.

From (29.1) one obtains the relation

$$M_k - R_k = b_k + b_{k+1} \qquad (k = 0, 1, \ldots) \tag{29.2}$$

from which the following theorem is derived.

Theorem 29.1. When the numbers M_k and R_k are finite, $M_k \geqq R_k$, $k=0,1,\ldots$.

Example 2: We shall presently see that the Betti numbers of the space of paths W joining two points A and B on an n-sphere, with $n > 2$, are

$$R_k = 1 \qquad (k \equiv 0 \text{ mod. } n-1) \tag{29.3)'}$$

$$R_k = 0 \qquad (k \not\equiv 0 \text{ mod. } n-1) \tag{29.3)''}$$

For example when $n = 3$ the Betti numbers of W are

R_0	R_1	R_2	R_3	R_4	R_5	...
1	0	1	0	1	0	...

Let M be any admissible coordinate manifold (cf. §7) which is homeomorphic with an n-sphere ($n > 2$). In accordance with Theorem 29.1, one can affirm that almost all pairs of points A and B are joined on M by a geodesic g_k of index k for each $k \equiv 0$ mod. $(n-1)$. According to the earlier theory, k is merely the number of conjugate points of A on the open arc of g_k, and the length of g_k, in the case of a non-degenerate pair A, B, becomes infinite with k. One can make a similar statement on replacing the length integral by any integral J with positive, and positive regular integrand.

<u>The case of a function</u> F <u>on an</u> m-<u>manifold</u> M. The metric space S can be taken as our m-manifold M, and we can suppose that F is a function of class C''' on M. If one supposes in addition that the differential critical points of F are non-degenerate (Cf. §17) then F and M satisfy our three axioms on F and S. In this case there will be no critical points of index $r > m$. For it follows from the properties of dimension that there is no topological r-disc in M for which $r > m$. One thus has

$$(29.4) \qquad 0 = M_{m+1} = M_{m+2} = \dots$$

A function F of this type is termed <u>non-degenerate</u>. It can be shown that any function F of class C''' on M can be approximated arbitrarily closely by a non-degenerate function F of class C''' in such a manner that the partial derivatives of F up to the third order approximate the corresponding partial derivatives of F_1. Note by way of example how x^3 is approximated by $x^3 - ex$ for e arbitrarily small and positive.

Example 3: The simplest example illustrating Theorem 29.1 is given by taking S as a torus in a space of coordinates (x, y, z) with the x-axis as the axis of the torus. If one sets $F = z$, the existence of four critical points on the torus with indices such that

$$M_0 = R_0 = 1 \qquad M_1 = R_1 = 2 \qquad M_2 = R_2 = 1$$

is immediately obvious.

The excess numbers $E_k = M_k - R_k$. Besides the critical points affirmed to exist in Theorem 29.1, there may exist other critical points. This will occur whenever any one of the numbers E_k, $k = 0, 1, \ldots$ are positive. Given a set of integers $R_k \geqq 0$ with $R_0 = 1$, it is possible to construct a space S with the given Betti numbers R_k and with a function F defined on S such that $M_k = R_k$ for each k. One may accordingly say that the critical points which are affirmed to exist in Theorem 29.1 are topologically necessary, given the numbers $R_0, R_1, \ldots$.

If, however, the space S is given rather than the numbers R_k, there may exist no admissible function F on S for which $M_k = R_k$ for every k. This statement is illustrated, rather than proved, by the following example.

Example 4: This example will show that if S is a coordinate m-manifold there need not exist any non-degenerate function of class C''' on S such that $M_k = R_k$, $k = 0, \ldots$. In particular let S be a 3-manifold M with the Betti numbers of the 3-sphere, but not the topological image of the 3-sphere§. Suppose (if possible) that F is a non-degenerate function of class C''' on M such that

$$M_k = R_k \qquad (k = 0, 1, \ldots, m) \tag{29.5}$$

§ For the existence of M see Seifert-Threlfall, Lehrbuch der Topologie, p.218

This would imply that $M_0 = 1$, $M_m = 1$ and that $M_k = 0$ for $k \neq 0$, or m. The only critical points of F on M would be the points of absolute minimum and absolute maximum of F on M. Let S_3 be a euclidean 3-sphere. On S_3 let f equal one of the euclidean coordinates of the space in which S_3 lies. We can suppose that the extreme values of f and F are the same, 0 and 1. If (29.5) held we could establish a homeomorphism between M and S_3 in which each level surface $F = c$ corresponds to the level surface $f = c$ $(0 < c < 1)$. The particular correspondence between the points of these surfaces could be explicitly defined by an appropriate use of the trajectories orthogonal to these surfaces, starting with a correspondence between these trajectories near the points of absolute minimum of f and F. This homeomorphism between M and S_3 is, however, contrary to our choice of M, so that (29.5) cannot hold.

Relations (29.1) can be used to prove the following theorem.

<u>Theorem 29.2.</u> <u>The Betti numbers of the space of paths</u> $W(A, B)$ <u>joining two distinct points</u> A <u>and</u> B <u>on an</u> n-<u>sphere</u> M, <u>with</u> $n > 2$, <u>are given by</u> (29.3).

Let the points A and B on the n-sphere be taken as any two distinct points not the extremities of a diameter of M. Let J be the integral of length on M. If one makes use of the Jacobian $M(s, s_0)$ of §11 whose zeros s define the conjugate points of s_0, one finds that the only points s conjugate to A coincide with the point A' directly opposite A on M. The multiplicity of A', as a conjugate point of A, is the nullity of the determinant $M(s, s_0)$ at the point s which determines A'. These multiplicities are all $n-1$.

Given any integer $i > 0$ there is a geodesic g leading from A to B which passes through A' exactly i times. If conjugate points are counted with their multiplicities there are $i(n-1)$ conjugate points of A on g prior to B. The index of g is accordingly $i(n-1)$. Moreover, these geodesics g are the only geodesics which join A to B on M. Thus

$$(29.6)' \qquad M_k = 1 \qquad (k \equiv 0 \text{ mod. } n-1)$$

$$(29.6)'' \qquad M_k = 0 \qquad (k \not\equiv 0 \text{ mod. } n-1)$$

If the values of M_k given by (29.6) are substituted in (29.1) and use be made of the fact that $R_o = 1$, and that R_k, a_k, b_k are positive on zero, it is found that the values of R_k are uniquely determined and are given by (29.3). This completes the proof of the theorem. Cf. Theorem 25.1.

The case n = 2. The theorem remains true even when $n = 2$. In this case each $R_k = 1$. The above method of proof fails because the substitution in (29.1) of the values M_k given by (29.6) does not determine the numbers R_k uniquely. A general proof of the theorem including the case $n = 2$ is given by Morse, op. cit., (Coll. Lect.), p.247.

The preceding theorem was established by proving that the numbers b_k are all zero, that is that every critical point is of linking type. We have the general theorem.

Theorem 29.3. *If each critical point of F on S is of linking type then the Betti numbers of S are given by the formulas*

$$(29.7) \qquad R_k = M_k$$

For example this theorem suffices to determine the Betti numbers of the torus as in Example 3. In addition this theorem has been used to determine the Betti numbers of the space of closed paths on an n-sphere, and the Betti numbers of the symmetric product of an n-sphere by itself. (Cf. Morse, op. cit., (Coll. Lect.), p.191.

Theorem 29.4. The Betti numbers of the space of paths on any orientable surface of genus $p > 0$ are all null except that $R_o = 1$.

The case $p = 1$. It is known that one can obtain a model for surfaces of genus $p = 1$ by identifying the opposite edges of a plane square. In the plane, and on this model for the torus

$$ds^2 = dx^2 + dy^2 ,$$

where (x, y) are rectangular coordinates in the plane. In this model there are no conjugate points; for the universal covering manifold is the whole plane, the geodesic straight lines, and the straight lines through a point never meet again. The Jacobian defining conjugate points never vanishes except at the initial point of the geodesic. Thus $M_k = 0$ for $k > 0$, so that $R_k = 0$ for $k > 0$.

The case $p > 1$. This case is similar to the torus except that one can get a model[§] for the surface as a polygon in a hyperbolic plane of constant negative curvature. The polygon has 4p sides which are arcs of non-euclidean straight lines, and these sides are identified in pairs to yield the model for the surface. The universal covering surface is the whole hyperbolic plane. J is the hyperbolic length, geodesics are the non-euclidean straight lines. Since the geometry is hyperbolic, geodesics through a point on the universal covering surface never meet again so that there are no conjugate points. As before $M_k = 0$ for $k > 0$, so that $R_k = 0$ for $k > 0$.

---

§ See, for example: Morse, A fundamental class of geodesics on any closed surface of genus $p > 1$. Trans Amer. Math. Soc. 26(1924) 25-60.

§30. The excess numbers $E_k = M_k - R_k$, when M_k and R_k are finite.

Staring with the relations

$$(30.1)' \qquad M_k = a_k + b_k \qquad (b_0 = 0)$$

$$(30.1)'' \qquad R_k = a_k - b_{k+1} \qquad (k = 0, 1, \dots)$$

we see that

$$(30.2) \qquad E_k = b_k + b_{k+1} \qquad (k = 0, 1, \dots)$$

From (30.2) it follows that

$$(30.3) \qquad E_0 - E_1 + E_2 - \dots (-1)^m E_m = (-)^m b_{m+1}$$

Hence (30.3) implies the inequalities

$$(30.4) \qquad \begin{array}{lcl} E_0 \geqq 0 & & E_0 \geqq 0 \\ E_0 - E_1 \leqq 0 & \text{or} & E_1 \geqq E_0 \\ E_0 - E_1 + E_2 \geqq 0 & & E_2 \geqq E_1 - E_0 \\ \dots\ \dots\ \dots & & \dots\ \dots\ \dots \end{array}$$

(with the signs $<$ and $\geq$ alternating). This means that if $E_0 > 0$ then $E_1 > 0$ in a compensating manner, and if this compensation is overdone so that $E_0 - E_1 > 0$, E_2 must compensate, and so on. The relations (30.4) can also be written in the form

$$(30.5) \qquad \begin{array}{c} M_0 \geqq R_0 \\ M_0 - M_1 \leqq R_0 - R_1 \\ M_0 - M_1 + M_2 \geqq R_0 - R_1 + R_2 \\ \dots\ \dots\ \dots\ \dots\ \dots\ \dots \end{array}$$

If S is a coordinate m-manifold $b_{m+1} = 0$, and one has the equality

$$(30.6) \qquad M_o - M_1 + M_2 - \dots (-)^m M_m = R_o - R_1 + R_2 - \dots (-)^m R_m$$

which goes back essentially to Kronecker and was used by Poincaré and Birkhoff. We state the following theorem.

<u>Theorem 30.1.</u> <u>Given the numbers</u> $R_k > 0$ $(k = 0, 1, \dots)$ <u>with</u> $R_o = 1$, <u>conditions</u> (30.5) <u>on the numbers</u> M_k <u>are equivalent to conditions</u> (30.1). <u>There are no other conditions on the numbers</u> M_k <u>and</u> R_k <u>alone.</u>

Given conditions (30.1), conditions (30.5) are implied. Conversely conditions (30.5) imply conditions (30.4) and permit $b_{m+1} \geqq 0$ to be defined by (30.3). Then (30.2) holds. One then defines a_k by using (30.1)". As a consequence

$$M_k = E_k + R_k = (b_k + b_{k+1}) + (a_k - b_{k+1}) = a_k + b_k$$

so that (30.1)' is valid.

In terms of the numbers $R_k \geqq 0$ with $R_o = 1$ there are no other conditions on the numbers M_k beyond the conditions (30.5). The conditions (30.5) imply (30.4), and conditions (30.4) imply that $E_k \geqq 0$ for each k and hence $M_k \geqq R_k \geqq 0$. With the numbers M_k and R_k given, satisfying (30.5) with $R_o = 1$, it is possible to construct a space S and a function F on S satisfying our axioms, and such that the type numbers of F are the given numbers M_k. The theorem follows.

<u>Theorem 30.2.</u> <u>The relations</u>

$$(30.7) \qquad E_{k-1} + E_{k+1} \geqq E_k \qquad (k = 1, 2, \dots)$$

<u>always hold.</u>

Relations (30.7) are an immediate consequence of (30.2).

§31. <u>The case of infinite M_k; R_k finite.</u> There can be infinite M_k even when S is a space of paths W(A, B). The following remarks can be made at once.

<u>The finite M_k.</u> If M_k and M_{k+1} are finite then as previously, using Lemma 28.3,

(31.1)' $$M_k = a_k + b_k$$

(31.1)" $$R_k = a_k - b_{k+1} ,$$

with the consequence

(31.2) $$M_k \geqq R_k$$

If M_k is finite for k = 0, 1, ..., m+1, then on setting $E_k = M_k - R_k$ the first m+1 <u>inequalities</u>

$$E_o \geqq 0$$

$$E_o - E_1 \leqq 0$$

$$E_o - E_1 + E_2 \geqq 0$$

$$\dots \quad \dots \quad \dots \quad \dots$$

follow from (31.1). If M_{r-1}, M_r, M_{r+1}, M_{r+2} are finite then (31.1) implies

$$E_{r-1} + E_{r+1} \geqq E_r$$

as previously, even if all other M_i are infinite.

<u>The numbers $M_k(g)$ and $P_k(g)$.</u> If g is an arbitrary critical point of F we set

$M_k(g)$ = The number of critical points of index k in X(g).

$a_k(g)$ = The number of these critical points of linking type.

$b_k(g)$ = The number of these critical points of non-linking type.

$P_k(g)$ = The Betti number of X(g).

For a fixed g these numbers are finite, and are zero for k greater than some integer N. The relations

(31.3)' $$M_k(g) = a_k(g) + b_k(g)$$

(31.3)" $$P_k(g) = a_k(g) + b_{k+1}(g)$$

are satisfied together with the relations

(31.4) $$E_k(g) = M_k(g) - P_k(g) = b_k(g) + b_{k+1}(g) \geqq 0$$

and all their algebraic consequences, as previously.

To establish the fundamental theorem, two lemmas are needed.

Lemma 31.1. *Given* k *there exists a critical point* g *of such high order in* (η) *that there is a minimal* k-*homology base* B_k *for* S *in* X(g). *For any such* g

(31.5) $$R_k = P_k(g) + q_k \qquad (q_k \geqq 0);$$

and there exists a minimal k-*homology base for* X(g) *composed of* B_k *and of* q_k k-*cycles* z_i $(i = 1, \ldots, q_k)$ *such that each* z_i *is bounding in* X(g'), *provided*[§] g' > g *is a critical point of sufficiently high order in* (η).

Since R_k is assumed finite the existence of a base B_k for S in some X(g) follows from Axiom II. (Cf. Theorem 26.3). Relation (31.5) then holds. As a consequence there will exist a minimal k-homology base for X(g) composed of B_k and q_k k-cycles u_i (i = 1, ..., k). Since B_k is a base for S there will exist a k-cycle v_i linearly dependent on the cycles of B_k, such that the k-cycle

$$z_i = u_i - v_i \qquad (i = 1, \ldots, q_k)$$

§ We use the sign > to indicate that the order of g' exceeds that of g in (η).

is bounding in S. It follows from Axiom II that for some critical point $g' > g$ of sufficiently high order, each z_i will bound in $X(g')$. It is clear that B_k and the q_k k-cycles z_i again form a minimal k-homology base for $X(g)$.

Note: The critical points g and g' affirmed to exist in the preceding lemma depend upon k and their order in (η) may become infinite as k becomes infinite.

Lemma 31.2. *Under the conditions of Lemma 31.1 there exist at least q_k critical points g_i of index k+1 with*

$$g < g_1 < g_2 < \dots < g_{q_k} \leqq g'$$

The k-cycles z_i of Lemma 31.1 satisfy no proper homology in $X(g)$ but are bounding in $X(g')$. If $q_k > 0$ there is accordingly a critical point g_1 of least order, with $g < g_1 \leqq g'$, such that some proper sum $a_i z_i = y_1$ is bounding in $X(g_1)$. With change of notation, if necessary, we can suppose that $a_1 \neq 0$. If g_0 is the immediate predecessor of g_1 in (η) then the cycles

$$y_1, z_2, \dots, z_{q_k}$$

are part of a minimal k-homology base in $X(g_0)$, while y_1 is bounding in $X(g_1)$. It follows from Theorem 28.5 that g_1 is of index k+1, and that the cycles

$$z_2, \dots, z_{q_k} \tag{31.7}$$

are part of a minimal k-homology base in $X(g_1)$.

If $g_1 = g'$ then $q_k = 1$, and the lemma is true. If $q_k > 1$, then $g_1 < g'$, and one can similarly infer the existence of a critical point g_2 of least order in (η) with $g_1 < g_2 < g'$, such that some proper sum $y_q = a_i z_i$ based on the cycles (31.7), is bounding in $X(g_2)$. One sees that g_2 is of index k+1. Continuing, one arrives at the existence of the critical points g_i of the lemma.

The k-th type number of a set of critical points is defined as the number of critical points of index k in the set.

The fundamental theorem in the case in which some of the numbers M_k are infinite is as follows.

Theorem 31.1. Suppose that the first r+1 Betti numbers

$$R_o R_1, \dots, R_r \qquad (r > 0)$$

of S are finite. Let H be an arbitrary finite subset of the critical points of F with type numbers,

$$n_o, n_1, \dots, n_r$$

and $n_k = 0$ for $k > r$. There exists a finite subset $K \supset H$ of the critical points of F with type numbers

$$m_o, m_1, \dots, m_r$$

and $m_k = 0$ for $k > r$, such that

$$
\begin{aligned}
&m_o \geqq R_o \\
(31.8) \qquad &m_o - m_1 \leqq R_o - R_1 \\
&\dots \quad \dots \quad \dots \quad \dots \\
&m_o - m_1 + \dots (-1)^r m_r \lesseqgtr R_o - R_1 + \dots (-1)^r R_r
\end{aligned}
$$

where the $>$ or $<$ sign in the final relation is to be chosen according as r is even or odd.

It is clear that Lemmas 31.1 and 31.2 can be satisfied for $k = 0,1,\dots,r$ by a single pair of critical points g and g' of arbitrarily high order in (η). Since g can be taken of arbitrarily high order in (η) and g' then determined, we can take g so that X(g) contains each of the critical points of the given set H. If one sets

(31.9) $$R_k = P_k(g) - q_k \qquad (k = 0, \ldots, r)$$

Lemma 31.2 affirms the existence of a set H_{k+1} of $q_k \geqq 0$ critical points g'' of index $k+1$, with $g < g'' \leqq g'$. Let H' be the subset of critical points of F in $X(g)$ with indices not exceeding r. We shall satisfy the theorem by setting

$$K = H' + \sum H_{k+1} \qquad (k = 0, \ldots, r-1)$$

If m_k is the k-th type number of K we have

(31.10) $$m_k = M_k(g) + q_{k-1} \qquad (k = 0, \ldots, r)$$

where $q_{-1} = 0$. On setting $e_k = m_k - R_k$ it follows from (31.9) and (31.10) that

(31.11) $$e_k = M_k(g) - P_k(g) + q_{k-1} + q_k \qquad (k = 0, 1, \ldots, r)$$

(31.12) $$= b_k(g) + b_{k+1}(g) + q_{k-1} + q_k$$

Hence

(31.13) $$e_0 - e_1 + e_2 - \ldots (-1)^k e_k = (-1)^k [b_{k+1}(g) + q_k] \qquad (k = 0, 1, \ldots, r)$$

The inequalities in the theorem follow from relation (31.13).

The inequalities (31.8) imply that $e_k \geqq 0$. Hence the corollary.

<u>Corollary 31.1.</u> <u>If the Betti numbers</u> R_k <u>are finite,</u> $M_k \geqq R_k$.

From (31.13) or from the inequalities (31.8) we have

$$e_{k-1} + e_{k+1} \geqq e_k \qquad (k = 1, 2, \ldots, r-1)$$

But the integer r in the theorem can be taken arbitrarily large and the subset H of critical points can be taken so as to include arbitrarily many of the M_k points of index k. Hence the Corollary.

<u>Corollary 31.2.</u> <u>Regardless of the finiteness of</u> M_{k-1}, M_k <u>and</u> M_{k+1}

$$E_{k-1} + E_{k+1} \geqq E_k$$

That is, if E_k is infinite, one at least of the numbers E_{k-1} or E_{k+1} must be infinite.

We depart from the assumption that R_k is finite and prove the theorem.

Theorem 31.2. *If R_k is infinite, M_k is infinite.*

If R_k is infinite it follows from Axiom II that $R_k[X(g)]$ must become infinite as the order of g in $[\eta]$ becomes infinite; for, each non-bounding k-cycle in S has a homologous k-cycle on some X(g). But we see from (31.4) that $M_k(g) \geqq R_k[X(g)]$. Since $M_k \geqq M_k(g)$ for each g, it follows that M_k is infinite if R_k is infinite.

§32. Normals from a point to a manifold. Let S be an m-dimensional regular manifold M of class C''' in an euclidean space of m+1 coordinates

$$(x_1, x_2, \ldots, x_{m+1}) = (x).$$

Let O be a point not on M. Then the distance from O to M will be a function F of the point on M of class C'''. Suppose that M is regularly represented neighboring a point P in terms of n parameters $(u_1, \ldots, u_m)$ in the form

$$x_i = X_i(u) \qquad (i = 1,\ldots,m+1)$$

If O has coordinates (a)

$$F^2 = (X_i - a_i)(X_i - a_i).$$

Since F is never zero, a critical point of F is a critical point of F^2, and satisfies the conditions

$$(32.1) \qquad (X_i - a_i)\frac{\partial X_i}{\partial u_j} = 0 \qquad (j = 1, \ldots, m)$$

Since the representation of M is assumed regular the matrix

$$\left\| \frac{\partial x_i}{\partial u_j} \right\| \qquad \begin{matrix}(i = 1,\ldots,m+1)\\(j = 1,\ldots,m)\end{matrix} \tag{32.2}$$

has the maximum rank m, and the conditions (32.1) are fulfilled if and only if the vector (X) - (a) is orthogonal to M at the point (X).

Thus a necessary and sufficient condition that a point P of M be a critical point of F is that the vector from O to the point P be normal to M at P.

To determine the condition that such a critical point be non-degenerate, and to find its index, let the critical point P be taken at the origin in the space (x), and let the coordinate m-plane $x_{m+1} = 0$ be tangent to M at P. After a suitable rotation of the $x_1,\ldots,x_m$ axes in the m-plane $x_{m+1} = 0$, M can be represented neighboring P in the form

$$x_{m+1} = \frac{a_i x_i x_i}{2} + \ldots + \qquad (i = 1,\ldots,m) \tag{32.3}$$

where the remainder vanishes with its first and second partial derivatives at the origin. The coefficients a_i are constants. For such of these constants as are not zero we set

$$r_i = \frac{1}{a_i} \qquad (i = 1,\ldots,m)$$

Proceeding by analogy with the theory of surfaces (m = 2), we call the point P_i on the x_{m+1} axis at which $x_{m+1} = r_i$ <u>a center of principal normal curvature of</u> M <u>belonging to</u> P. If $a_i = 0$ we say that the corresponding center P_i is at infinity. The relation of centers of curvature P_i to their base points P is naturally taken as invariant of any rotation or translation. In shorter terminology P_i is called a <u>focal point based on</u> P. This term arises from the fact that light rays emanating from M near P in directions orthogonal to M intersect the "caustic" at each point P_i.

We shall prove the following lemma.

Lemma 32.1. *A critical point P of F corresponding to a normal OP is non-degenerate if O coincides with no one of the m focal points P_i of P, and in case P is non-degenerate its index is the number of focal points of P between O and P.*

Suppose that M is represented near P as in (32.1), and that O lies at the point $x_{m+1} = c \neq 0$ on the x_{m+1} axis. Near P on M we have

$$F = [x_i x_i + (x_{m+1} - c)^2]^{\frac{1}{2}} \qquad (i = 1,\ldots,m)$$

If one uses (32.3), F takes the form

$$\begin{aligned} F &= [x_i x_i + \left(\frac{a_i x_i x_i}{2} - c\right)^2 + \ldots]^{\frac{1}{2}} \\ &= [c^2 + x_i x_i (1 - c a_i) + \ldots]^{\frac{1}{2}} \\ &= c[1 + \frac{x_i x_i}{c} (\frac{1}{c} - a_i) + \ldots]^{\frac{1}{2}} \\ &= c + \frac{x_i x_i}{2c} (\frac{1}{c} - a_i) + \ldots \end{aligned}$$

where the terms omitted are of class C''', and vanish with their first and second partial derivatives at the origin.

The quadratic form in this representation of F is seen to be degenerate if and only if $1 - c a_i = 0$, that is, if and only if O is at a focal point P_i of P. We may suppose that $c > 0$, and we then see that the coefficient involving a_i is negative if and only if $a_i \neq 0$, and if

$$\frac{1}{c} < a_i \qquad \text{or} \qquad c > r_i > 0$$

The lemma follows.

We have the following theorem.

Theorem 32.1. Suppose that O is not on the m-manifold M nor a focal point of M. Of the normals OP from O to points P on M let M_k be the number upon which there are k-focal points of P between O and P, and let R_k be the k-th Betti number of M. Then $M_k \geqq R_k$ and the excess numbers $E_k = M_k - R_k$ satisfy the inequalities (30.4) and the equality

$$E_o - E_1 + \dots (-1)^m E_m = 0.$$

We see that the number of normals from O to M is finite and at least

$$R_o + R_1 + \dots + R_m$$

These normals can be regarded as the positions of equilibrium of taut elastic cords stretching from O to M with their end points on M, free to move on M and passing freely through M at any point. The index of the corresponding critical points can be interpreted as an index of instability, complete stability arising in the case of index zero and maximum instability arising in the case of index m.

§33. The three levels of the theory, open fields. The theory of critical points needs further generality less than it needs effective extension and application. Basic problems such as the determination of the Betti numbers of the spaces of paths, and the generalization of this problem to the spaces of continuous images of r-discs with prescribed bounding topological (r-1)-spheres, etc. are typical. In the domain of analysis, an extension of the quadratic analysis of the first sections of these lectures to multiple integrals is needed, and problems in the large, such as the three body problem§,

§ A start on this has been made by Morse and Ewing, Variational theory in the large including the non-regular case. First and Second Paper, Ann. of Math. 44 (1943).

and a non-linear quantum theory based on an extension of the Schrödinger integral can and should be attacked.

The underlying critical point theory should be developed at three levels:

(1) A theory of non-degenerate functions.

(2) A theory of homology F-limits without reference to excess critical points.

(3) A theory of relation between critical groups in the form of isomorphisms.

Level (1). The first type of theory is illustrated by the present lectures. There is a sense in which such a theory is general. It is that the non-degenerate functions are both metrically and topologically dense among all admissible functions. This theory is at the simplest level.

Level (2). The second type of theory is illustrated by the author's "Functional topology and abstract variational theory" (Memorial). The fundamental theorem is that under appropriate conditions there is in each homology class a k-cycle at or under a minimum F-level, and at that level there is a homotopic critical point. Singular cycles ate not adequate for such a theory, Vietoris cycles are. The theory is based on two hypothesis the F-accessibility of the space and the upper reducibility of F. F-accessibility replaces the compactness of S and the lower-semi continuity of F of the classical minimum theory. Upper reducibility replaces the upper inequality in the definition of continuity.

Level (3). In a very general theory critical points are too numerous. One must find order in a maze. In general all type numbers M_k are infinite, critical points are replaced by critical sets of any complexity, and the classification of critical points by indices is replaced by a classification of critical sets by a countably infinite set of groups of relative cycles termed "caps" attached to each critical set. The relations between the type numbers M_k of the non-degenerate case are replaced by isomorphisms between groups of caps whose dimensions in the non-degenerate case reduce to the numbers M_k, R_k, a_k, b_k, etc. The theory can be brought to a finitary basis by the introduction of the notion of "span" of a cycle as in "Rank and span in functional topology", Ann. of Math., 41 (1940), 419-454. Essentially every cycle is ignored whose F-characteristics can be specified by limits which differ by less than e.

Some of the open fields are as follows.

(a) Quadratic analysis such as that in §§1 to 6, including multiple integrals and integral equations with symmetric kernels. Such an analysis is being developed by Morse and Transue. There is an associated non-linear analysis.

(b) Topological problems such as generalizations of the space of paths, and their relation to homotopy theory.

(c) Proof that the planetary orbits in the n-body problem are the reflection of the Betti numbers of the associated space of closed curves. Finding these Betti numbers.

(d) A non-linear quantum theory based on an extension of the Schrödinger integral.

(e) Related to (d) is a study of singular quadratic integrals whose Euler equations are of the Fuchsian type (Bessel's, Legendre's equation, etc.).

(f) The Dirichlet problem and minimal surface problem on curved manifolds.

(g) The mathematical meaning and causes of degeneracy, such as the phenomenon of ellipses on an ellipsoid reducing to the family of circles on a sphere. There is an underlying topological cause of great applicability.

(h) The meaning and frequency of non-degeneracy in problems of all types, in particular in minimal surface problems.

(i) The billiard ball problem from the point of view of the critical point theory.

(j) Critical point theory in isoperimetric problems, regarding an isoperimetric solution as a critical point on a boundary defined by the isoperimetric condition, and developing a theory of the relations of such boundary critical points with interior critical points. See for example the theory of pseudo-harmonic functions in Morse, Topological methods in the theory of functions of a complex variable, Annals of Math. Studies.

(k) Various accessory problems in analysis involving the analysis of transversality and critical boundary values.

(l) The search for the simplest non-degenerate function on a given manifold or space.

(m) The relation of critical point analysis to topological characteristics other than homology classes and Betti numbers.

APPENDIX

A SPECIAL PARAMETERIZATION OF CURVES[§]

By Marston Morse

1. Introduction. Parameterization of curves by means of arc length fails when the arc length is infinite. The present paper develops the properties of a special parameterization of curves which never fails to exist and which is most useful in applications. The special parameter is called a μ-length and is an extension of a function of sets defined by H. Whitney and applied by Whitney to families of simple non-intersecting curves. The curves employed in the present paper are general continuous images of a line segment. This necessitates a slight modification of the definition of Whitney taking order into account. The properties of μ-length developed here are directed largely towards applications in abstract variational theory. While many of Whitney's ideas go over, there are nevertheless certain sharp differences both in the proofs and in the results.

2. The μ-length. Let N be a space of points p, q, r with a metric which is not in general symmetric. That is, to each ordered pair p, q, of points of N there shall correspond a number denoted by pq such that pp = 0, pq > 0 if p $\neq$ q, and

$$pq \leq pr + rq \quad (1)$$

§ Marston Morse, Bull. Amer. Math. Soc., vol. 42(1936) pp.915-922.

* H. Whitney, Regular families of curves, Annals of Mathematics, vol.34(1933), pp.244-270. Also Proceedings of the National Academy of Sciences, vol.18 (1932), pp.275-278 and pp.340-342.

Appendix

We term pq the <u>distance</u> from p to q. Let $|pq|$ denote the maximum of pq and qp. We term $|pq|$ the <u>absolute distance</u> between p and q. We see that $|pq| = |qp|$ and that $|pq| \leqq |pr| + |rq|$. The points of n taken with the distance $|pq|$ form a metric space which we denote by $|N|$. We shall use the metric of N in defining μ-lengths. For other purposes, in particular in defining continuity on N and $|N|$, we shall use the metric of $|N|$.

Let t be a number on a closed interval* (0, a). Let f(p, t) be a single-valued (numerical) function of p and t for p on $|N|$ and t on (0, a). The function f(p, t) will be termed continuous at (q, τ) if f(p, t) tends to $f(q, \tau)$ as a limit as $|pq| + |t - \tau|$ tends to 0. For each value of t on (0, a) and a point p on $|N|$ let $\phi(p, t)$ be a point on $|N|$. The (point) function $\phi(p, t)$ will be termed continuous at (q, τ) if the distance $\phi(p, t)\phi(q, \tau)$ tends to 0 as a limit as $|pq| + |t - \tau|$ tends to 0.

Let p(t) be a continuous point function of t for t on (0, a). We term p = p(t) a <u>parameterized curve</u> λ (written p-curve) and regard two parameterized curves as <u>identical</u> if they are defined by the same point function p(t). We also say that λ is the continuous image on $|N|$ of (0, a). In general curves on $|N|$ will be denoted by Greek letters α, β, ..., while points on $|N|$ will be denoted by letters p, q, r, ...

---

* For the sake of simplicity we assume that a > 0 and that our p-curves do not reduce to points. One could however admit p-curves which reduce to points. The μ-lengths of such curves is zero and we would thus admit intervals (0, a) for which a = 0. For such exceptional curves Fréchet distance is defined in the obvious manner. Theorem 2 is obvious if ζ reduces to a point. In the proof of Theorem 4, in case λ_0 reduces to a point, (17) is an easy consequence of $\lambda\lambda_0 < \delta$ provided δ is sufficiently small. Otherwise the theorems and proofs hold as written even when the p-curves reduce to points.

Appendix

Let λ be a p-curve on $|N|$ given as the continuous image $p = p(t)$ of an interval $0 \leqq t \leqq a$. For $n \geqq 2$ let $t_1 \leqq t_2 \leqq \dots \leqq t_n$ be a set of n values of t on $(0, a)$ and let $(p_1, \dots, p_n) = S_n$ be the set of corresponding points p on λ. We term S_n and admissible set of n points on λ. Let the minimum of the numbers $p_i p_{i+1}$ as i ranges from 1 to n-1 be denoted by $d(S_n)$ and let the least upper bound of all such numbers $d(S_n)$ for a fixed n be denoted by $\mu_n(\lambda)$. Following Whitney we then set

$$(2) \qquad \mu_\lambda = \frac{\mu_2(\lambda)}{2} + \frac{\mu_3(\lambda)}{4} + \frac{\mu_4(\lambda)}{8} + \dots .$$

We term the μ_λ the μ-length of λ.

We enumerate certain properties of $\mu_n(\lambda)$ and μ_λ.

(a) $\mu_2(\lambda) = d$, the diameter of λ.

(b) $\mu_n(\lambda) \leqq d$, and $\mu(\lambda) \leqq d$.

(c) $\mu_n(\lambda)$ tends to 0 as n becomes infinite.

(d) $\mu_n(\lambda) \geqq \mu_{n+1}(\lambda)$.

(e) If $p(t)$ is not identically constant none of the numbers $\mu_n(\lambda)$ is 0.

Statements (a), (b), (c), and (e) are obvious. To establish (d) let S_{n+1} be any admissible set of n+1 points $p_1, \dots, p_{n+1}$ on λ, $(n > 1)$. There is an integer k such that $p_k p_{k+1} = d(S_{n+1})$. We shall form an admissible set S_n on λ by removing one point from S_{n+1}. If $k \neq 1$ we remove p_1. If $k = 1$ we remove p_{n+1}. In both cases the pair p_k, p_{k+1} remains and $d(S_n) = d(S_{n+1})$. The set of all numbers $d(S_{n+1})$ is thus a subset of the numbers $d(S_n)$. Hence (d) is true.

Let $\mu(t)$ be the μ-length of the p-curve on λ defined by $p = p(\tau)$ for τ on the interval $(0, t)$. We shall show that $\mu(t)$ has the following properties.

Appendix

(f) $\mu(t)$ <u>is a continuous, non-decreasing function of</u> t.

(g) $\mu(t)$ <u>is constant on each interval on which</u> p(t) <u>is constant</u>.

(h) $\mu(t)$ <u>is constant on no interval on which</u> p(t) <u>is not constant</u>.

It is clear that $\mu(t)$ is non-decreasing. To show that $\mu(t)$ is continuous let η be the p-curve defined by $p = p(t)$ for $0 \leqq t \leqq \tau < a$, and let e be an arbitrary positive number. There exists a positive number δ so small that

$$|p(t)\,p(\tau)| \leqq \frac{e}{2}, \qquad (\tau \leqq t \leqq \tau+\delta \leqq a). \tag{3}$$

Let ζ be the p-curve defined by $p = p(t)$ for $0 \leqq t \leqq \tau + \delta$. Let S_n^2 be an admissible set of n points on ζ. Let S_n^1 be an admissible set of n points on η obtained by replacing each point of S_n^2 for which $t > \tau$ by $p(\tau)$. No point of S_n^2 is thereby moved a distance greater than e/2. Hence

$$\begin{aligned} d(S_n{}^2) &\leqq d(S_n{}^1) + e, \\ \mu_n(\zeta) &\leqq \mu_n(\eta) + e, \\ \mu(\tau+\delta) &\leqq \mu(\tau) + e. \end{aligned} \tag{4}$$

Since $\mu(t)$ is non-decreasing, (4) implies continuity on the right. Continuity on the left is similarly established, and the proof of (f) is complete.

Statement (g) requires no proof. To establish (h) we assume that there are values τ and τ' of t on (0, a) with $\tau' > \tau$ such that p(0), $p(\tau)$, and $p(\tau')$ are distinct. Let h and k be the p-curves defined by p(t) for t on $(0, \tau)$ and $(0, \tau')$, respectively. We shall prove that $\mu_k > \mu_h$. To that end, let 2c be the minimum value of $p(t)p(\tau) + p(t)p(\tau')$ as t ranges on $(0, \tau)$. We observe that $c \neq 0$ since $p(\tau) \neq p(\tau')$. Let S_n be an admissible set of points $p_1, \ldots, p_n$ on h. We form S'_{n+1} on k by adding one point p_{n+1} to S_n as follows. If

$$p_n p(\tau) \geqq c \tag{5}$$

we add $p(\tau)$. If (5) does not hold we add $p(\tau')$, and note that

(6) $$p_n p(\tau') \geqq c$$

by virtue of our choice of c. We suppose N so large an integer that $d(S_n) < c$ for $n > N$. For such values of n it follows from (5) and (6) that $d(S'_{n+1}) = d(S_n)$, and hence

(7) $$\mu_{n+1}(k) \geqq \mu_n(h) .$$

Since $\mu_n(h)$ is not 0 and tends to 0 as n becomes infinite, for some value of $n > N$,

(8) $$\mu_n(h) < \mu_{n+1}(h),$$

and for such an n it follows from (7) and (8) that $\mu_{n+1}(k) > \mu_{n+1}(h)$. From this relation and from (2) it follows that $\mu_k > \mu_h$ as stated. The proof of (h) is complete.

3. Equivalent p-Curves. Let η and ζ be two p-curves given by the respective equations

(9) $$p = p(t) \qquad (0 \leqq t \leqq a),$$

(10) $$q = q(u) \qquad (0 \leqq u \leqq b).$$

Let H be a sense-preserving homeomorphism between the closed intervals (0, a) and (0, b). A homeomorphism of the nature of H will be termed admissible. Let $u = u(t)$ be the value of u corresponding to t under H and let D(H) be the maximum of the distances $|p(t)q[u(t)]|$ as t ranges over (0, a). The Fréchet distance $\eta\zeta$ between η and ζ is the greatest lower bound of the numbers D(H) as H ranges over the set of all admissible homeomorphisms H between η and ζ. We observe that $\eta\zeta = \zeta\eta \geqq 0$, and one readily proves that for any three p-curves η, ζ, λ, $\eta\lambda \leqq \eta\zeta + \zeta\lambda$.

We shall say that η is <u>derivable</u> from ζ if there exists a continuous non-decreasing function $u = u(t)$ which maps the closed interval $(0, a)$ onto the closed interval $(0, b)$ and such that $p(t) \equiv q[u(t)]$. Two p-curves which are derivable from the same p-curve η will be said to be <u>equivalent</u>.

Let $u(t)$ be the μ-length of η from the point 0 to the point t. Let $t(\mu)$ be the function inverse to $\mu(t)$. To each value of μ on the interval $0 \leqq \mu = \mu_n$ there corresponds a value or interval of values $t(\mu)$. We set

$$\text{(11)} \qquad p[t(\mu)] = r(\mu) , \qquad (0 \leqq \mu \leqq \mu_\lambda),$$

and observe that $r(\mu)$ is single-valued by virtue of (h), §2. We shall prove the following proposition.

(A) <u>The function</u> $r(\mu)$ <u>is continuous in</u> μ.

Let μ_o be a value of $\mu(t)$. Corresponding to μ_o, let $\tau_1 \leqq t \leqq \tau_2$ be the interval of values taken on by $t(\mu)$ at μ_o (τ_1 may equal τ_2). Corresponding to a positive constant e there exists a positive constant δ so small that

$$\text{(12)} \qquad |p(\tau_2)\, p(t)| < e , \qquad (\tau_2 \leqq t \leqq \tau_2 + \delta).$$

As t ranges over the interval in (12), $u(t)$ increases from μ_o to a value $\mu_1 > \mu_o$. From (12) we infer that $|r(\mu_o) r(\mu)| \leqq e$ for $\mu_o \leqq \mu \leqq \mu_1$. Thus $r(\mu)$ is continuous on the right at μ_o. Continuity on the left is established similarly, and the proof of (A) is complete. We shall now prove the following statement.

(B) <u>The curve</u> η <u>is derivable from the curve</u> $r = r(\mu)$, $(0 \leqq \mu \leqq \mu_n)$.

It follows from the definition of $r(\mu)$ in (11) that $p(t) \equiv r[\mu(t)]$, and the proof of (B) is complete. We term $r = r(\mu)$ a μ-<u>parameterization</u> of η and state the following theorem.

Appendix

Theorem 1. Two p-curves are equivalent if and only if they have the same μ-parameterization.

If η and ζ are equivalent, they are derivable from a common p-curve and will have μ-parameterizations identical with that of λ. Conversely, if η and ζ have a common μ-parameterization $r = r(\mu)$, η and ζ are both derivable from $r = r(\mu)$ in accordance with (B) and hence are equivalent. We shall prove the following theorem.

Theorem 2. The Fréchet distance between a p-curve ζ and a p-curve η derivable from ζ is null.

Suppose η and ζ have the representations (9) and (10), respectively. Suppose η is derivable from ζ under the substitution $u = u(t)$, so that $p(t) \equiv q[u(t)]$ for t on $(0, a)$. Let c be an arbitrarily small positive constant and consider the transformation

$$u \equiv [u(t) + ct]\left[\frac{b}{b + ca}\right], \qquad (0 \leqq t \leqq a). \tag{13}$$

This transformation establishes a homeomorphism between the closed intervals $(0, a)$ and $(0, b)$. Denote the right member of (13) by $\phi(t, c)$, and let λ_c be the p-curve

$$q = q[\phi(t, c)], \qquad (0 \leqq t \leqq a). \tag{14}$$

The Fréchet distance $\zeta\lambda_c = 0$. For under the transformation (13) corresponding points of ζ and λ_c are identical. To show that $\eta\zeta = 0$ we make use of the relation $\eta\zeta \leqq \eta\lambda_c + \lambda_c\zeta = \eta\lambda_c$. If c is sufficiently small, points on η and λ_c determined by the same values of t on $(0, a)$ are arbitrarily and uniformly near since $\phi(t, 0) \equiv u(t)$. Hence $\eta\lambda_c$ tends to 0 with c. But $\eta\zeta$ is independent of c and must then be 0. The proof of the theorem is complete.

We shall now prove Theorem 3.

Theorem 3. *If η and ζ are two p-curves for which $\eta\zeta < e$, then* $|\mu_\eta - u_\zeta| \leqq 2e$.

Let S_n be an admissible set of n points on η. There exists an admissible set S_n' of n points on ζ with distances from the correspondingly numbered points of S_n less than e; $|d(S_n') - d(S_n)| \leqq 2e$, and we infer that $|\mu_n(\zeta) - \mu_n(\eta)| \leqq 2e$. Upon referring to the definition (2) of μ-length we conclude that Theorem 3 holds as stated. We state the following corollary.

Corollary 1. *If* $\eta\zeta = 0$, $\mu_\eta = \mu_\zeta$.

4. Curves. *A class of equivalent p-curves will be called a curve class or a curve.*

Let α and β be two cruves. Let η and η' be p-curves in the class α, and ζ and ζ' p-curves in the class β. I say that

(15) $$\eta\zeta = \eta'\zeta'.$$

Relation (15) follows from the relation

$$\eta\zeta \leqq \eta\eta' + \eta'\zeta' + \zeta'\zeta.$$

For $\eta\eta' = 0$, since η and η' are at a distance 0 from their common μ-parameterized curve λ in accordance with Theorem 2. Similarly $\zeta\zeta' = 0$. Upon reversing the roles of $\eta\zeta$ and $\eta'\zeta'$ we infer that (15) holds as stated. We are accordingly led to the following statement and definition.

The distance between any two p-curve classes α and β equals the distance between any other two p-curves in the classes α and β, respectively, and will be taken as the distance $\alpha\beta$ between the curves α and β.

Let λ be an arbitrary curve with μ-length μ_λ. A pair (λ, μ) will be termed <u>admissible</u> if $0 \leqq \mu \leqq \mu_\lambda$. For admissible pairs (λ, μ) let $q(\lambda, \mu)$ be the point on λ which determines the μ-length μ on λ. The following theorem is fundamental.

<u>Theorem 4.</u> <u>The point function</u> $q(\lambda, \mu)$ <u>is continuous</u>* <u>in its arguments on the domain of admissible pairs</u> (λ, μ).

We shall prove (λ, μ) continuous at (λ_o, μ_o) understanding that (λ_o, μ_o) is admissible. Let e be an arbitrary positive constant. We shall show that there exists a positive constant δ such that if (λ, μ) is admissible and

$$\lambda\lambda_o < \delta, \qquad |\mu - \mu_o| < \delta, \tag{16}$$

then

$$|q(\lambda, \mu)q(\lambda_o, \mu_o)| < e. \tag{17}$$

To that end we shall subject δ to two conditions as follows:

(i) We take $\delta < e/2$. If $\lambda\lambda_o < \delta$, there will exist a homeomorphism T_δ between μ-parameterizations of λ and λ_o in which corresponding points have distances less than δ. If the point μ on λ thereby corresponds to μ_1 on λ_o,

$$|q(\lambda, \mu)q(\lambda_o, \mu_1)| < \frac{e}{2}. \tag{18}$$

(ii) The second condition on δ is that δ be so small that when $|\mu_1 - \mu_o| < 3\delta$,

$$|q(\lambda_o, \mu_1)q(\lambda_o, \mu_o)| < \frac{e}{2}. \tag{19}$$

This condition can be satisfied by virtue of the continuity of $q(\lambda_o, \mu)$ in μ.

---

* We have not yet shown that for two curves α and β, $\alpha\beta = 0$ only if $\alpha = \beta$. But this is not necessary to speak of continuity. The proof of our theorem will imply that $q(\alpha, \mu) = q(\beta, \mu)$ when $\alpha\beta = 0$.

With δ so chosen I say that (17) holds for admissible pairs (λ, μ) satisfying (16). We introduce T_δ and μ_1 on λ_o as in (i), and recall that

$$(20) \quad |q(\mu, \lambda)q(\mu_o, \lambda_o)| \leqq |q(\lambda, \mu)q(\lambda_o, \mu_1)| + |q(\lambda_o, \mu_1)q(\lambda_o, \mu_o)| .$$

The first term on the right of (20) is less than $e/2$ by virtue of (18) and the second term is likewise, provided (19) is applicable; that is, provided $|\mu_1 - \mu_o| < 3\delta$. But under T_δ a point μ on λ will correspond to a point μ_1 on λ_o such that $|\mu - \mu_1| < 2\delta$ in accordance with Theorem 3. Hence $|\mu_1 - \mu_o| < 3\delta$, (19) is applicable, and the right member of (20) is less than e. The proof of the theorem is complete. We conclude this section with the following theorem.

<u>Theorem 5.</u> <u>A necessary and sufficient condition that two p-curves</u> η <u>and</u> ζ <u>have the same</u> μ-<u>parameterization is that</u> $\eta\zeta = 0$.

To prove the condition necessary, let λ represent a μ-parameterized curve determined by η and ζ. Then $\eta\zeta \leqq \eta\lambda + \lambda\zeta$. But $\eta\lambda = \lambda\zeta = 0$ by virtue of Theorem 2. Hence the condition is necessary. To prove the condition sufficient suppose that $\eta\zeta = 0$ and let $p = p(\mu)$, $(0 \leqq \mu \leqq \mu_\eta)$, $q = q(\mu)$, $(0 \leqq \mu \leqq \mu_\zeta)$, be μ-parameterizations of η and ζ, respectively. By virtue of Corollary 1 to Theorem 3, $\mu_\eta = \mu_\zeta$. It then follows from Theorem 4 that $p(\mu)$ and $q(\mu)$ differ by a quantity arbitrarily small in absolute value. Hence $p(\mu) = q(\mu)$, and the condition is proved sufficient. Theorem 5 taken with Theorem 1 gives the following corollary.

<u>Corollary 2.</u> <u>A necessary and sufficient condition that two</u> p-<u>curves</u> η <u>and</u> ζ <u>be equivalent is that</u> $\eta\zeta = 0$.

Another way of stating the corollary is to say that for two curves α and β, $\alpha\beta = 0$ if and only if $\alpha = \beta$.

Singular Homologies

§5. The complex K(P). The theory reviewed here is taken with some modifications from S. Eilenberg, Singular homology theory, Annals of Math. 45 (1944) 407-446.

The complex K(P). Let P be a finite polyhedron with a fixed simplicial decomposition. The simplices of P taken without any ordering of their vertices will be called geometric simplices. These simplices will be non-degenerate, that is, the q+1 vertices of a q-simplex will lie in no (q-1)-plane.

A q-cell of K(P) is an ordered array $(v_o, \ldots, v_q)$ of vertices of P with possible repetitions, and with the condition that all the vertices in question lie on a geometric simplex of P. A q-cell without repetition of vertices is termed proper; with repetitions, degenerate. The free abelian group generated by the q-cells of K(P) will be denoted by $C^q(K)$. There are no relations given between q-cells of K(P); two q-cells are regarded as the same if and only if they are given by the same ordered set of vertices. The elements of $C^q(K)$ will be called q-chains of K(P). To distinguish chains formed of integral linear combinations of the generators, from chains formed with other coefficients (as defined later) the present chains may be called integral.

Given a cell, $u = (v_o, \ldots, v_q)$ of K(P), the boundary ∂u will be defined as the (q-1)-chain

$$\partial u = (-1)^i (v_o, \ldots, \cancel{v}_i, \ldots, v_q) \qquad (i = 0, 1, \ldots, q)^*$$

the vertex v_i being omitted as indicated. That $\partial\partial u = 0$ is seen from the relation

$$\begin{aligned}\partial\partial u &= (-1)^i(-1)^j (v_o, \ldots, \cancel{v}_j, \ldots, \cancel{v}_i, \ldots, v_q) && \text{(summed for } j < i)\\ &+ (-1)^i(-1)^{j-1} (v_o, \ldots, \cancel{v}_i, \ldots, \cancel{v}_j, \ldots, v_q) && \text{(summed for } j > i).\end{aligned}$$

* We sum over repeated indices, even if one is an exponent.

When $q = 0$ one sets $\partial u = 0$. We extend the definition of ∂z to the case where z is any chain $n_i u_i$ of $C^q(K)$ (u_i a q-cell of $K(P)$) by setting

$$\partial z = n_i \partial u_i$$

We term this extension of the meaning of ∂, definition by <u>linear extension</u>.

The chains of $C^q(K)$ with vanishing boundaries are called <u>cycles</u>. Chains c^q such that

$$c^q = \partial c^{q+1}$$

for some c^{q+1} are cycles, since $\partial\partial = 0$, and are called <u>bounding cycles</u>. The group of q-cycles of $C^q(K)$ mod. the subgroup of bounding q-cycles is called the q-dimensional homology group $H^q(K)$ of $K(P)$. The elements of $H^q(K)$ are called homology classes. Each q-cycle of $C^q(K)$ belongs to a unique homology class. Two q-cycles y and z in the same homology class are called homologous (written $y \sim z$ or $y - z \sim o$). The homology $w \sim 0$ is equivalent to the condition that w be a bounding cycle.

§6. <u>Singular chains in</u> X. Let a non-degenerate q-dimensional simplex $|s|$ be given in some euclidean space. The vertices of $|s|$ taken in a given order $p_o, \ldots, p_q$ define an <u>ordered</u> q-<u>simplex</u>

$$s = (p_o, \ldots, p_q).$$

The ordered (q-1)-simplex

$$s^i = (p_o, \ldots, \not{p}_i, \ldots, p_q)$$

is the face of s <u>opposite</u> p_i. Given tro ordered q-simplices s_1 and s_2 there is a unique affine mapping $B(s_1, s_2)$ of s_1 onto s_2 preserving the order of the vertices.

Let X be a topological space. By a singular q-simplex in X we understand a continuous mapping

$$T : s_1 \longrightarrow X$$

of an ordered q-simplex s into X. Two singular q-simplices

$$T_1 : s \longrightarrow X,$$

$$T_2 : s_2 \longrightarrow X$$

are called <u>equivalent</u> (notation $T_1 \equiv T_2$ provided $T_2\ B(s_1, s_2) = T_1$. One sees that the relation of equivalence is reflexive, symmetric, and transitive. Consequently the totality of all singular q-simplices in X is split into disjoint equivalence classes. We might define $C^q(X)$ as the free abelian group generated by these equivalence classes. For the present,§ $C^q(X)$ will be taken as the group with the singular q-simplices in X as generators, and the equivalences $T_1 \equiv T_2$ as relations. The elements of the group $C^q(X)$ are called <u>integral singular</u> q-<u>chains.</u>

<u>A chain transformation</u> T°. Suppose that we have a mapping

$$T : P \longrightarrow X$$

of the polyhedron P into X. If u_i is a proper q-cell of K(P), a chain $z = n_i u_i$ will be termed <u>proper</u>. Let $C^q(K)_p$ designate the subgroup of $C^q(K)$ composed of proper chains of $C^q(K)$. The mapping

$$T|u_i = T_i ,$$

defined by T over u_i gives a singular simplex in X which we denote by $T^\circ u_i$. We shall extend T° as a chain transformation

(6.1) $$T^\circ : C^q(K)_p \longrightarrow C^q(X) \qquad (q = 0, 1, \ldots)$$

by setting

$$T^\circ (n_i u_i) = n_i T^\circ u_i$$

---

§ This is largely a matter of notational convenience to avoid having a symbol both for a singular simplex and the corresponding equivalence class.

for every proper q-chain $n_i u_i$ of $C^q(K)$.

The boundary $\partial T^\circ z$ can be defined by setting

(6.2) $$\partial T^\circ z = T^\circ \partial z \qquad [z \in C^q(K_p)]$$

On applying (6.1) we have the relation

$$\partial\partial T^\circ z = T^\circ \partial\partial z = 0$$

Thus T° in (6.1) is a homomorphic mapping of $C^q(K)_p$ into $C^q(X)$ such that $\partial T^\circ = T^\circ \partial$.

We apply the preceding to the special case in which the polyhedron P is a non-degenerate geometric q-simplex $|s|$. If $T: s \longrightarrow X$ defines a singular q-simplex $T^\circ s$ in X then

(6.3) $$\partial T^\circ s = T^\circ \partial s = (-1)^i T^\circ s^i$$

in accordance with (6.2) and the definition of ∂s in §5. As previously $\partial\partial T^\circ s = 0$.

The boundary operator ∂ maps the singular q-simplex $T^\circ s$ into the singular (q-1)-chain $T^\circ \partial s$. We call a transformation of singular chains <u>admissible</u> only if the chain images of equivalent singular simplices are equivalent. The chain transformation ∂ as defined by (6.3) is admissible since the equivalence relation

$$T_1^\circ \, s_1 \equiv T_2^\circ \, s_2$$

implies the relation

$$T_1^\circ s_1^i \equiv T_2^\circ s_2^i$$

and accordingly the chain equality

(6.4) $$T_1^\circ \, \partial s_1 = T_2^\circ \, \partial s_2 .$$

Thus

$$\partial T_1^\circ s_1 = \partial T_2^\circ s_2$$

in accordance with (6.3) and (6.4). The set of singular simplices in X will be called a complex S(X).

§7. The deformation homomorphism $\bar{\bar{D}}$. Before coming to D we need certain preliminary constructions.

Given a finite simplicial polyhedron P let v be one of the vertices of P and P_v the closed star of simplices of P incident with v. Let σ be any q-cell of $K(P_v)$. Let $v\sigma$ denote the (q+1)-cell of $K(P_v)$ obtained by writing out the array of vertices defining σ and placing v in front of them all. If c is a q-chain $n_i\sigma_i$ of $K(P_v)$ one defines the join of v with c as

(7.1) $$vc = n_i(v\sigma_i) .$$

It is clear for q-cells and hence for q-chains that

(7.2) $$\partial(vc) = c - v\partial c$$

When c is a cycle $\partial(vc) = c$.

The prism A with base a. Let a be a non-degenerate geometric q-simplex and A a right prism of unit height with bases a and a'. We suppose that A lies in a (q+1)-dimensional euclidean space. Points x and x' of a and a' are said to correspond if x' projects orthogonally into x. Similarly a geometric sub-simplex b of a has a correspondent b' on a'. The subset of A which projects orthogonally into b is called the lateral face B with base b. Let v(b) denote the barycenter of B.

The prism A will be subdivided into geometric simplices as follows. The lateral faces B will be subdivided in the order of their dimensions by taking the simplices of B as the straight line join of v(b) with the respective simplices on the geometric boundary of B. In this process the bases a and a' are not subdivided. Let A now stand for the subdivided prism and B the subdivided lateral face with base b.

The base a of A is a polyhedron and the complex K(a) of §6 is well defined; similarly K(A). We shall define a chain homomorphism

$$(7.3) \qquad d: C^q[K_p(a)] \longrightarrow C^{q+1}[K_p(A)] \qquad (q = 1, 2, \ldots)$$

where the subscript p indicates limitation to proper chains. Let y be a q-cell and z a q-chain of $C^q[K_p(a)]$. The image dz of z will be defined by induction with respect to the dimension q in such a fashion that

$$(7.4) \qquad dy = v(y)[y' - y - d\partial y]$$

$$(7.5) \qquad \partial dz = z' - z - d\partial z$$

In (7.4) the notation of a "join", as developed in (7.1) and (7.2), is used, and in (7.5) z' is the image of z under the orthogonal projection of the base a' into the base a.

When q = 0, $\partial y = \partial z = 0$. If one defines dy by (7.4), $\partial dy = y' - y$ in accordance with (7.2). If $z = n_i y_i$ one sets

$$(7.6) \qquad dz = n_i dy_i \qquad (q = 0)$$

and (7.5) holds as stated. Assuming then that (7.4) and (7.5) hold for (q-1)-cells and chains respectively, let y and z be q-dimensional. One defines dy using (7.4), thus setting $dy = v(y)c$ where $c = y' - y - d\partial y$. But

$$\partial c = \partial y' - \partial y - \partial d\partial y = \partial y' - \partial y - [\partial y' - \partial y - d\partial\partial y] = 0$$

on using (7.5), so that c is a cycle. Hence $\partial y = \partial(vc) = c$ and (7.5) holds for $z = y$. On extending the definition of d as in (7.6), (7.5) holds in general. We have thus established the lemma.

Lemma 7.1. The chain homomorphism d of (7.3) satisfies the relation

$$\partial dz = z' - z - d\partial z \ . \tag{7.7}$$

Let I stand for the (time) interval $0 \leq t \leq 1$ and let the undivided prism A be represented as a product a x I. Suppose that a mapping

$$R : a \times I \longrightarrow X \tag{7.8}$$

is given. If s is an ordered q-simplex with the vertices of the geometric q-simplex a, then $R^\circ ds$ will be a singular (q+1)-chain in X. As seen in (6.2)

$$\partial R^\circ ds = R^\circ \partial ds$$

and it follows from the preceding lemma that

$$\partial(R^\circ ds) = R^\circ s' - R^\circ s - R^\circ(d\partial s) \tag{7.9}$$

The deformation D. Suppose that the image under R of a point (x, t) in a x I is R(x, t). For fixed t in I let

$$R^t : a \longrightarrow X$$

be a mapping in which $R^t(x) = R(x, t)$. Let

$$D : X \times I \longrightarrow X$$

be a deformation of X with D(x, t) the image of (x, t) in X x I. For fixed t let

$$D^t : X \longrightarrow X$$

be a mapping in which $D^t(x) = D(x, t)$. Let

$$T : s \longrightarrow X$$

define a singular simplex $T^\circ s$. Let a mapping $R: a \times I \longrightarrow X$ be defined by setting

$$R^t = D^t\, T\colon a \longrightarrow X \qquad (0 \leq t \leq 1)$$

for each fixed t. So defined R^t can be regarded as a "deformation" of the mapping T, since $R^0 = T$ and T is "replaced" at the time t by D^tT.

Given the singular q-simplex $\sigma = T\circ s$ and the deformation D the singular (q+1)-chain $R\circ ds$ is determined, provided $R^t = D^tT$. This mapping of q-cells $\sigma = T\circ s$ in $C^q(X)$ into (q+1)-chains in $C^{q+1}(X)$ can be extended linearly to define a homomorphism

$$(7.10) \qquad \bar{\bar{D}} : C^q(X) \longrightarrow C^{q+1}(X)$$

we designate the image of z in $C^q(X)$ by $\bar{\bar{D}}z$. In particular our notation implies that

$$R\circ ds = \bar{\bar{D}}\sigma \qquad (R^t = D^tT)$$

We write

$$(D^1T)\circ s = \bar{D}\sigma$$

and term $\bar{D}\sigma$ the <u>final image of σ</u> under D. Linearly extended over $C^q(X)$, $\bar{D}$ is a chain transformation

$$\bar{D} : C^q(X) \longrightarrow C^q(X)$$

which replaces a q-chain c in $C^q(X)$ by $\bar{D}c$ in $C^q(X)$. From (7.9) we have

$$(7.11) \qquad \partial\bar{\bar{D}}\sigma = \bar{D}\sigma - \sigma - \bar{\bar{D}}\partial\sigma$$

and by virtue of the linear extension of $\bar{D}$ and $\bar{\bar{D}}$

$$\partial\bar{\bar{D}}c = \bar{D}c - c - \bar{\bar{D}}\partial c$$

where $\bar{D}c$ is called the <u>final image</u> of c under D.

<u>The admissibility of $\bar{\bar{D}}$.</u> As previously indicated $\bar{\bar{D}}$ would not be admitted as an admissible homomorphism (7.10) unless it were constant over each equivalence class of singular-q-simplices. That is if $T_1\circ s_1 = T_2\circ s_2$ we must show,

subject to the equivalence relations, that

$$\bar{\bar{D}}T_1^\circ s_1 = \bar{\bar{D}}T_2^\circ s_2 \tag{7.12}$$

To that end let $B: s_1 \longrightarrow s_2$ be the unique linear mapping of s_1 into s_2. As previously we define mappings

$$R_1: s_1 \times I \longrightarrow X \qquad R_2: s_2 \times I \longrightarrow X$$

such that

$$R_1^t = D^t T_1: s_1 \longrightarrow X \quad R_2^t = D^t T_2: s_2 \longrightarrow X .$$

If $s_1 \longrightarrow s_2$ under B, and if I is transformed onto I under the identity, a mapping

$$s_1 \times I \longrightarrow s_2 \times I$$

is induced under which $R_1 = R_2$. The chain equality

$$R_1^\circ ds_1 = R_2^\circ ds_2$$

follows and this is equivalent to (7.12).

We have thus proved the following theorem.

<u>Theorem 7.1.</u> <u>Corresponding to a continuous deformation</u> D <u>of</u> X <u>on</u> X <u>there exists a homomorphic mapping</u>

$$\bar{\bar{D}}: C^q(X) \longrightarrow C^{q+1}(X)$$

<u>such that the image of a singular simplex depends only on the equivalence class and such that for any chain</u> c <u>in</u> $C^q(X)$

$$\partial\bar{\bar{D}}c = \bar{D}c - c - \bar{\bar{D}}\partial c$$

<u>where</u> $\bar{D}c$ <u>is the "final image" of</u> c <u>under</u> D. <u>If</u> c <u>is in a subset</u> $X_1 \subset X$ <u>then</u> $\bar{\bar{D}}c$ <u>is in the trajectory of</u> X_1 <u>under</u> D.

§8. <u>The barycentric subdivision</u> Bz <u>of a chain</u> z. What is needed here is to know that any singular cycle z is homologous on any subset of X in which it lies, to a suitably defined barycentric subdivision Bz.

Let a be a non-degenerate geometric q-simplex. The geometric barycentric subdivision βa of a is defined as follows. If a is 0-dimensional, let $\beta a = a$. Suppose βa has been defined for $q < k$. If a is a k-simplex βa shall be the polyhedron obtained by joining the barycenter of a to the (k-1)-simplices of the barycentric subdivision of the original boundary (k-1)-simplices.

We shall now define a chain transformation

$$B : K_p(a) \longrightarrow K_p(\beta a)$$

which carries each proper k-cell u of K(a) into a proper[§] k-chain

(8.1) $$Bu = \sum u_i \qquad (i = 1,\ldots,k+1)$$

where[*] $|u_1|, \ldots, |u_{k+1}|$ is the set of geometric simplices of $\beta|u|$, and where the linear extension of B to proper k-chains z of K(a) has the property that $\partial Bz = B\partial z$.

One defines Bu by an induction with respect to the dimension of u, setting Bu = u when u is a 0-cell. Suppose that Bz has been defined for all proper r-chains of K(a) for which $r < k$. Let u be a proper k-cell of K(a) with barycenter v. On using the vertex join of §7 one sets

(8.2) $$Bu = vB(\partial u)$$

and then extends B linearly over proper k-chains of K(a). We can show that

(8.3) $$\partial(Bu) = B(\partial u)$$

The proof of (8.3) is by an induction with respect to dimension proving at the same time

---

§ $K_p(a)$ is the complex of proper cells of K(a).

* As previously $|u|$ shall denote the geometric simplex of least dimension whose vertices include the vertices of u.

(8.4) $$\partial(Bz) = B(\partial z)$$

for any proper k-chain z in $\dot{K}(a)$. The starting point is the formula (Cf. (7.2)).

(8.5) $$\partial(Bu) = B(\partial u) - v\partial[B(\partial u)]$$

When u is 0-dimensional, $\partial u = 0$. The induction is automatic, using (8.4) and the relation

$$\partial[B(\partial u)] = B[\partial(\partial u)] = 0,$$

assumed valid when the dimension of ∂u is less than k.

<u>The barycentric subdivision of singular chains.</u> Let a mapping

$$T: s \longrightarrow X$$

define a singular k-simplex $\sigma = T^\circ s$. The barycentric subdivision $B\sigma$ of σ is a singular k-cell defined by the right member of the equation

(8.6) $$BT^\circ s = T^\circ Bs$$

One then extends B as a chain transformation§

$$B: S(X) \longrightarrow S(X)$$

by setting $B(n_i\sigma_i) = n_i B\sigma_i$. One can prove that

(8.6) $$\partial B\sigma = B\partial\sigma$$

as follows. We have

$$\begin{aligned}
\partial B\sigma &= \partial T^\circ Bs && \text{(by definition of } B\sigma) \\
&= T^\circ \partial Bs && \text{(commuting } T^\circ \text{ and } \partial) \\
&= T^\circ B\partial s && \text{(using (8.3))} \\
&= BT^\circ \partial s && \text{(by linear extension of (8.6))} \\
&= B\partial\sigma && \text{(by definition of } \partial\sigma)
\end{aligned}$$

§ That B is constant over any equivalence class follows from the fact that an affine transformation which carries s into a second non-degenerate ordered simplex s' carries the barycenter of s into that of s'.

We have proved the following lemma.

Lemma 8.1. The barycentric subdivision Bz of a singular k-chain is in any subset of X in which z lies, and has the property that $\partial Bz = B\partial z$.

The chain homomorphism ρz. To establish the necessary homologies (with moduli §9) between a singular k-chain z and its barycentric subdivision Bz we shall need the following theorem. Cf. Eilenberg, op. cit.

Theorem 8.1. There exist chain homomorphisms

$$\rho ; C^k(X) \longrightarrow C^{k+1}(X) \qquad (k = 0, 1, \ldots)$$

such that the image ρz of any k-chain of $C^k(X)$ is in any subset of X in which z lies and satisfies the relation

$$(8.7) \qquad \partial\rho z = Bz - z - \rho\partial z .$$

The prism A based on a. The proof depends upon a subdivision of the prism a x I used in §7. In addition to the subdivision of the lateral faces of this prism introduced in §7, we here subdivide the "top base" a', replacing a' by its barycentric division $\beta a'$ and then joining the barycenter of the prism a x I to the simplices on its subdivided boundary. The base a remains undivided. Let A denote the prism so subdivided. Exactly as in §7 we can prove the following lemma.

Lemma 8.2. There exists a chain homomorphism

$$d^*: C^k[K_p(a)] \longrightarrow C^{k+1}[K(A)] \qquad (k = 1, 2, \ldots)$$

which carries any proper k-chain z of $C^k[K(a)]$ into a proper (k+1)-chain d^*z of $C^{k+1}[K(A)]$ such that

$$(8.8) \qquad \partial d^*z = Bz' - z - d^*\partial z .$$

Here z' is the k-chain in $K(a')$ which projects orthogonally into z in $K(a)$.

Let π be the mapping

$$\pi : A \longrightarrow a$$

which projects each point of A orthogonally into a point of a. Let s be an ordered q-simplex such that $|s| = a$. Let

$$T : s \longrightarrow X$$

define a singular q-cell $\sigma = T^\circ s$. The product $R = T\pi$ maps $A \longrightarrow X$, and induces a chain homomorphism

$$\rho : C^q(X) \longrightarrow C^{q+1}(X)$$

in which the k-cell $\sigma = T^\circ s$ goes into the (k+1)-cell

$$\rho\sigma = R^\circ d^* x$$

Upon applying the chain transformation R° to the two members of (8.8) (setting $y = s$) we find that

(8.9) $$\partial\rho\sigma = B\sigma - \sigma - \rho\partial\sigma$$

In this application of R° to Bs' in (8.8) we have used the relations

$$R^\circ Bs' = R^\circ Bs = (T\pi)^\circ Bs = T^\circ Bs = B\sigma$$

The equality (8.7) follows from (8.9) on extending ρ linearly to k-chains of $C^k(X)$.

Theorem 8.1 follows.

§9. Singular k-chains over a group G. It will now be convenient to regard the equivalence classes of singular k-simplices in the topological space X as the k-cells σ in $S(X)$. Let $S^k(X)$ be the subset of such k-cells in $S(X)$. Let G be an abelian group with elements g. A function

(9.0) $$z : S^k(X) \longrightarrow G$$

is called a singular k-chain over G provided all but a finite set of values of z in G are null. The value $z(\sigma)$ may be called the coefficient of σ. If one adds k-chains z and w by adding coefficients $z(\sigma)$, $w(\sigma)$ for each σ in $S^k(X)$, one has an abelian group $C^k(X, G)$ "over G". A k-chain z is regarded as null if each coefficient is null. A k-chain may be given a representation

$$z = g_i\sigma_i \qquad (i = 1,\dots,n)$$

in which the terms written include all σ_i with non-null coefficients.

Equivalently and nearer the historic origins one could define a k-chain z in $C^k(X, G)$ by appropriately enumerating the above properties in inverse order, omitting the statement that z is a function of the type (9.0) since it is a consequence of the enumerated properties. The term "the value of z at σ", used in function theory, parallels the term "the coefficient of σ in z" used in a formal symbolic group approach.

Given a finite simplicial polyhedron P and the complex K(P) one can similarly define the groups $C^k(K, G)$ of k-chains of K(P) over G.

If σ is a k-cell in S(X), $\partial\sigma$ is an "integral" (k-1)-chain in $C^k(X)$ with coefficients 1, -1, or 0 in the group of integers. If one understands that

$$g(-\sigma) = -g\sigma$$

the definition

(9.1) $$\partial(g_i\sigma_i) = g_i(\partial\sigma_i)$$

is meaningful and yields a (k-1)-chain in $C^{k-1}(X, G)$. On using (9.1) and the fact that $\partial\partial\sigma_i = 0$ one finds that

$$\partial(\partial z) = \partial g_i(\partial\sigma_i) = g_i\partial\partial\sigma_i = 0$$

From the definition of sums and of ∂ it appears that $\partial(z + w) = \partial z + \partial w$ for arbitrary k-chains z and w.

The definition over G and in S(X) of bounding k-cycles, the homology group $H^k(X, G)$, homology classes, etc., is formally the same as for chains over the group I of integers.

The case of a field G. When G is a field, $C^k(X, G)$ is an abelian group with operators (See van der Waerden, Moderne Algebra I, p.132). In addition to the properties of k-chains already enumerated we assume that for a_i, b_i in G

$$a_i(b_i\sigma_i) = (a_ib_i)\sigma_i \qquad (i = 1,\ldots,n)$$

The only subgroups C of k-chains to be admitted are those for which az is in C when a is in G and z in C.

A base for an operator subgroup C is a set Z of elements such that any element of C is linearly dependent (using coefficients in G) upon some finite subset of elements of Z. The number of elements in a minimum§ base is called the rank or dimension of C. In the cases to be considered each group has a countable base, so that its dimension is either a finite integer or the smallest infinite cardinal. The dimension of the homology group $C^k(X, G)$ is called the Betti number R_k of X over G.

A special field. A very simple, finite field is the field GF(2) of integers mod. 2. In this case the boundary in K(P) of an ordered simplex s is

$$\partial s = \sum s^i$$

where the coefficient 1 has been identified with the unit in GF(2). The boundary of a linear combination z over GF(2), of distinct ordered k-simplices

§ For the cases at hand the existence of a minimum base is immediate.

of K(P), is the sum of the (k-1)-simplices which are incident with an odd number of k-simplices of z. In other words bounding relations, mod. 2 do not depend on the ordering of the vertices of the simplices.

Cycles mod. A. Let A be a subset of X. A k-chain w in X whose boundary ∂w is in A is called a k-cycle mod. A. If u is a (k-1)-chain in X one writes

$$u \sim 0 \qquad \text{(mod. A)}$$

if there exists a k-chain w such that

$$\text{(9.2)} \qquad \partial w = u - v \qquad \text{(with v in A).}$$

Since

$$\text{(9.3)} \qquad 0 = \partial\partial w = \partial u - \partial v$$

u is a (k-1)-cycle mod. A. We state a lemma of frequent use.

Lemma 9.1. If u is (k-1)-chain with $u \sim 0$ mod. A, then ∂u bounds in A.

This follows at once from (9.3) since v is a k-chain in A whose boundary equals ∂u, and v is in A.

An example. Lemma 9.1 can be illustrated as follows. Let X be half of a solid torus in the space of coordinates (x, y, z). More specifically let d be the circular disc

$$(x - 2)^2 + y^2 \leq 1 , \qquad z = 0$$

and let the (x, y) plane be revolved about the x axis through 180° so that d generates a solid half torus X on which $z \geq 0$. Let d' be the final image of d in X. Let X be subdivided into curvilinear simplices by first dividing the circular boundaries of d and d', then d and d', next the semi-torus S on the boundary of X, and finally X itself. Each simplex x of X may be regarded as the "carrier" of a singular simplex obtained by a 1 - 1 mapping of x into a euclidean simplex s. In (9.2) one can take w as the sum mod. 2

of the 3-dimensional singular simplices of X, u and v as similar sums mod. 2 over S and d + d' respectively. With A taken as the set on which $z < 1$, (9.2) holds mod. A and mod. 2. The conclusion of the lemma, that ∂u bounds mod. 2 in A, is obvious, since ∂u is the sum of the singular simplices of the circular boundaries of d and d', and bounds in the plane $z = 0$.

The carrier of a singular k-chain z. If $T: s \longrightarrow X$ is a singular cell σ in X the image of $|s|$ under X will be called the carrier $X(\sigma)$ of σ. The carrier of a chain $z = g_i \sigma_i$ $(i = 1,\ldots,n)$ in which the σ_i represent distinct equivalence classes and no $g_i = 0$ is defined as

$$X(z) = \text{Union } X(\sigma_i) \qquad (i = 1,\ldots,n)$$

Recall that the trajectory of a subset X_1 of X under a deformation D of X is the union of the trajectories of the points of X_1. It follows from the definition of the (k+1)-chain $\bar{\bar{D}}z$ that $\bar{\bar{D}}z$ is in the trajectory of X(z). With this understood we can state the following lemma.

Lemma 9.2. Let z be a singular cycle mod. A in X. If there exists a deformation D of X such that the final image $\bar{D}z$ of z is in A, and if the carrier of ∂z is deformed in A, then $z \sim 0$ mod. A.

In accordance with Theorem 8.1

$$\partial \bar{\bar{D}} z = \bar{D}z - z - \bar{\bar{D}} \partial z \tag{9.4}$$

By hypothesis $\bar{D}z$ is in A, and $\bar{\bar{D}}\partial z$ is in A since the carrier of ∂z is deformed in A. That $z \sim 0$ mod. A now follows from (9.4) on referring to the definition (9.2) of an homology $z \sim 0$ mod. A.

The following elementary lemma requires explicit statement.

Lemma 9.3. If a continuous mapping

$$S: X \longrightarrow Y$$

of a topological space X into a topological space Y is given, a chain homomorphism

$$S^*: C^k(X, G) \longrightarrow C^k(Y, G)$$

is induced in which a k-chain z in X is replaced by a k-chain S^*z in Y, while

$$\partial S^* z = S^* \partial z$$

To define S^* let $T: s \longrightarrow X$ be a singular k-cell, $\sigma = T^\circ s$ in X. The image of σ under S^* is defined by the right member of the equation

$$S^*(T^\circ s) = (ST)^\circ s,$$

and this mapping of cells shall be linearly extended to chains. We have

$$\partial S^* \sigma = \partial (ST)^\circ s = (ST)^\circ \partial s$$

$$= (-1)^i (ST)^\circ s^i = (-1)^i S^*(T^\circ s^i) = S^* \partial \sigma.$$

The linear extension of this relation leads to the equation $\partial S^* z = S^* \partial z$, and the proof of the lemma is complete.

Any chain homomorphism S^* of the type of the lemma in which $\partial S^* = S^* \partial$, carries homologous k-cycles in X into homologous k-cycles in Y. For a bounding relation $z = \partial w$ in X implies the relation

$$S^* z = S^* \partial w = \partial S^* w$$

in Y, so that $S^* z \sim 0$ if $z \sim 0$. Thus S^* induces a homomorphic mapping of the k-th homology group of X over G into the k-th homology group of Y over G.

The basic isomorphism β. If P is a finite simplicial polyhedron, K(P) is the complex of all ordered q-simplices s such that the vertices of s lie in some non-degenerate simplex of P. A singular q-simplex σ of S(P) will be termed the <u>projection</u> of a proper simplex s of K(P), if σ is defined by the identity mapping T = I of s onto s, and conversely s will be termed the projection of σ. A singular k-chain

$$z = g_i\sigma_i \qquad (i = 1,\ldots,n)$$

in which σ_i is the projection of a proper simplex s_i of K(P) will be said to be the projection in S(P, G) of the chain $g_i s_i$ in K(P, Q), and conversely $g_i s_i$ will be said to be the projection in S(P, G) of $g_i\sigma_i$ in K(P, G).

The following theorem is fundamental.

<u>Theorem 9.1.</u> <u>There exists a chain transformation</u>

$$\beta : K(P, G) \longrightarrow S(P, Q)$$

<u>in which</u>[§] <u>each proper</u> k-<u>chain of</u> K(P, G) <u>corresponds to its projection in</u> S(P, G), <u>and under which the</u> k-<u>th homology group of</u> K(P, G) <u>is mapped isomorphically onto the</u> k-<u>th homology group of</u> S(P, G). <u>There is a proper</u> k-<u>cycle in each homology class of</u> K(P, G) <u>with a projection in</u> S(P, G) <u>in the corresponding homology class of</u> S(P, G).

This theorem is a consequence of theorems in Eilenberg's paper, <u>op. cit.</u>

The following is the principal application which we shall make of the preceding theorem. Let a_{n+1} be a non-degenerate n+1-simplex ($n > 0$), G a field, and Q_n the simplicial polygon bounding a_{n+1}. Elementary algebraic methods suffice to show that the dimensions r_k of the respective k-th homology groups of $K(Q_n, G)$ are zero, except that $r_o = r_n = 1$. It follows from the preceding theorem that the Betti number R_k of $S(Q_n, G)$ is similarly zero except that

$$R_o = R_n = 1$$

§ This is not to imply that β is limited to the proper k-chains of K(P, G).

In particular there is a non-bounding 0-cycle c_o and a non-bounding n-cycle c_n in $S(Q_n, G)$ which form bases for the homology groups, $H^\circ(Q_n, G)$ and $H^n(Q_n, G)$ respectively.

Let s_{n+1} be an ordered n+1-simplex such that $|s_{n+1}| = a_{n+1}$ and let I be the identity mapping of a_{n+1} onto a_{n+1}. It is of interest, but not essential, to know that c_n can be taken as the singular n-cycle $I\partial s_{n+1}$. If $v \in Q_n$ is an arbitrary point and is regarded as an ordered 0-simplex, c_o can be taken as $I^\circ v$.

It follows from the preceding that if X is the topological image of an n-sphere ($n > 0$) then the Betti numbers of $S(X, G)$ are all zero except the numbers $R_o = R_n = 1$.

In general the Betti numbers which will appear in the later sections will be obtained with the aid of the critical point theory rather than by reference to a polyhedral complex. One may say that we approach the homology groups synthetically rather than analytically. The subspaces W_c of W are obtained by letting c increase from the absolute minimum value of the function J. One arrives at a knowledge of the homology groups of the spaces W_c by a synthesis of the topological changes observed in W_c as c increases and not by a simplicial decomposition and concurrent analysis of a given space W_c. The synthetic method has possibilities in topology which have not been fully exploited.

§10. Isomorphisms between homology groups. Let X be a topological space and A a subset of X termed an associated modulus. An n-cycle in X mod. A will be termed a relative (written rel.) n-cycle in X. Similarly, "bounding in X mod. A" is termed "rel. bounding in X", or "rel. homologous to zero".

If Y is a subset of X it will be convenient to associate $A \cap Y$ with Y as a modulus, so that a rel. cycle in Y is a rel. cycle in X, but in general not conversely.

Lemma[§] 10.1. Let X be a topological space with a modulus A, and Y a subspace of X with a modulus $A \cap Y$. Let U be an arbitrary rel. homology class of X, and U' the subclass of chains of U in Y. If

(a) each rel. cycle in X is rel. homologous in X to a rel. cycle in Y, and

(b) each rel. cycle in Y which is rel. bounding in X is rel. bounding in Y, then U' is a rel. homology class of Y, and the mapping $U \longrightarrow U'$ defines an isomorphism between the respective rel. homology groups of X and of Y.

We must establish the following.

(1) The class U' is not empty.

(2) Chains in U' are in the same rel. homology class of Y.

(3) The class U' is a complete rel. homology class of Y.

(4) The mapping $U \longrightarrow U'$ is 1 - 1.

(5) If V is a second rel. homology class of X then

(10.1) $$U' + V' = (U + V)'$$

The proof is as follows:

(1) Statement (1) is implied by (a).

(2) Let x and y be chains in U'. Then x and y are in U and $x - y$ is rel. bounding in X. By virtue of (b), $x - y$ is rel. bounding in Y so that (2) follows.

§ The lemmas of this section are not found in Eilenberg, *op. cit.* Similar lemmas in less abstract form were used in the early papers of the writer.

(3) If x is in U' and z is in the same rel. homology class of Y, then x - z is rel. bounding in Y, and hence rel. bounding in X. Thus z is both a chain of U and in Y. Hence (3) holds.

(4) The class U' cannot be a subclass of distinct classes U and V.

(5) To establish (10.1) it is sufficient to show that there is a chain common to both members of (10.1). Let x abd y be chains in U' and V' respectively. The mode of adding homology classes implies that x + y is in U'+ V'. But x + y is both in U + V and in Y. Hence x + y is in (U + V)', and (10.1) follows.

The proof of the lemma is complete.

If R_k is the k-th Betti number of a topological space X, there exists a set (z) of R_k k-cycles such that an arbitrary k-cycle is homologous in X to some finite sum $g_i z_i$ where z_i is in (z) and g_i in G. One then terms (z) a minimal homology base for k-cycles in X over G. Conversely, the number of cycles in a minimal homology base for k-cycles will be R_k. A minimal homology base for k-cycles in X mod. A is similarly defined.

The following lemma will be used in conjunction with Lemma 10.1.

Lemma 10.2. *If* K *is an* r-*disc with center* p, *the Betti number* R_k *of* K *mod.* K - p *is* δ_r^k . (k = 0, 1, ...).

The proof of this lemma will be left to the reader. Facts useful in the proof are that the Betti numbers of the spherical boundary of K are all zero except the 0-th and the (r-1)-st, which are 1; the Betti numbers of K are all zero except the 0-th; there exists a continuous deformation of K - p on itself into the spherical boundary of K.