The calculus of variations in the large

By Marston Morse, Cambridge, U. S. A.

Introduction. The theory to be presented here may be described as the analogue for functionals of the author's theory[9] of critical points[5] of a function f of n variables. The latter theory proceeds by virtue of three principal steps.

(1). The determination of the topological characteristics of the domain R of definition of the function.

(2). The assignment of type numbers to the critical sets.

(3). The determination of the relations between the topological characteristics of R and the type numbers of the critical sets of f.

Relative to step (1) we say simply that for functions (but not for functionals) an adequate topological theory is already at hand.

Relative to step (2) we remark that in case the function is analytic and the domain is a finite, analytic, regular n-manifold without boundary, the critical sets are necessarily finite in number. In case a critical set is an isolated point P and the hessian of f does not vanish at P (the non-degenerate case) type numbers M_k are assigned by the author to P as follows. Let k be the index of the quadratic form whose coefficients are the elements of the hessian of f at P. Corresponding to P we take all of the type numbers as zero except M_k and take M_k as 1. In the case of the general critical set ω, type numbers are assigned to ω in a way which is a generalization of the assignment in the non-degenerate case. If N_i is now the sum of the type numbers M_i assigned to the different critical sets ω of f, we have shown that

$$N_i \geqq R_i \tag{0.1}$$

where R_i is the ith connectivity (mod. 2) of the domain R. The assignment of type numbers to the general critical set ω is one which has the following remarkable property. If the function f is approximated sufficiently closely by a non-degenerate function F (one whose critical points are all non-degenerate), there will always appear, neighboring a critical set of f, a set of isolated, non-degenerate critical points of F whose type number sums ΣM_i are at least as great as the respective numbers M_i which we have assigned to ω. Moreover it can be shown that the function f can always be approximated by a non-degenerate function F.

The relations (0.1) establish the existence of the *topologically necessary critical points*. If $N_i > R_i$ for a particular i, we say that there exists an *excess* of critical points of index i. This excess is not arbitrary. In fact the numbers N_i and R_i must satisfy the relations

$$
(0.2)\qquad
\begin{aligned}
N_0 &\geqq R_0\\
N_0 - N_1 &\leqq R_0 - R_1\\
N_0 - N_1 + N_2 &\geqq R_0 - R_1 + R_2\\
&\dots\dots\dots\dots\\
N_0 - N_1 + \dots + (-1)^n N_n &= R_0 - R_1 + \dots + (-1)^n R_n.
\end{aligned}
$$

We call the relations (0.2) *limitations on the excess of critical points*[16]. These relations are complete in a sense upon which we shall not here enlarge.

We now turn to the calculus of variations. Corresponding to a given set of boundary conditions B, we shall regard a Jordan arc which satisfies the conditions B as a point in function space. The conditions B might, for example, require the end points of the arc to be fixed, or to lie on two end manifolds. The totality of the Jordan arcs satisfying conditions B will be termed the *functional domain Ω associated* with B. Step (1) now calls for new developments. We proceed by defining bounding on Ω, cycles on Ω, and the connectivities R_i of Ω ($i = 0, 1, \dots$). In contrast with ordinary topology we find that infinitely many of the connectivities of Ω are in general not null.

One naturally replaces the function f by the calculus of variations integral, and the critical points of f by critical extremals, that is, extremals satisfying the transversality conditions associated with the boundary problem B. We restrict ourselves to the analytic case. Basic critical sets ω are here maximal sets of critical extremals, mutually deformable into one another among extremals of ω. We solve the difficult problem of characterizing the non-degenerate critical point and of assigning type numbers to each critical set. In general there will be infinitely many critical sets. The value of the integral on extremals of a critical set ω will be constant. In the analytic case the critical values will be isolated.

Whereas the type numbers of a critical set ω of an ordinary function f consist of a set of $n + 1$ numbers, here the type numbers of ω consist of an infinite set of numbers

$$M_0, M_1, \dots$$

For any particular set ω all but a finite subset of these numbers M_i will be null. But the indices i of the non-null type numbers M_i in general will not be bounded for all critical sets ω of the problem. This corresponds to the fact that the indices i of non-null connectivities R_i of the function manifold Ω are in general not bounded, and that the relations (0.1) still hold. These relations are very powerful and will in general prove the existence of infinitely many critical extremals. The limitations (0.2) on the excess of critical extremals now become an infinite set of inequalities of the same form and produce a host of remarkable relations.

As a matter of pure topology, the connectivities of our functional domain reduce upon suitable specialization of the boundary conditions and the integral to the ordinary connectivities of an ordinary manifold, or if we please, to the connectivities of a product manifold. Other simple specializations are possible.

The present paper is necessarily limited. The special case of the periodic extremal is not treated here. References to that part of the theory which has been published are given at the end. A more complete treatment will be given in the Colloquium Lectures on "The Calculus of Variations" given in September, 1931 and to be published shortly. Other references are given at the end.

1. *An integral on a Riemannian space.* In this section we shall give a definition of a Riemannian space in the large and define an integral thereon.

Riemannian spaces as ordinarily defined are local affairs. It is necessary for us to add topological structure in the large. To that end we suppose that we have given an ordinary m-dimensional simplicial circuit K_m in an auxiliary Euclidean space on which the neighborhood of each point is well defined. (See Lefschetz ref. 13.) Our Riemannian space R will now be defined as follows. Its points and their neighborhoods shall be the one-to-one images of the respective points and their neighborhoods on K. Moreover K_m shall be of such nature that the neighborhood of each of its points is homeomorphic with the neighborhood of a point (x) in an Euclidean m-space of coordinates $(x) = (x_1, \ldots, x_m)$. With at least one such representation of the neighborhood of a point of R there shall be associated a positive definite Riemannian form

$$ds^2 = g_{ij}(x)\, dx^i\, dx^j \qquad (i, j) = (1, \ldots, m) \tag{1.1}$$

defining a metric[14] for the neighborhood. For simplicity we suppose that the coefficients $g_{ij}(x)$ are analytic. We term the coordinates (x) *admissible*. We also admit any other set of local coordinates (z) obtainable from admissible coordinates (x) by a transformation of the form

$$z^i = z^i(x), \tag{1.2}$$

defined by functions $z^i(x)$ analytic in (x) and possessing a non-vanishing Jacobian. We require that any two coordinate systems (x) and (z) which admissibly represent the neighborhood of the same point P must be related as in (1.2). We term (1.2) an admissible transformation of coordinates.

A set of points of R will be said to form a regular analytic n-manifold on R if the images of its points in any local coordinate system (x) is locally representable in the form

$$x^i = x^i(u_1, \ldots, u_n)$$

where the functions $x^i(u)$ are analytic in the parameters (u) and the functional matrix of the functions $x^i(u)$ is of rank n.

With each admissible coordinate system (x) we suppose that we have given a function

$$F(x, r) = F(x^1, \ldots, x^m, r^1, \ldots, r^m) \qquad (r) \pm (0)$$

serving to define an invariant integral

$$\int F(x, \dot{x})\, dt$$

in that system, where $\overline{x^i}$ represents the derivative of x^i with respect to a parameter t.

For at least one admissible coordinate system neighboring each point of R (and consequently for all such coordinate systems) we assume that the corresponding integrand is *positive* and *analytic* in (x) and (r) for $(r) \neq (0)$. We also assume that $F(x, r)$ is positively homogeneous of order 1 in the variables (r), as in the classical theory of Weierstrass. We further assume that the form

$$F_{r^i r^j}(x, r)\, \lambda^i \lambda^j \qquad (i, j = 1, \ldots, m) \tag{1.3}$$

is positive for unit vectors $(\lambda) \neq (r)$.

2. *The functional domain and its connectivities.* Let A^1 and A^2 be any two distinct points on R. The points A^1 and A^2 are to serve as the initial and final points respectively of admissible curves. Locally we represent A^1 and A^2 by sets of coordinates

$$(x) = (x^{11}, \ldots, x^{m1}) \qquad (x) = (x^{12}, \ldots, x^{m2})$$

respectively. Let B_r be an auxiliary simplicial r-cycle with $0 \leqq r < 2m$. Suppose that we have a set of pairs of points (A^1, A^2) homeomorphic with B_r. We understand that $A^1 \neq A^2$ and that A^1 and A^2 lie on R. For $r > 0$ we suppose that the neighborhood of each point of B_r can be represented as the one-to-one continuous image of a neighborhood of a point (a_0) in our auxiliary Euclidean r-space of r coordinates (a) and that the corresponding points (A^1, A^2) can be locally represented in the form

$$x^{is} = x^{is}(a) \qquad (i = 1, \ldots, m;\ s = 1, 2)$$

where the functions $x^{is}(a)$ are analytic in (a) for (a) near (a_0) and possess a functional matrix of rank r. We call this set of points (A^1, A^2) the *terminal manifold* Z.

Let γ be a sensed Jordan arc on R on which a parameter t runs from 0 to 1 inclusive. If the respective end points of γ determine a point (A^1, A^2) on Z, γ will be termed *topologically admissible.* When we are dealing with our integral we shall suppose that γ is of class D^1 as well as topologically admissible. We then term γ a *restricted* curve. Evaluated on restricted curves the integral defines our *basic functional* J.

The totality of topologically admissible curves will be termed the functional domain Ω *corresponding to the space* R *and terminal manifold cape.*

By a j-complex k_j on Ω we mean a j-parameter family of curves of Ω given as the continuous map of the product $c_j \times t$ of an auxiliary j-complex c_j, and the line segment $0 \leqq t \leqq 1$, with each curve γ corresponding to the product of a point of c_j and the line segment. As in the theory of singular complexes we regard points on these curves as distinct if they are images of distinct points on $c_j \times t$. In this family the curves corresponding to the product of a j-cell of c_j and the line segment t will define a family of curves which we call a k-cell of k_j. To *subdivide* k_j we subdivide c_j.

The boundary of k_j (mod. 2) shall be the $(j - 1)$-complex on Ω which is the image in the above map of the product of the boundary (mod. 2) of c_j and the line segment $0 \leqq t \leqq 1$. A j-complex on Ω will be termed a j-cycle if without boundary. A set of j-cycles on Ω is termed homologous to zero if the cycles of the set form the boundary of a $(j + 1)$-complex on Ω. Everything is to be understood mod. 2. We see that the usual operations with homologies are permissible.

By the *connectivity* P_j of Ω, $j = 0, 1, \ldots$ we mean the maximum number of j-cycles on Ω between which there is no homology, provided such a maximum exists. If no such maximum exists we say that P_j is infinite.

The connectivities of Ω are invariant under any topological transformation of R which carries admissible end points (A^1, A^2) into admissible end points. For purposes of pure topology the analyticity of R and Z is of course unessential.

3. *The function* $J(\Pi)$. It is well known that there exists a positive constant e, small enough to have the following properties. Any extremal arc E on which $J < e$ will give an absolute minimum to J relative to all sensed curves of class D^1 joining E's end points. On E the local coordinates of any point will be analytic functions of the local coordinates of the end points of E and of the distance of P along E from the initial end point Q of E, at least as long as E does not reduce to a point. The set of all extremal segments issuing from Q with $J < e$ will form a field covering a neighborhood of Q in a one-to-one manner, Q alone excepted. We now choose a positive constant ϱ less than e and make the following definition.

Any extremal segment on R for which J is less than ϱ will be called an elementary extremal.

An ordered set of $p + 2$ points

$$A^1, P^1, \ldots, P^p, A^2 \tag{3.1}$$

on R with (A^1, A^2) on the terminal manifold will be denoted by (Π). The points (3.1) will be called the vertices of (Π). It may be possible to join the successive points in (3.1) by elementary extremals. In such a case (Π) will be termed *admissible*. The resulting broken extremal will be denoted by $g(\Pi)$ and also termed admissible. The value of J taken along $g(\Pi)$ will be denoted by $J(\Pi)$.

Let a copy of R be provided for each point P^i. Points (Π) can be represented as points on the *product complex* Π whose factors are the auxiliary r-circuit B_r of § 2 and these p copies of R.

We shall regard $J(\Pi)$ locally as a function Φ of the parameters (α) locally representing its vertices A^1, A^2, and of the successive sets of coordinates (x) locally representing its vertices P^j. The function Φ will be analytic at least as long as the successive vertices remain distinct. A point (Π) with successive vertices distinct will be called a *critical point* of $J(\Pi)$ if all of the first partial derivatives of the function Φ are zero at that point.

If the successive vertices are distinct, the condition that $\Phi_{\alpha h}$ be null is seen to be

$$[F_{r^i}(x, \bar{x})\, x^{is}_{\alpha h}]_{s=1}^{s=2} = 0 \qquad (h = 1, \ldots, r) \tag{3.2}$$

where $(x, \bar{x})$ is to be evaluated on $g(\Pi)$ at the final end point of $g(\Pi)$ when $s = 2$, and at the initial end point of $g(\Pi)$ when $s = 1$. The r conditions (3.2), ($h = 1, \ldots, r > 0$), are called the *transversality conditions*.

The partial derivative of Φ with respect to the ith coordinate of a vertex (x) is seen to be

$$F_{r^i}(x, p) - F_{r^i}(x, q) \tag{3.3}$$

where (p) and (q) are the direction cosines at (x) of the elementary extremals of $g(\Pi)$ preceding and following (x) respectively. With the aid of (3.3) we could readily prove that a necessary and sufficient condition that an admissible point (Π), with successive vertices distinct, be a critical point, is that $g(\Pi)$ be a critical extremal, that is one satisfying the transversality conditions (3.2).

By a *critical set* σ of $J(\Pi)$ we understand any set of critical points on which $J(\Pi)$ is constant, and which is at a positive distance from other critical points of $J(\Pi)$. A critical set need not be connected. It is not necessarily closed since σ may have limit points (Π) whose successive vertices are not all distinct.

We shall regard a continuous family of critical extremals as *connected* if any curve of the family can be continuously deformed into any other curve of the family through the mediation of curves of the family. The set of all critical extremals on which $J < b$ make up at most a finite set of connected families of extremals on each of which J is constant.

4. *The connectivities of the domain* $J(\Pi) < b$. Let b be a non-critical value of J. Suppose the number, p + 2, of vertices in (Π) is such that

$$(p + 1)\,\varrho > b.$$

Understanding that points (Π) for which $J(\Pi)$ is evaluated are always restricted to admissible points (Π), we come to the problem of proving that the connectivities of the domain $J(\Pi) < b$ are finite. We shall accomplish this by deforming the domain

$J(\Pi) < b$ on itself onto a subdomain which is a subcomplex of the product complex Π.

In defining our deformations it will be convenient to term the value of J taken along a restricted curve g the *J-length* of the curve. Along g the J-length, measured from a fixed point to a variable point P and taken with a sign in accordance with the sense of measurement, will be termed the *J-coordinate* of P on g. If the J-coordinate of P is a differentiable function $h(t)$ of the time t, then P will be said to be moving at the time t on g at a *J-rate* equal to $h'(t)$.

We now define certain deformations of admissible points (Π) through admissible points (Π), which do not increase $J(\Pi)$ beyond its initial value, and which deform complexes of points (Π) continuously. Such deformations will be called J-deformations.

The deformation D_1. Let (Π) be an admissible point (Π). As the time t increases from 0 to 1, let the p vertices P^i of (Π) move along $g(\Pi)$ from their initial positions to a set of positions which divide g into $p+1$ successive segments of equal J-length, each vertex moving at a constant J-rate.

The deformation D_2. Let P^i be a vertex of (Π). Let h' and h'' be respectively the elementary extremals which join P^i to the preceding and following vertices (possibly coincident with P^i). As the time t increases from 0 to 1, let points H_t' and H_t'' start from P^i and move away from P^i respectively on h' and h'' at J-rates equal in absolute value to one half the J-lengths of h' and h''. Let H_t' and H_t'' be joined by an elementary extremal E_t. The deformation D_2 is hereby defined as one in which P is replaced at each time t by that point P_t which divides E_t into two segments of equal J-length.

That D_1 and D_2 are J-deformations follows readily. (Ref. 8, § 5.)

The two preceding deformations are designed to lower the value of $J(\Pi)$ in case $g(\Pi)$ is not an extremal. In case $g(\Pi)$ is an extremal but not a critical extremal, we must turn to the end conditions. The preceding deformations are independent of the local representation of the vertices deformed. To obtain the same result for our deformation of the end points we must employ the technique of tensor analysis.

With the terminal manifold $(r > 0)$ we now associate a metric defined by the quadratic form

$$ds^2 = \left(g^2_{ij}\frac{\partial x^{i2}}{\partial \alpha^h}\frac{\partial x^{j2}}{\partial \alpha^k} + g^1_{ij}\frac{\partial x^{i1}}{\partial \alpha^h}\frac{\partial x^{j1}}{\partial \alpha^k}\right) d\alpha^h\, d\alpha^k = a_{hk}(\alpha)\, d\alpha^h\, d\alpha^k \qquad (h, k = 1, \ldots, r) \tag{4.1}$$

where g^s_{ij} stands for $g_{ij}(x)$ evaluated for $x^i = x^{is}(\alpha)$. The quadratic form hereby defined is positive definite by virtue of the hypothesis that the matrix of partial derivatives of the functions $x^{is}(\alpha)$ has the maximum rank. Let a^{hk} be the cofactor of a_{hk} in the determinant $|a_{hk}|$ divided by that determinant.

Let (Π) be an admissible point neighboring an admissible point (Π_0) with successive vertices distinct. Let (z) represent the set of local coordinates of the intermediate vertices of (Π), and (a) the parameters determining the end vertices of (Π). The first and last elementary extremals of $g(\Pi)$ can be respectively represented by analytic functions in the form

$$x^i = \Phi^{i1}(t, a, z) \tag{4.2}$$
$$x^i = \Phi^{i2}(t, a, z) \qquad (i = 1, \ldots, m).$$

We suppose t has been so chosen on $g(\Pi)$ that it is a constant multiple of the arc length on $g(\Pi)$ and equals 0 and 1 at the respective ends of $g(\Pi)$. We now consider the differential equations

$$\frac{d\,a^k}{d\,\tau} = -\left[F_{r^i} \frac{\partial \overset{is}{x}(a)}{\partial a^h}\right]_{s=1}^{s=2} \overset{hk}{a}(a) = M^k(a, z) \qquad (h, k = 1, \ldots, r) \tag{4.3}$$

in which x^i and r^i in F_{r^i} are given by

$$x^i = \Phi^{i2}(1, a, z) \qquad r^i = \Phi_t^{i2}(1, a, z) \qquad s = 2$$
$$x^i = \Phi^{i1}(0, a, z) \qquad r^i = \Phi_t^{i1}(0, a, z) \qquad s = 1$$

for $s = 2$ and 1 respectively. For each set (z) we regard the equations (4.3) as differential conditions on the variables (a). Relative to non-singular analytic transformations of the parameters (a), the two members of (4.3) are contravariant vectors associated with the terminal manifold. Relative to similar transformations of the coordinates (x) of the end points the two members of (4.3) are invariant. The trajectories defined by (4.3) are accordingly independent of the local parameters (a) or coordinates (x) used.

Let the *distance* between two nearby points (Π) and (Π_0) be measured by the square root of the sum of the squares of the geodesic distances on R between corresponding vertices of (Π) and (Π_0). A point (Π) which is admissible and which determines elementary extremals of equal J-length will be termed *J-normal.* Note that the right members of (4.3) are not necessarily analytic if the successive vertices of (Π) are not distinct. Let ω be the set of all J-normal points (Π) for which $J < b$. Let μ be a positive constant so small that a point (Π) within a distance μ of some point of ω will be admissible and have successive vertices distinct. Let (a_0, z_0) represent such a point (Π) and let $\Psi^k(r)$ represent a solution of

$$\frac{d\,a^k}{d\,\tau} = M^k(a, z_0) \qquad (k = 1, \ldots, r),$$

which reduces to (a_0) for $\tau = 0$. If μ is sufficiently small, the point (Π) for which $(z) = (z_0)$ and $a^k = \Psi^k(\tau)$ will continue to be admissible and have its successive vertices distinct for $0 \leqq \tau \leqq \mu$ independently of the particular initial point (a_0, z_0) chosen.

The deformation D_3. Under D_3 a point (Π) which is at a distance $d < \mu$ from the set of all J-normal points on $J(\Pi) < b$ and which is represented by the point (a_0, z_0) when $t = 0$ shall be replaced at the time t $(0 \leqq t \leqq 1)$ by the point (Π) for which $(z) = (z_0)$ and

$$a^k = \Psi^k [(\mu - d)\, t] \qquad\qquad a^k(0) = a^k{}_0$$

while points for which $d \not< \mu$ shall remain fixed.

Under D_3 points (Π) for which $d < \mu$ are deformed so that

$$\text{(4.4)}\qquad \frac{dJ}{dt} = \frac{dJ}{da^k}\frac{da^k}{dt} = \left[F_{ri}\frac{\partial x^{is}}{\partial a^k}\right]_1^2 \left[F_{rj}\frac{\partial x^{js}}{\partial a^h}\right]_1^2 a^{hk}(a)\,(d - \mu) \leqq 0.$$

In particular if (Π) is a J-normal point we have $d = 0$. If $g(\Pi)$ is a non-critical extremal, the r components (3.2) are not all zero and $dJ/dt < 0$ since the elements a^{hk} are the coefficients of a positive definite form. Thus $J(\Pi)$ is never increased under D_3, and is actually decreased if (Π) is a J-normal point and $g(\Pi)$ a non-critical extremal.

Lemma 4. 1. The connectivities of the domain $J(\Pi) < b$ are finite.

To the points of the domain $J(\Pi) \leqq b$ we apply the product deformation (taking the factors in the order written).

$$T = D_1 D_2 D_1 D_3.$$

We note that the boundary of $J(\Pi) < b$ will consist of points (Π) for which $J(\Pi) = b$ and points at which $M(\Pi) = \varrho$ where $M(\Pi)$ is the maximum J-lenght of the elementary extremals of $g(\Pi)$. Since p has been chosen so that $(p+1)\,\varrho > b$ we see that points for which $M(\Pi) = \varrho$ will be deformed under D_1 into points for which $M(\Pi) < \varrho$. Moreover under D_2, $M(\Pi)$ is not increased, while D_3 carries admissible points into admissible points. Hence under T points for which $M(\Pi) = \varrho$ are carried into points which $M(\Pi) < \varrho$.

To show that T carries points for which $J(\Pi) = b$ into points for which $J(\Pi) < b$, we divide admissible points on the domain $J(\Pi) \leqq b$ into three classes as follows.

Class I shall contain those points (Π) which are deformed under D_1 into a point (Π_1) for which at least one elementary extremal is null.

Class II shall contain the remaining points (Π) for which $g(\Pi_1)$ is *not* an extremal.

Class III shall contain the remaining points (Π) for which $g(\Pi_1)$ is an extremal.

If (Π) belongs to Class I, we see that $J(\Pi)$ is decreased under D_1 and hence under T. If (Π) belongs to Class II, $g(\Pi_1)$ possesses an actual corner formed by two successive elementary extremals so that $J(\Pi)$ is decreased under $D_1 D_2$ and hence under T. If (Π) belongs to Class III, $J(\Pi)$ is not decreased under $D_1 D_2$ but is decreased under T by virtue of (4.4) unless $g(\Pi_1)$ is a critical extremal. But in such a case

$J(\Pi_1) < b$ since b is an ordinary value of J. Thus under T the domain $J(\Pi) < b$ is deformed on itself into a subdomain Σ whose boundary is distinct from that of $J(\Pi) < b$.

Let K be a subcomplex of the product complex Π that contains all of the points of Σ and is so finely subdivided as to consist wholly of points on $J(\Pi) < b$. We see that any cycle on $J(\Pi) < b$ is homologous to a cycle on K so that the connectivities of $J(\Pi) < b$ must be finite. The lemma is accordingly proved.

By a *critical value* of J we mean a value which J assumes on a critical extremal.

We can now prove the following theorem.

Theorem 4.1. *If a and b ($a < b$) are two ordinary values of J between which there are no critical values of J, the connectivities of the domains $J(\Pi) < a$ and $J(\Pi) < b$ will be equal.*

The analysis in the preceding proof shows that the application of the product deformation T^n to the domain $J(\Pi) < b$ for a sufficiently large power n will deform $J(\Pi) < b$ on itself into a subdomain on $J(\Pi) < a$. The theorem follows readily.

5. *The connectivities of the restricted domain $J < b$.* Let k_j be a j-complex on the functional domain Ω of § 2 represented by means of the product complex $c_j \times t$ of an auxiliary j-complex c_j and the line segment $0 \leqq t \leqq 1$. If k_j is composed of restricted curves on which the J-length of the curves from their initial points to the image of a point Q on $c_j \times t$ varies continuously with Q, k_j will be called a *restricted j-complex.* Employing restricted complexes and cycles only, one can now define the connectivities of Ω formally as before. We term these connectivities the *restricted connectivities.* We shall prove the following lemma.

Lemma 5.1. *The restricted connectivity R_i of the functional domain Ω equals the unrestricted connectivity P_i of Ω.*

Let k_j be an unrestricted complex on Ω represented by the product $c_j \times t$ as in § 2. Let P be a point on c_j and $k(P)$ the corresponding curve of k_j. Let p be a positive integer and let $k(P)$ be divided into $p + 1$ segments of equal variation of t. Let (Π) denote the point on Π determined by the successive ends of these segments of $k(P)$, and let h be any one of these segments. If p is sufficiently large (and we suppose it is), each point of h can be joined to the initial point of h by an elementary extremal λ, independently of the curve $k(P)$ of k_j under consideration.

We shall now deform $k(P)$ into $g(\Pi)$. Let τ represent the time during this deformation, $0 \leqq \tau \leqq 1$. At each time τ we suppose h divided into two segments λ and λ' in the ratio of τ to $1 - \tau$ with respect to the variation of t on h. For each value of τ we now replace the second of these segments of h by itself, while we replace the first λ by the elementary extremal μ which joins its end points. We make a point on λ which divides λ in a given ratio with respect to t, correspond to the point on μ which divides μ in the same ratio with respect to the variation of J, assigning to this point on μ the same value of t as its correspondent on λ bears. We denote this deformation by δ_1.

We need to subject the resulting curves $g\ (\Pi)$ to a further deformation δ_2 which does not change the curves $g\ (\Pi)$ except in parameterization. To that end we let the point t on $g\ (\Pi)$ move along $g\ (\Pi)$ to the point on $g\ (\Pi)$ which divides $g\ (\Pi)$ with respect to the J-length in the same ratio as t divides the interval (0,1), moving at a constant J-rate along $g\ (\Pi)$ equal to the J-length of the arc to be traversed. We term the resulting parameterization a *J-parameterization.* In it the parameter t runs from 0 to 1 and is proportional to the J-length of the arc preceding the point t.

We denote the product deformation $\delta_1\ \delta_2$ by Δ_p.

Under Δ_p the curve $k\ (P)$ is carried into a curve $g\ (\Pi)$ with a J-parameterization. If we associate $g\ (\Pi)$ so parameterized with the point P on c_j it appears that the totality of these curves $g\ (\Pi)$ forms are stricted complex $r\ (k_j)$ on Ω representable by $c_j \times t$.

If $k_j \to k_{j-1}$ we see that

$$r\ (k_j) \to r\ (k_{j-1}). \tag{5.1}$$

If k_j is a cycle on Ω, we see then that $r\ (k_j)$ will be a restricted j-cycle homologous to k_j on Ω. It follows that $P_i \leqq R_i$. In particular R_i must be infinite with P_i.

To show that $P_i = R_i$ we have merely to show that a restricted cycle z_j on Ω which bounds an unrestricted complex z_{j+1} on Ω necessarily bounds a restricted complex as well. To that end we suppose that

$$z_{j+1} \to z_j \qquad \text{on } \Omega$$

If the index p of Δ_p is sufficiently large, we can apply Δ_p to z_{j+1}. We then have

$$r\ (z_{j+1}) \to r\ (z_j). \tag{5.2}$$

But if z_j is deformed under Δ_p through the complex w_{j+1} we have (always mod. 2).

$$w_{j+1} \to z_j + r\ (z_j), \tag{5.3}$$

and hence from (5.2) and (5.3)

$$r\ (z_{j+1}) + w_{j+1} \to z_j.$$

Moreover the left complex is a restricted complex since z_j is restricted. Hence $P_i = R_i$ and the lemma is proved.

We shall now prove the following lemma.

Lemma 5.2. The restricted connectivities R_i' of the restricted domain $J < b$ equal the connectivities R_i'' of the subdomain $J\ (\Pi) < b$ of Π. The latter connectivities are thus independent of the number of vertices $(p + 2)$ of their points (Π), provided only that $(p + 1)\ \varrho > b$.

Let c_j be a complex on $J(\Pi) < b$. The set of broken extremals $g(\Pi)$ determined by points (Π) on c_j if given a 'J-parameterization' will afford a restricted complex c_j^p on Ω represented by the product $c_j \times t$. A restricted complex c_j^k derived from a complex c_j on Π in this manner will be called a *p-fold complex.*

We see that c_j^p will be a cycle on Ω if and only if c_j is a cycle on Π, and that c_j^p will bound among p-fold complexes on $J < b$ if and only if c_j bounds on $J(\Pi) < b$. The connectivities of the domain $J < b$ on Ω defined in terms of p-fold complexes for a fixed p will then equal the connectivities R_j'' of $J(\Pi) < b$.

But on the other hand any restricted j-cycle of Ω on $J < b$ can be deformed under Δ_p among restricted j-complexes of Ω on $J < b$ into a j-complex c_j^p on $J < b$, so that $R_i' \leqq R_i''$. To prove that $R_i' = R_i''$ one has merely to prove, as under the preceding lemma, that a p-fold cycle k_j on $J < b$ bounds a restricted complex on $J < b$ only if it bounds a p-fold complex on $J < b$.

To that end we suppose that k_j bounds a restricted complex k_{j+1} on $J < b$. As in (5.2) we have

$$r(k_{j+1}) \to r(k_j).$$

But the vertices of each curve $g(\Pi)$ of $r(k_j)$ lie on the curve $g(\Pi_1)$ from which it is deformed under Δ_p. We can deform $g(\Pi_1)$ into $g(\Pi)$ deforming k_j into $r(k_j)$ through a complex w_{j+1} by letting each vertex of (Π_1) move along $g(\Pi_1)$ to the corresponding vertex of $g(\Pi)$ at a J-rate equal to the J-length of the arc of $g(\Pi)$ to be traversed. We then have

$$r(k_{j+1}) + \dot{w}_{j+1} \to k_j$$

and the left hand complex is a p-fold complex.

Thus k_j bounds a p-fold complex on $J < b$ if it bounds a restricted complex on $J < b$. Hence $R_i' = R_i''$ and the lemma is proved.

The preceding lemma taken with Theorem 4.1 now gives us the following.

Theorem 5.1. *If a and b, a < b, are any two ordinary values of J between which there are no critical values of J, the restricted connectivities of the functional domains J < b and J < a are equal.*

6. *Spannable and critical cycles.* Any two points in our Riemannian space R will be said to possess a *J-distance* equal to the inferior limit of the J-lengths of restricted curves joining the two points. We see that this J-distance between the two points varies continuously with the points.

We shall now define the J-distance between two curves g_1 and g_2 of class D^1.

To that end let us regard points on g_1 and g_2 as *corresponding* if they divide g_1 and g_2 in the same ratio with respect to the J-lengths of their arcs. We now define the *J-distance d (g_1, g_2) between g_1 and g_2* as the maximum of the J-distances between corresponding points of g_1 and g_2 plus the absolute value of the difference between the J-lengths of g_1 and g_2.

If g_3 is a third curve of class D^1 we have the relation

$$d\,(g_1, g_3) \leqq d\,(g_1, g_2) + d\,(g_2, g_3).$$

With the aid of this relation we see that if g_2 is sufficiently near g_3, that is if $d\,(g_2, g_3)$ is sufficiently small, $d\,(g_1, g_2)$ will differ arbitrarily little from $d\,(g_1, g_3)$. This sort of thing is now well known, developed, for example, in the works of M. Fréchet.

Now let g_1 be any restricted curve. By a *neighborhood* of g_1 on Ω will be meant a set of restricted curves which includes all restricted curves within some small positive distance e of g_1. Let A be a subset of restricted curves on Ω. The curve g_1 will be called a *limit* curve of curves of A if there is a curve of A, other than g_1, in every neighborhood of g_1. The *boundary* of A is the set of restricted curves which are limit curves of curves both of A and $\Omega - A$. Open, closed, and compact sets A are now defined in the usual way. Particular examples of closed and compact sets A are restricted complexes and critical sets of extremals.

If A and B are any two sets of restricted curves, $d\,(A, B)$ will be defined as the inferior limit of the distances between curves of A and B. If A and B are closed and compact, $d\,(A, B)$ will be taken on by at least two curves of A and B. The distance $d\,(g_1, A)$ varies continuously with g_1, that is it changes arbitrarily little if g_1 is replaced by a restricted curve sufficiently near g_1.

Let σ be a critical set of extremals on which $J = c$. Let a and b be two ordinary values of J between which there is no critical value of J other than c. By a *neighborhood* N of σ on the functional domain Ω will be meant a set of restricted curves within a small positive distance of σ. We shall take N as open, that is such that when γ belongs to N all the restricted curves sufficiently near γ also belong to N. We admit only such neighborhoods of σ as consist of curves whose distances from other critical sets is bounded from zero. We also suppose that curves of N satisfy the condition

$$a < J < b.$$

It will be convenient to say that a restricted curve for which $J < c$ is *below the critical value c*.

We shall now introduce certain sets of j-cycles on Ω near the critical set σ. They serve to characterize the basic topological properties of the domain $J < c$ near σ. In defining these sets of cycles the phrase "a maximal set of j-cycles independent on etc." will be used as an abbreviation for the phrase, "a set of j-cycles which contains the maximum number of j-cycles independent with respect to bounding on etc.".

Let N_1 and N be neighborhoods of σ with $N_1 \subset N$.

A j-cycle on N_1 below c, bounding on N_1 but not bounding on N below c, will be called a *spannable* cycle corresponding to N and N_1 (written corr. $N\,N_1$). *A* j-cycle on N_1 independent on N of j-cycles below c will be called a *critical* cycle corr. $N\,N_1$. For suitable choices of N and N_1 we shall see that the following maximal sets of j-cycles exist (are finite).

The set $(a)_j$. *A maximal set of spannable j-cycles corr.* $N\,N_1$, *independent on* N *below* c.

The set $(c)_j$. *A maximal set of critical j-cycles corr.* $N\,N_1$, *independent on* N *of j-cycles on* N *below* c.

We now state a basic theorem.

Theorem 6. *There exists a fixed neighborhood* N^* *of* σ *and corresponding to each neighborhood* $N \subset N^*$ *a neighborhood* $M(N) \subset N$ *with the following property.*

Corresponding to any two choices of the pair of neighborhoods $N\,N_1$ *satisfying the conditions*

$$N \subset N^* \qquad N_1 \subset M(N) \tag{6.1}$$

there exist common maximal sets of spannable or critical k-cycles on any arbitrarily small neighborhood of σ.

We shall term three neighborhoods $N_0\,N\,N_1$ *admissible* if $N\,N_1$ satisfy (6.1), and $N_0\,N$ also satisfy (6.1) as an instance of $N\,N_1$ in (6.1). We admit only such pairs of neighborhoods $N\,N_1$ as belong to an *admissible triple*. We abbreviate the phrase "corresponding to an admissible pair of neighborhoods $N\,N_1$" by writing it in the form corr. $N\,N_1$.

7. *Classification of cycles of* Ω *on* $J < b$. Suppose now that σ is the set of all critical extremals at which $J = c$. A spannable k-cycle corr. $N\,N_1$ which does not bound below c will be termed *newly bounding* corr. $N\,N_1$, and a spannable cycle which bounds below c a *linkable* cycle corr. $N\,N_1$. We see that a maximal set $(a)_k$ of spannable k-cycles can be made up of a maximal set $(l)_k$ of linkable k-cycles independent on N below c, together with a maximal set $(b)_k$ of newly bounding k-cycles independent on N below c of linkable k-cycles.

Let l_{k-1} be any linkable $(k - 1)$ –cycle corr. $N\,N_1$. We have

$$\lambda_k'' \to l_{k-1} \qquad \text{below } c \tag{7.0}$$

where λ_k'' is a k-complex on Ω. But we also have

$$\lambda_k' \to l_{k-1} \qquad \text{on } N_1 \tag{7.1}$$

where λ_k' is again a complex on Ω. We now introduce the k-cycle

$$\lambda_k' + \lambda_k'' = \lambda_k \tag{7.2}$$

We say that λ_k links l_{k+1} corr. $N\,N_1$.

A set of k cycles $(\lambda)_k$ linking the respective $(k-1)$-cycles of a maximal set of linkable $(k-1)$-cycles corr. $N N_1$ will be called a *maximal set of linking k-cycles* corr. $N N_1$.

A k-cycle below c independent below c of the newly bounding k-cycles corr. $N N_1$ will be called an *invariant* k-cycle corr. $N N_1$. By a maximal set $(i)_k$ of invariant k-cycles corr. $N N_1$ is meant a set of such k-cycles, independent below c of newly bounding k-cycles corr. $N N_1$.

We now state the basic theorem.

Theorem 7.1. *A maximal set of restricted k-cycles on* $J < b$ *is afforded by maximal sets of linking, critical, and invariant k-cycles corr.* $N N_1$.

We also have the following.

In case c is an absolute minimum of J, and b separates c from the other critical values of J, then a maximal set of restricted k-cycles on $J < b$ *is afforded by the critical k-cycles corr.* $N N_1$.

8. *The existence of critical extremals.* Relative to the critical value c we shall call a k-cycle a *new* k-cycle if it is independent on $J < b$ of cycles on $J < a$. We shall consider maximal sets of new k-cycles independent on $J < b$ of k-cycles on $J < a$. By virtue of Theorem 7.1 such a maximal set will be afforded by a maximal set of critical k-cycles corr. $N N_1$ and linking k-cycles corr. $N N N_1$.

Relative to the critical value c we shall call a k-cycle *newly bounding* if it lies on $J < a$, is bounding on $J < b$, but non-bounding on $J < a$. We shall consider maximal sets of newly bounding k-cycles independent on $J < a$ but bounding on $J < b$. According to Theorem 7.1 such a maximal set will be afforded by a maximal set of newly bounding k-cycles corr. $N N_1$.

The number of cycles in a maximal set of new k-cycles depends upon more than the neighborhood of o. The same is true of the number of cycles in a maximal set of newly bounding $(k-1)$-cycles. It is a remarkable fact however that the *sum* of the two preceding numbers depends only upon the nature of J neighboring σ, in fact is the number of cycles in a maximal set of critical k-cycles, and a maximal set of spannable $(k-1)$-cycles. Moreover this sum is additive for the separate critical sets. We are thus led to the following definition.

The kth *type number* M *of a critical set* σ *shall be defined as the number of cycles in the corresponding maximal sets of critical k-cycles and spannable* $(k-1)$*-cycles corr.* $N N_1$.

The kth type number of a sum of critical sets will be defined as the sum of the respective type numbers of the component sets.

We state the following principal theorem.

Theorem 8. *Let* $N_0\, N_1 \ldots$ *be the type number sums for all critical sets of extremals,* *and let P_0 P_1 ... be the connectivities of the function space Ω. If the numbers N_i are finite they satisfy the infinite set of inequalities.*

$$\text{(A)}\qquad \begin{aligned} N_0 &\geqq P_0 \\ N_0 - N_1 &\leqq P_0 - P_1 \\ N_0 - N_1 - N_2 &\geqq P_0 - P_1 + P_2. \\ &\cdots\cdots \end{aligned}$$

If all of the integers N_1 are finite for $i < r + 1$ the first $r + 1$ relations in (A) still hold.

Further details and extensions will be given in the Colloquiom Lectures by the author.

1. *Tonelli,* Fondamenti di Calcolo delle Variazioni.

2. *Signorini,* Rendiconti del Circolo Matematico di Palermo, vol. 33 (1912) pp. 187—193. Here broken geodesics are used in the quest for a minimum.

3. *Birkhoff,* Dynamical Systems. American Mathematical Society Colloquium Publications, vol. 9. Birkhoff's minimum and minimax methods are developed here.

4. *MM. Lusternik et Schnirelmann,* Topological Methods in Variational Problems. Research Institute of Mathematics and Mechanics Gosizdat Moscow (1930). Here a topological theory of critical points of functions is developed. Specific results in the calculus of variations are confined mostly to the problem of the closed geodesic on the image of the 2-sphere.

5. *Morse,* Transactions of the American Mathematical Society, vol. 27 (1925), pp. 345—396.

6. *Morse,* Ibid vol. 30 (1928), pp. 213—274.

7. *Morse,* Ibid vol. 31 (1929), pp. 379—404.

8. *Morse,* Ibid vol. 32 (1930), pp. 599—631.

9. *Morse,* Ibid vol. 33 (1931), pp. 72—91.

10. *Morse,* Proceedings of the National Academy of Sciences, vol. 15 (1929), pp. 856—859. Extremal chords, and periodic extremals.

11. *Morse,* Annals of Mathematics, vol. 32 (1931), pp. 549—566.

12. *Morse,* Calculus of Variations. American Mathematical Society Colloquium Publications. Not yet published.

13. *Lefschetz,* Topology. American Mathematical Society Colloquium Publications.

14. *Eisenhart,* Riemannian Geometry.

15. *Osgood,* Funktionentheorie II.

16. *Ponincaré,* Liouville Journal 4th series, vol. 1 (1885) pp. 167—244. The equality in the case $n = 2$ was used by Poincaré.