E. Bompiani (Ed.)

CIME Summer Schools

Geometria del calcolo delle variazioni

23

Saltino, Italy 1961

Springer

FONDAZIONE CIME ROBERTO CONTI

E. Bompiani (Ed.)

Geometria del calcolo delle variazioni

Lectures given at the
Centro Internazionale Matematico Estivo (C.I.M.E.),
held in Saltino (Firenze), Italy,
August 21-30, 1961

Springer

FONDAZIONE CIME ROBERTO CONTI

C.I.M.E. Foundation
c/o Dipartimento di Matematica "U. Dini"
Viale Morgagni n. 67/a
50134 Firenze
Italy
cime@math.unifi.it

ISBN 978-3-642-10957-7 e-ISBN: 978-3-642-10959-1
DOI:10.1007/978-3-642-10959-1
Springer Heidelberg Dordrecht London New York

Reprint of the 1st ed. C.I.M.E., Ed. Cremonese, Roma, 1961
With kind permission of C.I.M.E.

Printed on acid-free paper

Springer.com

CENTRO INTERNATIONALE MATEMATICO ESTIVO
(C.I.M.E)

Reprint of the 1st ed.- Saltino, Italy, August 21-30, 1961

GEOMETRIA DEL CALCOLO DELLE VARIAZIONI

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C.I.M.E.)

H. BUSEMANN

THE SYNTHETIC APPROACH TO FINSLER SPACES IN THE LARGE

ROMA - Istituto Matematico dell'Università - 1961

THE SYNTHETIC APPROACH TO FINSLER SPACES IN THE LARGE

1. INTRODUCTION. CURVES AND SEGMENTS.

The geodesics of a Riemann space can be obtained as curves which are locally shortest connections or as the auto-parallel curves of the distinguished affine connexion of Levi-Civita. Parallel displacement along a curve maps the local geometry at one point of the curve isometrically on that at another.

Direct extension of the second method to Finsler spaces where the line element has the form

$$ds = F(x^1, \ldots, x^n, dx^1, \ldots, dx^n) = F(x, dx)$$

and $F(x; dx)$ satisfies certain standard conditions, is impossible because the local geometry of a Finsler space is Minkowskian and two n-dimensional Minkowski spaces are in general not isometric.

Nevertheless, generalizations of parallel displacement have played a major role in the theory of Finsler spaces in two different approaches. The space may be considered as a set of line elements rather than points to which local euclidean geometries are attached. Since this will be the topic of Professor Davies' lectures I will not dwell on it here.

The second approach is to consider the space as a point, hence locally Minkowskian space and to face the concomitant analytical difficulties as well as imperfection inherent to any concept of parallel displacement in Finsler spaces. This is the topic of Professor Wagner's lectures.[1)]

Thirdly, one may start from the definition of geodesic as a locally shortest join and avoid the analytical complications by not using analysis. Although this seems to contradict the very name "differential geometry" synthetic arguments partly topological and partly similar to those of euclidean geometry have proved very successful in particular when dealing with problems in the large. These lectures will give an introduction to the field. For simplicity we restrict ourselves to the case of symmetric distances $(F(x, dx) = F(x, -dx))$, but much of the material can and has been extended to the non-symmetric case, see Busemann [1] and Zaustinsky [1] .

Proofs are expected in these cycles. Some proofs in the present theory are long and the technicalities are uninteresting. In particular at the beginning, proofs become necessary only because the axioms are chosen as weak as possible, whereas all consequences of the axioms would have to be postulated, if they were not contained in the axioms. Doubts have been expressed regarding the power of synthetic methods in differential geometry. The only way of combatting the doubts is to exhibit the efficiency of these methods in many different areas. We therefore will outline the proofs only in those cases where they elucidate the reason for the superiority of geometric arguments.

---

1) Readers who should see these notes without those of Professors Davies and Wagner are referred to Cartan [1] and Rund [1] .

H.Busemann

We are interested in the intrinsic geometry in the large of complete Finsler spaces, and our method is axiomatic. The distance is for us not the infinitesimal distance given by a line element, but the finite intrinsic distance of two points in the manifold. Therefore our first axiom is :

I, <u>The space, R, is metric.</u>

The distance of two points x,y is denoted by xy and satisfies the standard conditions $xx = 0$, $xy = yx > 0$ for $x \neq y$ and $xy + yz \geqslant 0$.

A curve $x(t)$, $\alpha < t < \beta$ is a continuous map of the interval $[\alpha, \beta]$ in R . Its length is defined in the natural way: if D_t : $\alpha = t_o < t_1 < \dots < t_n = \beta$ is a partition of $[\alpha, \beta]$ we put

$$L(x, D_t) = \sum_{i=1}^{n} \left| x(t_{i-1}) - x(t_i) \right|$$

and

$$L(x) = L_\alpha^\beta (x) = \sup_{D_t} L(x, D_t). \quad \text{Notice } L(x) \geqslant x(\alpha)x(\beta).$$

If $\| D_t \| = \sup_i | t_i - t_{i-1} |$ then the usual theorem $L(x, D_t) \rightarrow L(x)$ for $\| D_t \| \rightarrow 0$ holds, see G , p.19 (G refers here and elsewhere to Busemann [2]). An important implication of this fact is the <u>additivity of arclength</u>

$$\sum_{i=1}^{n} L_{t_{i-1}}^{t_i} (x) = L_\alpha^\beta (x) \quad \text{for any partition } D_t .$$

Moreover, length is <u>lower semicontinuous</u> (G p.20) : If $x_\nu (t)$, $\alpha \leq t \leq \beta$, $\nu = o, 1, 2, \dots$ are curves and $x_\nu (t) \rightarrow x_o(t)$ for each t, then

(1.1) $$L(x_0) \leq \liminf L(x_\nu)$$

The curve $x(t)$ is rectifiable if $L(x)$ is finite. We can then introduce the arclength as parameter : $y(s)$, $0 < s < L(x)$, is the point $x(t)$ for which $s = L_0^t(x)$. A class of rectifiable curves $x(t)$ whose representations $y(s)$ in terms of arclength are identical is a geometric (rectifiable) curve C. The elements of the class are the parametrizations of C and the properties common to all parametrizations of C are the properties of C (initial point, end point, length are examples).

There are many important theorems for which a freedom in the choice of the parameter and hence the concept of geometric curve rather than parametrized curve is essential. The most important is this : We call a subset M of a metric space finitely compact if every bounded infinite set in M has an accumulation point in M , that is, if M satisfies the Bolzano-Weierstrass Theorem.Then the following selection theorem (G p.24) holds which is fundamental for much of the present theory, although we will not always mention the theorem explicitly.

(1.2) If M is a finitely compact subset of a metric space and $C_1, C_2, \ldots$ are geometric curves in R with $L(C_\nu) < \beta$ and whose initial points form a bounded sequence, then a suitable subsequence $C_{\nu_1}, C_{\nu_2}, \ldots$ of C_ν tends uniformly to a geometric curve C and

$$L(C) \leq \liminf L(C_{\nu_n})$$

The theorem means that parametrizations $x_n(t)$ of C_{ν_n}

and $x(t)$ of C, $\alpha \leq t \leq \beta$, exist such that $x_n(t)$ tends uniformly to $x(t)$.

A curve $x(t)$, $\alpha \leq t \leq \beta$, satisfying $L(x) = x(\alpha)x(\beta)$ is a shortest curve from $x(\alpha) = a$ to $x(\beta) = b$ and is called a <u>segment</u> $T(a, b)$ from a to b. If we want to specify that the segment is oriented from a towards b we use the notation $T^+(a, b)$. If $y(s)$ represents $T(a, b)$ in terms of arclength, then for $0 \leq s_1 < s_2 \leq ab = \beta$

$$ab \leq y(0)y(s_1) + y(s_1) + y(s_2)y(\beta) \leq$$

$$\leq L_0^{s_1}(y) + L_{s_1}^{s_2}(y) + L_{s_1}^{\beta}(y) = L_0^{\beta}(y) = ab$$

hence $y(s_1)y(s_2) = s_2 - s_1$, so that every subarc of a segment is a segment and $y(s) \rightarrow s$ maps the segment isometrically on a segment of the real axis, whence the name. Allowing a shift of the origin we call representation of a $T(a, b)$ curve $z(t)$ with

$$\alpha \leq t \leq \beta = \alpha + ab , \quad z(\alpha) = a , \quad z(\beta) = b \quad \text{and}$$

$$z(t_1)z(t_2) = \left| t_1 - t_2 \right| , \quad \alpha \leq t_i \leq \beta . \tag{1.3}$$

Let (xyz) indicate that $x \neq y$, $y \neq z$ and $xy + yz = xz$. Then

$$\text{(wxy) and (wyx) imply (xyz) and (wxz) .} \tag{1.4}$$

For

$$wz = wy + yz = wx + xy + yz > wx + xz > wz .$$

(Notice that (wxy) and (xyz) do not imply (wyz) or (wxz)).
The following trivial remark is often useful :

(1.5) <u>If</u> (xyz) <u>and</u> $T_1 = T(x,y)$ <u>and</u> $T_2 = T(y,z)$ <u>exist, then</u> $T_1 \cup T_2$ <u>is a</u> $T(x,z)$.

A segment connecting two given points will in general not exist. To insure their existence we add two axioms :

II <u>The space is finitely compact.</u>

III <u>For any two distinct points x,z a point y with (xyz) exists.</u>

It is easily seen that a segment T(x,y) <u>exists for any two points x,y</u> (G p.29). Finite compactness is much stronger than would be necessary for the existence of segments. However, I will insure in conjunction with the remaining two axioms that the space has those properties of finite dimensional spaces which are necessary to obtain differential geometric results, whether the axioms actually imply finite dimensionality is not known.

2. GEODESICS.

A <u>geodesic</u> is a locally isometric map of the entire real axis into the space R . This means that it can be represented in the form x(t), $-\infty < t < \infty$ and that for each real t_o a positive $\varepsilon(t_o)$ exists such that

$$x(t_1)x(t_2) = |t_1 - t_2| \quad \text{for} \quad |t_i - t_o| \leq \varepsilon(t_o) .$$

This trivially (G p.32) implies that t is arclength

$$L_{t_1}^{t_2}(x) = t_2 - t_1 \quad \text{for any} \quad t_1 < t_2$$

$x(t)$ and $y(t)$ represent the same geodesic if $\delta = +1$ and a real β exist such that

$$y(t) = x(\delta t + \beta) \quad \text{for all } t .$$

Frequently we will consider oriented geodesics. Then it is understood that in a representation $x(t)$ increasing t corresponds to traversal in the positive sense. A second representation $y(t)$ of the oriented geodesic has the form $y(t) = x(t + \beta)$.

A geodesic is a straight line if it is an isometric image in the large of the real axis, i.e. $x(t_1)x(t_2) = |t_1 - t_2|$ holds for any t_1, t_2.

A compact convex subset of a euclidean space satisfies axioms I, II, III but geodesics do not exist. We must have an axiom of prolongability. To include the usual objects of differential geometry the axiom must be a local requirement. We denote the open sphere with radius $\rho > 0$ about a point, i.e. the set of points x with $px < \rho$, by $S(p, \rho)$. The triangle inequality implies that

(2.1) $$S(p, \rho) \supset S(q, \rho - pq) \text{ if } \rho > pq .$$

We can then formulate our axiom as follows :

IV. For each point p there is a positive ρ_p such that for any two distinct points x,y in $S(p, \rho_p)$ a point z with (xyz) exists.

The function ρ_p may be erratic, however then IV implies the existence of a well behaved function satisfying the axiom. For put

$$\rho(p) = \sup \rho_p, \text{ where } \rho_p \text{ satisfies IV for fixed } p .$$

Then either $\rho(p) = \infty$, which means that z with (xyz) exists for any distinct x,y and hence $\rho(q) = \infty$ for any q .

$$(2.2) \quad \begin{cases} \rho(p) \quad , \quad \text{or} \quad 0 < \rho(p) < \infty \quad \text{and} \\ \quad |\rho(p) - \rho(q)| \leq pq \ . \end{cases}$$

The latter follows at once from (2.1), G p.33.

The existence of geodesics means in ordinary differential geometry that every line element lies on a geodesic. This implies that a segment, or shortest geodesic join, can be extended to a geodesic, and for us this will mean existence of geodesics, (G pp.34, 35)

(2.3) If Axioms I to IV hold and $x(t)$, $\alpha \leq t \leq \beta$, $\alpha < \beta$ represents a segment, then a geodesic $y(t)$ $(-\infty < t < \infty)$ exists such that $y(t) = x(t)$ for $\alpha \leq t \leq \beta$.

If $\rho(p) \equiv \infty$ then $y(t)$ may be chosen as a straight line.

According to a recent observation of Szenthe [1] it may happen that $y(t)$ extending $x(t)$ to a geodesic exists which is not a straight line even when $\rho(p) \equiv \infty$.

Axioms I to IV do not contain any uniqueness properties for segments or geodesics. The simplest example showing this is the (x_1, x_2)-plane the metric

$$xy = |x_1 - y_1| + |x_2 - y_2| \ .$$

If $x_1 < y_1$ and $x_2 < y_2$ then any curve $z(t) = (z_1(t), z_2(t))$ from x to y for which both $z_1(t)$ and $z_2(t)$ are non-decreasing will be (but not necessarily represent) a segment and monotone continuation will provide straight lines. Our final axiom will therefore be a uniqueness postulate. Observing that, in differential geometry, the shortest geodesic join is not necessarily unique, but that its continuation to a geodesic is, we require.

Axiom V. If (xyz_1), (xyz_2) and $yz_1 - yz_2$ then $z_1 = z_2$.

The spaces satisfying all five axioms are called G-spaces, the G alluding to geodesic. Unfortunately the word G-space has lately also been used in a different sense, where the G alludes to group. In a G-space the extension of a proper segment to a geodesic is unique. Therefore, if $\rho(p) \equiv \infty$ then all geodesics are straight lines and the space is called straight. There are many important straight spaces besides the euclidean and hyperbolic spaces. In the terminology of the calculus of variation all simple connected spaces without conjugate points are straight.

We do not postulate the local uniqueness of $T(x,y)$ because this important property is contained in the axioms.

(2.4) If (xyz) then $T(x,y)$ and $T(y,z)$ are unique.

For if two segments T_1, T_2 from y to z existed then $z_i \in T_i$ with $z_1 \neq z_2$ and $yz_1 = yz_2$ would exist and satisfy (xyz_i) by (1.4) contradicting V. In particular

(2.5) $T(x,y)$ is unique for $x,y \in S(p, \rho(p))$.

Regarding the $\varepsilon(t_0)$ occurring in the definition of a geodesic it can easily be proved (G p.38) that :

(2.6) If $x(t)$ represents a geodesic then it represents a segment for $|t - t_0| \leq \rho(x(t_0))$.

In the absence of differentiability we define a <u>lineal element at p</u> as a segment with center p and length min($\rho(p)/2,1$). The <u>multiplicity</u> of a geodesic at a point (on the geodesic) is the cardinal number of distinct lineal elements at p lying on the geodesic. (A lineal element represented twice, hence infinitely often, for different t-intervals if a representation counts only once).

<u>(2.7) The multiplicity of a geodesic at a point is finite or countable.</u>

For in a representation x(t) of a geodesic different lineal elements at the same point correspond to disjoint intervals of the t-axis. A similar simple argument shows (G p.44)

(2.8) A geodesic has an at most countable number of multiple points.

Here we use the usual terminology to call a point multiple if the multiplicity is greater than 1, simple if it is one. The geodesic is simple if all its points are simple. A standard argument (G p.45) shows

(2.9) $x(t)$ <u>represents a simple geodesic if and only if</u> $x(t_1) = x(t_2)$ <u>implies</u> $x(t_1 + t) = x(t_2 + t)$ <u>for all t</u>.

We eliminated the zero-dimensional G-spaces as trivial and want to do the same for one-dimensional spaces. A simple but most useful lemma is needed.

(2.10) <u>If</u> $x,y \in S(p,\rho)$ <u>then</u> $T(x,y) \subset S(p, 2\rho)$.

For let $w \in T(x,y)$. Then

$$\min(wx,wy) \le xy/2 \le (xp + py)/2 < \rho$$

and if $wx = \min(wx,wy)$ then

$$pw \leq px + xw < 2\rho .$$

Strangely enough this crude estimate is the best possible even on the sphere: If p and w are antipodal points on a sphere of radius 2, choose x and y on the same great circle through p and w with $ox = py = \pi + \varepsilon$, $0 < \varepsilon < \pi$. Then T(x,y) passes through w and $pw = \pi$. Whereas spheres are locally convex under the usual assumption for Finsler spaces, they are not necessarily so even in straight G-spaces.

A "great circle of length β" in a G-space is a geodesic isometric to a circle of length β. Its representation is distinguished by

$$x(t_1)x(t_2) = \min_{|\nu|=0,1,2} \left| t_1 - t_2 + \nu\beta \right| .$$

Straight lines and great circles contain with any two points x,y a segment T(x,y) and this property is characteristic:

(2.11) <u>If a geodesic G contains with any two points x,y a segment T(x,y) then it is a straight line or a great circle.</u>

First we show that G is simple. If G contained two lineal elements L_1, L_2 at the same point p, let a_i be an endpoint of L_i. Since $pa_i < \rho(p)/3$ the segment $T(a_1,a_2)$ is unique (see (2.5)) and lies therefore on G, moreover $T(a_1, a_2) \subset S(p, \rho(p))$, hence T(p,x) is unique for $x \in T(a_1,a_2)$ and lies on G, so that the multiplicity of G at p would not be countable.

If x(t) represents G then by (2.9) $x(t_1) = x(t_2)$ implies $x(t_1 + t) = x(t_2 + t)$ for all t. If $x(t_1) \neq x(t_2)$ for $t_1 < t_2$ then the arc $t_1 \leqslant t \leqslant t_2$ is the only arc in G from $x(t_1)$ to $x(t_2)$ and therefore by hypothesis a segment or $x(t_1)x(t_2) = |t_1 - t_2|$. Or there are two arcs on G from $x(t_1)$ to $x(t_2)$ and one of them must be a segment which leads to a great circle, for details see G p.46.

This proof shows that G-space containing two distinct line elements L_1, L_2 at p has at least dimension 2 because $\bigcup T(p,x)$, $x \in T(a_1,a_2)$ is homeomorphic to a triangle.

(2.12) <u>A one-dimensional G-space is a straight line or a great circle.</u>

For, a point of the space is center of at most one lineal element, therefore all geodesics are simple and no two geodesics intersect. If two different geodesics G_1, G_2 existed then for $p_i \in G_i$ a segment $T(p_1,p_2)$ would lie on a geodesic intersection G_1 and G_2.

Thus there is only one geodesic and this contains with any two points x,y a segment T(x,y) because the space has this property. The assertion now follows from (2.11).

Since zero- and one-dimensional G-spaces are trivial we will often tacitly <u>assume that the space has dimension at least two.</u> It can be proved that a two-dimensional G-space is a topological manifold (G pp.52-53), i.e. every point has a neighborhood homeomorphic to E^2. The corresponding problem for higher dimensions is open. As mentioned before, it is not known whether a G-space always has a finite dimension.

H.Busemann

3. SPACES IN WHICH THE GEODESIC THROUGH TWO POINTS IS UNIQUE.

A further corollary of (2.11) is

(3.1) If the geodesic through any two distinct points of a G-space is unique, then each geodesic is either a straight line or a great circle.

For if the geodesic G contains x and y ($x \neq y$) it must contain every T(x,y) because a geodesic containing T(x,y) exists and the geodesic through x and y is unique.

The statement (3.1) can be considerably improved. For this and other purposes we need the concept of universal covering space. The mapping α of the G-space R' on the G-space R is locally isometric if a positive function $\beta_{p'}$ exists such that α maps $S(p', \beta_{p'})$ isometrically on $S(p'\alpha, \beta_{p'})$. Putting $p'\alpha = p$ it can be proved (G p.171) that α maps $S(p', \rho(p)/2)$ isometrically on $S(p, \rho(p)/2)$ so that in contrast to the topological theory of covering spaces a $\beta_{p'}$ exists which is not only independent of the choice of p' in $p\alpha^{-1}$ but also of the space R' (and the mapping α). A consequence of this fact is that a locally isometric map of a compact G-space on itself is an isometry in the large or a motion (G p.172).

If a locally isometric map α of R' on the R exists then R' is a covering space of R . The cardinal number of points in $p\alpha^{-1}$ is independent of p and is usually called the number of sheets of R' (over R). This number is at most countable, because any two distinct points in $p\alpha^{-1}$ is at least $2\rho(p)$, (G pp.171,172).

If $x'(t)$ is a geodesic in R' then $x'(t)\alpha = x(t)$ is a geodesic in R . Conversely, <u>a geodesic in R can be lifted</u> : given a geodesic $x(t)$ in R and a point $p' \in x(t_0)\alpha^{-1}$ there is a unique geodesic $x'(t)$ in R' such that $x'(t_0) = p'$ and $x'(t)\alpha = x(t)$. In general it is not true that there is only one geodesic G' through p' with $G'\alpha = G$ where $x(t)$ represents G . For $x(t)$ may have a multiple point at $x(t_0)$ and different lineal elements of G at $x(t_0)$ may lead to different G'. However, <u>if</u> $x(t_0)$ <u>is a simple point of G, then G' is unique</u> (G p.169).

If $x(t)$, $\alpha \leq t \leq \beta$, represents a segment then the corresponding part of $x'(t)$ is a segment. The converse is obviously not true. If $T(x(\alpha), x(\beta))$ is unique then so is $T(x'(\alpha), x'(\beta))$ (G p.169).

Among the covering space of R' there is a simply connected one, which is unique up to isometries and is called <u>the universal covering space of R</u> (G. Section 28).

We now come to the improvement of (3.1). We observe first

(3.2) <u>If two distinct geodesics</u> G_1, G_2 <u>each contain with two points x,y a segment T(x,y) and have two common points a, a' then</u> G_1 <u>and</u> G_2 <u>are great circles of the same length and a, a' are antipodal on both, moreover</u> $G_1 \cap G_2 = a \cup a'$.

For both G_i contain segments T(a,a') and these are distinct, hence no point c with (abc) can exist, see (2.4) which proves that G_1, G_2 are great circles of the same length, (2.4) also shows that they cannot have any other common points.

A G-space is <u>sphere like</u> if the geodesics are great circles and the geodesics through a given point a all pass through a second point $a' \neq a$, which we call the antipode to a.

It follows from (3.2) that all geodesics through a given point a have the same length, moreover that a is the antipodal to a'. All geodesics have the same length, because for any two non-intersecting geodesics a third intersecting both exists.

A G-space is of the <u>elliptic type</u> if the geodesics are great circles of the same length and the geodesic through to distinct points is unique.

Identification of antipodal points in a spherelike space of dimension $\geqslant 2$ yields a space of the elliptic type. The argument is quite elementary and may be found in G p.129. We can now prove

(3.3) <u>If dim $R \geq 2$ and R is not simply connected, if, moreover each geodesic in R contains with any two points x,y a segment T(x,y), then R is of the elliptic type and has a spherelike space as two-sheeted universal covering space.</u>

Proof. We know from (2.11) that all geodesics in R are great circles or straight lines. Since R is not simply connected its universal covering space R' has at least two sheets.

We show first that a geodesic G' in R' contains with two points a', b' at least one segment T(a',b'). If G is the image of G' under the local isometry of R' on R, assume first that $a = a'\alpha$ and $b = b'\alpha$ are neither identical nor antipodal in G . If G' did not contain a T(a',b'), then a geodesic H' containing a T(a',b') would exist and G and $H'\alpha$ would be two different geodesics through a and b which contradicts (3.2) (G is simple, hence only one geodesic G' with $G'\alpha = G$ through a exists, see above). If a = b or a and b are antipodes on G , choose c' close to b' so that a and c are neither identical nor antipodal. Then G' contains a T(a',b') and by continuity also a T(a',b'). Thus all geodesics in R' are straight lines or great circles.

Next we prove that there is only one geodesic G' in R' over a given geodesic in R. Let both G_1' and G_2' lie over G and choose $a_i' \in G_i$ so that a_1 and a_2 are neither identical nor antipodal on G. The image H' of a geodesic H' containing a_1' and a_2' intersects G in a_1 and a_2. We conclude from (3.2) that G = H and hence $G_1' = H' = G_2'$.

It now follows that if $G'\alpha = G$ and $p \in G$ then $p\alpha^{-1} \subset G'$. For through every point $p' \in p\alpha^{-1}$ there is a G' with $G'\alpha = G$ and there is only one G' mapped on G . Therefore two distinct geodesics through p' both contain all points in $p\alpha^{-1}$ and by (3.2) there cannot be more than two. Thus every geodesic passing through one of these points passes through the other and the space R' is spherelike. R is then of the elliptic type.

(3.4) <u>A space in which the geodesic through two points is unique and which contains a great circle is not simply connected</u>.

This can be proved for general G-spaces (G pp.201, 202), but is so simple under a minimum of differentiability hypotheses, that we prove it only for this case.

The geodesics are all straight lines or great circles and defining the length of a straight line as infinite it is obvious that the length L(G) of a geodesic G through a fixed point p depends continuously on the geodesic. Choose p such that at least one geodesic through p is a great circle.

Consider an elliptic space E of the same dimension as R in which the geodesics have length 1. Choose a point p' in E and map the tangent space of R at p linearly on the tangent space of E at p'. To a line element L of R at p now corresponds a line element L' of E at p' and thus to the geodesics G_L through the geodesic $G_{L'}$ through L'. We map G_L in $G_{L'}$ such that

$$p'x' = \frac{1}{2} \cdot \frac{px}{1 + px} \cdot \frac{2 + L(G)}{L(G)}$$

The point x' is not uniquely determined by this relation, but we can determine it in a neighborhood of p' such that the map is topological there. Then it becomes topological everywhere, and is meaningful also for the great circles, because the antipodal point to p on G_L becomes the antipodal point on $G_{L'}$. If G_L is a straight line, then its image is $G_{L'}$ with the antipodal point to p omitted.

Thus R is mapped topologically as a subset E_R of E . A projective line cannot be contracted to a point in E , hence still less in E_R. Therefore the great circles in R through p , which are mapped on projective lines, cannot be contracted. The improvement of (3.1) is a corollary of (3.3,4) :

(3.5) <u>Theorem</u>. <u>If the geodesic through two distinct points of a G-space of dimension greater than 1 is unique, then R is either straight or R is of the elliptic type and has a spherelike space as two sheeted universal covering space.</u>

4. INVERSE PROBLEMS.

A two-dimensional G-space is a topological manifold and if the geodesic through 2 points is unique it is either homeomorphic to E^2 and straight or homeomorphic to p^2 and of the elliptic type. In view of the fact that in the Riemannian case the geodesics determine the metric up to trivial transformations except in a few cases (Liouville's Theorem), it might seem reasonable to look for all metrizations of E^2 as a straight G-space or of P^2 as a space of the elliptic type. The following considerations will show that their problems are too general to be interesting, but determining the curve systems which occur as sets of geodesics proves interesting. One of the principal consequences of our investigation is an insight into the enormous variety of Finsler metrics.

For a system S of curves in E^2 to be the geodesics of a straight space the following is obviously necessary : If e(x,y) is an auxiliary euclidean metrization of the plane then

a) Each curve in S can be represented in the form p(t) $(-\infty < t < \infty)$, $p(t_1) \neq p(t_2)$ for $t_1 \neq t_2$, and $e(p(o),p(t)) \to \infty$ for $|t| \to \infty$.

b) There is exactly one curve of S through two given distinct points of the plane.

We will show that these trivially necessary conditions are also sufficient.

(4.1) Theorem. Given a system S of curves in E^2 satisfying the conditions a) and b) then E^2 can be metrized (in a great variety of ways) as a G-space with the curves in S as geodesics.

For $a \neq b$ let $G(a,b)$ be the curve in S through a and b, and $T(a,b)$ the subarc of $G(a,b)$ with endpoints a,b. If x is an interior point of $T(a,b)$ we write $[a\ x\ b]$. Put $T(a,a) = a$. First one establishes some simple topological properties of S, (G pp.57,58), in particular the Axiom of Pasch : If $c \notin G(a,b)$ and $[apb]$ then an S-curve H through p intersects $T(a,c) \cup T(c,b)$. If $a_n \to a$ and $b_n \to b$ then $T(a_n,b_n) \to T(a,b)$ in the sense of Hausdorff's closed limit, and also $G(a_\nu,b_\nu) \to G(a,b)$ provided $a \neq b$.

Let G^+ be an oriented S-curve and $p \notin G^+$. If x traverses G^+ in the positive sense then the oriented S-curve $G^+(p,x)$ (p precedes x in the orientation) tends to an oriented S-curve A^+ called <u>the asymptote to G^+ through p</u>. For any $q \in A^+$ the asymptote to G^+ through q is also A^+. For example, if q follows p on A^+, then the segment $T(q,x)$ tends to a ray R. If R did not lie on A^+, then $G(p,r)$ with $r \in R - q$ would not intersect G^+ and A^+ cannot be the asymptote to G^+ through p. The case where q precedes p is similar (G pp.59, 60). We will see later that the asymptote relation is in general neither symmetric nor transitive.

Here we need only the following consequence of this construction : <u>a given S-curve G can be imbedded in a simple family of S-curves</u>, i.e. one with the property that every point of the plane lies on exactly one curve in the family (the asymptotes to an orientation of G form a simple family).

To construct our metric consider a simple family F of S-curves. Fix a point z in the plane, let L_a be the curve in F through a. Denote one side of L_z by H^+ the other by H^{-1} and put for any $L \in F$

$$t(L) = \begin{cases} e(z,L) & \text{if} \quad L \subset H^+ \\ -e(z,L) & \text{if} \quad L \subset H^- \\ 0 & \text{if} \quad L = L_z \end{cases}$$

Then $t(L)$ depends monotonically on L. Put

$$d(a,b) = \left| t(L_a) - t(L_b) \right| .$$

Then

$$d(a,b) = d(b,a) \quad \begin{cases} = 0 & \text{if} \quad L_a = L_b \\ > 0 & \text{if} \quad L_a \neq L_b \end{cases}$$

$d(a,b) + d(b,c) \geqslant d(a,c)$ with equality if $L_a = L_b$ or $L_b = L_c$ or L_b lies between L_a and L_c.

$$d(a,z) \leq e(a,z)$$

because $d(a,z) = \left| t(L_a) \right| = e(z,L_a) \leq e(z,a)$.

Choose a denumerable number of simple families $F_1, F_2, \dots$ such that $\bigcup F_i$ is dense among S-curves. For each F_i define the distance d_i (with the same point z) as d was defined for F and put

$$ab = \sum_{i=1} 2^{-i} d_i(a,b) .$$

This number is finite because

$$d_i(a,b) \leq d_i(a,z) + d_i(z,b) \leq e(a,z) + e(z,b) .$$

Clearly $ab = ba \geqslant 0$, and if $a \neq b$ then a S-curve separating a from b exists, hence also a curve in $\bigcup F_i$, say a curve in F_k. Then $d_k(a,b) > 0$ and $ab > 0$.

Trivially $ab + bc \geqslant ac$. We must show that the relations $\lceil axb \rceil$ and (axb) are equivalent. Let $\lceil axb \rceil$. For a given i either

$G(a,b) \in F_i$ and then $d_i(a,b) = d_i(a,x) = d_i(x,b) = 0$, or the curve in F_i through x in F_i lies between the curves through a and b, and then $d_i(a,x) + d_i(x,b) = d_i(a,b)$. Thus the latter relation holds for all i , which implies (axb).

We show next that $ax + xb > ab$ if a,x,b are distinct and $[axb]$ does not hold. If $x \in G(a,b)$ there is trivially a S-curve separating x from a and b . If $x \notin G(a,b)$ then $G(a',b')$ with $[aa'x]$ and $[bb'x]$ will separate x from a and b . Therefore a curve in some F_k will exist separating x from a and b . Then the curve in F_k through x is different from those through a and b and does not lie between them, hence

$$d_k(a,x) + d_k(x,b) > d_k(a,b)$$

and $ax + xb > ab$.

It remains to be shown that the Bolzano Weierstrass Theorem holds. I do not know whether this is always the case. But the parameter t(L) was largely arbitrary and by modifying it we can reach finite compactness (G p.62).

If the system S is well behaved we may replace the summation by an integration. We illustrate this by giving some interesting metrizations of the (x_1,x_2)-plane with the ordinary lines $ax_1 + bx_2 + c = 0$ as geodesics. Remember that the euclidean metric is by Beltrami's Theorem the only Riemannian metric with this property.

Let $g(t)$ $(t \geq 0)$ be continuous, non-negative, increasing with $g(t) \to \infty$ for $t \to \infty$. Put

$$f(x,\alpha) = \text{sign}(x_1\cos\alpha + x_2\sin\alpha)g(x_1\cos\alpha + x_2\sin\alpha), \frac{-\pi}{2} < \alpha \leq \frac{\pi}{2}.$$

Observe that $|x_1 \cos\alpha + x_2 \sin\alpha|$ is the euclidean distance of the line through the point $x = (x_1, x_2)$ with normal α from the origin $z = (0,0)$. Thus $f(x, \alpha)$ corresponds to our $t(L)$ in the general case and $|f(x, \alpha) - f(y, \alpha)|$ to $d(x,y)$. Integrating instead of summing we form the distance

$$\rho_g(x,y) = \int_{-\pi/2}^{\pi/2} |f(x,\alpha) - f(y,\alpha)| \, d\alpha ,$$

for which the ordinary lines are the geodesics. This distance is invariant under the euclidean rotations about z, which are therefore also motions, more specifically rotations, for ρ_g. We consider some special choices of g.

1) $g(t) = \log(1 + t)$. Then the distance becomes relatively small as we move out and a simple estimate shows that for any two parallel lines L_1, L_2 the distance $x_1 L_2 (x_1 \in L_1)$ tends to zero when x_1 traverses L_1 in either direction.

2) $g(t) = e^t$. The distance becomes large when moving out and $x_1 L_2 \to \infty$ if x_1 traverses L_1 in either direction.

3) $g(t) = t^\beta$, $\beta > 0$. (For $\beta = 1$ this gives the euclidean metric). Then with $\delta x = (\delta x_1, \delta x_2)(\delta > 0)$

$$f(\delta x, \alpha) = \delta^\beta f(x, \alpha) .$$

If $k > 0$ is given we can determine $\delta > 0$ such that $\delta^\beta = k$ and we have for any two points x,y

$$\rho_g(\delta x, \delta y) = k \rho_g(x,y) .$$

Thus $x' = \delta x$ is a dilation for the metric ρ_g with the given dilation factor k . Thus we have a metric with the ordinary lines as geodesics, all dilations from z and all rotations about z which is

<u>not euclidean</u>. Whereas the metrics in cases 1) and 2) are or can be made smooth everywhere, the metric 3) has in a well defined sense a singularity at z when $\beta \neq 1$, and this singularity cannot be eliminated without making the metric euclidean, see (10.3).

The superposition principle used to construct the metric in Theorem (4.1) can be used in many other ways. For example : define the function f(t) by

$$f(t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ \sqrt{t} & \text{for} \quad 0 \leq t \leq 1 \\ 1 & \text{for} \quad t > 1 \end{cases}$$

then with $e(x,y) = \left[(x_1 - y_1)^2 + (x_2 - y_2)^2\right]^{1/2}$

$$xy = e(x,y) + \left| f(x_1) - f(y_1) \right| + \left| f(x_2) - f(y_2) \right|$$

is a metrization of the (x_1, x_2)-plane with the ordinary lines as geodesics. Let z = (o,o), a = (2t,o), b = (t,t), c = (o,2t). Then (abc) and for $0 < t \leqslant 1/2$

$$za = zc = 2t + \sqrt{2t} \qquad zb = \sqrt{2}t + 2\sqrt{t}$$

$$zb - za = \sqrt{t}\,(2 - \sqrt{2})(1 - \sqrt{t})$$

and zb > za = zc for $0 < t \leq 1/2$. Therefore <u>no circle with radius ≤ 2 about z is convex</u>. This hinges on the singularity of t at 0. e mentioned that in smooth Finsler spaces small spheres are convex (G p.162).

The result for P^2 corresponding to (4.1) is :

(4.2) <u>Theorem</u>. <u>In the projective plane P^2 let a system S' of curves be given such that</u>

a') <u>Each curve in S' homeomorphic to a circle</u>

b') <u>There is exactly one curve of S' through given distinct points.</u>

<u>Then P^2 can be metrized as a G-space for which the curves in S' are the geodesics.</u>

This theorem was first proved by Skornyakov [1], a simpler proof is found in Busemann [4]. I will outline a proof of (4.2) and at the same time of

(4.3) <u>If P^n ($n \geq 2$) is metrized on a G-space with the projective lines as geodesics, then this metric can be extended to P^{n+1} so that the projective lines in P^{n+1} are the geodesics.</u>

In both problems we pass from P^n to the spherical space S^n. In (4.2) we then obtain a system of curves S on S^2 such that each curve is homeomorphic to a circle and any curve on S through a point a also passes through the antipodal point a' to a on S^2. We may assume that one S-curve is the equator A of S^2 and the ordinary great circles through the north and south poles belonging to A are S-curves. We metrize A as a great circle such that antipodal points of S^2 are antipodal on A.

In (4.3) on S^{n+1} a great S^n denoted by A and metrization of A as a G-space with the ordinary circles as geodesics is given. In both cases we denote by H and H' the two open hemispheres bounded by A. If $p \in H$ then its antipode $p' \in H'$. For any point $x \neq p,p'$ there is exactly one semi-great circle (S-curve in (4.2)) with endpoints p,p' and containing x. It intersects A in a point x_p. If also $y \neq p,p'$ we put $f_p(x,y) = x_p y_p$. Notice that $f_p(x,y) = xy$ for

$x,y \in A$. Since $x \rightarrow x_p$ maps antipodal points onto antipodal points on A , each semi-great circle not passing through p is with $f_p(x,y)$ as metric a segment of the same length β . (Figure 1). Put $0 = f_p(p',x) = f_p(p,x) = f_p(x,p') = f_p(x,p)$.

Let g(p) be any positive continuous function on H such that $\int_H g(p)dp = 1$ and put

$$xy = \int_H f_p(x,y)g(p)dp \; .$$

This integral exists as a Riemann integral, because $f_p(x,y)$ is bounded, continuous when x and y are different from p,p' and lower semicontinuous when x or y fall in p or p'. For any S-curve G : $\int_{G \cap H} f_p(x,y)g(p)dp = 0$, because G intersects each meridian in 2 points only. These facts yield very easily that each semi great circle is with the metric xy a segment of length β and the metric A is evidently the same as before.

The proof that $xy + yz > xz$ when x,y,z do not lie in this order on a semi great circle is very similar to the corresponding proof in (4.1).

The inverse problem has also been solved for the torus without conjugate points, i.e. a torus whose universal covering space (the plane) is straight. This problem differs from the preceding problems in that there are non-obvious necessary conditions. This will be discussed after the theory of parallels has been developed.