

# LOCAL METRIC GEOMETRY

BY

HERBERT BUSEMANN

Many *special* problems commonly ascribed to differential geometry have proved accessible to a purely geometric treatment without any differentiability assumptions. The present paper tries to develop systematically the foundations for a *general* geometric theory of Finsler spaces. It is naturally divided into two parts. The first (chapters I and II) covers the basic concepts such as line elements, extremals, their parametrizations and so forth; briefly all those facts which are usually derived from the theory of differential equations. The second part (chapters III and IV) shows that these concepts suffice to build a geometric theory on them. It is largely inspired by Cartan [1]<sup>(1)</sup> which sets forth admirably the simple geometric facts underlying the results of Riemannian geometry.

A brief résumé of the results follows:

Chapter I contains the general tools and is believed to prove useful also for other fields. After defining metric spaces so as to include the case where *the distance is not symmetric* (the spaces occurring in the calculus of variations are in general of this type) §1 discusses the concept of finite compactness. In §2 the subsets of a metric space are metrized by a distance similar but superior to the known Hausdorff distance inasmuch as it is not restricted to bounded sets. §3 defines a similar metric for the motions of a metric space. These distances will serve to replace local coordinates or other parameters in analytical arguments.

Chapter II contains the fundamental definitions and facts. First (§4), the basic Axioms A, B, C, D for a Finsler space are given, and their implications for the local geometry are discussed. Then (§5), extremals are defined, in a rather unorthodox way, as classes of segments. It is shown that they have parametric representations with the properties commonly used for the definition. Next (§6), convergence of extremals as point sets, sets of line-elements, and curves is analyzed. In the symmetric case ( $yx = xy$ ) it is natural to consider non-oriented segments and extremals (called geodesics). They have many additional properties (§§7 and 8)<sup>(2)</sup> which may all be traced to the fact that two different geodesics have at most countably many common points.

Chapter III presents a theory of *parallelism for rays* in unbounded spaces. Prolongation of a segment beyond its end point leads either to the absolutely

---

Presented to the Society, August 14, 1944; received by the editors November 3, 1943.

(<sup>1</sup>) Numbers in brackets refer to the references cited at the end of the paper.

(<sup>2</sup>) These sections are an elaboration of B [1, chapter I, §§2 and 3].

conjugate point to the initial point (§9) beyond which prolongation is not possible or to an infinite ray, which suggests the idea of a point at infinity. The main problem is to determine to which extent, or under which conditions, this point has the customary properties<sup>(3)</sup> (§§10 and 11).

Chapter IV treats several problems of differential geometry proper. Since the concept of a covering space is essential for all of them, §§12 and 13 supply a theory of covering spaces for Finsler spaces. They include some new results on *locally isometric mappings*. §14 contains a variety of applications. An essential part of Cartan [1, chapters III, V and VI] is devoted to geometric characterizations, among Riemann spaces, of the *spaces which are locally euclidean, hyperbolic or spherical*. §§15 and 16 characterize the same spaces among general symmetric Finsler spaces. The "axiome du plan" of Cartan (loc. cit. p. 123) is evidently not sufficient for Finsler spaces, that is, the configurational hypothesis must be supplemented by a metric condition. It is shown that a Finsler space is locally euclidean, hyperbolic or spherical when, locally, the bisectors of two points are linear. The second axiom of Cartan, "l'axiome de libre mobilité" (loc. cit. p. 124) proves to be a sufficient condition also under the present general hypothesis. The following even weaker form of this axiom will be used here: every point  $p$  of the space has a neighborhood  $px < \rho(p)$  such that for any two congruent triples<sup>(4)</sup>  $p, a, b$  and  $p, a', b'$  in this neighborhood an isometric mapping of the neighborhood on itself exists, which leaves  $p$  fixed and carries  $a$  into  $a'$  and  $b$  into  $b'$ .

If either this or the bisector property hold in the large the space is a finite dimensional euclidean, hyperbolic or spherical space.

*Notations.* Small Roman letters, except  $e$ , denote points. Roman capitals are used for point sets. Gothic letters indicate parametrized point sets like curves, geodesics, . . . . Small Greek letters stand for real numbers. Greek capitals, except  $\Gamma, \Delta, H$  denote transformations. Products of transformations are to be read from left to right.  $\Gamma, \Delta, H$  are reserved for groups of transformations.

## I. FINITELY COMPACT SPACES

**1. Nonsymmetric metric spaces. Finite compactness.** The natural distance  $xy$  of the calculus of variations is in general not symmetric, that is,  $xy$  may be different from  $yx$ . To include this case, the present paper uses the term "*metric space*" for any set  $R$  with the following four properties:

(<sup>3</sup>) Cartan [1, Note III] and B [1, chapter III, §4] study parallelism in symmetric spaces in which all geodesics are isometric to straight lines.

(<sup>4</sup>) That is,  $pa = pa' < \rho$ ,  $pb = pb' < \rho$  and  $ab = a'b'$ . The literature on this problem is very extensive, compare Weyl [1]. The present paper furnishes the first treatment of the local problem without differentiability assumptions. The problem in the large for the euclidean and hyperbolic cases was first discussed by the author (see B [1]), and for the spherical cases by G. Birkhoff [1]. However the assumptions of G. Birkhoff are somewhat different; while relaxing the topological conditions, he considerably strengthens the mobility hypothesis.

(1.1)  $xy$  is defined for all ordered pairs of points in  $R$  and is non-negative.

(1.2)  $xy = 0$  if and only if  $x = y$ .

(1.3)  $xy + yz \geq xz$  for any three points  $x, y, z$ .

(1.4)  $xx, \rightarrow 0$  if and only if  $x, x \rightarrow 0$ .

The ordinary metric spaces for which  $xy = yx$  will be distinguished from the others as *symmetric* metric spaces. Due to (1.4) the distinction is purely metric and not topological if the convergence  $x, \rightarrow x$  is defined by  $xx, \rightarrow 0$ . For put

$$(1.5) \quad \sigma(x, y) = \max(xy, yx), \quad \sigma^*(x, y) = \min(xy, yx).$$

Then (1.4) shows that any one of the four relations  $xx, \rightarrow 0$ ,  $x, x \rightarrow 0$ ,  $\sigma(x, x), \rightarrow 0$ ,  $\sigma^*(x, x), \rightarrow 0$  implies the three others. Moreover  $\sigma(x, y)$  satisfies the triangle inequality  $\sigma(x, y) + \sigma(y, z) \geq \sigma(x, z)$  and is therefore a symmetric metrization of  $R$  equivalent to  $xy$ .

The numbers  $AB$ ,  $\sigma(A, B)$ ,  $\sigma^*(A, B)$  are defined as the greatest lower bounds of  $xy$ ,  $\sigma(x, y)$ ,  $\sigma^*(x, y)$  respectively, where  $x$  ranges over  $A$  and  $y$ , independently, over  $B$ . Finally  $S'(A, \rho)$ ,  $S''(A, \rho)$ ,  $S(A, \rho)$ , and  $S^*(A, \rho)$ ,  $\rho > 0$ , will denote the sets of those points  $x$  for which, respectively,  $Ax < \rho$ ,  $xA < \rho$ ,  $\sigma(A, x) < \rho$ , and  $\sigma^*(A, x) < \rho$ . By (1.4) the four sets  $\{S'(x, \rho)\}$ ,  $\{S''(x, \rho)\}$ ,  $\{S(x, \rho)\}$ , and  $\{S^*(x, \rho)\}$ , where  $x$  ranges over  $R$  and  $\rho$  over the positive real numbers, form equivalent sets of neighborhoods (bases) of  $R$ . Obviously

$$(1.6) \quad S(x, \rho) = S'(x, \rho) \cap S''(x, \rho), \quad S^*(x, \rho) = S'(x, \rho) \cup S''(x, \rho).$$

Menger developed a general theory of arclength<sup>(5)</sup>. The following is a brief summary of Menger's results for spaces which satisfy axioms (1.1) to (1.4) to the extent in which they will be needed in the present paper.

The arclength of the continuous curve  $c$ , parametrically represented as  $c(\tau)$ ,  $\alpha \leq \tau \leq \beta$ , is the number

$$(1.7) \quad \lambda(c) = \sup_{[\tau]} \sum_{\kappa=0}^{\mu-1} c(\tau_{\kappa})c(\tau_{\kappa+1})$$

where  $[\tau]$  traverses all subdivisions  $\tau_0 = \alpha < \tau_1 < \dots < \tau_{\mu-1} < \tau_{\mu} = \beta$  of the interval  $(\alpha, \beta)$ . When  $\lambda(c)$  is finite  $c$  is called *rectifiable*. Let  $c(\tau', \tau'')$  denote the subarc  $\tau' \leq \tau \leq \tau''$  of  $c$ . Obviously

$$(1.8) \quad c(\tau')c(\tau'') \leq \lambda(c(\tau', \tau'')).$$

The arclength is additive, that is, the relation

$$(1.9) \quad \lambda(c(\tau', \tau'')) + \lambda(c(\tau'', \tau''')) = \lambda(c(\tau', \tau'''))$$

holds for any three numbers  $\tau', \tau'', \tau'''$  with  $\alpha \leq \tau' < \tau'' < \tau''' \leq \beta$ .

<sup>(5)</sup> Compare Menger [1, 2]. Menger's hypotheses do not exactly cover the present case, but his arguments carry over.

To introduce the arclength  $\sigma$  as parameter on the rectifiable curve  $c$  means the transition from the given parametrization  $c(\tau)$  to the parametrization  $c'(\sigma)$ ,  $0 \leq \sigma \leq \lambda(c)$ , defined in the following manner:  $c'(\sigma)$  is the point  $c(\tau_0)$  for which  $\sigma = \lambda(c(\alpha, \tau_0))$  (the point  $c(\tau_0)$  is unique, but there may be a whole interval of values  $\tau$  for which  $c(\tau) = c(\tau_0)$ ).

For all topological and metric considerations two rectifiable curves  $c_1, c_2$  do not differ essentially when their arclength parametrizations  $c'_1(\sigma)$  and  $c'_2(\sigma)$  are identical, that is, when  $c'_1(\sigma) = c'_2(\sigma)$  for  $0 \leq \sigma \leq \lambda(c_1) = \lambda(c_2)$ . Such curves are therefore considered *equivalent*. It is easily seen that a curve is homotopic to its arclength representation.

(1.10) *The arclength is lower semicontinuous: If  $c_\nu: c_\nu(\tau)$ ,  $\alpha \leq \tau \leq \beta$ ,  $\nu = 0, 1, 2, \dots$ , are continuous curves with  $c_\nu(\tau) \rightarrow c_0(\tau)$  for every  $\tau$  then  $\lambda(c_0) \leq \liminf \lambda(c_\nu)$ .*

A continuous curve  $c: c(\tau)$ ,  $\alpha \leq \tau \leq \beta$ ,  $c(\alpha) = a$ ,  $c(\beta) = b$  is called a *segment*  $\mathfrak{s}(a, b)$  from  $a$  to  $b$  if  $\lambda(c) = ab$ . By (1.3), (1.8), (1.9):

(1.11)  $c(\tau')c(\tau'') = \lambda(c(\tau', \tau''))$  when  $c$  is a segment and  $\alpha \leq \tau' < \tau'' \leq \beta$ , so that each subarc  $c(\tau', \tau'')$  of a segment is a segment.

Any curve  $c'(\tau)$  which is equivalent to  $c$  and for which

$$c'(\tau_1)c'(\tau_2) = \tau_2 - \tau_1 \quad \text{for} \quad \tau_1 < \tau_2$$

is called a *representation of the segment*  $c = \mathfrak{s}(a, b)$ . Any segment represented by a subarc of  $c'(\tau)$  is called a *subsegment* of  $\mathfrak{s}(a, b)$ . In particular, the parametrization  $c'(\sigma)$  in terms of the arclength  $\sigma$  is a representation of  $c$ .

(1.12) If  $c_1(\tau)$  and  $c_2(\tau)$  represent the same segment, then  $c_1(\tau) = c_2(\tau + \alpha)$  for a suitable real  $\alpha$ .

The further discussion of segments will be simplified by the notation  $(abc)$  which expresses that the points  $a, b, c$  are different and that  $ab + bc = ac$ . This relation has the following important property:

(1.13)  $(abc)$  and  $(acd)$  if and only if  $(abd)$  and  $(bcd)$ .

For if  $(abd)$  and  $(bcd)$  then

$$ad = ab + bd = ab + bc + cd \geq ac + cd \geq ad$$

hence  $ab + bc = ac$  and  $ac + cd = ad$ . The converse is established similarly. A consequence of (1.13) is

(1.14) If  $(abc)$  and segments  $\mathfrak{s}(a, b)$  and  $\mathfrak{s}(b, c)$  exist, then  $\mathfrak{s}(a, b)$  continued by  $\mathfrak{s}(b, c)$  is a segment  $\mathfrak{s}(a, c)$ .

The following trivial estimates will be used several times.

(1.15) If  $a, b \in S(x, \alpha)$  and  $(acb)$  then  $c \in S(x, 3\alpha)$ . Hence each  $\mathfrak{s}(a, b) \subset S(x, 3\alpha)$ . If the space is symmetric then  $\mathfrak{s}(a, b) \subset S(x, 2\alpha)$ .

These facts follow from

$$cx \leq cb + bx < ab + bx \leq ax + xb + bx < 3\alpha$$

and

$$xc \leq xa + ac < xa + ab \leq xa + ax + xb < 3\alpha$$

in the general case, and from  $xc < \min(ac, bc) + \alpha$  in the symmetric case.

Following Menger the metric space  $R$  will be called *convex* when for any ordered pair of points  $a, c$  a point  $b$  with  $(abc)$  exists. If for every pair  $a, b$  a segment  $\mathfrak{s}(a, b)$  exists the space is obviously convex. Menger [3] showed that the converse holds for complete spaces, in particular:

(1.16) *If the metric space  $R$  is finitely compact and convex, then any point  $a$  can be connected to any point  $b$  by a segment  $\mathfrak{s}(a, b)$ .*

The notion of finite compactness needs some clarification for non-symmetric spaces: A set  $A$  is said to be *bounded* when a number  $\rho > 0$  and a point  $a$  exist such that  $A \subset S^*(a, \rho)$ . The space  $R$  is *finitely compact*<sup>(6)</sup>, if every bounded infinite set  $A$  contains a sequence of points,  $\{x_\nu\}$ , which converges to a point  $x$  (not necessarily in  $A$ ).

(1.17) If in a finitely compact space  $R$  the set  $A \subset S'(a, \rho')$  then  $A \subset S''(a, \rho'')$  for a suitable  $\rho''$ .

Otherwise  $x_\nu \in A$  would exist with  $x_\nu a \rightarrow \infty$ . Since  $R$  is finitely compact  $\{x_\nu\}$  contains a subsequence  $\{x_\mu\}$  which converges to a point  $x$ . Then  $x_\mu a \leq x_\mu x + xa$  would imply that  $\{x_\mu a\}$  is bounded.

(1.18) *Let  $A$  be a bounded set in the finitely compact space  $R$ . For a given  $\epsilon > 0$  a  $\delta > 0$  can be found such that  $ab < \delta$  and  $a, b \in A$  imply  $ba < \epsilon$ .*

Otherwise points  $a_\nu$  and  $b_\nu$  in  $A$  would exist with  $a_\nu b_\nu \rightarrow 0$  and  $b_\nu a_\nu \geq \epsilon$ . Since  $A$  is bounded  $\{a_\nu\}$  contains a subsequence  $\{a_\mu\}$  which tends to a point  $a$ . Then  $ab_\mu \leq aa_\mu + a_\mu b_\mu \rightarrow 0$ , hence, by (1.4),  $b_\mu a \rightarrow 0$  and  $b_\mu a_\mu \leq b_\mu a + aa_\mu \rightarrow 0$ .

In a finitely compact space the rectifiable curves have the following important compactness property:

(1.19) *In the finitely compact space  $R$  let  $c_\nu$ ,  $\nu = 1, 2, \dots$ , be a continuous curve with initial point  $a_\nu$ . If  $\{\lambda(c_\nu)\}$  and  $\{a_\nu\}$  are bounded then curves  $c_\nu(\tau)$ ;  $0 \leq \tau \leq \alpha$ , equivalent to  $c_\nu$  exist, such that a suitable subsequence  $\{c_\mu(\tau)\}$  of  $\{c_\nu(\tau)\}$  converges uniformly for  $0 \leq \tau \leq \alpha$  to a continuous curve  $c$  (with  $\lambda(c) \leq \liminf \lambda(c_\mu)$  by (1.10)).*

The literature does not seem to cover the present case. Therefore a proof will be given. By the hypothesis and (1.8) the set  $Uc_\nu$  is bounded. If  $\lambda(c_\mu) \rightarrow 0$  for a suitable subsequence  $\{c_\mu\}$  the assertion follows from (1.8). Otherwise we introduce  $\tau = \sigma_\nu / \lambda(c_\nu)$  as parameter on  $c_\nu$ , where  $\sigma_\nu$  is the arclength on  $c_\nu$ , and obtain a curve  $c_\nu(\tau)$ ,  $0 \leq \tau \leq 1$ , equivalent to  $c_\nu$ . For every fixed  $\tau$  the sequence  $\{c_\nu(\tau)\}$  is bounded and contains a converging subsequence. By a well known procedure a subsequence  $\{c_\mu\}$  is obtained such that  $\{c_\mu(\tau)\}$  converges for every  $\tau$  of the form  $\kappa 2^{-\rho}$  to a limit  $c(\tau)$ , where  $\rho = 1, 2, 3, \dots$  and  $\kappa = 0, 1, \dots, 2^\rho$ .

According to (1.18) choose for a given  $\epsilon > 0$  a number  $\delta > 0$  such that  $ab < \delta$  and  $a, b \in Uc_\mu$  imply  $ba < \epsilon/3$ , then  $\rho$  such that  $\lambda(c_\mu) \cdot 2^{-\rho} < \min(\epsilon/3, \delta)$ , and

(6) In Riemannian geometry finitely compact spaces are called normal, compare Cartan [1, p. 65].

finally  $\mu_0$  such that  $c_{\mu_1}(\kappa 2^{-\rho})c_{\mu_2}(\kappa 2^{-\rho}) < \epsilon/3$  for  $\kappa = 0, 1, 2, \dots, 2^\rho$  and  $\mu_1, \mu_2 > \mu_0$ .

For every  $\tau$ ,  $0 \leq \tau \leq 1$ , there is a  $\kappa$  with  $0 \leq \kappa 2^\rho - \tau < 2^{-\rho}$ . When  $\mu_1, \mu_2 > \mu_0$ ,  $c_{\mu_1}(\tau)c_{\mu_2}(\tau) \leq c_{\mu_1}(\tau)c_{\mu_1}(\kappa 2^{-\rho}) + c_{\mu_1}(\kappa 2^{-\rho})c_{\mu_2}(\kappa 2^{-\rho}) + c_{\mu_2}(\kappa 2^{-\rho})c_{\mu_2}(\tau) < \epsilon/3 + \epsilon/3 + \epsilon/3$ , because  $c_\mu(\tau)c_\mu(\kappa 2^{-\rho}) < 2^{-\rho}\lambda(c_\mu) < \min(\epsilon/3, \delta)$  and  $c_{\mu_2}(\tau)c_{\mu_2}(\kappa 2^{-\rho}) < \delta$  implies  $c_{\mu_2}(\kappa 2^{-\rho})c_{\mu_2}(\tau) < \epsilon/3$ . This shows that  $c_\mu(\tau)$  tends to a point  $c(\tau)$ , and that the convergence is uniform for  $0 \leq \tau \leq 1$ .

An important consequence of (1.8), (1.10), (1.19) is:

(1.20) *If in a finitely compact space  $a_v \rightarrow a$ ,  $b_v \rightarrow b$ , and segments  $\mathfrak{s}(a_v, b_v)$  exist, then a suitable subsequence  $\{\mathfrak{s}(a_\mu, b_\mu)\}$  of  $\{\mathfrak{s}(a_v, b_v)\}$  converges to a segment  $\mathfrak{s}(a, b)$ .*

Finite compactness is not a topologically invariant concept. But the spaces which can be metrized such that they become finitely compact may easily be characterized topologically. In the following theorem all terms are used in the same sense as in Lefschetz [1, chapter I] (quoted as L).

(1.21) THEOREM. *A Hausdorff space  $R$  possesses a finitely compact (symmetric) metrization if and only if it is locally compact and has a countable base.*

The necessity is obvious; for if  $R$  is finitely compact, then  $\bar{S}(a, \rho)$  is a compactum, hence separable [L, (45.1)] so that  $R = \bigcup_{\kappa=1}^{\infty} \bar{S}(a, \kappa)$  is separable and has therefore a countable base [L, (43.7)].

Conversely, let  $R$  be a locally compact Hausdorff space with a countable base. If  $R$  is compact, metrizability is a well known fact (see L, (46.4)). If  $R$  is locally compact but not compact, a point  $b$  can be added to  $R$  such that  $R_1 = R + b$  is a compact Hausdorff space [L, (29.3) and (37.1)]. The open sets of  $R_1$  are (1) the open sets  $G$  in  $R$  and (2) the sets  $R_1 - F$ , where  $F$  is a compact set in  $R$  (see the proof of L, (29.3)).  $R_1$  has also a countable base. For there exists a base  $\{B_v\}$  of  $R$  such that  $\bar{B}_v$  is compact in  $R$  [L, (29.2)]. Then  $b$  lies in no  $\bar{B}_v$ , so that  $C_\mu = R_1 - \bigcup_{v=1}^{\mu} \bar{B}_v$  contains  $b$ . The  $B_v$  and  $C_\mu$  form a base of  $R_1$ , that is, every open set in  $R_1$  is sum of  $B_v$  and  $C_\mu$ . An open set  $G$  in  $R$  is sum of  $B_v$ 's. Consider an open set  $R_1 - F$  of the second type. Since  $F$  is compact it is covered by a finite number of  $B_v$ , hence  $F \subset \bigcup_{\kappa=1}^{\mu} B_\kappa$  for a suitable  $\mu$ , so that  $R_1 - F \supset C_\mu \supset b$ . The set  $R - F$  is open in  $R$ , hence  $R_1 - F = (R - F) + b$  is sum of  $C_\mu$  and suitable  $B_v$ 's.

As a compact Hausdorff space with a countable base,  $R_1$  can be metrized [L, (46.4)]. Let  $\rho(x, y)$  be a symmetric metrization of  $R_1$ . Because  $\rho^{-1}(x, b)$  is defined and continuous on  $R$

$$(1.22) \quad xy = \rho(x, y) + |\rho^{-1}(x, b) - \rho^{-1}(y, b)|$$

is a symmetric metrization of  $R$ . Let  $\{x_v\}$  be a bounded sequence in  $R$ , that is,  $x_v p < \alpha$  for a suitable  $p \in R$  and  $\alpha > 0$ . Since  $R_1$  is compact,  $\{x_v\}$  contains a subsequence  $\{x_\lambda\}$  which converges to a point  $x$  in  $R_1$ . The definition (1.22) yields that  $\rho^{-1}(x_\lambda, b) - \rho^{-1}(p, b) < \alpha$  or  $\rho(x_\lambda, b) > [\alpha + \rho^{-1}(p, b)]^{-1}$ , hence  $x \neq b$ , or  $x \in R$ , which proves (1.21).

**2. Topology of subsets.** The purpose of this section is to metrize the subsets of a metric space in such fashion that the most important features of the known Hausdorff distance are preserved, and that the distance is defined for any two sets. Since at no occasion will this topology be applied to non-symmetric spaces,  $xy = yx$  is assumed throughout this section.

The Hausdorff distance  $\rho_h(G, H)$  of the two non-empty sets  $G, H$  is the greatest lower bound of all numbers  $\alpha$  for which simultaneously  $S(H, \alpha) \supset G$  and  $S(G, \alpha) \supset H$ . For unbounded sets  $\rho_h$  is in general not defined, for instance, not for two intersecting lines of the euclidean plane, which is precisely the type of sets with which this paper will be most concerned. Observe that

$$(2.1) \quad \rho_h(G, H) = \sup_{x \in R} |xG - xH|,$$

whenever  $\rho_h(G, H)$  is defined. For,

$$(2.2) \quad xG = xH \text{ for all } x \in R, \text{ if and only if } \overline{G} = \overline{H},$$

so that the two sides of (2.1) vanish only simultaneously (compare Hausdorff [1, p. 146]). Let  $\rho_h(G, H) > 0$  and  $0 < \epsilon < \rho_h(G, H)$ . Then either

$$G \not\subset S(H, \rho_h(G, H) - \epsilon) \quad \text{or} \quad H \not\subset S(G, \rho_h(G, H) - \epsilon).$$

Take the first case. Then  $G$  contains a point  $z$  such that

$$\rho_h(G, H) - \epsilon \leq zH = zH - zG \leq \sup_{x \in R} |xG - xH|.$$

Next choose  $z$  in  $H$  with  $xz < xH + \epsilon$ . There is a point  $y$  in  $G$  with  $yz < \rho_h(G, H) + \epsilon$  because  $H \subset S(G, \rho_h(G, H) + \epsilon)$ . Hence

$$\begin{aligned} xG - xH &\leq xy - xH \leq xz + zy - xH < xH + \epsilon + \rho_h(G, H) + \epsilon - xH \\ &= \rho_h(G, H) + 2\epsilon. \end{aligned}$$

Similarly  $xH - xG < \rho_h(G, H) + 2\epsilon$  is seen, which proves (2.1).

The relation (2.1) suggests defining a *distance for arbitrary non-empty sets* as follows: Fix a point  $p$  and put for any two sets  $G, H$

$$(2.3) \quad \rho_p(G, H) = \sup_{x \in R} |xG - xH| \cdot \exp(-px).$$

For the discussion of this function the following three trivial relations will be useful:

$$(2.4) \quad \tau \cdot \exp(-\tau) < 1 \quad \text{for all real } \tau.$$

$$(2.5) \quad |xG - yG| \leq xy.$$

$$(2.6) \quad \exp(-px) \leq \exp(-px) \cdot \exp(qp + px - qx) = \exp(-qx) \cdot \exp(pq).$$

The number  $\rho_p(G, H)$  is always finite because by (2.5)

$$|xG - xH| \leq xG + xH \leq pG + pH + 2px$$

hence by (2.4)

$$(2.7) \quad \rho_p(G, H) < pG + pH + 2.$$

The relations (2.2) and (2.3) yield that

$$(2.8) \quad \rho_p(G, H) = \rho_p(H, G) \geq 0 \text{ and } \rho_p(G, H) = 0 \text{ if and only if } \bar{G} = \bar{H}.$$

To prove the triangle inequality

$$(2.9) \quad \rho_p(G, H) + \rho_p(H, K) \geq \rho_p(G, K)$$

for any three sets  $G, H, K$ , observe that for a given  $\epsilon > 0$  a point  $z$  exists with

$$\begin{aligned} \rho_p(G, H) - \epsilon &< |zG - zH| \exp(-pz) \\ &\leq (|zG - zH| + |zH - zK|) \exp(-pz) \\ &\leq \rho_p(G, H) + \rho_p(H, K). \end{aligned}$$

For an arbitrary point  $q$  let

$$\rho_q(G, H) = \sup_{x \in R} |xG - xH| \cdot \exp(-qx)$$

then by (2.6)

$$\rho_q(G, H) \leq \sup_{x \in R} |xG - xH| \cdot \exp(-px) \exp(pq) \leq \rho_p(G, H) \exp(pq)$$

hence

$$(2.10) \quad \rho_p(G, H) \exp(-pq) \leq \rho_q(G, H) \leq \rho_p(G, H) \exp(pq)$$

which shows that  $\rho_p$  and  $\rho_q$  are topologically equivalent.

(2.1) and (2.3) show that  $\rho_h(G_r, G) \rightarrow 0$  implies  $\rho_p(G_r, G) \rightarrow 0$ . The converse is obviously not true. It is not even necessarily true when the sets  $G_r$  and  $G$  are bounded (individually). However, clearly  $\rho_h(G_r, G) \rightarrow 0$  when  $\rho_p(G_r, G) \rightarrow 0$  and  $\bigcup G_r$  and  $G$  are bounded.

*Example.*  $R^1$  consists of the numbers  $0, 1, 2, \dots$  with  $|\nu - \mu|$  as distance;  $G$  consists of 0 and  $G_r$  of the numbers 0 and  $\nu$ . When  $p=0$  then

$$\begin{aligned} \rho_p(G, G_\mu) &= \sup_{\nu=0, 1, 2, \dots} (\nu - \min(\nu, |\mu - \nu|)) \exp(-\nu) \\ &= \sup_{2\nu > \mu} (\nu - |\mu - \nu|) \exp(-\nu) \\ &\leq \sup_{2\nu > \mu} (2\nu + \mu) \exp(-\nu) \leq \sup_{2\nu > \mu} 4\nu \exp(-\nu). \end{aligned}$$

Hence  $\rho_p(G, G_\mu) \rightarrow 0$  for  $\mu \rightarrow \infty$ , whereas  $\rho_h(G, G_\mu) \rightarrow \infty$ .

Notice for later applications that

$$(2.11) \quad \{\rho_p(G_r, G_\mu)\} \text{ is bounded if and only if } \{pG_\mu\} \text{ is bounded.}$$

The necessity is obvious because

$$\rho_p(G_\mu, G_1) \geq pG_\mu - pG_1$$



and the sufficiency follows from (2.7).

The existence of the closed limit  $FlG_\nu$ <sup>(7)</sup> is connected to convergence in terms of  $\rho_p$  in a very similar fashion as to convergence in terms of  $\rho_h$ .

(2.12) If  $\rho_p(G_\nu, G) \rightarrow 0$  then  $FlG_\nu$  exists and equals  $\bar{G}$ .

For let  $x \in \bar{G}$ . Then

$$|xG_\nu - xG| \cdot \exp(-px) = xG_\nu \exp(-px) \rightarrow 0.$$

Hence  $y_\nu \in G_\nu$  exists with  $xy_\nu \rightarrow 0$ , so that  $\bar{G} \subset Fl \inf G_\nu$ . Next let  $y \in Fl \sup G_\nu$ . Then  $y_\nu \in G_{\mu_\kappa}$  with  $y_\nu \rightarrow y$  exists and

$$\exp(py)\rho_p(G_{\mu_\kappa}, G) \geq y\bar{G} - yG_{\mu_\kappa} \geq y\bar{G} - y_\nu G_{\mu_\kappa} - yy_\nu = y\bar{G} - yy_\nu$$

so that  $y\bar{G} = 0$  and  $Fl \sup G_\nu \subset \bar{G}$ .

(2.13) THEOREM. If  $R$  is finitely compact and  $FlG_\nu = G \neq 0$  then  $\rho_p(G_\nu, G) \rightarrow 0$ .

For an indirect proof let the subsequence  $\{G_\nu^*\}$  of  $\{G_\nu\}$  satisfy the inequality  $\rho_p(G_\nu^*, G) \geq 4\alpha > 0$ . Then a point  $x_\nu$  with

$$(2.14) \quad |x_\nu G_\nu^* - x_\nu G| \cdot \exp(-px_\nu) > 2\alpha$$

exists for large  $\nu$ . If  $z \in G$  then  $G_\nu^*$  contains a point  $z_\nu$  with  $z_\nu \rightarrow z$  and

$$|x_\nu G_\nu^* - x_\nu G| \leq x_\nu z_\nu + x_\nu z \leq 2px_\nu + pz_\nu + pz$$

and by (2.14),  $(2px_\nu + pz_\nu + pz) \exp(-px_\nu) > 2\alpha$ , so that  $\{px_\nu\}$  is bounded.

(2.14) implies that  $\{G_\nu^*\}$  contains either a subsequence  $\{G'_\nu\}$  with

$$(2.15) \quad x'_\nu G'_\nu - x'_\nu G > 2\alpha$$

or a subsequence  $\{G''_\nu\}$  with

$$(2.16) \quad x''_\nu G - x''_\nu G''_\nu > 2\alpha$$

(or subsequences of both types), where  $x'_\nu, x''_\nu$  signify the elements in  $\{x_\nu\}$  with the same subscript as  $G'_\nu, G''_\nu$  have in  $\{G_\nu^*\}$ . Let (2.15) hold and choose  $y_\nu$  in  $G$  such that  $x'_\nu y_\nu < x'_\nu G + \alpha$ , so that  $\{y_\nu\}$  is bounded and has an accumulation point  $y$ . As a closed limit  $G$  is closed and contains  $y$ . On the other hand we obtain from the last inequality and (2.5), (2.15)

$$y_\nu G'_\nu + \alpha > x'_\nu G'_\nu - x'_\nu y_\nu + \alpha > x'_\nu G'_\nu - x'_\nu G > 2\alpha$$

which yields  $yG'_\nu > \alpha/2$  for large  $\nu$ , so that  $G \not\subset Fl \inf G_\nu$ .

Finally let (2.16) hold and choose  $t_\nu$  in  $G''_\nu$  with  $x''_\nu t_\nu < x''_\nu G''_\nu + \alpha$ . The sequence  $\{t_\nu\}$  is bounded, because  $\{x_\nu\}$  and  $\{z_\nu\}$  are bounded. By (2.5), (2.16)

$$t_\nu G + \alpha > x''_\nu G - x''_\nu t_\nu + \alpha > x''_\nu G - x''_\nu G''_\nu > 2\alpha.$$

<sup>(7)</sup> The terminology and notations of Hausdorff [1, pp. 146-150] are used. The closed limit superior  $Fl \sup G_\nu$  (closed limit inferior  $Fl \inf G_\nu$ ) of  $\{G_\nu\}$  consists of those points  $x$  for which every neighborhood  $S(x, \epsilon) > 0$  contains points of infinitely many  $G_\nu$  (of all but a finite number of  $G_\nu$ ). If  $Fl \sup G_\nu = Fl \inf G_\nu$  this set is the closed limit  $FlG_\nu$ .

Hence an accumulation point  $t$  of  $\{t_v\}$  would not lie in  $G$  although  $t \in Fl \sup G_v$ .

This theorem yields the following result.

(2.17) *If  $R$  is finitely compact then every bounded sequence  $\{G_v\}$  of non-empty sets (that is,  $\{\rho_p(G_v, G_\mu)\}$  is bounded) contains a subsequence  $\{G_\kappa\}$  which converges metrically to a set  $G$ , that is,  $\rho_p(G_\kappa, G) \rightarrow 0$ .*

For, since  $R$  is separable,  $\{G_v\}$  contains a subsequence  $\{G_\kappa\}$  and  $R$  a set  $G$  such that  $Fl G_\kappa = G$  (see Hausdorff [1, p. 147]).  $G_\kappa$  is not empty and contains by (2.11) a point  $z_\kappa$  such that  $\{pz_\kappa\}$  is bounded. Hence  $\{z_\kappa\}$  has an accumulation point  $z \in Fl \sup G_v = G$ , so that  $G$  is not empty. (2.13) yields now (2.17).

Combining the results (2.2), (2.12), (2.13), (2.17) leads to

(2.18) **THEOREM.** *The non-empty closed sets of a finitely compact symmetric space  $R$  form with the metric  $\rho_p$  a finitely compact metric space.*

This theorem in conjunction with Theorem (1.21) yields an interesting fact on the closed sets of a locally compact Hausdorff space with a countable base, whose formulation may be left to the reader.

**3. Topology of motions.** The usual terminology will be used for mappings:  $\Phi$  is a mapping of the set  $A$  in the set  $B$  if every point  $x$  of  $A$  determines a unique point  $y = x\Phi$  in  $B$ . For a subset  $T$  of  $A$  we put  $T\Phi = \bigcup_{x \in T} x\Phi$ . In general  $A\Phi \subset B$ ; only if  $A\Phi = B$  then  $\Phi$  is a mapping of  $A$  on  $B$ .

A mapping  $\Phi$  of a (not necessarily symmetric) metric space  $R$  on itself is called a motion if it preserves distances:

$$(3.1) \quad xy = x\Phi y\Phi \quad \text{for all } x, y \text{ in } R.$$

Obviously  $\Phi$  is a topological mapping of  $R$  on itself and  $\Phi^{-1}$  is also a motion. Moreover the motion  $\Phi$  followed by the motion  $\Psi$  is again a motion  $\Phi\Psi$ . Hence the motions which  $R$  admits form a group  $\Gamma$ .

The purpose of this section is to provide  $\Gamma$  with a topology which agrees with the ordinary notions of continuity in cases like the euclidean space. (3.1) applied to  $x, y$  and  $y, x$  yields (compare (1.5))

$$(3.2) \quad \sigma(x, y) = \sigma(x\Phi, y\Phi),$$

so that a motion of  $R$  for the metric  $xy$  is at the same time a motion for  $R$  with  $\sigma(x, y)$  as metric. The converse is not necessarily true, so that the group  $\Gamma_\sigma$  of motions of  $R$  with  $\sigma(x, y)$  as metric may contain  $\Gamma$  as a proper subgroup.

*Example.*  $R^2$ . In the  $E^\lambda$  with rectangular coordinates  $\xi_1, \dots, \xi_\lambda$  let  $K$  be a closed convex surface which contains  $(0, \dots, 0)$  in its interior. Every ray  $\eta_\kappa = \tau\xi_\kappa$ ,  $\xi_\kappa \neq 0$ ,  $\tau \geq 0$ , intersects  $K$  in exactly one point  $\eta(\xi)$ . Putting generally  $|\alpha| = [\sum \alpha_\kappa^2]^{1/2}$  and defining  $\phi(\xi) = |\xi|/|\eta(\xi)|$ ,  $\phi(0) = 0$ , the function

$$\rho(\xi_1, \xi_2) = \phi(\xi_2 - \xi_1) \quad (*)$$

will satisfy the requirements (1.1)–(1.4), moreover  $E^\lambda$  will be a finitely com-

(\*) On this metric compare Minkowski [1, §§17, 18].

pact space  $R^2$ , which admits as motions all translations of the  $E^\lambda$ . The space  $R^2$  is not symmetric unless 0 is the center of  $K$ . Therefore the symmetry  $\xi\Phi = -\xi$  in 0, or more generally the symmetry  $\xi\Phi = 2\zeta - \xi$  in an arbitrary point  $\zeta$ , will not be a motion of  $R^2$  unless 0 is the center of  $K$ . However, these symmetries are motions for  $R$  with the symmetric metric

$$\sigma(\xi_1, \xi_2) = \max (\phi(\xi_2 - \xi_1), \phi(\xi_1 - \xi_2)).$$

By topologizing  $\Gamma_\sigma$  the group  $\Gamma$  is also topologized. For bounded spaces the number  $\sup_{x \in R} \sigma(x\Phi, x\Psi)$  yields a natural topology. In unbounded spaces this number is not always defined. A similar device as in §2 may be applied: fix a point  $p$  and put

$$(3.3) \quad \rho_p(\Phi, \Psi) = \sup_{x \in R} [\sigma(x\Phi, x\Psi)] \exp (-\sigma(p, x)).$$

For the sake of brevity set  $\exp (-\sigma(a, b)) = \phi(a, b)$ . Since

$$\begin{aligned} \sigma(x\Phi, x\Psi) &\leq \sigma(x\Phi, p) + \sigma(p, x\Psi) = \sigma(x, p\Phi^{-1}) + \sigma(p\Psi^{-1}, x) \\ &\leq 2\sigma(p, x) + \sigma(p, p\Phi^{-1}) + \sigma(p, p\Psi^{-1}) \end{aligned}$$

(2.4) yields

$$(3.4) \quad \rho_p(\Phi, \Psi) \leq 2 + \sigma(p, p\Phi^{-1}) + \sigma(p, p\Psi^{-1}),$$

so that  $\rho_p(\Phi, \Psi)$  is always finite. Obviously  $\rho_p(\Phi, \Psi) = \rho_p(\Psi, \Phi) \geq 0$  and  $\rho_p(\Phi, \Psi) = 0$  if and only if  $\sigma(x\Phi, x\Psi) = 0$  for all  $x$ , or  $\Phi = \Psi$ . The proof of the triangle inequality

$$(3.5) \quad \rho_p(\Phi, \Psi) + \rho_p(\Psi, \Omega) \geq \rho_p(\Phi, \Omega)$$

for any three elements of  $\Gamma_\sigma$  is practically identical with the proof of (2.9). Let for any point  $q$

$$\rho_q(\Phi, \Psi) = \sup_{x \in R} [\sigma(x\Phi, x\Psi) \cdot \phi(q, x)].$$

Then (2.6) shows that

$$\rho_q(\Phi, \Psi) \leq \sup_{x \in R} (x\Phi, x\Psi) \cdot \phi(p, x) \cdot \phi^{-1}(p, q) \leq \rho_p(\Phi, \Psi) \phi^{-1}(p, q)$$

hence

$$(3.6) \quad \rho_p(\Phi, \Psi) \phi(p, q) \leq \rho_q(\Phi, \Psi) \leq \rho_p(\Phi, \Psi) \phi^{-1}(p, q)$$

which shows that  $\rho_p$  and  $\rho_q$  are topologically equivalent metrizations of  $\Gamma_\sigma$ . Henceforth it is understood that  $\Gamma_\sigma$  is topologized in this fashion. Notice the following important properties of  $\rho_p$ .

(3.7)  $\rho_p(\Phi, \Psi) = \rho_p(\Phi\Omega, \Psi\Omega)$  for any  $\Phi, \Psi, \Omega$  because  $\sigma(x\Phi\Omega, x\Psi\Omega) = \sigma(x\Phi, x\Psi)$  for all  $x$ .

(3.8)  $\rho_p(\Phi, \Psi) = \rho_{p\Omega^{-1}}(\Omega\Phi, \Omega\Psi)$  because  $y = x\Omega^{-1}$  traverses  $R$  when  $x$  does and therefore

$$\begin{aligned}\sup_x \sigma(x\Phi, x\Psi)\phi(p, x) &= \sup_x \sigma(x\Omega^{-1}\Omega\Phi, x\Omega^{-1}\Omega\Psi)\phi(p\Omega^{-1}, x\Omega^{-1}) \\ &= \sup_y \sigma(y\Omega\Phi, y\Omega\Psi)\phi(p\Omega^{-1}, y).\end{aligned}$$

(3.9) The sequence  $\{\Phi_\nu\}$  is bounded (that is,  $\rho_p(\Phi_\nu, \Phi_\mu) < \alpha$ ) if and only if for a suitable (or every) point  $a$  the set  $\{\sigma(a\Phi_\nu, a\Phi_\mu)\}$  is bounded.

If  $\rho_p(\Phi_\nu, \Phi_\mu) < \alpha$ , then  $\sigma(a\Phi_\nu, a\Phi_\mu) < \alpha \cdot \phi^{-1}(p, a)$  for every  $a$ . Let  $\sigma(a\Phi_\nu, a\Phi_\mu) < \alpha$ . Then for any point  $x$  by (2.4), (2.6)

$$\begin{aligned}\sigma(x\Phi_\nu, x\Phi_\mu)\phi(p, x) &\leq [\sigma(x\Phi_\nu, a\Phi_\nu) + \sigma(a\Phi_\nu, a\Phi_\mu) + \sigma(a\Phi_\mu, x\Phi_\mu)] \\ \cdot \phi(a, x)\phi^{-1}(a, p) &\leq [2\sigma(a, x)\phi(a, x) + \alpha]\phi^{-1}(a, p) \leq [2 + \alpha]\phi^{-1}(a, p).\end{aligned}$$

With the help of these facts the influence of the properties of  $R$  on the topology of  $\rho_p$  in  $\Gamma_\sigma$  may be discussed.

(3.10) THEOREM. *If  $R$  is complete then  $\Gamma_\sigma$  is complete.*

It is to be shown that if  $\lim_{\mu, \nu} \rho_p(\Phi_\mu, \Phi_\nu) = 0$  then a motion  $\Phi$  in  $\Gamma_\sigma$  exists with  $\rho_p(\Phi_\mu, \Phi) \rightarrow 0$ . For every point  $y$  the sequence  $\{y\Phi_\nu\}$  is a fundamental sequence because

$$\sigma(y\Phi_\mu, y\Phi_\nu) \leq \rho_p(\Phi_\mu, \Phi_\nu)\phi^{-1}(p, y).$$

Since  $R$  is complete,  $y\Phi_\mu$  converges to a point  $y_0$  and it follows from

$$\sigma(x, y) = \sigma(x\Phi_\mu, y\Phi_\mu) \rightarrow \sigma(x_0, y_0)$$

that  $x \rightarrow x_0$  is a mapping  $\Phi_0$  of  $R$  in itself which preserves distances.  $\Phi_0$  will be a motion if it maps  $R$  on itself, or if for a given point  $q$  of  $R$  a point  $\bar{q}$  with  $\bar{q}\Phi_0 = q$  exists.

Because of (3.6), (3.7), (3.8)

$$\begin{aligned}\sigma(q\Phi_\mu^{-1}, q\Phi_\nu^{-1})\phi(p, q) &\leq \rho_p(\Phi_\mu^{-1}, \Phi_\nu^{-1}) = \rho_p(E, \Phi_\nu^{-1}\Phi_\mu) \\ &\leq \rho_{p_0}(E, \Phi_\nu^{-1}\Phi_\mu) \cdot \phi^{-1}(p, p_0) = \rho_{p_0\Phi_\mu^{-1}}(\Phi_\nu, \Phi_\mu)\phi^{-1}(p, p_0) \\ &\leq \rho_p(\Phi_\mu, \Phi_\nu)\phi^{-1}(p, p_0)\phi^{-1}(p, p_0\Phi_\mu^{-1}).\end{aligned}$$

Since  $\sigma(p_0, p_0\Phi_\mu^{-1}) = \sigma(p\Phi_\mu, p_0) \rightarrow 0$ , it follows that  $\{q\Phi_\mu^{-1}\}$  is a fundamental sequence, hence converges to a point  $\bar{q}$  with

$$\sigma(q, \bar{q}\Phi) = \lim \sigma(q, \bar{q}\Phi_\mu) = \lim \sigma(q\Phi_\mu^{-1}, q) = 0.$$

Finally it must be shown that  $\rho_p(\Phi_\mu, \Phi) \rightarrow 0$ . For a given  $\epsilon > 0$  choose  $\alpha$  such that for every  $x$  and  $\mu, \nu > \alpha$ ,

$$\sigma(x\Phi_\mu, x\Phi_\nu)\phi(p, x) \leq \rho_p(\Phi_\mu, \Phi_\nu) < \epsilon.$$

For  $\nu \rightarrow \infty$  this becomes  $\sigma(x\Phi_\mu, x\Phi)\phi(p, x) \leq \epsilon$  for  $\mu > \alpha$  and all  $x$ , so that  $\rho_p(\Phi_\mu, \Phi) \leq \epsilon$  for  $\mu > \alpha$ .

In the theory of transformations one frequently uses a different topology which is determined by the definition

$$(3.11) \quad \Phi = \lim \Phi_\nu \text{ if } \sigma(y\Phi_\nu, y\Phi) \rightarrow 0 \text{ for every } y.$$

In sufficiently general spaces this limit concept may differ from  $\rho_p(\Phi, \Phi_\nu) \rightarrow 0$ .

*Example.*  $R^3$  consists of countably many points  $x_1, x_2, \dots$  with the metric  $x_\nu x_\mu = 1$  for  $\nu \neq \mu$  and  $x_\nu x_\nu = 0$ , so that  $R^3$  is locally compact and separable, but not finitely compact. Every 1-1-mapping of  $R^3$  is a motion.  $\rho_{xy}(\Phi, \Psi) = 1$  or  $\exp(-1)$  according to whether  $x_\nu \Phi x_\nu \Psi = 1$  or 0, hence all elements of  $\Gamma_\sigma$  are isolated. But with the topology (3.11) no isolated elements exist. It suffices to show that the identity  $E$  is not isolated. Define  $\Phi_\nu$  by  $x_\mu \Phi_\nu = x_\mu$  for  $\mu \neq \nu$ ,  $\nu + 1$ ,  $x_\nu \Phi_\nu = x_{\nu+1}$ ,  $x_{\nu+1} \Phi_\nu = x_\nu$ . Then  $\lim_\nu x_\mu \Phi_\nu = x_\mu$  for every  $\mu$ , hence  $\lim \Phi_\nu = E$  in the sense of (3.11).

The two topologies coincide for a finitely compact space  $R$ .

(3.12) *In a finitely compact space  $\rho_p(\Phi_\nu, \Phi) \rightarrow 0$  if and only if  $\lim y\Phi_\nu = y\Phi$  for every  $y \in R$ .*

Since for every  $y$

$$\phi^{-1}(p, y)\rho_p(\Phi_\nu, \Phi) \geq \sigma(y\Phi_\nu, y\Phi)$$

the relation  $\rho_p(\Phi_\nu, \Phi) \rightarrow 0$  implies  $y\Phi_\nu \rightarrow y\Phi$ . This much is true in any metric space.

Conversely, if  $y\Phi_\nu \rightarrow y\Phi$  for every  $y$  did not imply  $\rho_p(\Phi_\nu, \Phi) \rightarrow 0$ , then a sequence of points  $y_\nu$  and a subsequence  $\{\Phi_\nu^*\}$  of  $\{\Phi_\nu\}$  would exist such that

$$\sigma(y_\nu \Phi_\nu^*, y_\nu \Phi) \cdot \phi(p, y_\nu) > \eta > 0.$$

Since  $R$  is finitely compact it may be assumed that either  $p y_\nu \rightarrow \infty$  or  $y_\nu$  converges to a point  $y$ . The first case is impossible because

$$\begin{aligned} \sigma(y_\nu \Phi_\nu^*, y_\nu \Phi) \phi(p, y_\nu) &\leq [\sigma(y_\nu \Phi_\nu^*, p \Phi_\nu^*) + \sigma(p \Phi_\nu^*, p \Phi) + \sigma(p \Phi, y_\nu \Phi)] \phi(p, y_\nu) \\ &\leq [2\sigma(p, y_\nu) + \sigma(p \Phi_\nu^*, p \Phi)] \phi(p, y_\nu) \rightarrow 0. \end{aligned}$$

The second case is excluded by the relations  $\sigma(y_\nu, y) \rightarrow 0$ ,  $\sigma(y\Phi_\nu^*, y\Phi) \rightarrow 0$  and

$$\begin{aligned} \sigma(y_\nu \Phi_\nu^*, y_\nu \Phi) \phi(p, y_\nu) &\leq \sigma(y_\nu \Phi_\nu^*, y\Phi_\nu^*) + \sigma(y\Phi_\nu^*, y\Phi) + \sigma(y\Phi, y_\nu \Phi) \\ &= 2\sigma(y, y_\nu) + \sigma(y\Phi_\nu^*, y\Phi). \end{aligned}$$

The most important fact for the present paper is

(3.13) **THEOREM.** *If  $R$  is finitely compact then  $\Gamma_\sigma$  and  $\Gamma$  are finitely compact.*

**Proof.** By (3.9) the sequence  $\{x\Phi_\nu\}$  is bounded for every  $x$  if  $\{\Phi_\nu\}$  is bounded. Let  $\{\rho_\nu\}$  be a countable set of points which is everywhere dense in  $R$ . Since  $\{p_1\Phi_\nu\}$  is bounded there is a subsequence  $\{1\nu\}$  of  $\{\nu\}$  such that  $\{p_1\Phi_{1\nu}\}$  tends to a point  $p_1$ . In  $\{1\nu\}$  select a subsequence  $\{2\nu\}$  such that  $\{p_2\Phi_{2\nu}\}$  tends to a point  $p_2$ , and so forth. The diagonal sequence  $\Psi_\nu = \Phi_{\nu\nu}$  satisfies the relation

$$(3.14) \quad \lim_{\nu \rightarrow \infty} p_\mu \Psi_\nu = \bar{p}_\mu \quad \text{for all } \mu.$$

Since  $\Psi_\nu$  is a motion

$$(3.15) \quad \sigma(p_\lambda, p_\mu) = \sigma(\bar{p}_\lambda, \bar{p}_\mu) \quad \text{for all } \lambda, \mu.$$

Hence, when  $\{p_{\nu_k}\}$  is a fundamental sequence, then  $\{\bar{p}_{\nu_k}\}$  is fundamental, so that the mapping  $p_k \rightarrow \bar{p}_k$  can be extended to a mapping  $\Psi$  of  $R$  in itself with

$$(3.16) \quad x\Psi_\nu \rightarrow x\Psi \quad \text{and} \quad \sigma(x, y) = \sigma(x\Psi, y\Psi) \quad \text{for all } x, y.$$

$\Psi$  will be a motion if for every point  $q$  in  $R$  a point  $\bar{q}$  with  $\bar{q}\Psi = q$  exists. Since  $q\Psi_\nu^{-1}p = qp\Psi_\nu \rightarrow qp\Psi$  the sequence  $\{q\Psi_\nu^{-1}\}$  is bounded. Let  $q\Psi_\lambda^{-1} \rightarrow \bar{q}$  then

$$q\bar{q}\Psi = \lim q\bar{q}\Psi_\lambda = \lim q\Psi_\lambda^{-1}\bar{q} = 0.$$

Hence  $\Psi$  is a motion.  $\rho_p(\Psi_\nu, \Psi) \rightarrow 0$  by (3.12), (3.16), so that  $\Gamma_\sigma$  is finitely compact. As a consequence of (3.12) the group  $\Gamma$  is a closed subset of  $\Gamma_\sigma$ , hence  $\Gamma$  is also finitely compact.

This theorem is the main justification for the metric  $\rho_p(\Phi, \Psi)$ . The choice of the function  $\phi(p, q)$  is, of course, largely arbitrary, for instance  $[\sigma^2(p, x) + 1]^{-1}$  would yield the same results. Other functions, like the known

$$\rho(\Phi, \Psi) = \sup_x \sigma(x\Phi, x\Psi) [1 + \sigma(x\Phi, x\Psi)]^{-1},$$

might seem more adequate, in particular because

$$\rho(\Phi, \Psi) = \rho(\Phi\Omega, \Psi\Omega) = \rho(\Omega\Phi, \Omega\Psi).$$

But (3.13) does not hold for  $\rho(\Phi, \Psi)$ , even for the euclidean plane  $E^2$  with the euclidean metric. Namely  $\rho(\Phi, \Psi) = 1$ , for any two different rotations  $\Phi, \Psi$  about a point  $p$ , which shows that the motions of  $E^2$  with  $\rho(\Phi, \Psi)$  as metric do not form a finitely compact space; moreover convergence of motions does not agree with the intuition. But

$$\rho_p(\Phi, \Psi) = \rho_p(E, \Psi\Phi^{-1}) = \sup_{\tau \geq 0} [2\tau^2(1 - \cos \theta) \exp(-\tau)] = 8e^{-2}(1 - \cos \theta),$$

where  $\theta$  is the angle of the rotation  $\Psi\Phi^{-1}$ , so that  $\rho_p(\Phi, \Psi) \rightarrow 0$  and therefore  $\rho_q(\Phi, \Psi) \rightarrow 0$  for any  $q$ , whenever  $\theta \rightarrow 0$ .

Some remarks on subgroups of  $\Gamma_\sigma$  will conclude this section.

(3.17) If  $\Delta$  is a subgroup of  $\Gamma_\sigma$ , then the closure  $\bar{\Delta}$  of  $\Delta$  is a subgroup of  $\Gamma_\sigma$ .

It is to be shown that the relations  $\Phi_\nu, \Psi_\nu \in \Delta$ ,  $\Phi, \Psi \in \Gamma_\sigma$  and  $\rho_p(\Phi_\nu, \Phi) \rightarrow 0$ ,  $\rho_p(\Psi_\nu, \Psi) \rightarrow 0$  imply  $\Phi\Psi^{-1} \in \bar{\Delta}$ . This follows from (3.6), (3.7), (3.8) and

$$\begin{aligned} \rho_p(\Phi_\nu\Psi_\nu^{-1}, \Phi\Psi) &= \rho_p(\Phi_\nu, \Phi\Psi^{-1}\Psi_\nu) \leq \rho_p(\Phi_\nu, \Phi) + \rho_p(\Phi, \Phi\Psi^{-1}\Psi_\nu) \\ &= \rho_p(\Phi_\nu, \Phi) + \rho_{p^*}(\Psi, \Psi_\nu), \end{aligned}$$

where  $p^* = p\Phi\Psi^{-1}$ .

With any subgroup  $\Delta$  of  $\Gamma$  and any point  $a$  of  $R$  we associate the orbit  $T(\Delta, a)$  of  $a$  under  $\Delta$ , which consists of those points  $x$  for which a motion  $\Phi$  in  $\Delta$  with  $x = a\Phi$  exists. The sets  $T(\Delta, a)$  have obviously the following properties:

(3.18) Either  $T(\Delta, a) = T(\Delta, b)$  or  $T(\Delta, a) \cap T(\Delta, b) = 0$ .

(3.19)  $xT(\Delta, b) = T(\Delta, a)y = T(\Delta, a)T(\Delta, b)$  for  $x \in T(\Delta, a)$  and  $y \in T(\Delta, b)$ .

(3.20)  $T(\Delta, a)\Omega = T(\Omega^{-1}\Delta\Omega, a\Omega)$  for any  $\Omega$  in  $\Gamma$ , in particular

(3.21)  $T(\Delta, a)\Omega = T(\Delta, a)$  if  $\Omega \in \Delta$ .

(3.22) If  $R$  is finitely compact, then  $T(\bar{\Delta}, a)$  is finitely compact.

For a proof let  $\{x_\nu\}$  be a bounded sequence in  $T(\bar{\Delta}, a)$ . Then  $\Phi_\nu \in \bar{\Delta}$  exists with  $x_\nu = a\Phi_\nu$ . By (3.9), (3.13) the sequence  $\{\Phi_\nu\}$  is bounded and contains a subsequence  $\{\Phi_\lambda\}$  which converges to a motion  $\Phi \in \bar{\Delta}$ . Then  $x_\lambda = a\Phi_\lambda \rightarrow a\Phi \in T(\bar{\Delta}, a)$ , q.e.d.

These remarks lead to the following fact which is useful for the construction of covering spaces (see §13).

(3.23) *Let  $\Delta$  be a closed group of motions of the finitely compact space  $R$ . Then the orbits  $T(\Delta, a) = a^*$  form with the metric  $a^*b^* = T(\Delta, a)T(\Delta, b)$  a finitely compact space  $R^*$ , which is convex when  $R$  is convex.*

**Proof.** For every point  $x \in T(\Delta, a)$  there exists by (3.19), (3.22) a point  $y \in T(\Delta, b)$  with  $xy = T(\Delta, a)T(\Delta, b)$ . Hence by (3.18),  $T(\Delta, a)T(\Delta, b) = 0$  if and only if  $T(\Delta, a) = T(\Delta, b)$  or  $a^* = b^*$ . To prove the triangle inequality choose for a given point  $x \in T(\Delta, a)$  a point  $y \in T(\Delta, b)$  with  $xy = T(\Delta, a)T(\Delta, b)$  and then  $z \in T(\Delta, c)$  with  $yz = T(\Delta, b)T(\Delta, c)$ . Then

$$a^*b^* + b^*c^* = xy + yz \geq xz \geq T(\Delta, a)T(\Delta, c) = a^*c^*.$$

Let  $a_\nu^* \in S'(a_0^*, \rho)$ . If  $a_\nu^* = T(\Delta, a_\nu')$ ,  $\nu = 0, 1, 2, \dots$ , put  $a_0 = a_0'$  and choose  $a_\nu \in T(\Delta, a_\nu')$  with  $a_0a_\nu = a_\nu^*a_0^*$ , so that  $a_\nu \in S'(a_0, \rho)$ . Hence a subsequence  $\{a_\mu\}$  of  $\{a_\nu\}$  tends to a point  $b$ . Then  $a_\mu b \geq T(\Delta, a_\mu')T(\Delta, b) \geq a_\mu^*b^*$ , and  $a_\mu^* \rightarrow b^*$ .

Finally, if  $R$  is convex and  $a^* = T(\Delta, a) \neq b^* = T(\Delta, b)$  are given, choose  $a' \in T(\Delta, a)$ ,  $b' \in T(\Delta, b)$  with  $a'b' = a^*b^*$  and let  $(a'c'b')$ ,  $c^* = T(\Delta, c')$ . Then  $a^*b^* \leq a^*c^* + c^*b^* \leq a'c' + c'b' = a'b' = a^*b^*$  which proves  $(a^*c^*b^*)$ .

The following fact is important for abelian groups:

(3.24) *If every motion  $\Phi$  of a group  $\Delta$  of motions of the metric space  $R$  satisfies the relation*

$$(3.25) \quad xx\Phi = yy\Phi \quad \text{for any } x, y \text{ in } R$$

*then the correspondence  $\Phi \rightarrow a\Phi$  maps, for every  $\rho_p$  and  $a$ , the group  $\Delta$  isometrically on the orbit  $T(\Delta, a)$  with the metric  $\sigma$ . The group  $\Delta$  is simply transitive on  $T(\Delta, a)$  and  $\bar{\Delta}$  also satisfies (3.25).*

*If  $\Delta$  is transitive in  $R$  (that is,  $T(\Delta, a) = R$ ) then it is simply transitive and closed.*

(3.25) holds for transitive abelian  $\Delta$ .

**Proof.** Any two motions  $\Phi, \Psi$  in  $\Delta$  and any two points  $p, x$  in  $R$  satisfy the relations

$$p\Phi p\Psi = p p\Psi \Phi^{-1} = x x\Psi \Phi^{-1} = x\Phi x\Psi \quad \text{and} \quad p\Psi p\Phi = x\Psi x\Phi,$$

hence  $\sigma(p\Phi, p\Psi) = \sigma(x\Phi, x\Psi)$  and

$$\rho_p(\Phi, \Psi) = \sigma(p\Phi, p\Psi) = \sigma(a\Phi, a\Psi),$$

which shows that the mapping  $\Phi \rightarrow a\Phi$  of  $\Delta$  on  $T(\Delta, a)$  is isometric.

Now let  $\Phi_v \in \Delta$  and  $\Phi_v \rightarrow \Phi \in \bar{\Delta}$ . Since  $xx\Phi_v = yy\Phi_v$  for all  $v$  and  $x, y$  it follows from the proof of (3.12) that  $xx\Phi = yy\Phi$ , so that  $\Phi$  satisfies (3.25).

Let the two motions  $\Phi$  and  $\Psi$  in  $\Delta$  carry  $p$  into  $q$ . Then  $0 = pq\Phi^{-1} = pp\Psi\Phi^{-1} = xx\Psi\Phi^{-1} = x\Phi x\Psi$  for every  $x$  so that  $\Phi = \Psi$ , which shows that  $\Delta$  is simply transitive on  $T(\Delta, a)$ , in particular simply transitive in  $R$  if  $T(\Delta, a) = R$ . In that case  $\bar{\Delta} = \Delta$  because  $\bar{\Delta}$  has also property (3.25) and is therefore simply transitive.

Finally let  $\Delta$  be transitive and abelian. If  $x, y$  are arbitrary in  $R$ , then  $\Psi \in \Delta$  exists with  $x\Psi = y$ . The relation (3.25) holds for any motion  $\Phi$  in  $\Delta$  because

$$xx\Phi = x\Psi x\Phi\Psi = x\Psi x\Psi\Phi = yy\Phi.$$

## II. EXTREMALS AND GEODESICS

**4. Segments in  $E$ -spaces.** The basic axioms for a space  $R$  with extremals, which will be assumed throughout the remaining part of this paper, are the following.

- A.  $R$  is metric, that is,  $R$  satisfies the conditions (1.1), (1.2), (1.3), (1.4).
- B.  $R$  is finitely compact.
- C.  $R$  is convex.

D. Every point  $x$  of  $R$  has a neighborhood  $S(x, \alpha)$  with this property: For any two distinct points  $a, b$  in  $S(x, \alpha)$  and any given  $\epsilon > 0$  there are positive numbers  $\delta_\kappa(a, b) \leq \epsilon$ ,  $\kappa = 1, 2$ , for which points  $a_1$  and  $b_2$  with

$$a_1a + ab = a_1a \quad \text{and} \quad a_1a = \delta_1,$$

$$ab + bb_2 = ab_2 \quad \text{and} \quad bb_2 = \delta_2$$

exist and are unique.

(1.16) shows that a zero-dimensional space which satisfies Axioms A, B, C, D consists of exactly one point. Therefore it will be assumed throughout that the space has at least dimension 1.

Although D implies only uniqueness of the prolongation in the small, the four axioms together contain uniqueness in the large.

(4.1) THEOREM. If  $(abc_1)$ ,  $(abc_2)$ , and  $bc_1 = bc_2$  then  $c_1 = c_2$ ; if  $(a_1bc)$ ,  $(a_2bc)$ , and  $a_1b = a_2b$  then  $a_1 = a_2$ .



A *proof* is obtained by repeated application of (1.13): the points  $x$  for which the two relations

$$bx + xc_1 = bc_1 \quad \text{and} \quad bx + xc_2 = bc_2$$

hold simultaneously form a compact set, which contains therefore a point  $p$  for which  $bx$  reaches its maximum  $bp$  (possibly  $b=p$ ). If the assertion were not true then  $p \neq c_1, c_2$  and  $(apc_\kappa)$ ,  $\kappa=1, 2$ . For  $b=p$  the relation  $(apc_\kappa)$  is the hypothesis, for  $b \neq p$  it follows from  $(abc_\kappa)$ ,  $(bpc_\kappa)$ , and (1.13).

Let  $S(p, \alpha)$  satisfy Axiom D. Since the neighborhoods  $S'(a, \rho)$  and  $S''(a, \rho)$  are equivalent to the  $S(a, \rho)$ , there are numbers  $\alpha' > 0$  and  $\alpha'' > 0$  such that  $S'(p, \alpha') \subset S(p, \alpha)$  and  $S''(p, \alpha'') \subset S(p, \alpha)$ . Put  $\epsilon = (1/2) \min(\alpha', pc_1 = pc_2)$ . Since a segment  $\mathfrak{s}(p, c_\kappa)$  exists (compare (1.16)),  $\kappa=1, 2$ , there is for every  $0 < \delta \leq \epsilon$  a point  $q_\kappa$  with  $(pq_\kappa c_\kappa)$  and  $pq_\kappa = \delta$ . The points  $q_1$  and  $q_2$  are different. This follows for  $p=b$  from the definition of  $p$ . For  $p \neq b$  the relations  $(pq_\kappa c_\kappa)$ ,  $(bpc_\kappa)$ , and (1.13) yield  $(bpq_\kappa)$ , hence  $bq_\kappa > bp$  and  $(bq_\kappa c_\kappa)$  so that  $q_1 = q_2$  is excluded by the definition of  $p$ .

Moreover  $(apc_\kappa)$ ,  $(pq_\kappa c_\kappa)$ , and (1.13) imply  $(apq_\kappa)$ . Finally let  $(arp)$  and  $rp < \alpha''$ . Then  $(rpq_\kappa)$  by  $(apq_\kappa)$  and (1.13). Now  $r, p \in S(p, \alpha)$  and  $q_1 \neq q_2$  contradict the fact that  $S(p, \alpha)$  satisfies Axiom D.

This proves the first part of (4.1). The second part is established similarly.

Among the various consequences of (4.1) we notice first:

(4.2) *If  $(abc)$  then the segments  $\mathfrak{s}(a, b)$  and  $\mathfrak{s}(b, c)$  are unique.  $\mathfrak{s}(a, b) \cup \mathfrak{s}(b, c)$  is the only segment from  $a$  to  $c$  that contains  $b$ .*

For two different segments from  $b$  to  $c$  would contain points  $c_1$  and  $c_2$  respectively with  $c_1 \neq c_2$  and  $bc_1 = bc_2$ . But (1.13) would yield  $(abc_\kappa)$  which contradicts (4.1). The uniqueness of  $\mathfrak{s}(a, b)$  is obtained in the same way from the second part of (4.1). The last statement follows from the fact that any segment from  $a$  to  $c$  through  $b$  contains subsegments  $\mathfrak{s}(a, b)$  and  $\mathfrak{s}(b, c)$ .

The center of a segment  $\mathfrak{s}' = \mathfrak{s}(a, b)$  is the point  $c$  for which  $ac = cb = ab/2$ . If two segments  $\mathfrak{s}_1$  and  $\mathfrak{s}_2$  have the same initial point and the same center  $c$  then  $\mathfrak{s}_1 = \mathfrak{s}_2$ . For if  $b_\kappa$  is the end point of  $\mathfrak{s}_\kappa$  then  $(acb_\kappa)$  and  $cb_1 = cb_2 = ac$ , hence  $b_1 = b_2$  and by (4.2) there is only one segment  $\mathfrak{s}(a, b)$  through  $c$ . (However different segments may have the same initial and end points.)

This fact yields a very simple metrization for the segments of a space which satisfies Axioms A to D. For any two segments  $\mathfrak{s}_1, \mathfrak{s}_2$  put

$$(4.3) \quad \sigma(\mathfrak{s}_1, \mathfrak{s}_2) = 1/2[\sigma(a_1, a_2) + \sigma(c_1, c_2)]$$

where  $a_\kappa$  is the initial point of  $\mathfrak{s}_\kappa$  and  $c_\kappa$  its center.

Obviously this function satisfies the axioms for a symmetric metric space. Moreover:

(4.4) *The segments of a space  $R$  which satisfies Axioms A to D, including the points of  $R$  as improper segments, form with the metric (4.3) a finitely compact space  $R_s$ .*

For if  $\sigma(\mathfrak{s}_\nu, \mathfrak{s}_\mu) < \alpha$ , then  $\sigma(a_\nu, a_\mu) < 2\alpha$  and  $\sigma(c_\nu, c_\mu) < 2\alpha$ . Hence a subse-

quence  $\{\lambda\}$  of  $\{\nu\}$  exists such that  $a_\lambda$  and  $c_\lambda$  converge to points  $a$  and  $c$  respectively. The end point  $b_\lambda$  of  $\mathfrak{s}_\lambda$  converges to a point  $b$ . For  $\{b_\lambda\}$  is bounded and every accumulation point  $b$  satisfies the relations  $(acb)$  and  $ac=cb$ , so that  $b$  is unique by (4.1). Using (1.20) it is seen that  $\mathfrak{s}_\lambda$  tends to the (by (4.2) unique) segment from  $a$  to  $b$  which contains  $c$ .

A space which satisfies Axioms A, B, C, D and with a topology for its segments equivalent to (4.3) will be called an *E-space*.

In many applications it is important to dispose of neighborhoods such that every segment with end points in this neighborhood can be extended in either direction a given number of times by its own length. To prove the existence of such neighborhoods consider for a given real  $\lambda \geq 2$  and a given point  $x$  the number  $\eta_\lambda(x)$  defined as the least upper bound of 0 and the numbers  $\beta > 0$  (termed admissible for  $\lambda$  and  $x$ ) such that every segment  $\mathfrak{s}(a, b)$  with end points in  $S(x, \beta)$  is a co-central subsegment of a segment  $\mathfrak{s}(a_1, b_1)$  of length  $\lambda\beta$ . (Two segments are co-central if they have the same center. The restriction  $\lambda \geq 2$  is necessary because a segment  $\mathfrak{s}(a, b)$  with end points in  $S(x, \beta)$  may have any length less than  $2\beta$ .)

It will first be shown that

$$(4.5) \quad \eta_\lambda(x) > 0 \quad \text{for every } x \in R \quad \text{and} \quad \lambda \geq 2.$$

Let  $S(x, 2\alpha)$  satisfy Axiom D. By (4.2) the segment  $\mathfrak{s}(y, z)$  is unique for any two points in  $S(x, 2\alpha)$ . Let  $y, z \in S(x, \alpha)$ . The points  $z^*$  with  $yz + zz^* = yz^*$  and  $\sigma(y, z^*) \leq \alpha$  form a compact set  $C$  which contains a point  $z'$  at which  $zz^*$  reaches its maximum  $zz'$ . Then  $\sigma(x, z') = \alpha$ . For if  $\sigma(x, z') < \alpha$ , Axiom D would imply the existence of a point  $z''$  with  $(yz'z'')$ . Now (1.13) yields  $(yzz'')$  and  $(zz'z'')$  hence  $zz'' > zz'$ , so that  $\sigma(x, z'') > \alpha$  by the definition of  $z'$ . But a segment  $\mathfrak{s}(z', z'')$  would contain a point  $z'''$  with  $0 < \sigma(z', z''') < \alpha - \sigma(x, z')$ . Then  $(yzz''')$ ,  $(zz'z''')$ , and  $\sigma(x, z''') \leq \sigma(x, z') + \sigma(z', z''') \leq \alpha$ , which contradicts the definition of  $z'$ .

Next consider the set  $C_1$  formed by the points  $y^*$  which satisfy the relations

$$y^*y + yz' = y^*z' \quad \text{and} \quad \sigma(x, y^*) \leq \alpha.$$

Since  $y, z' \in S(x, 2\alpha)$  the same argument as before shows that  $C_1$  contains a point  $y'$  with  $\sigma(x, y') = \alpha$ . The result obtained so far is this: If  $S(x, 2\alpha)$  satisfies D then for any two points  $y, z$  in  $S(x, \alpha)$  the unique segment  $\mathfrak{s}(y, z)$  is subsegment of a segment  $\mathfrak{s}(y', z')$  with  $\sigma(x, y') = \sigma(x, z') = \alpha$ .

There is a number  $\bar{\alpha} > 0$  such that

$$(4.6) \quad S'(x, \bar{\alpha}) \subset S(x, \alpha) \quad \text{and} \quad S''(x, \bar{\alpha}) \subset S(x, \alpha).$$

Put  $\beta = 2\bar{\alpha}/(\lambda + 2)$  and let  $a, b$  be points of  $S(x, \beta) = S'(x, \beta) \cap S''(x, \beta)$ . Consider a segment  $\mathfrak{s}(a', b')$  which contains  $\mathfrak{s}(a, b)$  as subsegment and for which  $\sigma(x, a') = \sigma(x, b') = \alpha$ . Then  $a'x \geq \bar{\alpha}$  and  $xb' \geq \bar{\alpha}$  because of (4.6). Hence, if  $c$  denotes the center of  $\mathfrak{s}(a, b)$ ,

$$a'c > a'a \geq a'x - ax \geq \bar{\alpha} - \beta = (\lambda + 2)\beta/2 - \beta = \lambda\beta/2,$$

$$cb' > bb' \geq xb' - xb \geq \bar{\alpha} - \beta = \lambda\beta/2.$$

Therefore  $\mathfrak{s}(a', b')$  contains a subsegment  $\mathfrak{s}(a_1, b_1)$  which contains  $\mathfrak{s}(a, b)$  as co-central subsegment and for which  $a_1c = cb_1 = \lambda\beta/2$ . This proves that  $\beta$  is admissible for  $\lambda$  and  $x$ , so that  $\eta_\lambda(x) > 0$ .

The other properties of  $\eta_\lambda(x)$  which will be used here are much easier to establish:

$$(4.7) \quad \eta_\mu(x) \leq \eta_\lambda(x) \leq \mu\eta_\mu(x)/\lambda \quad \text{for } \mu \geq \lambda.$$

Any  $\beta$  admissible for  $\mu$  and  $x$  is also admissible for  $\lambda$  and  $x$ , which proves the first inequality. To prove the second let  $\beta > 0$  be admissible for  $\lambda$  and  $x$  and put  $\bar{\beta} = \lambda\beta/\mu$ . If  $a, b \in S(x, \bar{\beta})$  then  $a, b \in S(x, \beta)$ . Hence a segment  $\mathfrak{s}(a_1, b_1)$  of length  $\lambda\beta = \mu\bar{\beta}$  exists which contains  $\mathfrak{s}(a, b)$  as co-central subsegment. This shows that  $\bar{\beta}$  is admissible for  $\mu$  and  $x$ , hence  $\eta_\mu(x) \geq \lambda\eta_\lambda(x)/\mu$ .

(4.8) If  $\eta_{\lambda_0}(x_0) = \infty$  for one point and  $x_0$  and one  $\lambda_0 \geq 2$ , then  $\eta_\lambda(x) = \infty$  for all points  $x$  of  $R$  and all real  $\lambda \geq 2$ .

By (4.7),  $\eta_\lambda(x_0) = \infty$  for every real  $\lambda \geq 2$  if  $\eta_{\lambda_0}(x_0) = \infty$ . If  $\eta_{\lambda_0}(x_0) = \infty$ , then  $S(x_0, \eta_{\lambda_0}(x_0)) = R$ , hence  $\eta_{\lambda_0}(x) = \infty$  for every  $x$  in  $R$ , which proves (4.8).

If  $\eta_\lambda(x) < \infty$ , then  $\eta_\lambda(x)$  is itself admissible for  $\lambda$  and  $x$ , and for any two points  $x, y$  in  $R$

$$(4.9) \quad |\eta_\lambda(x) - \eta_\lambda(y)| \leq \sigma(x, y).$$

If  $a, b \in S(x, \eta_\lambda(x))$ , then for large  $\nu$  also  $a, b \in S(x, \beta_\nu)$  where  $\beta_\nu = \eta_\lambda(x) - 1/\nu$ . Since  $\beta_\nu$  is admissible for  $\lambda$  and  $x$ , a segment  $\mathfrak{s}(a_\nu, b_\nu)$  of length  $\lambda\beta_\nu$  exists which contains  $\mathfrak{s}(a, b)$  as co-central subsegment. It is easily seen that  $\mathfrak{s}(a_\nu, b_\nu)$  tends to a segment  $\mathfrak{s}(a', b')$  of length  $\lambda\eta_\lambda(x)$  which contains  $\mathfrak{s}(a, b)$  as co-central subsegment, so that  $\eta_\lambda(x)$  is admissible.

(4.9) is trivial when  $\eta_\lambda(x) \leq \sigma(x, y)$  and  $\eta_\lambda(y) \leq \sigma(x, y)$ . Let therefore  $\eta_\lambda(x) > \sigma(x, y)$ . Then  $S(x, \eta_\lambda(x)) \supset S(y, \eta_\lambda(x) - \sigma(x, y))$  hence  $\eta_\lambda(y) \geq \eta_\lambda(x) - \sigma(x, y)$ . If  $\eta_\lambda(y) \leq \sigma(x, y)$  then  $\eta_\lambda(x) \geq \eta_\lambda(y) - \sigma(x, y)$  and (4.9) holds. If finally  $\eta_\lambda(y) > \sigma(x, y)$ , then (4.9) follows from  $S(x, \eta_\lambda(y)) \supset S(x, \eta_\lambda(y) - \sigma(x, y))$ .

(4.7), (4.8), (4.9) show that  $\eta_\lambda(x)$  is continuous in  $x$  and  $\lambda$  (jointly). However, only the continuity in  $x$  for fixed  $\lambda$  will be used. It is now possible in a variety of ways to define a set of segments which play the same rôle as the oriented line-elements or directions in the differentiable case. In order to avoid a separate treatment of the case  $\eta_\lambda(x) \equiv \infty$ , it is convenient to put

$$(4.10) \quad \eta(x) = \min(\eta_6(x), 1)$$

and to define as *direction* with origin  $x$  any segment of length  $\eta(x)$  with  $x$  as initial point.

Because of (4.5) and (4.9)<sup>(9)</sup>,  $\eta(x)$  is a positive and continuous function

<sup>(9)</sup>  $\eta(x)$  satisfies the relation  $|\eta(x) - \eta(y)| \leq \sigma(x, y)$  because the following is obviously true: If  $\alpha_1(x)$  and  $\alpha_2(x)$  are positive functions in  $R$  and  $|\alpha_k(x) - \alpha_k(y)| \leq \sigma(x, y)$  then  $|\min(\alpha_1(x), \alpha_2(x)) - \min(\alpha_1(y), \alpha_2(y))| \leq \sigma(x, y)$ .

of  $x$ ; hence the directions of  $R$  form with the metric (4.3) a finitely compact subspace  $R_d$  of  $R_*$  (compare (4.4)). Observe that because of  $\eta(x) \leq 1$  a sequence of directions is bounded if and only if their origins form a bounded sequence in  $R$ .

The following consequence of the preceding considerations will be used in the next section:

(4.11) If  $\mathfrak{s} = \mathfrak{s}(a, b)$  is any segment, then segments  $\mathfrak{s}(a_1, a)$  and  $\mathfrak{s}(b, b_1)$  exist such that  $a_1a = \eta(a)$ ,  $bb_1 = \eta(b)$  and  $(a_1ax)$  for points  $x$  of  $\mathfrak{s}$  close to  $a$ ,  $(ybb_1)$  for points  $y$  of  $\mathfrak{s}$  close to  $b$ .

An  $E$ -space is topologically *homogeneous* in the following sense:

(4.12) THEOREM. *If two points  $p$  and  $q$  of an  $E$ -space  $R$  and an  $\epsilon > 0$  are given, then  $S(q, \epsilon)$  contains an open set  $Q \supset q$  which admits a topological mapping on  $S(p, \eta(p))$  under which  $q$  goes into  $p$ .*

First let  $q = p$ . Choose  $\epsilon' > 0$  such that  $S'(p, \epsilon') \subset S(p, \min(\epsilon, \eta(p)))$ . Then  $\epsilon' \leq \epsilon$ . Map  $p$  on itself and any point  $x \neq p$  of  $S(p, \eta(p))$  on the point  $x'$  of the (because of  $px < \eta(p)$  unique) segment  $\mathfrak{s}(p, x)$  for which  $px' = px \cdot \epsilon' / \eta(p)$ . The mapping  $x \rightarrow x'$  is topological and maps  $S(p, \eta(p))$  on an open subset  $Q$  of  $S'(p, \epsilon')$ .

Next consider two different points  $p$  and  $q$  with  $\sigma(p, q) < \eta(p)/2$ . Determine  $p'$  and  $q'$  with  $(p'pq)$ ,  $(p'qq')$ ,  $0 < p'p = qq' = \delta$  and  $\sigma(p', p) < pq/8$ ,  $\sigma(q, q') < pq/8$ . These points exist because by (4.9),  $\eta(p') \geq \eta(p) - \sigma(p, p') \geq 15\eta(p)/16 > p'q + qq'$ . For the same reason  $S''(q', \delta/2) \subset S(p', \eta(p'))$  if  $\delta$  is sufficient small. Then the segment  $\mathfrak{s}(p', x)$  is unique for  $x \in S''(q', \delta/2)$  and  $p'x \geq p'q' - xq' = p'q + qq' - xq' > p'q + \delta/2 = pq + 3\delta/2$ . Let  $c_x(\tau)$ ,  $0 \leq \tau \leq p'x$ , represent  $\mathfrak{s}(p', x)$ , and consider the two open subsegments  $\mathfrak{s}_x^1$  and  $\mathfrak{s}_x^2$  of  $\mathfrak{s}(p', x)$  defined respectively by  $\delta/2 < \tau < 3\delta/2$  and  $p'q - \delta/2 = pq + \delta/2 < \tau < pq + 3\delta/2$ . Then the sets  $P' = \bigcup_x \mathfrak{s}_x^1$  and  $Q' = \bigcup_x \mathfrak{s}_x^2$ , where  $x$  traverses  $S''(q', \delta/2)$ , are open and contain  $p$  and  $q$  respectively.

For  $q' \rightarrow q$  or  $\delta \rightarrow 0$  the set  $Q'$  shrinks to  $Q$ . Hence  $\delta$  may be chosen such that it satisfies the previous restrictions and  $Q' \subset S(q, \epsilon)$ .

Define  $\Omega$  as the topological mapping of  $P'$  on  $Q'$  which associates for every  $x$  in  $S''(q', \delta/2)$  and every  $\alpha$  between 0 and  $\delta$  the point  $c_x(\delta/2 + \alpha)$  with the point  $c_x(pq + \delta/2 + \alpha)$  (that is,  $\mathfrak{s}_x^1$  is mapped isometrically on  $\mathfrak{s}_x^2$ ).

Since  $P' \supset S(p, \eta)$  for sufficiently small  $\eta > 0$ , the first part of this proof yields a topological mapping  $\Omega_1$  of  $S(p, \eta(p))$  on a subset  $P$  of  $S(p, \eta)$ , under which  $p$  remains fixed. Under  $\Omega$  the point  $p$  goes into  $q$  and  $P$  into a subset  $Q$  of  $Q'$ . Hence  $\Omega_1\Omega$  maps  $S(p, \eta(p))$  topologically on  $Q$  and  $p$  goes into  $q$ .

If (4.12) holds for the points  $p$  and  $q$  and for the points  $q$  and  $u$  then it also holds for the points  $p$  and  $u$ . For then topological mappings  $\Omega'$  of  $S(p, \eta(p))$  on an open subset  $Q$  of  $S(q, \eta(q))$  and  $\Omega''$  of  $S(q, \eta(q))$  on an open subset  $U''$  of  $S(u, \epsilon)$  exist, under which  $p$  and  $q$  go into  $q$  and  $u$  respectively.  $\Omega'\Omega''$  maps  $p$  on  $u$  and the set  $S(p, \eta(p))$  topologically on an open

subset  $U$  of  $U''$ . Since  $\inf \eta(x) > 0$  for  $x \in S(p, 2\sigma(p, q))$  it follows easily that (4.12) holds for any two points  $p$  and  $q$ . Notice the corollary:

(4.13) *An  $E$ -space has at every point the same dimension.*

5. **Extremals.** The proper segments of an  $E$ -space will now be divided into classes: the extremals. Two proper segments  $\mathfrak{s}'$  and  $\mathfrak{s}''$  are said to belong to the same *extremal*  $\mathfrak{E}$  if segments  $\mathfrak{s}_1 = \mathfrak{s}'$ ,  $\mathfrak{s}_2, \dots, \mathfrak{s}_{\mu-1}$ ,  $\mathfrak{s}_\mu = \mathfrak{s}''$  exist such that  $\mathfrak{s}_\nu$  and  $\mathfrak{s}_{\nu+1}$ ,  $\nu = 1, 2, \dots, \mu-1$ , have a common proper subsegment. The class axioms are obviously satisfied. The extremal which contains the segment  $\mathfrak{s}$  will be denoted by  $\mathfrak{E}(\mathfrak{s})$ .

The continuous curve  $x(\tau)$ ,  $-\infty < \tau < \infty$ , is said to *represent*  $\mathfrak{E}(\mathfrak{s})$  if it has the following two properties:

PROPERTY I. *For every real  $\tau_0$  there is a  $\delta(\tau_0) > 0$  such that  $x(\tau)$  represents a segment for  $\tau_0 - \delta \leq \tau \leq \tau_0 + \delta$ .*

PROPERTY II. *A segment  $\mathfrak{s}'$  belongs to  $\mathfrak{E}(\mathfrak{s})$  if and only if numbers  $\tau_0$  and  $\tau_1$  exist such that  $x(\tau)$  represents  $\mathfrak{s}'$  for  $\tau_0 \leq \tau \leq \tau_1$ .*

An immediate consequence of Property I and the definition of arc length is:

(5.1) *If  $x(\tau)$  represents an extremal, then  $\tau_2 - \tau_1$  is the length of the subarc  $\tau_1 \leq \tau \leq \tau_2$  of  $x(\tau)$ . Hence  $x(\tau_1)x(\tau_2) \leq \tau_2 - \tau_1$ , and the equality sign holds if and only if the arc represents a segment.*

Property II implies that a continuous curve represents at most one extremal. To prove the existence of representations, several auxiliary results are needed.

(5.2) Let  $x(\tau)$  have Property I and let  $x'(\tau)$ ,  $\alpha' \leq \tau \leq \beta'$ , represent a segment. If  $\alpha \leq \alpha' < \beta' \leq \beta$  and  $x(\tau) = x'(\tau)$  for  $\alpha' \leq \tau \leq \beta'$ , then  $x(\tau) = x'(\tau)$  for  $\alpha \leq \tau \leq \beta$ .

**Proof.** Consider the set  $V$  of those values  $\tau_0$  in  $(\alpha', \beta)$  for which  $x(\tau) = x'(\tau)$  when  $\alpha \leq \tau \leq \tau_0$ . It is to be shown that  $V = (\alpha', \beta)$ . Obviously  $V$  is closed in  $(\alpha', \beta)$  and contains the number  $\tau_1 \geq \alpha$  if it contains a  $\tau_2 > \tau_1$ . Hence it suffices to see that  $V$  is open in  $(\alpha', \beta)$ . Let  $V$  contain  $\tau_0$  with  $\alpha' < \tau_0 < \beta$ . Then  $x(\tau)$  and  $x'(\tau)$  represent segments for  $\tau_0 - \delta \leq \tau \leq \tau_0 + \delta$  for a suitable positive  $\delta \leq \min(\tau_0 - \alpha', \beta - \tau_0)$ . These segments have the same initial point  $x(\tau_0 - \delta) = x'(\tau_0 - \delta)$  and the same center  $x(\tau_0) = x'(\tau_0)$ , hence their terminal points  $x_0(\tau_0 + \delta)$  and  $x'_0(\tau_0 + \delta)$  coincide (compare the preceding section), which shows that  $V$  is open in  $(\alpha', \beta)$ . The values  $\tau$  in  $(\alpha, \beta')$  are treated in the same way.

(5.3) If the continuous curve  $x(\tau)$  has Property I and if the subarc  $\alpha \leq \tau \leq \beta$  of  $x(\tau)$  represents a proper segment  $\mathfrak{s}'$ , then any segment  $\mathfrak{s}''$  which has a common proper subsegment  $\bar{\mathfrak{s}}$  with  $\mathfrak{s}'$  is represented by a suitable subarc of  $x(\tau)$ .

**Proof.** Let  $x(\tau)$  represent  $\bar{\mathfrak{s}}$  for  $\bar{\alpha} \leq \tau \leq \bar{\beta}$ . If  $y''(\tau)$  represents  $\mathfrak{s}''$  for  $\alpha'' \leq \tau \leq \beta''$ , then a subarc  $\alpha^* \leq \tau \leq \beta^*$  of  $y(\tau)$  also represents  $\bar{\mathfrak{s}}$ . By (1.12) there is a real  $\lambda$  such that  $y''(\tau) = x(\tau + \lambda)$  for  $\alpha^* \leq \tau \leq \beta^*$ . The function  $y(\tau) = x(\tau + \lambda)$ ,  $-\infty < \tau < \infty$ , still has Property I. Now it follows from (5.2)

that  $y(\tau) = y''(\tau)$  for  $\alpha'' \leq \tau \leq \beta''$ , hence  $x(\tau)$  represents  $\mathfrak{s}''$  for  $\alpha'' - \lambda \leq \tau \leq \beta'' - \lambda$ .

Repeated application of (5.3) yields that a curve  $x(\tau)$  with Property I represents with  $\mathfrak{s}$  every segment of the extremal  $\mathfrak{E}(\mathfrak{s})$ . Conversely, every segment  $\mathfrak{s}'$  represented by a subarc of  $x(\tau)$  belongs to  $\mathfrak{E}(\mathfrak{s})$ . For let  $\mathfrak{s}$  and  $\mathfrak{s}'$  be represented by the subarcs  $\tau_0 - \delta_0 \leq \tau \leq \tau_0 + \delta_0$  and  $\tau' - \delta' \leq \tau \leq \tau' + \delta'$  of  $x(\tau)$  respectively, and assume  $\tau_0 < \tau'$ .

By Property I we may associate with every  $\tau^*$  between  $\tau_0$  and  $\tau'$  an open interval  $\tau^* - \delta^* < \tau < \tau^* + \delta^*$  such that the arc  $x(\tau)$  represents a proper segment when  $|\tau^* - \tau| \leq \delta^*$ . With  $\tau_0$  and  $\tau'$ , we associate the intervals  $|\tau_0 - \tau| < \delta_0$  and  $|\tau' - \tau| < \delta'$  respectively. The Heine-Borel Theorem yields a finite number of values  $\tau_1 < \tau_2 < \dots < \tau_{\mu-1}$  such that the corresponding open intervals cover the closed interval  $(\tau_0, \tau')$ , and that none of the intervals is completely contained in another. If  $\mathfrak{s}_\nu$  designates the segment represented by  $x(\tau)$  for  $|\tau_\nu - \tau| \leq \delta_\nu$  ( $\delta_\nu$  is the  $\delta^*$  belonging to  $\tau^* = \tau_\nu$ ), then  $\mathfrak{s}_0 = \mathfrak{s}$ ,  $\mathfrak{s}_1, \mathfrak{s}_2, \dots, \mathfrak{s}_{\mu-1}, \mathfrak{s}_\mu = \mathfrak{s}'$  have the property that  $\mathfrak{s}_\nu$  and  $\mathfrak{s}_{\nu+1}$  contain a proper subsegment. This shows that  $x(\tau)$  has Property II. Hence we have

(5.4) THEOREM. *A continuous curve  $x(\tau)$ ,  $-\infty < \tau < \infty$ , represents an extremal if (and only if) it has Property I.*

The definition of  $\eta(x)$  yields

(5.5) *If  $x(\tau)$  represents an extremal, then  $x(\tau)$  represents a segment for  $|\tau - \tau_0| \leq \eta(x(\tau_0))$ .*

In order to establish the existence of a representation for a given extremal  $\mathfrak{E}(\mathfrak{s}_0)$ , it suffices, therefore, to show that a continuous curve  $x(\tau)$  with Property I exists, a subarc of which represents  $\mathfrak{s}_0$ . Let  $\mathfrak{s}_0 = \mathfrak{s}(a_0, a_1)$  and, according to (4.11),  $\mathfrak{s}_1 = \mathfrak{s}(a_1, a_2)$  the segment with initial point  $a_1$  and of length  $a_1 a_2 = \eta(a_1)$ , such that  $(y a_1 a_2)$  for points  $y$  on  $\mathfrak{s}_0$  sufficiently close to  $a_1$ . By induction define  $\mathfrak{s}_\nu = \mathfrak{s}(a_\nu, a_{\nu+1})$  as the segment of length  $a_\nu a_{\nu+1} = \eta(a_\nu)$  such that  $(y a_\nu a_{\nu+1})$  for points  $y$  of  $\mathfrak{s}_{\nu-1}$  close to  $a_\nu$ . Then  $\sum_{\nu=0}^{\infty} a_\nu a_{\nu+1} = \infty$ ; otherwise  $a_0 a_\mu \leq \sum_{\nu=0}^{\mu-1} a_\nu a_{\nu+1}$  would be bounded and  $a_\mu a_{\mu+1} = \eta(a_\mu) \rightarrow 0$ , which contradicts the fact that  $\eta(x)$  is positive and continuous in  $R$ .

Similarly define  $\mathfrak{s}_{-1}$  as segment  $\mathfrak{s}(a_{-1}, a_0)$  of length  $a_{-1} a_0 = \eta(a_0)$  and such that  $(a_{-1} a_0 x)$  for points of  $\mathfrak{s}_0$  close to  $a_0$ , and, by induction, define  $\mathfrak{s}_{-\nu}$  as the segment  $\mathfrak{s}(a_{-\nu}, a_{-\nu+1})$  of length  $a_{-\nu} a_{-\nu+1} = \eta(a_{-\nu+1})$  such that  $(a_{-\nu} a_{-\nu+1} x)$  for points  $x$  of  $\mathfrak{s}_{-\nu+1}$  close to  $a_{-\nu+1}$ . By the same argument as before it is seen that  $\sum_{\nu=0}^{\infty} a_{-\nu} a_{-\nu+1} = \infty$ .

Now define the function  $x(\tau)$ ,  $-\infty < \tau < \infty$ , as follows:  $x(\tau)$  represents (1) the segment  $\mathfrak{s}_0$  for  $0 \leq \tau \leq a_0 a_1$ , (2) the segment  $\mathfrak{s}_\mu$ ,  $\mu = 1, 2, \dots$ , for  $\sum_{\nu=0}^{\mu-1} a_\nu a_{\nu+1} \leq \tau \leq \sum_{\nu=0}^{\mu} a_\nu a_{\nu+1}$ , (3) the segment  $\mathfrak{s}_{-1}$  for  $-a_{-1} a_0 \leq \tau \leq 0$ , and (4) the segment  $\mathfrak{s}_{-\mu}$ ,  $\mu = -2, -3, \dots$ , for  $-\sum_{\nu=-\mu}^{-1} a_\nu a_{\nu+1} \leq \tau \leq -\sum_{\nu=-\mu+1}^{-1} a_\nu a_{\nu+1}$ .

$x(\tau)$  has obviously Property I for every  $\tau$  different from one of the values  $0, a_0 a_1, \sum_{\nu=0}^{\mu} a_\nu a_{\nu+1}$ , and  $-\sum_{\nu=-\mu}^{-1} a_\nu a_{\nu+1}$ ,  $\mu = 1, 2, \dots$ . To see that Property I

holds, for instance at  $\tau = a_0 a_1$ , observe that it follows from (4.2) and  $(y a_1 a_2)$  for  $y \in \mathfrak{s}_0$  close to  $a_1$  that the segments  $\mathfrak{s}(y, a_1)$  and  $\mathfrak{s}(a_1, a_2)$  are unique and coincide therefore with the subsegments  $\mathfrak{s}(y, a_1)$  of  $\mathfrak{s}_0$  and with  $\mathfrak{s}_1$  respectively; moreover  $\mathfrak{s}(y, a_1) \cup \mathfrak{s}(a_1, a_2)$  is a segment (see (4.2)), hence  $x(\tau)$  represents a segment for  $a_0 a_1 - a_1 y \leq \tau \leq a_0 a_1 + a_1 a_2$ .

The existence of a representation yields together with (5.2), (5.3), (5.4) the following theorems:

(5.6) *Let  $x'(\tau)$ ,  $\alpha \leq \tau \leq \beta$ , represent the segment  $\mathfrak{s}$ ; there is one and only one representation  $x(\tau)$  of an extremal (namely  $\mathfrak{E}(\mathfrak{s})$ ) such that*

$$x(\tau) = x'(\tau) \quad \text{for } \alpha \leq \tau \leq \beta.$$

(5.7) **THEOREM.** *Every extremal possesses a representation.  $x(\tau)$  and  $y(\tau)$  represent the same extremal if and only if a number  $\beta$  exists such that*

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

An extremal  $\mathfrak{E}(\mathfrak{s})$  is called a *directed straight line*, if one, and hence by (5.7) every, representation  $x(\tau)$  of  $\mathfrak{E}(\mathfrak{s})$  satisfies the relation  $x(\tau_1)x(\tau_2) = \tau_2 - \tau_1$  for any real  $\tau_1 < \tau_2$ .

The example of the cylinder shows that such extremals may occur without all extremals being directed straight lines. It is easily seen that

(5.8) *All extremals of an E-space are directed straight lines if and only if  $\eta_\lambda(x) \equiv \infty$  (or, by (4.8),  $\eta_\lambda(x) = \infty$ , for one pair  $x, \lambda$ ).*

**6. Convergence of extremals.** The next problem which presents itself after extremals have been defined regards the *topology of the set of all extremals* in a given E-space. On the one hand there exists an intuitive concept of convergence for the extremals of the euclidean, spherical, and other simple spaces, with which any general definition has to agree if it claims to be natural. On the other hand this concept leads, already, to difficulties when applied to certain very simple analytic Riemann spaces. The problem does not seem to admit a general solution. The distance  $\bar{\sigma}(\mathfrak{E}_1, \mathfrak{E}_2)$ , to be defined presently, will agree with the intuition but, in complicated cases, may vanish for two different extremals  $\mathfrak{E}_1, \mathfrak{E}_2$ .

Call  $E[\mathfrak{s}]$  the point set which carries the extremal  $\mathfrak{E}(\mathfrak{s})$ . Every point  $x$  of  $E[\mathfrak{s}]$  is origin of at least one direction (see the end of §4) which belongs to  $\mathfrak{E}(\mathfrak{s})$ . Let  $\mathfrak{E}\{\mathfrak{s}\}$  denote the set of all directions in  $\mathfrak{E}(\mathfrak{s})$ . In this way there corresponds to every extremal a non-empty subset  $\mathfrak{E}\{\mathfrak{s}\}$  of elements in  $R_d$ , which, by the way, is easily seen to be a continuous curve without multiple points. Two such curves are either identical or disjoint. Since the sets  $\mathfrak{E}\{\mathfrak{s}\}$  in  $R_d$  are, in general, simpler than the sets  $E[\mathfrak{s}]$  in  $R$  and  $\mathfrak{E}(\mathfrak{s})$  in  $R_*$ , they will be used for the distance of extremals.

Every point  $p$  of  $R_d$  (that is, every direction in  $R$ ) determines a distance  $\rho_p = \bar{\rho}$  as defined in §2. In particular we put for any two extremals  $\mathfrak{E}(\mathfrak{s}_1)$  and  $\mathfrak{E}(\mathfrak{s}_2)$

$$(6.1) \quad \begin{aligned} \bar{\sigma}(\mathfrak{E}(\mathfrak{s}_1), \mathfrak{E}(\mathfrak{s}_2)) &= \bar{\rho}(\mathfrak{E}\{\mathfrak{s}_1\}, \mathfrak{E}\{\mathfrak{s}_2\}) \\ &= \sup_{d \in R_d} |\sigma(d, \mathfrak{E}\{\mathfrak{s}_1\}) - \sigma(d, \mathfrak{E}\{\mathfrak{s}_2\})| \exp(-\sigma(p, d)). \end{aligned}$$

It follows from (2.2) that

(6.2)  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_1), \mathfrak{E}(\mathfrak{s}_2)) = 0$  if and only if  $\bar{\mathfrak{E}}\{\mathfrak{s}_1\} = \bar{\mathfrak{E}}\{\mathfrak{s}_2\}$  (that is, if the closures of  $\mathfrak{E}\{\mathfrak{s}_1\}$  and  $\mathfrak{E}\{\mathfrak{s}_2\}$  in  $R_d$  coincide).

An extreme case where this happens for two different extremals is the following: if  $\bar{\mathfrak{E}}\{\mathfrak{s}\} = R_d$  the extremal  $\mathfrak{E}(\mathfrak{s})$  is what is known as transitive with respect to the directions in  $R$ . Clearly the distance of any two such transitive extremals vanishes. But the distance may vanish even for two extremals which do not have this property (see  $R^5$  in §8).

According to §2 any point  $p$  in  $R$  determines for any two sets  $G, H$  in  $R$  the distance

$$\rho_p(G, H) = \sup_{x \in R} |\sigma(x, G) - \sigma(x, H)| \cdot \exp(-\sigma(p, x)),$$

in particular also the distance  $\rho_p(E[\mathfrak{s}_1], E[\mathfrak{s}_2])$  of the point sets which carry  $\mathfrak{E}(\mathfrak{s}_1)$  and  $\mathfrak{E}(\mathfrak{s}_2)$ . This distance may vanish even if  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_1), \mathfrak{E}(\mathfrak{s}_2)) \neq 0$ .

This is obviously the case when  $\mathfrak{E}(\mathfrak{s}_1)$  and  $\mathfrak{E}(\mathfrak{s}_2)$  are the two different directed straight lines carried by the same straight line in euclidean geometry. But in symmetric spaces ( $xy=yx$ ) another distance is more natural (compare §8). A more convincing example is the space  $R^2$  in §3, where 0 is not the center of  $K$ . The point sets which carry the extremals are the euclidean straight lines (Minkowskian geometry, see Minkowski [1, chapter I]). Since, in general,  $xy \neq yx$  it would not be natural to identify the two extremals carried by the same straight line.

Therefore  $\rho_p(E[\mathfrak{s}_v], E[\mathfrak{s}]) \rightarrow 0$  does not imply  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_v), \mathfrak{E}(\mathfrak{s})) \rightarrow 0$ . But the converse holds:

(6.3) When  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_v), \mathfrak{E}(\mathfrak{s})) \rightarrow 0$  then  $\rho_p(E[\mathfrak{s}_v], E[\mathfrak{s}]) \rightarrow 0$  and  $\bar{E}[\mathfrak{s}] = Fl E[\mathfrak{s}_v]$ .

**Proof.** (2.12), (2.13) show that  $Fl \mathfrak{E}\{\mathfrak{s}_v\} = \bar{\mathfrak{E}}\{\mathfrak{s}\}$  and that the last two relations are equivalent. The set  $\bar{E}[\mathfrak{s}]$  is easily seen to consist of those points in  $R$  which are origins of directions in  $\bar{\mathfrak{E}}\{\mathfrak{s}\}$ . Let  $x \in \bar{E}[\mathfrak{s}]$  and  $\mathfrak{s}' \in \bar{\mathfrak{E}}\{\mathfrak{s}\}$  with origin  $x$ . Then  $\mathfrak{s}'_v \in \mathfrak{E}\{\mathfrak{s}_v\}$  exists with  $\sigma(\mathfrak{s}', \mathfrak{s}'_v) \rightarrow 0$ . If  $x'_v$  denotes the origin of  $\mathfrak{s}'_v$ , then  $\sigma(x'_v, x) \leq 2\sigma(\mathfrak{s}', \mathfrak{s}'_v) \rightarrow 0$  (see (4.3)); hence  $x \in Fl \inf E[\mathfrak{s}_v]$  and  $\bar{E}[\mathfrak{s}] \subset Fl \inf E[\mathfrak{s}_v]$ .

Next let  $y \in Fl \sup E[\mathfrak{s}_v]$ , so that for a suitable subsequence  $\{\lambda\}$  of  $\{v\}$  the set  $E[\mathfrak{s}_\lambda]$  contains a point  $y_\lambda$  with  $y_\lambda \rightarrow y$ . The sequence  $\{\mathfrak{s}_\lambda\}$  is bounded because  $\{y_\lambda\}$  is bounded (compare (4.10)); hence it contains a subsequence converging to a direction  $\mathfrak{s}''$  with origin  $y$ , where  $\mathfrak{s}'' \in \bar{\mathfrak{E}}\{\mathfrak{s}\}$  because  $Fl \sup \mathfrak{E}\{\mathfrak{s}_v\} = \bar{\mathfrak{E}}\{\mathfrak{s}\}$ . Therefore  $y \in \bar{E}[\mathfrak{s}]$  and  $Fl \sup E[\mathfrak{s}_v] \subset \bar{E}[\mathfrak{s}]$ .

The discussion of convergence of extremals in terms of their representations is based on the following simple fact.



(6.4) Let  $x_\nu(\tau)$ ,  $\nu=0, 1, 2, \dots$ , represent extremals. If  $\tau_\nu \rightarrow \tau_0$  and  $x_\nu(\tau_\kappa) \rightarrow x_0(\tau_\kappa)$  for fixed  $\kappa=1, 2, 3, \dots$ , then  $x_\nu(\tau_0) \rightarrow x_0(\tau_0)$  and  $x_\nu(\tau_\nu) \rightarrow x_0(\tau_0)$ .

For a proof observe that (compare (1.5))

$$(6.5) \quad \sigma^*(x_0(\tau_0), x_\nu(\tau_0)) \leq \sigma(x_0(\tau_0), x_0(\tau_\kappa)) + \sigma(x_0(\tau_\kappa), x_\nu(\tau_\kappa)) + \sigma^*(x_\nu(\tau_\kappa), x_\nu(\tau_0)).$$

By (5.1) any representation  $y(\tau)$  of an extremal satisfies the inequality

$$\sigma^*(y(\tau_1), y(\tau_2)) \leq |\tau_1 - \tau_2|.$$

Hence, if  $\epsilon > 0$  is given, first  $\kappa$  may be chosen (compare (1.4)) such that  $\sigma(x_0(\tau_0), x_0(\tau_\kappa)) < \epsilon/3$  and  $|\tau_\kappa - \tau_0| < \epsilon/3$ , then by hypothesis  $\nu_0$  so large that  $\sigma(x_0(\tau_\kappa), x_\nu(\tau_\kappa)) < \epsilon/3$  for  $\nu > \nu_0$ . This shows that  $\sigma^*(x_0(\tau_0), x_\nu(\tau_0)) \rightarrow 0$ , hence also  $\sigma(x_0(\tau_0), x_\nu(\tau_0)) \rightarrow 0$ . Moreover

$$\begin{aligned} \sigma^*(x_0(\tau_0), x_\nu(\tau_\nu)) &\leq \sigma(x_0(\tau_0), x_\nu(\tau_0)) + \sigma^*(x_\nu(\tau_0), x_\nu(\tau_\nu)) \\ &\leq \sigma(x_0(\tau_0), x_\nu(\tau_0)) + |\tau_0 - \tau_\nu|, \end{aligned}$$

which proves the second part of (6.4).

It is now easy to prove

(6.6) THEOREM. Let  $x_\nu(\tau)$  represent an extremal  $\mathfrak{E}(\mathfrak{s}_\nu)$ ,  $\nu=1, 2, \dots$ . If  $\{x_\nu(\tau)\}$  converges for every  $\tau$  of an interval  $\alpha \leq \tau \leq \beta$ ,  $\alpha < \beta$ , then  $x_\nu(\tau)$  converges uniformly on every bounded set  $W$  of the real axis to a limit  $x(\tau)$  which represents an extremal  $\mathfrak{E}(\mathfrak{s})$ . Moreover  $\overline{E}[\mathfrak{s}] \subset Fl \inf E[\mathfrak{s}_\nu]$ , and  $\overline{\mathfrak{E}}\{\mathfrak{s}\} \subset Fl \inf \mathfrak{E}\{\mathfrak{s}_\nu\}$ .

**Proof.** Put  $\gamma = (\alpha + \beta)/2$ . According to the hypothesis  $q_\nu = x_\nu(\gamma)$  tends to a point  $q$ . Then  $x_\nu(\tau)$  represents a segment for  $\gamma - \eta(q_\nu) \leq \tau \leq \gamma + \eta(q_\nu)$ , and  $\eta(q_\nu) > \eta(q)/2$  for large  $\nu$  because  $\eta(x)$  depends continuously on  $x$ . Hence  $x'(\tau) = \lim x(\tau)$  represents a segment for

$$(6.7) \quad |\tau - \gamma| \leq 1/2 \min(\eta(p), \beta - \alpha).$$

By (5.6) there is a representation  $x(\tau)$  of an extremal  $\mathfrak{E}(\mathfrak{s})$  with  $x(\tau) = x'(\tau)$  in (6.7).

Consider the set  $V$  of all values  $\bar{\tau} \geq \gamma$  such that  $x_\nu(\tau) \rightarrow x(\tau)$  for  $\gamma \leq \tau \leq \bar{\tau}$ . The set  $V$  contains  $\beta$  and with any  $\bar{\tau}_1$  every  $\bar{\tau}$  with  $\gamma \leq \bar{\tau} \leq \bar{\tau}_1$ . By virtue of (6.4),  $V$  is closed. But it follows from (5.5) and the continuity of  $\eta(x)$  that  $V$  is open in the set  $\bar{\tau} \geq \gamma$ , so that  $V$  comprises all  $\bar{\tau} \geq \gamma$ . Similarly it is seen that  $x_\nu(\tau) \rightarrow x(\tau)$  for all  $\tau \leq \gamma$ .

(6.4) shows in addition that this convergence is continuous everywhere. Hence it is uniform on every compact set, in particular on the closure  $\overline{W}$  of the given set  $W$ . The relation  $E[\mathfrak{s}] \subset Fl \inf E[\mathfrak{s}_\nu]$  follows from  $x_\nu(\tau) \rightarrow x(\tau)$ .

To prove  $\overline{\mathfrak{E}}\{\mathfrak{s}\} \subset Fl \inf \mathfrak{E}\{\mathfrak{s}_\nu\}$  let the arbitrary direction  $\mathfrak{s}' \in \overline{\mathfrak{E}}\{\mathfrak{s}\}$  be represented by  $x(\tau)$  for  $\tau_0 \leq \tau \leq \tau_0 + \eta(x(\tau_0)) = \tau_0 + 2\delta$ , and denote by  $\mathfrak{s}'_\nu$  the direction in  $\mathfrak{E}\{\mathfrak{s}_\nu\}$  represented by  $x_\nu(\tau)$  for  $\tau_0 \leq \tau \leq \tau_0 + \eta(x_\nu(\tau_0)) = \tau_0 + 2\delta_\nu$ . According to the definition (4.3),  $2\sigma(\mathfrak{s}, \mathfrak{s}_\nu) = \sigma(x(\tau_0), x_\nu(\tau_0)) + \sigma(x(\tau_0 + \delta), x_\nu(\tau_0 + \delta_\nu))$ . The

first term on the right side tends to 0 because  $x_\nu(\tau_0) \rightarrow x(\tau_0)$ , the second term because  $\eta(x_\nu(\tau_0)) \rightarrow \eta(x(\tau_0))$ , hence  $\tau_0 + \delta_\nu \rightarrow \tau_0 + \delta$ , moreover  $x_\nu(\tau_0 + \delta_\nu) \rightarrow x(\tau_0 + \delta)$  for every  $\kappa$  (compare (6.4)).

A corollary of these considerations is

(6.8) *The extremals  $\mathfrak{E}(\mathfrak{s}_\nu)$ ,  $\nu = 0, 1, 2, \dots$ , have representations  $x_\nu(\tau)$  with  $x_\nu(\tau) \rightarrow x_0(\tau)$  for every  $\tau$  if and only if  $\mathfrak{E}(\mathfrak{s}_\nu)$  contains a proper segment  $\mathfrak{s}'_\nu$  with  $\mathfrak{s}'_\nu \rightarrow \mathfrak{s}'_0$ .*

The following theorem establishes a relation between  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_\nu), \mathfrak{E}(\mathfrak{s}_0)) \rightarrow 0$  and  $x_\nu(\tau) \rightarrow x_0(\tau)$ .

(6.9) **THEOREM.** *If  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_\nu), \mathfrak{E}(\mathfrak{s}_0)) \rightarrow 0$ , then representations  $x_\nu(\tau)$  of  $\mathfrak{E}(\mathfrak{s}_\nu)$ ,  $\nu = 0, 1, 2, \dots$ , exist such that  $x_\nu(\tau) \rightarrow x_0(\tau)$  for every  $\tau$ .*

*If  $x_\nu(\tau)$  represents the extremal  $\mathfrak{E}(\mathfrak{s}_\nu)$  and  $x_\nu(\tau)$  tends uniformly on the whole  $\tau$ -axis to  $x_0(\tau)$  (that is,  $\sup_\tau \sigma(x_\nu(\tau), x_0(\tau)) = \epsilon_\nu \rightarrow 0$ ), then  $x_0(\tau)$  represents an extremal  $\mathfrak{E}(\mathfrak{s}_0)$  for which  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_\nu), \mathfrak{E}(\mathfrak{s}_0)) \rightarrow 0$ .*

*The converse of either statement is false.*

**Proof.** Put  $\mathfrak{E}\{\mathfrak{s}_\nu\} = \mathfrak{E}_\nu$ . Let  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_\nu), \mathfrak{E}(\mathfrak{s}_0)) \rightarrow 0$  and  $\mathfrak{d}_0 \in \mathfrak{E}_0$ . Choose  $\mathfrak{d}_\nu \in \mathfrak{E}_\nu$  such that  $\sigma(\mathfrak{d}_0, \mathfrak{E}_\nu) + 1/\nu > \sigma(\mathfrak{d}_0, \mathfrak{d}_\nu)$ . Then by (6.1)

$$\begin{aligned} \exp \sigma(\mathfrak{p}, \mathfrak{d}_0) \sigma(\mathfrak{E}(\mathfrak{s}_\nu), \mathfrak{E}(\mathfrak{s}_0)) &\geq |\sigma(\mathfrak{d}_0, \mathfrak{E}_\nu) - \sigma(\mathfrak{d}_0, \mathfrak{E}_0)| \\ &= \sigma(\mathfrak{d}_0, \mathfrak{E}_\nu) \geq \sigma(\mathfrak{d}_0, \mathfrak{d}_\nu) - 1/\nu \end{aligned}$$

which shows that  $\sigma(\mathfrak{d}_0, \mathfrak{d}_\nu) \rightarrow 0$ . The first part of (6.9) follows now from (6.8).

The proof of the second part is simple in the symmetric case  $xy = yx$  but rather long in the general case. Let a positive  $\epsilon \leq 1$  be given. The hypothesis  $x_\nu(\tau) \rightarrow x_0(\tau)$  and (6.6) show that  $\sigma(\mathfrak{p}, \mathfrak{E}_\nu)$  is bounded. Since for any direction  $\mathfrak{d}$

$$|\sigma(\mathfrak{d}, \mathfrak{E}_\nu) - \sigma(\mathfrak{d}, \mathfrak{E}_0)| \leq 2\sigma(\mathfrak{d}, \mathfrak{p}) + |\sigma(\mathfrak{p}, \mathfrak{E}_\nu) - \sigma(\mathfrak{p}, \mathfrak{E}_0)|$$

it follows that  $\mu_0$  can be chosen such that

$$(6.10) \quad |\sigma(\mathfrak{d}, \mathfrak{E}_\nu) - \sigma(\mathfrak{d}, \mathfrak{E}_0)| \exp(-\sigma(\mathfrak{p}, \mathfrak{d})) < \epsilon$$

for  $\sigma(\mathfrak{p}, \mathfrak{d}) > \mu_0$  and all  $\nu$ .

Let now  $\sigma(\mathfrak{p}, \mathfrak{d}) \leq \mu_0$  and choose  $\mathfrak{w}_0 \in \mathfrak{E}_0$  such that  $\sigma(\mathfrak{d}, \mathfrak{w}_0) - \sigma(\mathfrak{d}, \mathfrak{E}_0) \leq \epsilon/3$ . Let  $\mathfrak{w}_0$  be represented by  $x_0(\tau)$  for  $\tau_0 \leq \tau \leq \tau_0 + 2\delta_0$  and call  $\mathfrak{w}_\nu$  the direction in  $\mathfrak{E}_\nu$  represented by  $x_\nu(\tau)$  for  $\tau_0 \leq \tau \leq \tau_0 + \eta(x_\nu(\tau_0)) = \tau_0 + 2\delta_\nu$ . The inequalities

$$\sigma(\mathfrak{p}, \mathfrak{w}_0) \leq \sigma(\mathfrak{p}, \mathfrak{d}) + \sigma(\mathfrak{d}, \mathfrak{w}_0) \leq 2\sigma(\mathfrak{p}, \mathfrak{d}) + \sigma(\mathfrak{p}, \mathfrak{E}_0) + 1$$

and  $\eta(x) \leq 1$  show that the points  $x_0(\tau_0 + \delta_\nu)$  belong to a bounded set  $A$  which depends on  $\epsilon$  but not on  $\nu$  or  $\mathfrak{d}$ . By (1.18) there is a  $\delta > 0$  such that  $\sigma^*(a, b) < \delta$  and  $a, b \in A$  imply  $\sigma(a, b) < 2\epsilon/3$ . Now

$$\begin{aligned} 2\sigma(\mathfrak{w}_\nu, \mathfrak{w}_0) &= \sigma(x_\nu(\tau_0), x_0(\tau_0)) + \sigma(x_\nu(\tau_0 + \delta_\nu), x_0(\tau_0 + \delta_0)) \\ &< 2\epsilon_\nu + \sigma(x_0(\tau_0 + \delta_\nu), x_0(\tau_0 + \delta_0)) \end{aligned}$$

and by (5.1) and (4.9)

$$\begin{aligned} \sigma^*(x_0(\tau_0 + \delta_\nu), x_0(\tau_0 + \delta_0)) &\leq |\delta_0 - \delta_\nu| = (1/2) |\eta(x_0(\tau_0)) - \eta(x_\nu(\tau_0))| \\ &< (1/2) \sigma(x_0(\tau_0), x_\nu(\tau_0)) < \epsilon_\nu/2. \end{aligned}$$

If  $\nu_0$  is such that  $2\epsilon_\nu < \epsilon/3$  and  $\epsilon_\nu/2 < \delta$  for  $\nu > \nu_0$ , then for  $\sigma(p, b) \leq \mu_0$  and  $\nu > \nu_0$

$$(6.11) \quad \sigma(b, \mathfrak{E}_\nu) - \sigma(b, \mathfrak{E}_0) < \sigma(b, w_\nu) - \sigma(b, w_0) + \epsilon/3 \leq \sigma(w_\nu, w_0) + \epsilon/3 < \epsilon.$$

Next choose  $v_\nu \in \mathfrak{E}_\nu$  such that  $\sigma(b, v_\nu) - \sigma(b, \mathfrak{E}_\nu) < \epsilon/3$ . Let  $v_\nu$  be represented by  $x_\nu(\tau)$  for  $\tau_\nu \leq \tau \leq \tau_\nu + \eta(x_\nu(\tau_\nu)) = \tau_\nu + 2\delta'_\nu$  and call  $u_\nu$  the direction represented by  $x_0(\tau)$  for  $\tau_\nu \leq \tau \leq \tau_\nu + 2\delta''_\nu$ . As before it follows from

$$2\sigma(v_\nu, u_\nu) \leq 2\epsilon_\nu + \sigma(x_0(\tau_\nu + \delta'_\nu), x_0(\tau_\nu + \delta''_\nu))$$

that for  $\sigma(p, b) \leq \mu_0$  and for  $\nu > \nu_1$

$$(6.12) \quad \sigma(b, \mathfrak{E}_0) - \sigma(b, \mathfrak{E}_\nu) < \epsilon.$$

The three relations (6.10), (6.11), (6.12) prove the second part of (6.9).

The example  $R^4$  in §8 will show that the converse of the first part is false. The converse of the second part is easily refuted with the help of the following remark:

(6.13) *If the extremal  $\mathfrak{E}(\mathfrak{s}_\nu)$  is a directed straight line,  $\nu = 1, 2, \dots$ , and contains a segment  $\mathfrak{s}'_\nu$  which tends to a proper segment  $\mathfrak{s}_0$ , then  $\mathfrak{E}(\mathfrak{s}_0)$  is a straight line and  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_\nu), \mathfrak{E}(\mathfrak{s}_0)) \rightarrow 0$ .*

By (6.8),  $\mathfrak{E}(\mathfrak{s}_\nu)$  has a representation  $x_\nu(\tau)$ ,  $\nu = 0, 1, 2, \dots$ , with  $x_\nu(\tau) \rightarrow x_0(\tau)$ . Since, for  $\tau_1 < \tau_2$ ,  $\tau_2 - \tau_1 = x_\nu(\tau_1)x_\nu(\tau_2) \rightarrow x_0(\tau_1)x_0(\tau_2)$ , the extremal  $\mathfrak{E}(\mathfrak{s}_0)$  is a straight line, and  $\mathfrak{E}\{\mathfrak{s}_0\} \subset Fl \inf \mathfrak{E}\{\mathfrak{s}_\nu\}$ , because of (6.7), (6.8). The relation  $\mathfrak{E}\{\mathfrak{s}_0\} \supset Fl \inf \mathfrak{E}\{\mathfrak{s}_\nu\}$  is an immediate consequence of  $x_\nu(\tau) \rightarrow x_0(\tau)$  and the fact that each point of a directed straight line is origin of exactly one direction.

The extremals of the euclidean plane are all directed straight lines hence  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_\nu), \mathfrak{E}(\mathfrak{s}_0)) \rightarrow 0$  whenever  $\sigma(\mathfrak{s}_\nu, \mathfrak{s}_0) \rightarrow 0$ . When the  $\mathfrak{s}_\nu$  are directions with the same initial and different end points, the convergence  $x_\nu(\tau) \rightarrow x_0(\tau)$  is not uniform.

The following consequence of (6.6), (6.9) throws some light on the case where the distance between two different extremals vanishes.

(6.14) If  $\bar{\sigma}(\mathfrak{E}(\mathfrak{s}_1), \mathfrak{E}(\mathfrak{s}_2)) = 0$ , and a representation  $x_1(\tau)$  of  $\mathfrak{E}(\mathfrak{s}_1)$  and an  $\epsilon > 0$  are given, then a representation  $x_2(\tau)$  of  $\mathfrak{E}(\mathfrak{s}_2)$  exists such that  $\sigma(x_1(\tau), x_2(\tau)) < \epsilon$  for  $|\tau| < 1/\epsilon$ .

**7. Segments in  $G$ -spaces. Convexity of spheres.** In a space  $R$  which satisfies Axioms A, B, C, D and is symmetric, that is,  $xy = yx$ , not all the preceding agreements are natural; at any rate, they must be supplemented by others which take the symmetry into account.

If  $x(\tau)$ ,  $\alpha \leq \tau \leq \beta$ , represents a segment  $\mathfrak{s}(a, b)$  and  $xy = yx$ , then  $x(\beta - \tau)$ ,  $0 \leq \tau \leq \beta - \alpha$ , will represent a segment  $\mathfrak{s}(b, a)$ . The point sets carrying the two

segments are identical. As in elementary geometry it is frequently convenient to identify the two segments  $\mathfrak{s}(a, b)$  and  $\mathfrak{s}(b, a)$ .

A *non-oriented segment*  $t(a, b) = t(b, a) = t$  is a pair of oriented segments  $\mathfrak{s}_1 = \mathfrak{s}(a, b)$  and  $\mathfrak{s}_2 = \mathfrak{s}(b, a)$  carried by the same point set. A representation of  $t$  is by definition either a representation of  $\mathfrak{s}_1$  or of  $\mathfrak{s}_2$ . Hence by (1.12)

(7.1)  $x(\tau)$ ,  $\alpha \leq \tau \leq \beta$ , represents a segment  $t$  if and only if  $x(\tau_1)x(\tau_2) = |\tau_1 - \tau_2|$  for  $\alpha \leq \tau_1, \tau_2 \leq \beta$ . The function  $y(\tau)$  represents the same segment  $t$  if and only if

$$(7.1a) \quad y(\eta\tau + \gamma) = x(\tau) \quad \text{for} \quad \alpha \leq \tau \leq \beta$$

and a suitable real  $\gamma$  and  $\eta = \pm 1$ .

A consequence of (7.1) is

(7.2) *A point set  $T$  of a  $G$ -space carries a segment  $t$  if and only if  $T$  can be mapped isometrically on an interval of the real axis.  $T$  carries at most one segment.*

As distance between the two segments  $t_1$  and  $t_2$  carried by  $T_1$  and  $T_2$  respectively we might therefore use the number  $\rho_p(T_1, T_2)$  defined in §2. On the other hand, if  $t_\kappa$  is the pair  $(\mathfrak{s}'_\kappa, \mathfrak{s}''_\kappa)$  of oriented segments, the definition (4.3) suggests

$$(7.3) \quad \sigma(t_1, t_2) = (1/2) \min [\sigma(\mathfrak{s}'_1, \mathfrak{s}'_2) + \sigma(\mathfrak{s}''_1, \mathfrak{s}''_2), \sigma(\mathfrak{s}'_1, \mathfrak{s}''_2) + \sigma(\mathfrak{s}''_1, \mathfrak{s}'_2)]$$

as distance.  $\sigma$  satisfies the metric axioms for  $\sigma(t_1, t_2) = \sigma(t_2, t_1) \geq 0$  is obvious. Since  $\mathfrak{s}'_1 = \mathfrak{s}'_2$  and  $\mathfrak{s}''_1 = \mathfrak{s}''_2$  are equivalent to  $\mathfrak{s}'_1'' = \mathfrak{s}'_2''$  and  $\mathfrak{s}'_1 = \mathfrak{s}'_2$  respectively,  $\sigma(t_1, t_2) = 0$  is equivalent to  $t_1 = t_2$ . The triangle inequality follows from inequalities of this type

$$\begin{aligned} \sigma(\mathfrak{s}'_1, \mathfrak{s}'_2'') + \sigma(\mathfrak{s}'_1'', \mathfrak{s}'_2) + \sigma(\mathfrak{s}'_2, \mathfrak{s}'_3) + \sigma(\mathfrak{s}'_2', \mathfrak{s}'_3'') \\ \geq \sigma(\mathfrak{s}'_1'', \mathfrak{s}'_3) + \sigma(\mathfrak{s}'_1, \mathfrak{s}'_3'') \geq 2\sigma(t_1, t_3). \end{aligned}$$

(2.12) and (2.13) show that the two distances are topologically equivalent: let  $T_\nu$  carry  $t_\nu$ ,  $\nu = 0, 1, 2, \dots$ . Then  $\rho_p(T_\nu, T_0) \rightarrow 0$  if and only if  $\sigma(t_\nu, t_0) \rightarrow 0$ . The distance (7.3) will be used in preference to  $\rho_p$  because it is easy to calculate and because (by (4.4)) the segments form with it a finitely compact space  $R_t$ .

With  $\rho_p$  the segments of the real axis do not form a finitely compact space. For let  $T_\nu$  be the set  $0 \leq \tau \leq \nu$  and  $p = 0$ . Then  $\{\rho_p(T_\nu, T_\mu)\}$  is bounded by (2.11), but  $\{T_\nu\}$  contains obviously no subsequence which tends to a segment.

A space which satisfies the Axioms A, B, C, D and is symmetric ( $xy = yx$ ) and with topologies for its oriented and non-oriented segments equivalent to  $R_d$  and  $R_t$  respectively will be called a *G-space*.

An important fact which has no analogue in *E-spaces* is the following:

(7.4) *If  $T_\kappa$  carries the segment  $t_\kappa$ ,  $\kappa = 1, 2$ , then  $T_1 \cap T_2$  is either empty, or exactly one point, or consists of two points which are the end points of both  $t_1$  and  $t_2$ , or  $T_1 \cap T_2$  is carrier of a proper segment.*

**Proof.** Let  $T_1 \neq T_2$  and let  $T_1 \cap T_2$  contain the two distinct points  $a$  and  $b$ , where  $a$  is no end point of  $t_1$  say. Then  $t(a, b)$  is unique by (4.2). Hence the carrier of  $t(a, b)$  is contained in both  $T_1$  and  $T_2$ , whence it easily follows that  $T_1 \cap T_2$  carries a segment. If  $a$  and  $b$  are the end points both of  $t_1$  and  $t_2$  it follows from  $T_1 \neq T_2$  and (4.2) that  $T_1 \cap T_2$  contains no other point.

Just as the subspace  $R$  of  $R_s$  formed by the directions of  $R$  is more important than  $R_s$  itself, so the subspace  $R_e$  of  $R_t$  formed by the line elements of  $R$  will prove more important than  $R_t$ . A *line element* with origin  $x$  is a non-oriented segment with center  $x$  and length  $2\eta(x)$ . The space  $R_e$  is finitely compact since  $\eta(x)$  is continuous; and, since  $\eta(x) \leq 1$ , a sequence of line elements is bounded if and only if the sequence of their origins is bounded.

In  $G$ -spaces the spherical neighborhoods  $S(x, \rho)$ ,  $S'(x, \rho)$  and  $S''(x, \rho)$  coincide.  $S(x, \rho)$  is called the interior of the sphere  $K(x, \rho)$  which consists by definition of the points  $y$  with  $xy = \rho$ . Since segments exist the closure  $\bar{S}(x, \rho)$  of  $S(x, \rho)$  equals  $S(x, \rho) \cup K(x, \rho)$ . The uniqueness of small segments enables us to define convexity for small spheres.

The sphere  $K(x, \rho)$  of radius  $\rho \leq \eta_4(x)$  is called *weakly convex* if  $t(x, y) \subset \bar{S}(x, \rho)$  for any two points  $x, y \in K(x, \rho)$ , and *strictly convex* if  $t(x, y) - x - y \subset S(x, \rho)$  for  $x, y \in K(x, \rho)$ .

This terminology will be justified by three simple facts which show that convexity has similar properties as in euclidean geometry.

(7.5) If  $K(p, \rho)$  is weakly convex and  $xp < \rho$ ,  $yp = \rho$ , and  $(xzy)$ , then  $zp < \rho$ .

For consider the segment  $t(a, b)$  of length  $4\rho$  which is co-central with and contains  $t(x, y)$ ; ( $t(a, b)$  exists because  $\rho \leq \eta_4(x)$ ). Choose the notation such that  $(axy)$  and  $(xyb)$ . Then  $ap \geq ay - yp > 2\rho - \rho = \rho$ , hence  $t(a, x)$  contains a point  $a_1$  on  $K(p, \rho)$ , and  $t(a_1, y)$  contains  $t(x, y)$  and it follows from the convexity of  $K(p, \rho)$  that  $pz \leq \rho$ .

Every point of  $U = S(x, \rho - px)$  lies in  $S(p, \rho)$ . According to the result just established  $t(u, y) \subset \bar{S}(p, \rho)$  for  $u \in U$ . Since  $\bigcup_{u \in U} t(u, y) - y$  is an open set and lies in  $S(p, \rho)$  it cannot contain a point of  $K(p, \rho)$ , which proves that  $t(x, y) - y \subset S(p, \rho)$ .

(7.6) If  $K(p, \rho)$  is weakly convex and  $x, y \in K(p, \rho)$  then either  $t(x, y) \subset K(p, \rho)$  or  $t(x, y) - x - y \subset S(p, \rho)$ .

(7.7) If  $K(p, \rho)$  is weakly convex and  $x, y \in S(p, \rho)$  then  $t(x, y) \subset S(p, \rho)$ .

Namely, the convexity of  $K(p, \rho)$  and  $x, y \in K(p, \rho)$  imply that  $t(x, y) \subset \bar{S}(p, \rho)$ . If  $t(x, y)$  contains a point  $z$  of  $S(p, \rho)$  it follows from (7.5) that  $t(x, z) - x$  and  $t(y, z) - y$  lie in  $S(p, \rho)$  which proves (7.6).

If  $x$  and  $y$  are different points in  $S(p, \rho)$  consider the segment  $t(a, b)$  of length  $4\rho$  co-central with and containing  $t(x, y)$ . The proof of (7.5) shows that  $t(a, b)$  contains a segment  $t(a_1, b_1)$  which in turn contains  $t(x, y)$  and has its end points on  $K(p, \rho)$ . Hence (7.7) follows from (7.6).

Next it will be shown that the convexity of the spheres with a certain radius  $\rho$  implies the convexity of the spheres with smaller radii.

(7.8) Let  $\eta = \inf_{x \in F} \eta_4(x) > 0$  and let the spheres  $K(p, \rho)$  be weakly convex for  $p \in F$  and a fixed positive  $\rho \leq \eta$ . Then each sphere  $K(q, \sigma)$  with  $0 < \sigma < \rho$  and  $\bar{S}(q, \rho - \sigma) \subset F$  is strictly convex.

**Proof.** If  $K(q, \sigma)$  were not strictly convex, points  $x, y, z$  with  $xq = yq = \sigma$ ,  $(xzy)$ , and  $qz \geq \sigma$  would exist. Since  $x, y \in S(q, \rho)$  and  $K(q, \rho)$  is weakly convex it follows from (7.7) that  $qz < \rho$ . Let  $(pqz)$  and  $pz = \rho$  ( $p$  exists because  $\rho \leq \eta_4(p)$ ). Then  $pq = pz - qz \leq \rho - \sigma$ , hence  $px < pq + qx \leq pq + \sigma \leq \rho$  and, similarly,  $py < \rho$ . Since  $p \in \bar{S}(q, \rho - \sigma) \subset F$  the sphere  $K(p, \rho)$  is by hypothesis weakly convex and (7.7) yields  $t(x, y) \subset S(p, \rho)$  which contradicts  $zp = \rho$ .

The most important property of convexity is its relation to perpendicularity. In any  $E$ -space the point  $f$  is called a *foot* of the point  $p$  on the set  $A$  (and any  $\mathfrak{s}(p, f)$  *perpendicular* to  $A$  at  $f$ ) if  $f \in A$  and  $pf = pA$ .

(7.9) If in an  $E$ -space  $(pxf)$  and  $f$  is a foot of  $p$  on  $A$ , then  $f$  is the unique foot of  $x$  on  $A$  (or: the perpendicular to  $A$  from  $x$  is unique and coincides with  $\mathfrak{s}(x, f)$ ).

For if  $g \in A$  then

$$xg \geq pg - px \geq pf - px = xf$$

which shows that  $f$  is a foot of  $x$  on  $A$  and that  $xg = xf$  only for  $(pxg)$ . But then  $f = g$  by (4.1).

The following two facts interrelate convexity of spheres with perpendicularity in  $G$ -spaces.

(7.10) If the spheres  $K(p, \sigma)$  are strictly convex for  $0 < \sigma < \bar{\eta} \leq \eta_4(p)$  then  $p$  has exactly one foot on every segment that contains points of  $S(p, \bar{\eta})$ .

Since segments are compact sets,  $p$  has at least one foot on any segment. If  $p$  had two different feet  $f_1$  and  $f_2$  on a segment  $t$  with  $t \cap S(p, \bar{\eta}) \neq \emptyset$ , then  $pf_1 = pf_2 = \sigma < \bar{\eta}$ , and  $px \geq pf_i$  for  $(f_1xf_2)$ , which contradicts the strict convexity of  $K(p, \sigma)$ .

(7.11) Let the subset  $F$  of  $R$  contain with  $x$  the sphere  $S(x, \rho(x))$ , where  $0 < \rho(x) \leq \eta_4(x)$  and  $|\rho(x) - \rho(y)| \leq xy$ . If every point  $x$  of  $F$  has exactly one foot on any segment that contains points of  $S(x, \rho(x))$  then  $K(x, \sigma)$  is strictly convex for  $x \in F$  and  $0 < \sigma < \rho(x)$ .

If (7.11) were not true, a point  $q$  in  $F$  and points  $a, b, c$  would exist such that  $qa = qc = \sigma < \rho(q)$ ,  $(abc)$  and  $qb \geq \sigma$ . Since the foot  $f$  of  $q$  on  $t(a, c)$  is unique it is different from  $a, b$  and  $c$ ; assume  $(afb)$ . If  $(quc)$ , then  $u \in S(q, \rho(q)) \subset F$ . By hypothesis  $uc = qc - qu < \rho(q) - qu \leq \rho(u)$  and  $u$  has exactly one foot  $f(u)$  on  $t(a, c)$ . Therefore  $f(u)$  varies continuously, when  $u$  traverses  $t(q, c)$  from  $q$  to  $c$ . Since  $f(q) = f$  and  $f(c) = c$ , there is a point  $u_0$  on  $t(q, c)$  such that  $f(u_0) = b$ . But

$$ub \geq qb - qu \geq qc - qu = uc$$

which contradicts the fact that  $b$  is the unique foot of  $u_0$  on  $t(a, c)$ .

**8. Geodesics.** If  $x(\tau)$  represents an extremal of a  $G$ -space, then  $y(\tau) = x(-\tau)$  will also represent an extremal. As in the case of segments it is fre-

quently convenient to identify these two extremals. The same procedure as in the preceding section may be followed. Since, however, these new objects have quite a few properties which extremals do not have, in general, an independent definition analogous to the definition of extremals is preferable.

The proper non-oriented segments of a  $G$ -space are divided into classes, called *geodesics*. Two proper segments  $t'$  and  $t''$  belong to the same class if segments  $t_1=t'$ ,  $t_2, \dots, t_{\mu-1}$ ,  $t_\mu=t''$  exist such that  $t_\nu$  and  $t_{\nu+1}$ ,  $\nu=1, \dots, \mu-1$ , have a common proper subsegment. By (7.4) it suffices that the intersection  $T_\nu \cap T_{\nu+1}$  of the carriers of  $t_\nu$  and  $t_{\nu+1}$  carries a proper segment. The geodesic which contains the segment  $t$  will be denoted by  $\mathfrak{G}(t)$ .

A continuous curve  $x(\tau)$ ,  $-\infty < \tau < \infty$ , is said to represent  $\mathfrak{G}(t)$  if it has Property I in §5 and

PROPERTY II'. *A segment  $t'$  belongs to  $\mathfrak{G}(t)$  if and only if numbers  $\tau_0$  and  $\tau_1$  exist such that  $x(\tau)$  represents  $t'$  for  $\tau_0 \leq \tau \leq \tau_1$ .*

By this definition  $x(\tau)$  represents a geodesic if and only if it represents an extremal. Hence (5.4), (5.6) hold for geodesics as well as for extremals and (5.7) becomes with the help of (7.2)

(8.1) *Every geodesic possesses a representation.  $x(\tau)$  and  $y(\tau)$  represent the same geodesic if and only if a real  $\beta$  and  $\eta = \pm 1$  exist such that*

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

The point set which carries the geodesic  $\mathfrak{G}(t)$  will be denoted by  $G[t]$ . The following fact is basic for the properties which geodesics have in addition to those of the extremals.

(8.2) THEOREM. *If  $\mathfrak{G}(t)$  and  $\mathfrak{G}(t')$  are different geodesics, then  $G[t] \cap G[t']$  is empty, finite, or countable.*

It will be shown that if  $G[t] \cap G[t']$  is nondenumerable, then  $\mathfrak{G}(t) = \mathfrak{G}(t')$ . Let  $x(\tau)$  and  $x'(\tau)$  represent  $\mathfrak{G}(t)$  and  $\mathfrak{G}(t')$  respectively. By (5.2),  $x(\tau)$  represents a line element for  $|\tau - \tau_0| \leq \eta(x(\tau_0))$ . Choose  $\tau_\nu$ ,  $\nu=1, 2, \dots$ , such that the intervals  $|\tau - \tau_\nu| \leq \eta(x(\tau_\nu))$  cover the  $\tau$ -axis. At least one of the corresponding line elements in  $\mathfrak{G}(t)$ , say  $t_\lambda$ , is carried by a set  $T_\lambda$  which contains a nondenumerable subset  $V$  of  $G[t] \cap G[t']$ . Next choose  $\tau'_\nu$ ,  $\nu=1, 2, \dots$ , such that the intervals  $|\tau - \tau'_\nu| \leq \eta(x'(\tau'_\nu))$  cover the  $\tau$ -axis. At least one of the corresponding line elements in  $\mathfrak{G}(t')$ , say  $t'_\mu$ , is carried by a set  $T'_\mu$  which contains a nondenumerable subset  $W$  of  $V$ . Now (7.4) and  $T_\lambda \cap T'_\mu \supset W$  imply that  $t_\lambda$  and  $t'_\mu$  have a common subsegment, hence

$$\mathfrak{G}(t) = \mathfrak{G}(t_\lambda) = \mathfrak{G}(t'_\mu) = \mathfrak{G}(t').$$

As a corollary we obtain the following improvement on (5.6).

(8.3) *If  $T'$  carries the proper segment  $t'$  and  $T' \subset G[t]$  then  $\mathfrak{G}(t') = \mathfrak{G}(t)$ .*

Every point  $p$  of  $G[t]$  is origin of at least one line element in  $\mathfrak{G}(t)$ . The cardinal number  $\mu$  of the different line elements in  $\mathfrak{G}(t)$  with origin  $p$  is called

the *multiplicity* of  $p$  as point of  $\mathfrak{G}(t)$  or  $G[t]$ ; the point  $p$  is called a *simple point* of  $\mathfrak{G}(t)$  when  $\mu=1$  and *multiple* when  $\mu>1$ .

(8.4) *The multiplicity  $\mu$  of a point  $p$  of a geodesic  $\mathfrak{G}(t)$  is at most denumerable.*

Let  $t_1$  and  $t_2$  be two different line elements in  $\mathfrak{G}(t)$  with origin  $p$ . Let  $x(\tau)$  represent  $\mathfrak{G}(t)$ . By Property II' there are values  $\tau_1$  and  $\tau_2$  such that  $x(\tau)$  represents  $t_\kappa$  for  $|\tau - \tau_\kappa| \leq \eta(p)$  ( $p = x(\tau_\kappa)$ ). Because of (4.1) the carriers of  $t_1$  and  $t_2$  intersect only at  $p$ , so that the intervals  $|\tau - \tau_\kappa| \leq \eta(p)$  of the  $\tau$ -axis are disjoint, which proves (8.4).

(8.5) *A geodesic has at most denumerably many multiple points.*

**Proof.** Let  $x(\tau)$  represent a geodesic  $\mathfrak{G}$ . It suffices to show that for an arbitrary real  $\tau_0$  the arc  $|\tau - \tau_0| \leq \eta(x(\tau_0))/2$  contains at most denumerably many multiple points of  $\mathfrak{G}$ .

Put  $x(\tau_0) = p$  and denote the segment represented by  $x(\tau)$  for  $|\tau - \tau_0| \leq \eta(p)/2$  by  $t_3$ . If  $a_1$  and  $a_2$  are two different multiple points of  $\mathfrak{G}$  on  $t_3$ , then  $a_\kappa$ ,  $\kappa=1, 2$ , is origin of a line element  $t'$  which has no common point with  $t_3$  except  $a_\kappa$ . Since  $\eta(a_\kappa) \geq \eta(p) - pa_\kappa \geq \eta(p)/2$  the line element  $t'_\kappa$  contains a co-central subsegment  $t_\kappa$  of length  $\eta(p)$ .

The segments  $t_\kappa$  are unique segments between their end points. By (7.4) they intersect in at most one point, or they contain a common proper subsegment. The second case is impossible. For by the definition (4.10) of  $\eta(p)$ , this common subsegment is co-central subsegment of a segment  $t^*$  of length  $6\eta(p)$ . Comparing the lengths of  $t^*$ ,  $t_1$ ,  $t_2$  shows that  $t^*$  must contain both  $t_1$  and  $t_2$ , in particular  $a_1$  and  $a_2$ . By (7.4),  $t^*$  and  $t_3$  have a common proper subsegment which contains  $a_1$  and  $a_2$ , consequently also proper subsegments of  $t_1$  and  $t_2$ . But  $t_\kappa$  and  $t_3$  intersect only at  $a_\kappa$ .

Let  $x(\tau)$  represent  $t_\kappa$  for  $|\tau - \tau_\kappa| \leq \eta(p)/2$ . Since  $t_1$  and  $t_2$  intersect in at most one point it follows that the intervals  $|\tau - \tau_\kappa| < \eta(p)/2$  of the real axis are disjoint. Hence  $t_3$  contains at most denumerably many different multiple points of  $G$ .

A geodesic without multiple points is called *simple*. The simple geodesics may be characterized both by their shape and the properties of their representations.

(8.6) *Either of the following properties (a) or (b) is necessary and sufficient that the geodesic  $\mathfrak{G}(t)$  be simple:*

(a)  *$G[t]$  is homeomorphic to a circle or is a one-to-one continuous map of a straight line.*

(b) *Let  $x(\tau)$  represent  $\mathfrak{G}(t)$ . If  $x(\tau_1) = x(\tau_2)$  then  $x(\tau_1 + \tau) = x(\tau_2 + \tau)$  for all  $\tau$ .*

(a) is sufficient because it implies that any point  $p$  of  $\mathfrak{G}(t)$  is center of exactly one line element. To see that (b) is sufficient let  $t_1$  and  $t_2$  be line elements in  $\mathfrak{G}(t)$  with the same origin  $p$ , and let  $x(\tau)$  represent  $t_\kappa$  for  $|\tau - \tau_\kappa| \leq \eta(p)$ . Then  $x(\tau_1) = x(\tau_2)$ , hence  $x(\tau_1 + \tau) = x(\tau_2 + \tau)$  for all  $\tau$  and  $t_1 = t_2$ .

Conversely let  $\mathfrak{G}(t)$  be simple. If no pair  $\tau_1 \neq \tau_2$  exists with  $x(\tau_1) = x(\tau_2)$ , then  $\tau \rightarrow x(\tau)$  is a one-to-one and continuous mapping of the  $\tau$ -axis on  $G[t]$ ;



moreover (b) holds. If  $x(\tau_1) = x(\tau_2) = p$  and  $\tau_1 \neq \tau_2$  then  $x(\tau)$  represents a line element  $t_\kappa$  for  $|\tau - \tau_\kappa| \leq \eta(p)$ ,  $\kappa = 1, 2$ . The simplicity of  $\mathfrak{G}(t)$  implies  $t_1 = t_2$ , hence it follows from (7.2) that for  $|\tau| \leq \eta(p)$  either  $x(\tau_1 + \tau) = x(\tau_2 + \tau)$  or  $x(\tau_1 + \tau) = x(\tau_2 - \tau)$ . By (5.6) and (8.1) either of these relations holds for all  $\tau$  if it holds for  $|\tau| \leq \eta(p)$ . The second equation is impossible since it would imply that for  $\tau_0 = (\tau_1 + \tau_2)/2$  and all  $\tau$

$$x(\tau_0 + \tau) = x(\tau_1 + (\tau_2 - \tau_1)/2 + \tau) = x(\tau_2 - (\tau_2 - \tau_1)/2 - \tau) = x(\tau_0 - \tau)$$

which contradicts Property I for  $\tau = \tau_0$ . Hence (b) is necessary.

This completes also the proof for the necessity of (a). For let  $\eta$  be the smallest positive number for which  $x(\tau_1 + \eta) = x(\tau_1)$ . Then  $x(\tau + \eta) = x(\tau)$  for all  $\tau$  and  $x(\tau_1) \neq x(\tau_2)$  when  $\tau_1 - \tau_2$  is no integral multiple of  $\eta$ . Therefore  $x(\tau) \rightarrow \exp(i2\pi\eta^{-1}\tau)$  maps  $G[t]$  topologically on the unit circle of the complex plane.

The most familiar examples of simple geodesics are the straight lines of the euclidean geometry and the great circles of the spherical. Generally a geodesic  $\mathfrak{G}$  will be called a *straight line* when  $x(\tau_1)x(\tau_2) = |\tau_1 - \tau_2|$  for any real  $\tau_1, \tau_2$  and at least one (and then every) representation  $x(\tau)$  of  $\mathfrak{G}$ , and a *great circle* when a positive number  $\eta$  exists such that

$$x(\tau_1)x(\tau_2) = \min_{\nu=0, \pm 1, \pm 2, \dots} |\tau_1 - \tau_2 + \nu\eta| \quad \text{for all real } \tau_1, \tau_2.$$

In the first case  $x(\tau) \rightarrow \tau$  maps  $\mathfrak{G}$  isometrically on the real axis, in the second on the circle of radius  $\eta/2\pi$  with the length of the shorter arc as distance. Therefore  $\eta$  is defined as the length of  $\mathfrak{G}$ .

These two types of geodesics have the property to be *straight*, that is, by definition, to contain with any two points  $x, y$  at least one segment  $t(x, y)$ .

(8.7) THEOREM. *The straight lines and the great circles are the only straight geodesics.*

**Proof.** (1) A straight geodesic  $\mathfrak{G}(t)$  is simple. For an indirect proof let  $t_1$  and  $t_2$  be two different line elements in  $\mathfrak{G}(t)$  with the same origin  $p$ . Let  $q$  be any point on  $t_1$  and not on  $t_2$ . For  $x_1 \neq x_2$  on  $t_2$  the segments  $t(q, x_1)$  and  $t(q, x_2)$  are unique and have  $q$  as only common point. Otherwise (4.2) would imply that one of the segments contains the other, hence  $q$  would lie on  $t_2$ . Since the segments  $t(q, x)$  with  $x \in t_2$  are unique and have their end points on  $G[t]$  it follows from the hypothesis that they are contained in  $\mathfrak{G}(t)$ . But then the multiplicity of  $q$  as a point of  $\mathfrak{G}(t)$  would be nondenumerable which contradicts (8.4).

(2) Let  $x(\tau)$  represent  $\mathfrak{G}(t)$ . If  $x(\tau_1) \neq x(\tau_2)$  for any real  $\tau_1 \neq \tau_2$ , then the subarc  $\tau_1 \leq \tau \leq \tau_2$  of  $x(\tau)$  is the only arc of  $G[t]$  from  $x(\tau_1)$  to  $x(\tau_2)$  and must therefore be a segment  $t(x(\tau_1), x(\tau_2))$ , hence  $x(\tau_1)x(\tau_2) = |\tau_1 - \tau_2|$  by (5.1).

The only other possibility is according to the proof of (8.6) that a smallest  $\eta > 0$  exists such that  $x(\tau + \eta) = x(\tau)$  for all  $\tau$ . If  $\tau_1$  and  $\tau_2$  are given determine

the integer  $\nu$  such that  $|\tau_1 - \tau_2 + \nu\eta|$  is minimal. Then  $|\tau_1 - \tau_2 + \nu\eta| \leq \eta/2$ . Take the case  $\tau_1 - \tau_2 + \nu\eta \leq 0$ . Since  $x(\tau_2) = x(\tau_2 - \nu\eta)$  the arcs  $\tau_1 \leq \tau \leq \tau_2 - \nu\eta$  and  $\tau_2 - \nu\eta \leq \tau \leq \tau_1 + \eta$  are the only arcs in  $G[t]$  from  $x(\tau_1)$  to  $x(\tau_2)$ . By the hypothesis at least one of them must be a segment. Since

$$\tau_2 - \nu\eta - \tau_1 \leq \eta/2 \quad \text{and} \quad \tau_1 + \eta - \tau_2 + \nu\eta \geq \eta - \eta/2 = \eta/2$$

it follows from (5.1) that the first arc is a segment and that  $x(\tau_1)x(\tau_2) = |\tau_1 - \tau_2 + \nu\eta|$ .

A corollary of the last result is

(8.8) *A one-dimensional G-space is either a straight line or a great circle.*

The space contains at least two points and therefore at least one geodesic. No point  $p$  can be the center of two different line elements  $t_1$  and  $t_2$ . For if  $a_x$  is an end point of  $t_x$  then  $t(a_1, a_2)$  is unique, and all segments  $t(p, x)$  with  $x \in t(a_1, a_2)$  are unique and form a set homeomorphic to a 2-simplex. Hence all geodesics of the space are simple and disjoint. Since the space is convex there is only one geodesic, which is straight because there is a segment connecting any two distinct points.

If the geodesic through two distinct points of a  $G$ -space is unique, each geodesic contains by (5.6) with any two points  $a, b$  each segment  $t(a, b)$  and is therefore straight. This statement can be considerably improved (compare B [2, Theorem 3] and (14.5) of the present paper).

It is natural to base a distance for geodesics on the line elements rather than on the directions. The set of all line elements contained in the geodesic  $\mathcal{G}(t)$  will be denoted by  $\mathcal{G}\{t\}$ . Observe that

(8.9)  $\mathcal{G}\{t\} = \bar{\mathcal{G}}\{t\}$  if and only if  $G[t] = \bar{G}[t]$ . ( $\bar{\mathcal{G}}\{t\}$  is the closure of  $\mathcal{G}\{t\}$  in  $R_*$  with respect to the topology (7.3).)

Since the necessity is obvious let  $G[t] = \bar{G}[t]$  and let  $\{t_r\}$  be a sequence of line elements in  $\mathcal{G}\{t\}$  which converges to a line element  $t^*$ . The carrier  $T^*$  of  $t^*$  is contained in  $G[t]$  because  $G[t]$  is closed, whence by (8.3),  $\mathcal{G}(t^*) = \mathcal{G}(t)$  and  $t^* \in \mathcal{G}\{t\}$ .

Fix now a line element  $q$  and define in analogy to (6.1) as distance of two geodesics  $\mathcal{G}(t_1)$  and  $\mathcal{G}(t_2)$  the number

$$(8.10) \quad \begin{aligned} \bar{\sigma}(\mathcal{G}(t_1), \mathcal{G}(t_2)) &= \bar{\rho}(\mathcal{G}\{t_1\}, \mathcal{G}\{t_2\}) \\ &= \sup_{l \in R_*} |\sigma(l, \mathcal{G}\{t_1\}) - \sigma(l, \mathcal{G}\{t_2\})| \exp(-\sigma(q, l)). \end{aligned}$$

The theory of §6 carries over without change to geodesics, but some important additions are possible. Notice in particular the following case in which the converse of (6.3) holds.

(8.11) THEOREM. *If  $G[t] = \bar{G}[t]$  and  $\rho_p(G[t_r], G[t]) \rightarrow 0$  then  $\bar{\sigma}(\mathcal{G}(t_r), \mathcal{G}(t)) \rightarrow 0$ .*

**Proof.** Put  $\mathcal{G}\{t_r\} = \mathcal{G}_r$ ,  $\mathcal{G}\{t\} = \mathcal{G}$  and assume for an indirect proof that for

a subsequence  $\{\mu\}$  of  $\{\nu\}$  the inequality  $\bar{\sigma}(\mathfrak{G}(t_\mu), \mathfrak{G}(t)) \geq 4\alpha > 0$  holds. Then a line element  $t_\mu^*$  exists with

$$3\alpha < |\sigma(t_\mu^*, \mathfrak{G}_\mu) - \sigma(t_\mu^*, \mathfrak{G})| \exp(-\sigma(q, t_\mu^*)) \\ \leq (\sigma(q, \mathfrak{G}_\mu) + \sigma(q, \mathfrak{G}) + 2\sigma(q, t_\mu^*)) \exp(-\sigma(q, t^*))$$

so that  $\sigma(q, t_\mu^*)$  is bounded. For a suitable subsequence  $\{t_\lambda^*\}$  of  $\{t_\mu^*\}$  and the corresponding subsequence  $\{\mathfrak{G}_\lambda\}$  of  $\{\mathfrak{G}_\nu\}$  at least one of the following two relations holds:

$$(8.12) \quad \begin{aligned} \sigma(t_\lambda^*, \mathfrak{G}_\lambda) - \sigma(t_\lambda^*, \mathfrak{G}) &> 3\alpha \quad \text{or} \\ \sigma(t^*, \mathfrak{G}) - \sigma(t_\lambda^*, \mathfrak{G}_\lambda) &> 3\alpha. \end{aligned}$$

In the second case choose  $u_\lambda$  in  $\mathfrak{G}_\lambda$  such that  $\sigma(t_\lambda^*, u_\lambda) < \sigma(t_\lambda^*, \mathfrak{G}_\lambda) + \alpha$ . As in the proof of (2.16) it is seen that  $\sigma(u_\lambda, \mathfrak{G}) > 2\alpha$ . The sequence  $\{u_\lambda\}$  is bounded because  $\{t_\mu^*\}$  is bounded and contains a subsequence which converges to a line element  $u$  whose carrier  $U$  is contained in  $Fl \sup G[t]$ . But (2.13) and  $\rho_p(G[t], G[t]) \rightarrow 0$  yield  $G[t] = \bar{G}[t] = Fl \sup G[t]$ , so that  $U \in G[t]$  and, as a consequence of (8.3), also  $u \in \mathfrak{G}\{t\}$ , which contradicts  $\sigma(u_\lambda, \mathfrak{G}) \geq 2\alpha$ .

In the case (8.12) choose  $u_\lambda$  in  $\mathfrak{G}$  such that  $\sigma(t_\lambda^*, u_\lambda) < \sigma(t_\lambda^*, \mathfrak{G}) + \alpha$ . Then  $\{u_\lambda\}$  will again be bounded and  $\sigma(u_\lambda, \mathfrak{G}_\lambda) \geq 2\alpha$ . A suitable subsequence of  $\{u_\lambda\}$  tends to a line element  $u$  in  $\mathfrak{G}$  because  $\mathfrak{G} = \bar{\mathfrak{G}}$  by (8.9). Every point  $q$  of  $u$  is origin of exactly one line element  $u(q)$  which has with  $u$  a common proper subsegment, and  $\sigma(u(q), u) < \alpha$  when  $q$  belongs to a sufficiently small proper co-central subsegment  $u^*$  of  $u$ .

For every point  $q$  of  $u^*$  there is a point  $q_\lambda$  in  $G[t_\lambda]$  such that  $q_\lambda \rightarrow q$ . Let  $v_\lambda$  be a line element in  $\mathfrak{G}_\lambda$  with origin  $q$ . Then

$$\sigma(u(q), v_\lambda) \geq \sigma(u, v_\lambda) - \sigma(u, u(q)) \geq \sigma(u, \mathfrak{G}_\lambda) - \sigma(u, u(q)) > \alpha.$$

Hence any accumulation line element  $v(q)$  of  $\{v_\lambda\}$  is different from the line element  $u(q)$ . But the carrier  $V(q)$  of  $v(q)$  lies in  $Fl \sup G[t] = G[t]$ , hence  $v(q) \in \mathfrak{G}$  (again by (8.9)), so that every point  $q$  of  $u^*$  has multiplicity not less than 2, which contradicts (8.5).

The following facts show the agreement of the topology  $\bar{\sigma}$  with the intuition in simple cases:

(8.13) If the geodesic  $\mathfrak{G}(t_r)$  is straight and  $t_r$  tends to a proper segment  $t$ , then  $\mathfrak{G}(t)$  is straight and  $\bar{\sigma}(\mathfrak{G}(t_r), \mathfrak{G}(t)) \rightarrow 0$ .

A proof is readily obtained from the arguments used for (6.13).

(2.18), (6.8), (6.9) and (8.12), (8.13) yield

(8.14) *If all geodesics of a G-space are straight, then the topologies  $\bar{\sigma}$  and  $\rho_p$  are equivalent and metrize the set of all geodesics in such way that they form a finitely compact space.  $\bar{\sigma}(\mathfrak{G}_\nu, \mathfrak{G}) \rightarrow 0$  if and only if  $\mathfrak{G}_\nu$  contains a segment, which tends to a proper segment in  $\mathfrak{G}$ .*

The cylinder and the torus with euclidean metrics are simple examples which will confirm the various negative statements of §6 (which also apply to con-

vergence of geodesics). The points of the cartesian  $(\xi, \eta)$ -plane  $P$  are divided into classes by two different methods:

$R^4$ :  $(\xi_1, \eta_1)$  and  $(\xi_2, \eta_2)$  belong to the same class if and only if  $\xi_1 = \xi_2$  and  $\eta_1 - \eta_2$  is an integer.

$R^5$ :  $(\xi_1, \eta_1)$  and  $(\xi_2, \eta_2)$  belong to the same class if and only if  $\xi_1 - \xi_2$  and  $\eta_1 - \eta_2$  are integers.

The distance between two classes  $p$  and  $q$  is defined in both cases as

$$pq = \min [(\xi_1 - \xi_2)^2 + (\eta_1 - \eta_2)^2]^{1/2}$$

where  $(\xi_1, \eta_1)$  and  $(\xi_2, \eta_2)$  traverse independently their respective classes  $p$  and  $q$ . In both cases the metric is locally euclidean so that Axiom D holds and the space is a  $G$ -space. Moreover it follows from Property I that the geodesics correspond in both cases to the euclidean straight lines in  $P$ .

In case of the cylinder  $R^4$  two straight lines  $\mathfrak{L}_1$  and  $\mathfrak{L}_2$  in  $P$  correspond to the same geodesic in  $R^4$  if and only if  $\mathfrak{L}_1$  can be transformed into  $\mathfrak{L}_2$  by a translation  $\xi' = \xi$ ,  $\eta' = \eta + \nu$ ,  $\nu$  an integer. Call  $\mathfrak{G}_\kappa$  the geodesic corresponding to the line  $\eta = \kappa\xi$ ,  $\kappa = 1, 2, \dots$ . With  $\beta_\kappa = (1 + \kappa^2)^{-1/2}$  both

$$x_\kappa(\tau) = (\beta_\kappa\tau, \kappa\beta_\kappa\tau) \quad \text{and} \quad x'_\kappa(\tau) = (\beta_\kappa\tau + 1, \kappa\beta_\kappa\tau)$$

are representations of  $\mathfrak{G}_\kappa$ . For  $\kappa \rightarrow \infty$ , the function  $x_\kappa(\tau)$  tends to  $(0, \tau)$ , which represents a geodesic  $\mathfrak{G}_0$ , and  $x'_\kappa(\tau)$  tends to  $(1, \tau)$  which represents a different geodesic  $\mathfrak{G}_1$ . The geodesics  $\mathfrak{G}_0$  and  $\mathfrak{G}_1$  are great circles. (Hence if the existence of representations  $x_\nu^*(\tau)$  of  $\mathfrak{G}_\nu^*$ ,  $\nu = 0, 1, 2, \dots$ , with  $x_\nu^*(\tau) \rightarrow x_0^*(\tau)$  were taken as the definition of  $\mathfrak{G}_\nu^* \rightarrow \mathfrak{G}_0^*$ , the geodesics of  $R^4$  would not form an  $L$ -space.) If  $G_\nu$  denotes the carrier of  $\mathfrak{G}_\nu$ , then  $\text{Fl sup } G_\nu = R^4$  and  $\rho_p(G_\nu, R^4) \rightarrow 0$  hence, by the triangle inequality,  $\rho_p(G_\nu, H_0) \rightarrow \rho_p(H_0, R^4) > 0$ . This shows that the converse of the first part of (6.9) does not hold, and also that in (8.13) the assumption that the  $\mathfrak{G}(t_\nu)$  are straight cannot be replaced by the requirement that  $\mathfrak{G}(t)$  be straight.

Moreover the carrier of every geodesic is closed in  $R^4$ , hence  $\rho_p$  and  $\bar{\sigma}$  are equivalent and metrize the set  $R_G$  of all geodesics in  $R^4$ . In the topology  $\rho_p$  it is easily seen that the geodesics  $\xi = \text{const.}$  form in  $R_G$  a set homeomorphic to the real axis, which is not connected with its complement.

In the case of the torus  $R^5$  two straight lines  $\mathfrak{L}_1$  and  $\mathfrak{L}_2$  in  $P$  correspond to the same geodesic in  $R^5$  if and only if  $\mathfrak{L}_1$  can be carried into  $\mathfrak{L}_2$  by a translation  $\xi' = \xi + \mu$ ,  $\eta' = \eta + \nu$ , where  $\mu$  and  $\nu$  are integers. Therefore the geodesics corresponding to the lines  $\mathfrak{L}_1: \eta = \pi\xi$  and  $\mathfrak{L}_2: \eta = \pi\xi + 1/2$  are different. For a given  $\epsilon > 0$  integers  $\mu$  and  $\nu$  can be found such that  $|\mu - \nu\pi - 1/2| < \epsilon$ . Then  $\mathfrak{L}'_1: \eta - \mu = \pi(\xi - \nu)$  or  $\eta = \pi\xi + (\mu - \nu\pi)$  also corresponds to  $G_1$ . Points  $(\xi, \eta_1)$  and  $(\xi, \eta_2)$  on  $\mathfrak{L}'_1$  and  $\mathfrak{L}_2$  respectively have distance not greater than  $|\eta_1 - \eta_2| < \epsilon$ . Put  $\beta = (1 + \pi^2)^{-1/2}$ . Then  $x(\tau) = (\beta\tau, \pi\beta\tau + (\mu - \nu\pi))$  and  $y(\tau) = (\beta\tau, \pi\beta\tau + 1/2)$  are representations of  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$ . But  $\sup x(\tau)y(\tau) < \epsilon$ . By varying  $\epsilon$  we conclude from (6.9) that  $\bar{\sigma}(\mathfrak{G}_1, \mathfrak{G}_2) = 0$ . Since all line elements of  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  are

parallel in  $P$  the geodesics are not transitive with respect to all line elements. If  $G_1, G_2$  denote the carriers of  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$ , then  $\bar{G}_1 = \bar{G}_2 = R^5$ , so that  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  are transitive with respect to the points. Similarly the carrier  $G_3$  of the geodesic  $\mathfrak{G}_3$  corresponding to  $y = -\pi x$  is transitive with respect to the points, or  $\bar{G}_3 = R^5$ . Hence  $\rho_p(G_1, G_3) = 0$ , but obviously  $\sigma(l_1, l_3) > \alpha > 0$  for any two line elements in  $\mathfrak{G}_1$  and  $\mathfrak{G}_3$  respectively so that  $\bar{\sigma}(\mathfrak{G}_1, \mathfrak{G}_3) > 0$ . This shows that  $\rho_p$  and  $\bar{\sigma}$  are not equivalent and that the condition  $G[t] = \bar{G}[t]$  in (8.11) is essential.

### III. RAYS AND CO-RAYS

**9. Absolutely conjugate points.** A real theory of conjugate points cannot be developed without differentiability conditions. The discussion will therefore be restricted to some remarks on absolute conjugacy, which will be needed later on. They are based on the concept of a half extremal.

The curve  $x^+(\tau)$ ,  $0 \leq \tau < \infty$  (or  $x^-(\tau)$ ,  $-\infty < \tau \leq 0$ ), of an  $E$ -space is called a *positive (negative) half extremal with origin  $x$* , if a representation  $x'(\tau)$  of an extremal exists such that  $x'(\tau) = x^+(\tau)$  for  $0 \leq \tau < \infty$  ( $x'(\tau) = x^-(\tau)$  for  $-\infty < \tau \leq 0$ ). Notice that the parametrization  $x^+(\tau)$  (or  $x^-(\tau)$ ) is part of the definition.

The half extremals  $x^+(\tau)$  and  $x^-(\tau)$  will be denoted by  $\mathfrak{x}^+$  and  $\mathfrak{x}^-$  respectively, the subarc  $0 \leq \tau \leq \gamma$  of  $x^+(\tau)$  by  $\mathfrak{x}^+(\gamma)$ , and the subarc  $-\gamma \leq \tau \leq 0$  of  $x^-(\tau)$  by  $\mathfrak{x}^-(\gamma)$ .

The extremal  $\mathfrak{G}(\mathfrak{d})$  which contains a given direction  $\mathfrak{d}$  with origin  $p$  has exactly one representation  $x(\tau)$  such that  $x(\tau)$  represents  $\mathfrak{d}$  for  $0 \leq \tau \leq \eta(p)$ . The half extremals  $x^+(\tau) = x(\tau)$  for  $\tau \geq 0$  and  $x^-(\tau) = x(\tau)$  for  $\tau \leq 0$  are said to have  $\mathfrak{d}$  as initial direction.

The positive and negative half extremals are metrized by means of their initial directions. If  $\mathfrak{x}_1^+$ ,  $\mathfrak{x}_1^-$  and  $\mathfrak{x}_2^+$ ,  $\mathfrak{x}_2^-$  have the initial directions  $\mathfrak{d}_1$  and  $\mathfrak{d}_2$  respectively we put

$$(9.1) \quad \sigma(\mathfrak{x}_1^+, \mathfrak{x}_2^+) = \sigma(\mathfrak{x}_1^-, \mathfrak{x}_2^-) = \sigma(\mathfrak{d}_1, \mathfrak{d}_2).$$

Thus both the positive and the negative half extremals form each a finitely compact space isometric with  $R_d$ .

A consequence of (6.6), (6.8) is

(9.2)  $\sigma(\mathfrak{x}^+, \mathfrak{x}^+) \rightarrow 0$  (or  $\sigma(\mathfrak{x}^-, \mathfrak{x}^-) \rightarrow 0$ ) if and only if  $x^+(\tau)$  (or  $x^-(\tau)$ ) tends uniformly to  $x^+(\tau)$  (or  $x^-(\tau)$ ) in every interval  $0 \leq \tau \leq \alpha < \infty$  ( $-\infty < -\alpha \leq \tau \leq 0$ ).

The numbers  $\gamma(\mathfrak{x}^+)$  and  $\gamma(\mathfrak{x}^-)$  are defined as the least upper bounds of those values  $\tau$  for which  $\mathfrak{x}^+(\tau)$  and  $\mathfrak{x}^-(\tau)$  respectively are segments. If  $\gamma(\mathfrak{x}^+)$  is finite  $\mathfrak{x}^+(\gamma)$  is itself a segment, and so is  $\mathfrak{x}^-(\gamma)$  if  $\gamma(\mathfrak{x}^-)$  is finite. The points  $x^+(\gamma(\mathfrak{x}^+))$  and  $x^-(\gamma(\mathfrak{x}^-))$  are called the *absolutely conjugate points to the origin  $p$  on  $\mathfrak{x}^+$  or  $\mathfrak{x}^-$* . If  $\gamma(\mathfrak{x}^+) = \infty$  then  $x^+(\tau)$  is a segment for every  $\tau$ , hence  $x^+(\tau_1)x^+(\tau_2) = \tau_2 - \tau_1$  for  $0 \leq \tau_1 < \tau_2$ . In that case  $\mathfrak{x}^+$  is called a *positive ray*. If  $\gamma(\mathfrak{x}^-) = \infty$  the half extremal  $\mathfrak{x}^-$  is a *negative ray*. The ray  $\mathfrak{y}^+$  with the representation  $y^+(\tau) = x^+(\tau + \alpha)$ ,  $\alpha \geq 0$ , is called a *sub-ray* of  $\mathfrak{x}^+$ . We see

(9.3)  $\mathfrak{x}^+$  (or  $\mathfrak{x}^-$ ) is a ray if and only if its origin has no conjugate point on  $\mathfrak{x}^+$  (or  $\mathfrak{x}^-$ ).

When a ray exists the space is obviously unbounded. The converse holds too and is contained in

(9.4) Let  $p, \rightarrow p$  and  $p, q, \rightarrow \infty$ . Denote by  $S'_\nu$  ( $S''_\nu$ ) the carriers of an arbitrary segment  $\mathfrak{s}'_\nu$  ( $\mathfrak{s}''_\nu$ ) from  $p_\nu$  to  $q_\nu$  (from  $q_\nu$  to  $p_\nu$ ) and by  $\mathfrak{x}^+_\nu$  ( $\mathfrak{x}^-_\nu$ ) the half extremal with  $\mathfrak{x}^+_\nu(p_\nu, q_\nu) = \mathfrak{s}'_\nu$  ( $\mathfrak{x}^-_\nu(q_\nu, p_\nu) = \mathfrak{s}''_\nu$ ). Then the closed limit  $FlS'_\lambda$  ( $FlS''_\lambda$ ) of any converging subsequence  $\{S'_\lambda\}$  of  $\{S'_\nu\}$  ( $\{S''_\lambda\}$  of  $\{S''_\nu\}$ ) carries a positive ray  $\mathfrak{x}^+$  (negative ray  $\mathfrak{x}^-$ ) and  $\sigma(\mathfrak{x}^+_\lambda, \mathfrak{x}^+) \rightarrow 0$  ( $\sigma(\mathfrak{x}^-_\lambda, \mathfrak{x}^-) \rightarrow 0$ ).

**Proof.** By (2.11), (2.13), (2.18),  $\{S'_\nu\}$  always contains a subsequence  $\{S'_\lambda\}$  which converges to a closed set  $T$ , that is,  $\rho_p(S'_\lambda, T) \rightarrow 0$  and  $FlS'_\lambda = T$ . If  $\tau > 0$  is given and  $\lambda$  is sufficiently large,  $S'_\lambda$  contains a point  $t_\lambda$  with  $p_\lambda t_\lambda = \tau$ . The sequence  $\{t_\lambda\}$  is bounded and converges to a point  $t$ . Otherwise  $\{t_\lambda\}$  would contain two subsequences  $\{t^1_\mu\}$  and  $\{t^2_\mu\}$  which tend to different limits  $t^1$  and  $t^2$ . For a suitable subsequence  $\{t_k\}$  of  $\{t^1_\mu\}$  the center  $c$  of  $\mathfrak{s}'_k$  will converge to a point  $c$ . Then  $c \in Fl \sup S'_\lambda = Fl \inf S'_\lambda$ , hence  $S'_\lambda$  contains a point  $c_\lambda$  which tends to  $c$ , in particular  $c'_\lambda \rightarrow c$ . But then  $(pct^1)$ ,  $(pct^2)$ ,  $ct^1 = ct^2$  and  $t^1 \neq t^2$  which contradicts (4.1). Put  $t = x^+(\tau)$  and  $x^+(0) = p$ . Then  $x^+(\tau_1)x^+(\tau_2) = \tau_2 - \tau_1$  for  $0 \leq \tau_1 < \tau_2$  so that  $x^+(\tau)$  is a ray whose carrier is  $T$ . If  $\mathfrak{x}^+$  is defined as in (9.4) it follows from (9.1) that  $\sigma(\mathfrak{x}^+_\lambda, \mathfrak{x}^+) \rightarrow 0$ .

According to (1.17),  $p, q, \rightarrow \infty$  and  $p, \rightarrow p$  imply  $q, p, \rightarrow \infty$ . Hence  $\{S''_\nu\}$  may be treated in the same way as  $\{S'_\nu\}$ .

(9.3) and (9.4) yield

(9.5) If  $R$  is bounded every half extremal contains an absolutely conjugate point to its origin. When  $R$  is unbounded, every point is origin of at least one positive and one negative ray.

(9.6) The set  $C^+_p$  ( $C^-_p$ ) of all absolutely conjugate points to  $p$  on the different positive (negative) half extremals with origin  $p$  consists of the points  $y \neq p$  for which no point  $z$  with  $(pyz)$  (or  $(zyp)$ ) exists.

Let  $y = x^+(\gamma(\mathfrak{x}^+))$  be the conjugate point to  $p = x^+(0)$  on the positive half extremal  $x^+(\tau)$  and assume for an indirect proof that a point  $z$  with  $(pyz)$  exists. Then the segments  $\mathfrak{s}(p, y)$  and  $\mathfrak{s}(y, z)$  are unique and constitute a segment  $\mathfrak{s}(p, z)$ . Let  $\mathfrak{y}^+$  be the half extremal with  $\mathfrak{y}^+(pz) = \mathfrak{s}(p, y) \cup \mathfrak{s}(y, z)$ . Then  $\mathfrak{y}^+(py) = \mathfrak{x}^+(py)$  hence  $y^+(\tau) = x^+(\tau)$  for all  $\tau \geq 0$ , and  $\gamma(\mathfrak{x}^+)$  would not be the last value for which  $x^+(\tau)$  is a segment.

Conversely let for the point  $y \neq p$  no  $z$  with  $(pyz)$  exist and designate by  $\mathfrak{x}^+$  a half extremal with  $\mathfrak{x}^+(py) = \mathfrak{s}(p, y)$ . Then  $y$  is clearly absolutely conjugate to  $p$  on  $\mathfrak{x}^+$ .

(9.6) shows that in symmetric spaces, in particular in  $G$ -spaces,  $C^+_p = C^-_p$ . Absolute conjugacy is in general not symmetric, that is,  $q$  may belong to  $C^+_p$  and  $C^-_p$  but  $p$  neither to  $C^+_q$  nor to  $C^-_q$ .

**Example.**  $R^6$ . In  $E^3$  with rectangular coordinates  $\xi, \eta, \zeta$  consider the surface  $R^6$  defined by the equations

$$\xi^2 + \eta^2 = 1 \quad \text{for } \zeta \geq 0 \quad \text{and} \quad \xi^2 + \eta^2 + \zeta^2 = 1 \quad \text{for } \zeta < 0,$$

with the ordinary geodesic distance as distance. Then  $R^6$  is a  $G$ -space. Every half geodesic with origin  $p = (0, 0, -1)$  is a ray, hence  $C_p^+ = C_p^-$  is empty, but  $p$  is conjugate to every point  $q = (\xi, \eta, \zeta)$  with  $\zeta < 0$ .

From (9.6) and (4.2) the following sufficient condition for symmetry is obtained:

(9.7)  $q \in C_p^+$  and  $p \in C_q^-$  if two segments from  $p$  to  $q$  exist.

The set  $C_p^+$  is in general a "pathological" set, even in smooth Riemann spaces. This may be seen from the fact that  $R$  is the closure of  $R - C_p^+$ . (Every point  $z$  of  $C_p^+$  can be connected to  $p$  by a segment; if  $(pyz)$  then  $y \in R - C_p^+$ .) All that can be stated in general regarding  $\gamma(\mathfrak{x}^+)$  is the obvious inequality

$$(9.8) \quad \gamma(\mathfrak{x}^+) \geq \eta_2(x^+(0)) \geq \eta(x^+(0))$$

and that  $\gamma(\mathfrak{x}^+)$  is an upper semicontinuous function of  $\mathfrak{x}^+$ .

For if  $\mathfrak{x}^+$  is given and  $\sigma(\mathfrak{x}_r^+, \mathfrak{x}^+) \rightarrow 0$ , then  $\mathfrak{x}_r^+(\gamma(\mathfrak{x}_r^+))$  is a segment, hence  $\mathfrak{x}^+(\limsup \gamma(\mathfrak{x}_r^+))$  is a segment, so that  $\gamma(\mathfrak{x}^+) \geq \limsup \gamma(\mathfrak{x}_r^+)$ .

This lack of properties of  $C_p^+$  and  $\gamma(\mathfrak{x}^+)$  is the reason why absolute conjugacy is inaccessible to the analytic methods of the calculus of variations, and therefore not studied, although it is geometrically significant. It is also the reason why convergence of extremals cannot conveniently be based on the distance of the sets  $\mathfrak{E}(\mathfrak{s}_1)$  and  $\mathfrak{E}(\mathfrak{s}_2)$  in the space  $R_s$  of all segments.

The case of a  $G$ -space in which  $C_p^+$  consists of only one point  $p'$  is important because every point of a sphere has this property.

The implications of  $C_p^+ = p'$  for a single point  $p$  are these:

(9.9) *If in a  $G$ -space  $C_p^+ = p'$ , then  $(pxp')$  for every  $x \neq p, p'$  (and conversely). Hence  $C_{p'}^+ = p$  and the space is compact. Each geodesic  $\mathfrak{G}$  through  $p$  or  $p'$  consists of segments  $t(p, p')$  and is either simple, or both  $p$  and  $p'$  are multiple points of  $\mathfrak{G}$ .*

**Proof.** The union  $F$  of all rays with origin  $p$  is a proper subset of  $R$  because  $p'$  does not lie in  $F$ . By (9.4),  $F$  is a closed set. Let  $x \in R - F - p'$ . A half extremal  $\mathfrak{x}^+$  with  $\mathfrak{x}^+(px) = \mathfrak{s}(p, x)$  is no ray because  $x \in R - F$ . The relations (9.3) and  $C_p^+ = p'$  yield that  $p'$  is the conjugate point to  $p$  on  $\mathfrak{x}^+$ , hence  $(pxp')$ . The union  $F'$  of all segments  $t(p, p')$  is closed and  $R = F' \cup F$ . If  $F$  were not empty then  $F' \cap F = p$  (by (7.4)). The proof of (8.8) shows that every point  $p$  of an at least two-dimensional  $G$ -space is interior point of a set homeomorphic to a euclidean triangle. Hence  $F' \cap F = p$  shows that  $R$  is one-dimensional. But then  $F \neq 0$  and  $F' \neq 0$  contradict (8.8). This shows  $R = F'$  or  $(pxp')$  for  $x \neq p, p'$ .

If conversely  $(pxp')$  for  $x \neq p, p'$ , then (9.6) shows that  $x$  is not conjugate to  $p$  and  $p'$  is conjugate to  $p$ , hence  $C_p^+ = p'$ . Any half extremal  $\mathfrak{x}^+$  with origin  $p$  starts with a segment  $t_1$  from  $p$  to  $p'$  and, because  $C_{p'}^+ = p$ , it continues with a segment  $t_2$  from  $p'$  to  $p$ , then it continues with a segment  $t_3$  from  $p$  to  $p'$  and so forth. If  $t_3 = t_1$  then the geodesic which carries  $\mathfrak{x}^+$  is simple. If  $t_3 \neq t_1$  then it follows from  $t_3 \neq t_2$  that both  $p$  and  $p'$  are multiple points of  $G$ .

A  $G$ -space is called *sphere like* if all its geodesics are great circles and if all geodesics which pass through a given point  $p$  pass through a fixed point  $p' \neq p$ , the antipodal point to  $p$ . It can easily be proved that a sphere like space  $R$  shares the following properties with an ordinary sphere (compare the statement (8) of B [2] and (14.3) of the present paper):  $pp'$  is constant and  $2pp'$  the common length of all geodesics.  $p'$  is the absolutely conjugate point to  $p$  on every half extremal with origin  $p$ , hence  $C_p^+ = p'$ . The mapping  $p \rightarrow p'$  of  $R$  on itself is an isometry. The geodesic through  $p$  and  $q$  is unique when  $q \neq p, p'$ .

(9.9) yields the following characterization of sphere like spaces:

(9.10) THEOREM. *A  $G$ -space  $R$  is sphere like if and only if every point  $p$  has exactly one absolutely conjugate point  $p'$ .*

The necessity was mentioned before. Let therefore  $C_p^+$  consist of one point  $p'$  for every  $p \in R$ . According to (9.9) an arbitrary geodesic through  $p$  passes also through  $p'$ . Therefore it is only to be shown that every geodesic  $\mathfrak{G}$  is a great circle. By (8.5),  $\mathfrak{G}$  has a simple point  $q$ . Since  $q' = C_q^+$  it follows from (9.9) that  $\mathfrak{G}$  is simple. If  $a, b$  are any two points of  $\mathfrak{G}$ , then  $G$  consists by (9.9) of two segments  $t(a, a')$ . One of them must contain  $b$ . Hence  $\mathfrak{G}$  contains a segment  $t(a, b)$ , that is, is straight and by (8.7) a great circle.

10. **The function  $\alpha(r, p)$ .** Parallelism between rays in unbounded  $E$ - and  $G$ -spaces will be the next subject. Surprisingly many results can be obtained under these general conditions, which are also new in the differentiable case. The main tool is a certain metric function which will be discussed in the present section.

Let  $r(\tau)$  represent a positive ray  $r$ , that is,  $r(\tau_1)r(\tau_2) = \tau_2 - \tau_1$  for  $\tau_2 \geq \tau_1 \geq 0$ . Any point  $p$  of the space and any three values  $0 \leq \tau_1 < \tau_2 < \tau_3$  satisfy the relation

$$(10.1) \quad \begin{aligned} r(\tau_1)p &\geq \tau_3 - \tau_1 \\ &- pr(\tau_3) \left\{ \begin{array}{l} > \tau_2 - \tau_1 - pr(\tau_2) \text{ unless} \\ = \tau_2 - \tau_1 - pr(\tau_2) \text{ if} \end{array} \right\} (pr(\tau_2)r(\tau_3)) \quad \text{or} \quad p = r(\tau_2). \end{aligned}$$

For

$$r(\tau_1)p \geq r(\tau_1)r(\tau_3) - pr(\tau_3) = \tau_2 - \tau_1 + r(\tau_2)r(\tau_3) - pr(\tau_3) \geq \tau_2 - \tau_1 - pr(\tau_2),$$

since  $pr(\tau_2) \geq pr(\tau_3) - r(\tau_2)r(\tau_3)$ ; the equality sign holds only for  $(pr(\tau_2)r(\tau_3))$  or  $p = r(\tau_2)$ .

Applying (10.1) with  $\tau_1 = 0$  it is seen that  $\tau - pr(\tau)$  is a bounded and non-decreasing function of  $\tau$  so that  $\lim_{\tau \rightarrow \infty} (\tau - pr(\tau))$  and

$$(10.2) \quad \alpha(p, r) = \lim_{\tau \rightarrow \infty} (\tau - pr(\tau))$$

exist and are finite. Obviously



$$(10.3) \quad \alpha(r(\tau_0), r) = -\tau_0 \text{ for every } r \geq 0,$$

$$(10.4) \quad \alpha(p, r) - \alpha(q, r) = \lim_{r \rightarrow \infty} (pr(r) - qr(r)) \leq pq$$

hence

$$|\alpha(p, r) - \alpha(q, r)| \leq \sigma(p, q),$$

so that  $\alpha(p, r)$  is a continuous function of  $p$ .

The variation of  $\alpha(p, r)$  as a function of  $r$  depends on the origin and the "point at infinity" of  $r$ . Only the latter will prove important. To obtain a precise meaning, we define "distance at infinity" of any two unbounded sets  $U, V$  as the number

$$(10.5) \quad \sigma_\infty(U, V) = \lim_{\eta \rightarrow \infty} \sigma(U - U \cap S(p, \eta), V - V \cap S(p, \eta))$$

where  $p$  is an arbitrary point. The limit (10.4) exists but may be infinite, because the distance after the limit sign is an increasing function of  $\eta$ . Moreover  $\sigma_\infty(U, V)$  does not depend on  $p$  because  $S(p, \eta) \subset S(q, \eta + pq)$  for any  $p, q$  hence

$$\begin{aligned} \sigma(U - U \cap S(p, \eta), V - V \cap S(p, \eta)) \\ \leq \sigma(U - U \cap S(q, \eta + pq), V - V \cap S(q, \eta + pq)). \end{aligned}$$

If  $\sigma_\infty(r_1, r_2)$  designates the distance at infinity of the carriers of  $r_1$  and  $r_2$  (or of "the points at infinity" of  $r_1$  and  $r_2$ ) the following relation will hold for any two points  $p, q$  (the second point serves to eliminate the influence of the origins, as may be seen in the case where  $r_1$  is a subray of  $r_2$ ):

$$(10.6) \quad |\alpha(p, r_1) + \alpha(q, r_2) - \alpha(q, r_1) - \alpha(p, r_2)| \leq 2\sigma_\infty(r_1, r_2).$$

Namely there are sequences  $r_\nu^1 \rightarrow \infty$  and  $r_\nu^2 \rightarrow \infty$  such that  $\sigma(r_1(r_\nu^1), r_2(r_\nu^2)) \rightarrow \sigma_\infty(r_1, r_2)$ . By (10.4) the left side of (10.6) is the limit of

$$|[pr(r_\nu^1) - pr(r_\nu^2)] - [qr(r_\nu^1) - qr(r_\nu^2)]| \leq 2\sigma(r_1(r_\nu^1), r_2(r_\nu^2)).$$

A consequence of (10.6) is

$$(10.7) \text{ If } \sigma_\infty(r_1, r_2) = 0, \text{ then } \alpha(x, r_1) - \alpha(x, r_2) \text{ is constant.}$$

The loci  $\alpha(x, r) = \text{const.}$  are called *limit spheres with the central ray*  $r$ . Since  $\alpha(x, r)$  is defined for every  $x$  and single-valued, there is exactly one limit sphere through a given point  $p$  with a given (positive) central ray  $r$ . Its equation is

$$(10.8) \quad \alpha(x, r) = \alpha(p, r)$$

and it will be denoted by  $L''(r, p)$ . Since  $\alpha(x, r)$  is continuous in  $x$ , limit spheres are closed sets.

The name limit sphere derives from the fact that  $L''(r, p)$  is in hyperbolic geometry (from which the word limitsphere or horosphere is taken) always

the limit of spheres. Also in the present general case  $L''(r, p)$  is closely related to limits of spheres. To see this denote by  $K''(r, pr)$  or  $K''(r, p)$  the locus of those points  $x$  for which  $xr = pr$ . A first relation between  $K''(r, p)$  and  $L''(r, p)$  is

(10.9) If  $p_r \rightarrow p$  and  $\tau_r \rightarrow \infty$ , then  $Fl \sup K''(r(\tau_r), p_r) \subset L''(r, p)$ .

For a subsequence  $\{\lambda\}$  of  $\{\nu\}$  let  $y_\lambda \in K(r(\tau_\lambda), p_\lambda)$  and  $y_\lambda \rightarrow y$ . Then by (10.4)

$$\begin{aligned} |\alpha(y, r) - \alpha(p, r)| &= \lim_{\tau \rightarrow \infty} |yr(\tau) - pr(\tau)| \\ &\leq \lim |p_\lambda r(\tau_r) - y_\lambda r(\tau_\lambda)| + \lim p p_\lambda + \lim y y_\lambda = 0, \end{aligned}$$

hence  $y \in L''(r, p)$  according to (10.8), which proves (10.9).

The interior of  $K''(r, p)$  is by definition the set  $S''(r, pr)$  defined in §1. It consists of those points  $x$  for which  $xr < pr$ . Correspondingly the sets of points  $x$  for which  $\alpha(x, r) < \alpha(p, r)$  and  $\alpha(x, r) > \alpha(p, r)$  are called the *interior* and *exterior* of  $L''(r, p)$  respectively.

(10.10) If  $\tau_1 < \tau_2$  then  $K''(r(\tau_2), r(\tau_1))$  and its interior lie, except for  $r(\tau_1)$ , in the interior of  $L''(r, r(\tau_1))$ .

For (10.1) and (10.4) show that for  $\tau > \tau_2$

$$\begin{aligned} (10.11) \quad \alpha(r(\tau_1), r) - \alpha(x, r) &= \lim (r(\tau_1)r(\tau) - xr(\tau)) \geq \tau - \tau_1 - xr(\tau) \\ &\geq \tau_2 - \tau_1 - xr(\tau_2). \end{aligned}$$

If  $x$  is in the interior of  $K''(r(\tau_2), r(\tau_1))$  then  $\tau_2 - \tau_1 > xr(\tau_2)$ , hence  $\alpha(r(\tau_1), r) > \alpha(x, r)$ . Let  $x \in K''(r(\tau_2), r(\tau_1))$  or  $\tau_2 - \tau_1 = xr(\tau_2)$ . Because of (10.1) the equality sign holds in (10.11) only for  $(xr(\tau_2)r(\tau))$  or  $x = r(\tau_2)$ ; in this case (4.1) yields  $x = r(\tau_1)$ , which proves (10.10).

If  $L''(r, p)$  contains a point of  $r$ , the fact of hyperbolic geometry mentioned above holds:

$$(10.12) \quad Fl K''(r(\tau_r), r(\tau)) = L''(r, r(\tau))$$

for any point  $r(\tau)$  of  $r$  and any sequence  $\tau_r \rightarrow \infty$ .

(10.9) reduces (10.12) to  $Fl \inf K''(r(\tau_r), r(\tau)) \supset L''(r, r(\tau))$ . Let  $x \in L''(r, r(\tau))$ . Then (10.10) yields  $xr(\tau_r) \geq \tau_r - \tau$  for  $\tau_r > \tau$ . Hence a segment  $\mathfrak{s}(x, r(\tau_r))$  contains a point  $x_r$  with  $x_r r(\tau_r) = \tau_r - \tau$  or  $x_r \in K(r(\tau_r), r(\tau))$ . Now  $\alpha(x, r) = \alpha(r(\tau), r)$  (compare (10.8)) and (10.3) imply  $xr(\tau_r) - \tau_r \rightarrow \tau$  hence  $xx_r = xr(\tau_r) - x_r r(\tau_r) \rightarrow 0$ , and  $x \in Fl \inf K''(r(\tau_r), r(\tau))$ .

For the study of parallelism between negative rays  $\bar{r}$  the function

$$\lim_{\tau \rightarrow -\infty} (x^-(\tau)p + \tau)$$

must be introduced. The developments are entirely parallel and will therefore be omitted.

11. Co-rays. As before let  $r(\tau)$ ,  $0 \leq \tau < \infty$ , represent a positive ray  $r$ . A

positive ray  $q$  with origin  $p$  is a co-ray from  $p$  to  $r$  if its carrier is the closed limit of the carriers of a sequence of segments  $\mathfrak{s}(p_\nu, r(\tau_\nu))$  with  $p_\nu \rightarrow p$  and  $\tau_\nu \rightarrow \infty$ . (9.4) implies that each sequence of segments  $\mathfrak{s}(p_\mu, r(\tau_\mu))$  with  $p_\mu \rightarrow p$  and  $\tau_\mu \rightarrow \infty$  contains a subsequence which tends to a positive ray with origin  $p$ , so that every point  $p$  is origin of at least one co-ray to a given ray  $r$ . (9.4) shows also that the positive rays form a closed subset of the set of all positive half extremals.

(11.1) The co-rays to a given positive ray  $r$  form a closed subset in the set of all positive half extremals.

For let  $q_\nu$  be a co-ray to  $r$  and  $\sigma(q_\nu, q) \rightarrow 0$ . Then the origin  $q_\nu$  of  $q_\nu$  tends to the origin  $q$  of  $q$ . Since  $q_\nu$  is a co-ray to  $r$ , it follows from (9.6) that a positive half extremal  $\mathfrak{x}_\nu^+$  and a value  $\tau_\nu > \nu$  exist with  $\sigma(q_\nu, \mathfrak{x}_\nu^+(0)) < 1/\nu$ ,  $\mathfrak{x}_\nu^+(x_\nu^+(0)r(\tau_\nu)) = \mathfrak{s}(x_\nu^+(0), r(\tau_\nu))$  and  $\sigma(\mathfrak{x}_\nu^+, q_\nu) < 1/\nu$ . Then  $x_\nu^+(0) \rightarrow q$  and  $\sigma(\mathfrak{x}_\nu^+, q) \rightarrow 0$ , whence it easily follows that  $q = Fl_{\mathfrak{x}_\nu^+}(x_\nu^+(0)r(\tau_\nu))$ .

The relation of the co-rays to  $r$  with the function  $\alpha(x, r)$  is established by

(11.2) THEOREM. *The half extremal  $\mathfrak{x}^+$  is a co-ray to  $r$  from  $x^+(0)$  if and only if*

$$(11.3) \quad \alpha(x^+(\tau'), r) - \alpha(x^+(\tau''), r) = \tau' - \tau'' \quad \text{for } \tau', \tau'' \geq 0.$$

Only the necessity will be proved now, the sufficiency will follow later from (11.6). Let  $\mathfrak{x}^+$  be a co-ray to  $r$  and, without restriction,  $\tau' > \tau''$ . By (9.6) there is a sequence of half extremals  $\{\mathfrak{x}_\nu^+\}$  and a sequence  $\tau_\nu \rightarrow \infty$  such that  $\mathfrak{x}_\nu^+(x_\nu^+(0)r(\tau_\nu))$  is a segment from  $x_\nu^+(0)$  to  $r(\tau_\nu)$  and  $\sigma(\mathfrak{x}_\nu^+, \mathfrak{x}^+) \rightarrow 0$ . Because of (9.2),  $x_\nu^+(\tau') \rightarrow x^+(\tau')$  and  $x_\nu^+(\tau'') \rightarrow x^+(\tau'')$ . For large  $\nu$  (since  $x_\nu^+(0) \rightarrow x^+(0)$ ),  $x_\nu^+(0)r(\tau_\nu) > \tau'$ , hence  $(x_\nu(\tau'')x_\nu(\tau')r(\tau_\nu))$  and

$$\begin{aligned} \sigma(x_\nu^+(\tau'), x^+(\tau')) + \sigma(x_\nu^+(\tau''), x^+(\tau'')) \\ \geq |x^+(\tau'')r(\tau_\nu) - x^+(\tau')r(\tau_\nu) - [x_\nu^+(\tau'')r(\tau_\nu) - x_\nu^+(\tau')r(\tau_\nu)]| \\ = |x^+(\tau'')r(\tau_\nu) - x^+(\tau')r(\tau_\nu) - (\tau' - \tau'')| \rightarrow 0. \end{aligned}$$

The relation (11.3) follows now from (10.4).

Next observe that

$$(11.4) \quad \alpha(p, r) - \alpha(q, r) = pq \text{ if and only if } q \text{ is a foot of } p \text{ on } L''(r, q).$$

For the definition of "foot" see §7. If  $x \in L''(r, q)$ , then  $\alpha(x, r) = \alpha(q, r)$ . If  $\alpha(p, r) - \alpha(q, r) = pq$  it follows from (10.4) that  $px \geq \alpha(p, r) - \alpha(q, r) = pq$  so that  $q$  is a foot of  $p$  on  $L''(r, q)$ .

Conversely let  $q$  be a foot of  $p$  on  $L''(q, r)$  and consider on a co-ray  $p$  from  $p$  to  $r$  the point  $p(\tau_0)$  with  $\tau_0 = \alpha(p, r) - \alpha(q, r)$ . By (11.3),  $\alpha(p, r) - \alpha(p(\tau_0), r) = \tau_0$  hence  $\alpha(p(\tau_0), r) = \alpha(q, r)$  or  $p(\tau_0) \in L''(q, r)$ . Moreover  $\tau_0 = pp(\tau_0) \geq pq$  because  $q$  is a foot of  $p$ . The definition of  $\tau_0$  and (10.4) show that  $pq = \tau_0$ , q.e.d.

A consequence of (11.3), (11.4) and (7.8) is

(11.5) If  $p(\tau)$  represents a co-ray from  $p(0)$  to  $r$ , and  $0 < \tau_1 < \tau_2$ , then  $p(\tau_2)$  is a foot of  $p(0)$ , and the only foot of  $p(\tau_1)$  on  $L''(r, p(\tau_2))$ .

The intuitive idea of a co-ray from  $p$  to  $r$  is that of an infinite segment which connects  $p$  to the point at infinity of  $r$ . If  $r$  is a point at finite distance,  $\mathfrak{s}(p, r)$  is not necessarily unique; the Example  $R^6$  in §9 shows the same for co-rays. If  $q$  is an interior point of  $\mathfrak{s}(p, r)$ , then  $\mathfrak{s}(q, r)$  is unique and coincides with a subsegment of  $\mathfrak{s}(p, r)$ . (11.5) implies the remarkable fact that this result extends to co-rays.

(11.6) THEOREM. *If  $p$  is a co-ray from  $p$  to  $r$  and  $q$  is any point of  $p$  different from  $p$ , then the co-ray from  $q$  to  $r$  is unique and coincides with the sub-ray of  $p$  with origin  $q$ .*

For let  $p(\tau)$  represent  $p$  and  $q = p(\tau_1)$ ; let  $q(\tau)$  represent a co-ray  $q$  from  $q$  to  $r$ . For  $\tau_2 > \tau_1$ , (11.3) yields

$$\alpha(p(\tau_1), r) - \alpha(p(\tau_2), r) = \tau_2 - \tau_1 = \alpha(q(0), r) - \alpha(q(\tau_2 - \tau_1), r),$$

hence  $q(\tau_2 - \tau_1) \in L''(r, p(\tau_2))$ . By (11.5),  $q(\tau_2 - \tau_1)$  is a foot and  $p(\tau_2)$  the only foot of  $p(\tau_1) = q(0)$  on  $L''(r, p(\tau_2))$ , so that  $q(\tau_2 - \tau_1) = p(\tau_2)$ , which shows that  $q$  is a sub-ray of  $p$ .

It is now easy to prove that (11.4) is sufficient for  $x^+(\tau)$  to be a co-ray to  $r$ . For,  $\mathfrak{x}^+$  is a ray because (5.1) and (10.4) yield for  $\tau' > \tau''$

$$\tau' - \tau'' \geq x^+(\tau'')x^+(\tau') \geq \alpha(x^+(\tau''), r) - \alpha(x^+(\tau'), r) = \tau' - \tau''.$$

By (11.4) and (7.9),  $x^+(\eta + \tau)$  is for positive  $\eta$  and  $\tau$  the only foot of  $x^+(\eta)$  on  $L''(r, x^+(\eta + \tau))$ . Therefore (11.5) shows that the ray  $x_q^+(\tau) = x^+(\eta + \tau)$ ,  $0 \leq \tau < \infty$ , is a co-ray from  $x^+(\eta)$  to  $r$ . Since  $\sigma(\mathfrak{x}_q^+, \mathfrak{x}^+) \rightarrow 0$  for  $\eta \rightarrow 0$ , it follows from (11.1) that  $\mathfrak{x}^+$  is a co-ray to  $r$ .

Let  $q$  be any co-ray to  $r$ . The union of the carriers of all co-rays to  $r$  which contain  $q$  as sub-ray carries by (11.1) either a co-ray to  $r$  or a directed straight line (which in case of ambiguity, as in the euclidean geometry, is to be oriented such that  $q(\tau) = x(\tau)$  for  $\tau \geq 0$  and a suitable representation  $x(\tau)$  of the directed line). This ray or directed straight line is an *asymptote*  $a(q)$  to  $r$ . Every positive subray of  $a(q)$  is a co-ray to  $r$ . (11.6) yields

(11.7) *Every point of the space is either origin of at least one, or interior point of exactly one, asymptote to  $r$ .*

The points  $p$  and  $q$  lie on one asymptote to  $r$ , if and only if either  $p$  lies on a co-ray from  $q$  to  $r$ , or  $q$  on a co-ray from  $p$  to  $r$ . Therefore (11.3) yields:

(11.8)  *$p$  and  $q$  lie on one asymptote to  $r$  if and only if  $\alpha(p, r) - \alpha(q, r) = pq$  or  $\alpha(q, r) - \alpha(p, r) = qp$ .*

It should be noticed that, unless  $q = r$ , the carrier of  $a(q)$  is in general not identical with the union of the carriers of all rays that contain  $q$  as sub-ray. In an important case, namely when  $\eta_\lambda(x) = \infty$  (compare (5.8)), it can be proved that the two sets coincide:

(11.9) THEOREM. *If all extremals of  $R$  are directed straight lines, then all*

*asymptotes to a given ray  $r$  are straight lines. Any two asymptotes to  $r$  are either disjoint or identical.*

**Proof.** Let  $q$  be a co-ray from  $q$  to  $r$ . The representation  $q(\tau)$  of  $q$  may be extended to negative  $\tau$ , such that  $q(\tau)$  represents a directed straight line  $q^*$ . Since  $q$  is a co-ray to  $r$ , there is a sequence  $\tau_\nu \rightarrow \infty$  and a sequence  $q_\nu \rightarrow q$  such that the half extremal  $x_\nu^+$  with  $x_\nu^+(q_\nu r(\tau_\nu)) = s(q_\nu, r(\tau_\nu))$  (this segment is unique because  $\eta_\lambda(x) = \infty$ ) tends to  $q$ . The extremal  $x_\nu(\tau)$  with  $x_\nu(\tau) = x_\nu^+(\tau)$  for  $\tau \geq 0$  is by hypothesis a straight line. This implies for fixed  $\tau' > \tau''$  and large  $\nu$  that  $\tau' - \tau'' = x_\nu(\tau'')r(\tau_\nu) - x_\nu(\tau')r(\tau_\nu)$  hence

$$\sigma(q(\tau'), x_\nu(\tau')) + \sigma(q(\tau''), x_\nu(\tau'')) \geq |\tau' - \tau'' - [q(\tau'')r(\tau_\nu) - q(\tau')r(\tau_\nu)]|.$$

By (9.2) and (6.8),  $x_\nu(\tau') \rightarrow q(\tau')$  and  $x_\nu(\tau'') \rightarrow q(\tau'')$ , hence by (10.4)

$$\alpha(q(\tau''), r) - \alpha(q(\tau'), r) = \tau' - \tau''.$$

(11.2) yields now that the sub-ray  $q_{\tau_0}(\tau) = q(\tau - \tau_0)$ ,  $\tau \geq 0$ , of  $q^*$  is for every  $\tau_0$  a co-ray to  $r$ , hence  $q^*$  is an asymptote to  $r$ . The second part of (11.9) follows now from (11.7).

The exterior of a limit sphere may be empty as example  $R^6$  shows. But (11.10). *The interior of a limit sphere  $L''(r, p)$  is not empty and is connected.* Observe first that

$$(11.11) \quad S''(r, rp) \supset S''(r', rp) \quad \text{if} \quad (pr'r)$$

because for  $x \in S''(r', r'p)$  the inequality  $xr < xr' + r'r < pr' + r'p = pr$  holds. If  $L''(r, p) = L''(r, r(\tau_0))$  then (10.10) shows that the interior of  $L''(r, p)$  equals  $\bigcup_\nu S''(r(\tau_0 + \nu), \nu)$ , and is therefore not empty and by (11.11) connected. If no  $\tau_0 \geq 0$  exists such that  $L''(r, p) = L''(r, r(\tau_0))$  then  $\alpha(p, r) > 0$  (otherwise (10.3) with  $\tau_0 = -\alpha(p, r)$  would yield  $\alpha(r(\tau_0), r) = \alpha(p, r)$  or  $p \in L''(r, r(\tau_0))$ ). Hence  $L''(r, r(0))$  and its interior lie in the interior of  $L''(r, p)$ , and it suffices to show that every point  $q$  in the interior of  $L''(r, p)$  but not in the interior of  $L''(r, r(0))$  can be connected to a point in the interior of  $L''(r, r(0))$  by a segment which lies in the interior of  $L''(r, p)$ .

On a co-ray  $q(\tau)$  from  $q$  to  $r$ , consider the point  $q(\tau_1)$  with  $\tau_1 = \alpha(q, r) + 1$ . By the definition of the interior of  $L''(r, p)$  and (11.2)

$$\alpha(p, r) > \alpha(q, r) \geq \alpha(q(\tau), r)$$

and  $\alpha(q, r) - \alpha(q(\tau_1), r) = \tau_1 = \alpha(q, r) + 1$ , or  $\alpha(q(\tau_1), r) = -1$  so that  $q(\tau_1)$  is in the interior of  $L''(r, r(0))$  and the segment  $0 \leq \tau \leq \tau_1$  of  $q(\tau)$  lies in the interior of  $L''(r, p)$ .

The next question which arises after (11.6) is whether the relation between ray and co-ray is symmetric, that is, whether  $r$  is a co-ray to  $q$  when  $q$  is a co-ray to  $r$ . B [1, chapter I, 3] contains an example which shows that even in a two-dimensional  $G$ -space in which all geodesics are straight lines,

the relation need not be symmetric. This means that a ray  $r$  and a co-ray  $q$  to  $r$  may not be looked at as "having the same point at infinity," or that points at infinity without definite rays as carriers have in general no meaning. There do not seem to be any simple, necessary *and* sufficient conditions for symmetry in the present general case. The following is a sufficient condition which is also necessary when the space is a Finsler space with the usual differentiability properties (compare the end of this section).

(11.12) *If  $\alpha(x, r) - \alpha(x, q)$  is constant, then every co-ray to one of the two rays  $q, r$  is a co-ray to the other.* (Since  $q$  is a co-ray to itself this implies that  $q$  is a co-ray to  $r$ , and  $r$  a co-ray to  $q$ .)

The hypothesis yields  $\alpha(x, r) - \alpha(y, r) = \alpha(x, q) - \alpha(y, q)$  for any two points  $x, y$ . If  $x^+(\tau)$  is a co-ray to  $r$ , then by (11.2)

$$\tau' - \tau'' = \alpha(x^+(\tau''), r) - \alpha(x^+(\tau'), r) = \alpha(x^+(\tau''), q) - \alpha(x^+(\tau'), q)$$

so that by the other part of (11.2),  $x^+(\tau)$  is a co-ray to  $q$ , which proves (11.12). Notice the following corollary of (11.12) and (10.7) which partly justifies the idea of points at infinity.

(11.13) *If  $\sigma_\infty(r, q) = 0$ , then every co-ray to one of the two rays  $q, r$  is a co-ray to the other.*

An interesting application of (11.6) and (11.13) is

(11.14) THEOREM. *If the origin  $q$  of the ray  $q$  is an interior point of the ray  $r$  and  $\sigma_\infty(q, r) = 0$ , then  $q$  is a sub-ray of  $r$ .*

*If the two straight lines  $\mathfrak{G}, \mathfrak{H}$  of a  $G$ -space intersect and  $\sigma_\infty(\mathfrak{G}, \mathfrak{H}) = 0$ , then  $\mathfrak{G} = \mathfrak{H}$ .*

The euclidean metric shows that  $\sigma_\infty(r, q) = 0$  is not necessary for  $r$  and  $q$  to be co-rays to each other. Neither is (11.12) necessary without differentiability conditions because it implies  $L''(r, x) = L''(q, x)$  for all  $x$ . Any Minkowski metric  $R^2$  (see §3) in which  $K$  is not everywhere differentiable contains rays  $q, r$  which are co-rays to each other, and for which  $L''(r, x)$  never coincides with  $L''(q, x)$  (see B [1, p. 39, statement 21]).

In the euclidean and hyperbolic geometries, and more generally in every simply connected Riemann space of nonpositive curvature, the condition  $\sigma_\infty(r, q) < \infty$  is necessary and sufficient for  $r$  and  $q$  to be co-rays to each other (compare Cartan [1, Note III, §16, pp. 263–265]). *In general  $G$ -spaces it is neither*, even if all geodesics are straight lines, nor is the following stronger condition sufficient:

(11.15) The distances  $\sigma(x, q)$  and  $\sigma(y, r)$  are uniformly bounded for all  $x$  on  $r$  and  $y$  on  $q$ .

The following example  $R^7$  will confirm these statements.  $R^7$  will be a two-dimensional  $G$ -space with  $\eta_\lambda(x) = \infty$ , and will contain straight lines  $\mathfrak{G}$  and  $\mathfrak{H}$  with this strange property:

$\mathfrak{G}$  and  $\mathfrak{H}$  contain rays  $q_1$  and  $r_1$  respectively which are co-rays to each

other, but  $\sigma_\infty(q_1, r_1) = \infty$ . The rays  $q_2$  on  $\mathfrak{G}$  and  $r_2$  on  $\mathfrak{H}$  opposite to  $q_1$  and  $r_1$  respectively satisfy (11.15) but neither is a co-ray to the other.

$R^7$  is the subset  $\xi > 0$ ,  $0 < \eta < 1$ , of the cartesian  $(\xi, \eta)$ -plane. With  $\psi(\eta) = -\eta^{-1} + (1 - \eta)^{-1}$  the distance of the two points  $a_1 = (\xi_1, \eta_1)$  and  $a_2 = (\xi_2, \eta_2)$  is defined by

$$a_1 a_2 = [(\xi_1 - \xi_2)^2 + (\eta_1 - \eta_2)^2]^{1/2} + |\xi_1^{-1} - \xi_2^{-1}| + |\psi(\eta_1) - \psi(\eta_2)|.$$

Since the last two terms depend, respectively, on  $\xi$  and  $\eta$  only, it is immediately seen that the geodesics of  $R^7$  are the intersections of the euclidean straight lines of the  $(\xi, \eta)$ -plane with the domain  $R^7$  and straight lines in the sense of  $G$ -spaces.

Consider any two geodesics  $\mathfrak{G}$ ,  $\mathfrak{H}$  of the form  $\mathfrak{G}: \eta = \alpha\xi$ ,  $\mathfrak{H}: \eta = \beta\xi$ ,  $0 < \eta < 1$ ,  $0 < \alpha < \beta$ . Call  $q_1$  and  $r_1$  the sub-rays  $0 < \eta \leq 1/2$  of  $\mathfrak{G}$  and  $\mathfrak{H}$ , and  $q_2$ ,  $r_2$  the opposite rays. Then  $q_1$  and  $r_1$  are co-rays to each other with origins  $q = (\alpha/2, 1/2)$ ,  $r = (\beta/2, 1/2)$ . Let  $q_\nu = (\xi_{1\nu}, \eta_{1\nu})$ ,  $r_\nu = (\xi_{2\nu}, \eta_{2\nu})$  and  $qq_\nu \rightarrow \infty$ ,  $rr_\nu \rightarrow \infty$ . Then  $\xi_{1\nu} \rightarrow 0$ ,  $\xi_{2\nu} \rightarrow 0$ , and  $q_\nu r_\nu = |\xi_{1\nu}^{-1} - \xi_{2\nu}^{-1}| + |(\alpha\xi_{1\nu})^{-1} - (\beta\xi_{2\nu})^{-1}| + \epsilon_\nu$ ,  $\epsilon_\nu \rightarrow 0$ . The right side can be bounded only if  $|\xi_{1\nu}^{-1} - \xi_{2\nu}^{-1}|$  is bounded, but then the last term is not bounded because  $\alpha < \beta$ . Hence  $q_\nu r_\nu \rightarrow \infty$  and  $\sigma_\infty(q_1, r_1) = \infty$ .

Neither of the rays  $q_2$ ,  $r_2$  is a co-ray to the other. But  $q_\eta = (\eta\alpha^{-1}, \eta) \in q_2$  and  $r_\eta = (\eta\beta^{-1}, \eta) \in r_2$  for  $1/2 \leq \eta < 1$  and

$$q_\eta r_\eta = (\alpha^{-1} - \beta^{-1})\eta + (\beta - \alpha)\eta^{-1} < (\alpha^{-1} - \beta^{-1}) + 2(\beta - \alpha),$$

which shows that  $q_2$  and  $r_2$  satisfy (11.15).

The condition (11.15) has therefore in general  $E$ -spaces or  $G$ -spaces no bearing on the ray co-ray relation. This is emphasized by the fact that two rays  $q$  and  $r$  may satisfy the condition  $\sigma_\infty(q, r) = 0$  (and therefore be co-rays to each other), without satisfying (11.15).

An example is furnished by the surface of revolution:

$$R^8: \quad \xi^2 + \eta^2 = \zeta^{-1} + \exp \zeta \cdot \sin^2 \zeta, \quad \zeta > 0,$$

in the cartesian  $(\xi, \eta, \zeta)$ -space with the ordinary geodesic distance as distance.

The meridians of  $R^8$  are geodesics and straight lines in the sense of  $G$ -spaces. Call  $\mathfrak{G}_1$  the meridian  $\eta = 0$ ,  $\xi < 0$ , and  $\mathfrak{G}_2$  the meridian  $\eta = 0$ ,  $\xi > 0$ , of  $R^8$ , and  $r_\kappa$  the sub-ray  $\zeta \geq 1$  of  $\mathfrak{G}_\kappa$ . The ray  $r_\kappa$  contains the points  $q_{\kappa\nu} = ((-1)^\kappa(\nu\pi)^{-1}, 0, \nu\pi)$  with  $q_{1\nu} q_{2\nu} \leq \pi(\nu\pi)^{-1} = \nu^{-1}$ . Hence  $\sigma_\infty(r_1, r_2) = 0$ . But  $r_\kappa$  also contains the points

$$p_{\kappa\nu} = ((-1)^\kappa[(\nu\pi + \pi/2)^{-1} + \exp(\nu\pi + \pi/2)], 0, \nu\pi + \pi/2)$$

with  $p_{1\nu} \mathfrak{G}_2 = p_{2\nu} \mathfrak{G}_1 > \exp \nu\pi$ , so that neither  $xr_2$  is bounded for  $x \in r_1$  nor  $yr_1$  for  $y \in r_2$ .

In the study of negative rays and their co-rays, the spheres  $K'(r, p)$  (defined as the loci  $rx = rp$ ) and the corresponding limit spheres  $L'(r, p)$  play the

role of  $K''(r, p)$  and  $L''(r, p)$ . These sets coincide in  $G$ -spaces. Therefore (11.5) may be strengthened as follows:

(11.16) *If in a  $G$ -space  $\alpha$  is an asymptote to the ray  $r$  and  $p$  is on  $\alpha$  but not the origin of  $\alpha$ , then any point  $z$  of  $\alpha$  is the unique foot of  $p$  on  $L''(r, z)$ .*

If  $z$  follows  $p$  on  $\alpha$  the assertion is contained in (11.5). If  $z$  precedes  $p$ , let  $t$  be any point on  $\alpha$  that follows  $p$ , and  $z^*$  a foot of  $p$  on  $L''(r, z)$  (such a foot  $z^*$  exists because  $L''(r, z)$  is a closed set). Then by (11.2) and (10.4)

$$zt = \alpha(z, r) - \alpha(t, r) = \alpha(z^*, r) - \alpha(t, r) \leq z^*t \leq z^*p + pt \leq zp + pt = zt$$

( $z^*p \leq zp$  because  $z^*$  is a foot of  $p$  on  $L(r, z)$  and  $pz^* = z^*p$ ,  $pz = zp$ ). Hence ( $z^*pt$ ) and  $z^*p = zp$ ; by (4.1),  $z^*$  coincides with  $z$ .

It is easily seen that in euclidean geometry positive sub-rays  $q$  and  $r$  of oppositely oriented parallel straight lines satisfy the relation:

$$(11.17) \quad \alpha(x, r) + \alpha(x, q) \text{ is constant.}$$

This leads to the question whether (11.17) is in general  $G$ -spaces connected with the existence of parallels. Observe first that (11.17) implies  $\alpha(x, r) = \alpha(y, r) = \alpha(y, q) - \alpha(x, q)$  for all  $x, y$  and therefore  $L''(r, y) = L''(q, y)$  for all  $y$ , and that the interior of  $L''(r, y)$  coincides with the exterior of  $L''(q, y)$  and conversely. Moreover:

(11.18) If the relation (11.17) holds for two rays  $r, q$  of a  $G$ -space, then each asymptote to one of the rays is a directed straight line whose opposite orientation is an asymptote to the other ray.

For since the interiors of  $L''(r, p)$  and  $L''(q, p)$  are disjoint, co-rays  $p_1$  and  $p_2$  from  $p$  to  $r$  and  $q$  respectively have no common point but  $p$ . Any point  $p_1$  of  $p_1$  has, by (11.16),  $p$  as foot on  $L''(q, p)$ , hence it follows from (11.2), (11.4) that  $p$  lies on a co-ray from  $p_1$  to  $q$  and from (11.6) that  $p_2$  is a sub-ray of this co-ray. Hence the union of the carriers of  $p_1$  and  $p_2$  carries a straight line  $\mathfrak{G}$ , which, with the orientation for which  $p_2$  is a positive sub-ray, is an asymptote to  $q$ . In the same way it is seen that the opposite orientation of  $\mathfrak{G}$  is an asymptote to  $r$ .

Now let  $\mathfrak{G}$  be a straight line and  $r_1, r_2$  two "opposite" sub-rays of  $\mathfrak{G}$  (more exactly, but less intuitively, we require that neither of the rays is a sub-ray of the other). A straight line  $\mathfrak{H}$  is called a parallel to  $\mathfrak{G}$  if its two orientations are asymptotes to  $r_1$  and  $r_2$ . (This definition is obviously independent of the choice of the rays  $r_1, r_2$  on  $\mathfrak{G}$ .) (11.18) yields

(11.19) THEOREM. *In a  $G$ -space let  $\alpha(x, r_1) + \alpha(x, r_2)$  be constant for the two opposite sub-rays  $r_1, r_2$  of the straight line  $\mathfrak{G}$ ; then each asymptote to  $r_1$  or  $r_2$  carries a parallel to  $\mathfrak{G}$ .*

It is worth noticing that under differentiability assumptions usual in Finsler spaces, the converse of this theorem holds. Namely if every asymptote to  $r_1$  or  $r_2$  is a parallel to  $\mathfrak{G}$ , then these parallels are extremals transversal to



the two families of hypersurfaces  $L''(\tau_1, x)$  and  $L''(\tau_2, x)$ . As soon as the differential equations expressing this property of the  $L''(\tau, x)$  are such that they have unique solutions,  $L''(\tau_1, x)$  must coincide with  $L''(\tau_2, x)$  for every  $x$ , whence it easily follows that (11.17) holds. In general  $G$ -spaces the converse of (11.19) does not hold, for example, in a Minkowski space  $R^2$  in which  $K$  is not everywhere differentiable.

#### IV. COVERING SPACES AND THE SPACES OF CONSTANT CURVATURE

**12. Extremals and segments in covering spaces.** Some important geometric problems require, for an adequate formulation, the concept of a covering space. Covering spaces for general spaces have been widely used by topologists. A theory of covering spaces for  $E$ -spaces leads however, naturally, to metric questions in addition to topological ones. The present section contains a brief discussion of these problems. Applications will be found in §§14 and 16.

The space  $\bar{R}$  is called a *covering space* of the  $E$ -space  $R$  if it has the following two properties:

- (1)  $\bar{R}$  satisfies Axioms A, B, C.
- (2) There is a locally isometric mapping  $\Phi$  of  $\bar{R}$  on  $R$ .

The last requirement means more explicitly this: every point  $\bar{x}$  of  $\bar{R}$  has a neighborhood  $S(\bar{x}, \alpha(\bar{x}))$  such that

$$(12.1) \quad \bar{y}\Phi\bar{z}\Phi = \bar{y}\bar{z} \quad \text{for} \quad \bar{y}, \bar{z} \in S(\bar{x}, \alpha(\bar{x})).$$

(12.1) implies that  $\Phi$  is a continuous mapping of  $\bar{R}$  on  $R$ . The mapping  $\Phi$  is in general not unique; for instance, if  $\Psi$  is a motion of  $\bar{R}$  and  $X$  a motion of  $R$ , then  $\Psi\Phi X$  is also a locally isometric mapping of  $\bar{R}$  on  $R$ . The space  $\bar{R}$  together with the definite mapping  $\Phi$  will be denoted by  $R_\Phi$ .

The image  $\bar{A}\Phi$  of a set  $\bar{A}$  in  $R_\Phi$  will be called  $A$ , in particular  $x$  is the image of  $\bar{x}$ . When  $A = \bar{A}\Phi$  the set  $\bar{A}$  is said to lie over  $A$ . The mapping of the sphere  $S(\bar{x}, \beta)$  in  $\bar{R}$  on the set  $S(x, \beta)\Phi$  in  $R$  under  $\Phi$  will be denoted by  $\Phi(\bar{x}, \beta)$ . If  $\Phi(\bar{x}, \beta)$  is one-to-one then  $\Phi^{-1}(x, \beta)$  is defined and maps, by (12.1),  $S(x, \beta)\Phi$  topologically on  $S(\bar{x}, \beta)$ . In particular  $\Phi^{-1}(x, \alpha(\bar{x}))$  is defined and maps  $S(x, \alpha(\bar{x}))$  isometrically on  $S(\bar{x}, \alpha(\bar{x}))$ . Some simple properties of covering spaces will now be discussed.

(12.2) *A covering space of an  $E$ -space is an  $E$ -space.*

Since  $R_\Phi$  satisfies A, B, C, it has only to be shown that Axiom D holds. Let  $S(\bar{x}, \alpha)$  satisfy (12.1). There is a  $\beta \leq \alpha$  such that  $S(x, \beta)$  satisfies Axiom D in  $R$ .

By (1.4) there is an  $\eta > 0$  such that  $U_\nu(S'(y, \eta) + S''(y, \eta)) \subset S(x, \beta)$  when  $y$  traverses  $S(x, \beta/2)$ . Then  $S(\bar{x}, \beta/2)$  satisfies Axiom D in  $\bar{R}$ . For if  $\bar{a}, \bar{b}$  in  $S(\bar{x}, \beta/2)$  and  $\epsilon > 0$  are given, positive numbers  $\delta_\kappa \leq \min(\epsilon, \eta)$ ,  $\kappa = 1, 2$ , can be found such that unique points  $a_1$  and  $b_2$  with  $(a_1\bar{a}b)$ ,  $(\bar{a}b\bar{b}_2)$  and  $a_1a = \delta_1$ ,  $\bar{b}\bar{b}_2 = \delta_2$  exist.  $a_1$  and  $b_2$  lie in  $S(x, \beta)$ , since  $a$  and  $b$  lie in  $S(x, \beta/2)$  and  $\delta_\kappa \leq \eta$ .

If  $\bar{a}_1 = a_1 \Phi^{-1}(\bar{x}, \alpha)$ ,  $\bar{b}_2 = b_2 \Phi^{-1}(\bar{x}, \alpha)$  then  $(\bar{a}_1 \bar{a} \bar{b})$ ,  $(\bar{a} \bar{b} \bar{b}_2)$  and  $\bar{a}_1 \bar{a} = \delta_1$ ,  $\bar{b} \bar{b}_2 = \delta_2$  because  $\Phi^{-1}(\bar{x}, \alpha)$  is an isometry.

Because of (12.1) the continuous curve  $x(\tau) = \bar{x}(\tau) \Phi$  has Property I of §4 when  $\bar{x}(\tau)$  has Property I. Therefore (5.4) yields

(12.3) If  $\bar{x}(\tau)$  represents an extremal in  $R_\Phi$ , then  $\bar{x}(\tau) \Phi = x(\tau)$  represents an extremal in  $R$ .

If the extremals  $\mathfrak{E}_\Phi$  in  $R_\Phi$  and  $\mathfrak{E}$  in  $R$  have representations  $\bar{x}(\tau)$  and  $x(\tau)$  for which  $\bar{x}(\tau) \Phi = x(\tau)$  then  $\mathfrak{E}_\Phi$  is said to lie over  $\mathfrak{E}$ .

(12.4) *There is exactly one extremal  $\mathfrak{E}_\Phi$  in  $R_\Phi$  over a given extremal  $\mathfrak{E}$  in  $R$  and through a given point  $\bar{x}$  over a point  $x$  of  $\mathfrak{E}$ .*

**Proof.** Let  $x(\tau)$  represent  $\mathfrak{E}$  and  $x(\tau_0) = x$ , and let  $S(\bar{x}, \alpha)$  satisfy (12.1). For a suitable positive  $\beta < \alpha$  the arc  $x(\tau)$ ,  $|\tau - \tau_0| \leq \beta$ , represents a segment. Hence  $x(\tau) \Phi^{-1}(\bar{x}, \alpha)$  represents a segment in  $R_\Phi$  for  $|\tau - \tau_0| \leq \beta$ . By (5.6) there is an extremal  $\mathfrak{E}_\Phi$  in  $R_\Phi$  with a representation  $\bar{x}(\tau)$  for which  $\bar{x}(\tau) = x(\tau) \Phi^{-1}(\bar{x}, \alpha)$  for  $|\tau - \tau_0| \leq \beta$ . By (12.3),  $\mathfrak{E}_\Phi$  lies over the extremal  $\bar{x}(\tau) \Phi$  in  $R$ . Since  $\bar{x}(\tau) \Phi = x(\tau)$  for  $|\tau - \tau_0| \leq \beta$  it follows from (5.6) that  $\bar{x}(\tau) \Phi = x(\tau)$  for all  $\tau$ , so that  $\mathfrak{E}_\Phi$  lies over  $\mathfrak{E}$ .

Let  $\bar{y}(\tau)$  represent any extremal through  $\bar{x}$  over  $\mathfrak{E}$ . Then  $\bar{y}(\tau) \Phi$  represents  $\mathfrak{E}$ , hence a number  $\beta$  exists such that  $\bar{y}(\tau) \Phi = x(\tau + \beta)$  for all  $\tau$  (compare 5.7), or  $\bar{y}(\tau - \beta) \Phi = x(\tau)$ . This yields for  $|\tau - \tau_0| \leq \alpha$  the relation  $\bar{x}(\tau) = x(\tau) \Phi^{-1}(\bar{x}, \alpha) = \bar{y}(\tau - \beta) \Phi \Phi^{-1}(\bar{x}, \alpha) = \bar{y}(\tau - \beta)$ . Hence  $\bar{x}(\tau) = \bar{y}(\tau - \beta)$ , so that  $\bar{y}(\tau)$  represents also  $\mathfrak{E}_\Phi$ .

(12.5) The image  $y(\tau) = \bar{y}(\tau) \Phi$  of a continuous curve  $\bar{y}(\tau)$ ,  $\beta_1 \leq \tau \leq \beta_2$ , has the same length as  $\bar{y}(\tau)$ .

For there are values  $\tau_1, \dots, \tau_\mu$ ,  $\beta_1 \leq \tau_\lambda \leq \beta_2$ , such that the neighborhoods  $S(\bar{y}(\tau_\lambda), \alpha(\bar{y}(\tau_\lambda)))$  cover the curve  $\bar{y}(\tau)$ . If the values  $\beta_1 = \tau'_1 < \tau'_2 < \dots < \tau'_\nu = \beta_2$  are such that  $\bar{y}(\tau'_\lambda)$  and  $y(\tau'_{\lambda+1})$  belong to one  $S(\bar{y}(\tau_\lambda), \alpha(\bar{y}(\tau_\lambda)))$  then

$$\sum_{\lambda} \bar{y}(\tau'_\lambda) \bar{y}(\tau'_{\lambda+1}) = \sum_{\lambda} y(\tau'_\lambda) y(\tau'_{\lambda+1})$$

hence the limits of these sums are equal.

(12.6)  $\bar{x}\bar{y} \geq xy$  for any  $\bar{x}, \bar{y}$  in  $R_\Phi$ .

For  $\bar{x}\bar{y} = \inf \lambda(\bar{z}(\tau))$  where  $\bar{z}(\tau)$  traverses all continuous curves in  $\bar{R}$  from  $\bar{x}$  to  $\bar{y}$ . Because of (12.5),  $xy \leq \inf \lambda(\bar{z}(\tau) \Phi) = \inf \lambda(\bar{z}(\tau))$ .

(12.7) If  $\bar{x}(\tau)$ ,  $\alpha \leq \tau \leq \beta$ , is a continuous curve whose image  $\bar{x}(\tau) \Phi = x(\tau)$  represents a segment then  $\bar{x}(\tau)$  represents a segment and  $x(\alpha)x(\beta) = \bar{x}(\alpha)\bar{x}(\beta)$ .

This follows from (12.5), (12.6) because  $x(\alpha)x(\beta) = \lambda(x(\tau)) = \lambda(\bar{x}(\tau)) \geq \bar{x}(\alpha)\bar{x}(\beta) \geq x(\alpha)x(\beta)$ .

A consequence of (12.4), (12.7) is

(12.8) *For a given segment  $\mathfrak{s}(a, b)$  in  $R$  and a given point  $\bar{a}$  over  $a$  ( $\bar{b}$  over  $b$ ) there is exactly one segment  $\mathfrak{s}(\bar{a}, \bar{b})$  in  $R_\Phi$  over  $\mathfrak{s}(a, b)$  and  $ab = \bar{a}\bar{b}$ .*

It is now possible to make statements about the extent to which  $\Phi$  is a topological mapping. Using the notations of (9.6) denote by  $D_p^+$  and  $D_p^-$  the

complements of the sets  $C_p^+$  and  $C_p^-$  respectively.  $D_p^+$  consists of  $p$  and those points  $y$  for which a point  $z$  with  $(pyz)$  exists.

(12.9) For a given point  $\bar{p}$  in  $R_\Phi$  there is exactly one set  $\bar{D}(\bar{p}) \supset \bar{p}$  such that  $\Phi$  maps  $\bar{D}(\bar{p})$  topologically on  $D_p^+$ . If  $\bar{p}_1$  and  $\bar{p}_2$  are two different points over  $p$ , then  $\bar{D}(\bar{p}_1) \cap \bar{D}(\bar{p}_2) = 0$ .

*Note.* The example of the elliptic plane with the sphere as covering space shows (12.9) to be the best possible result in the sense that  $\bar{D}(\bar{p})$  is in general not a proper subset of a set  $F$  which  $\Phi$  maps topologically on  $F\Phi$ .

By (12.1) and (9.8) there is a  $\beta > 0$  such that  $S(p, 2\beta) \subset D_p^+$  and  $\Phi(\bar{p}, 2\beta)$  is an isometry. If  $y \in D_p^+ - S(p, 2\beta)$ , the segment  $\mathfrak{s}(p, y)$  is unique; therefore the point  $y_\beta$  with  $(py_\beta y)$  and  $py_\beta = \beta$  is unique.

The image  $\bar{y}$  of  $y$  is defined as the end point of the, by (12.8) unique, segment  $\mathfrak{s}(\bar{p}, \bar{y})$  over  $\mathfrak{s}(p, y)$  which starts at  $\bar{p}$ . By the definition of this term  $z\Phi^{-1}(\bar{p}, \bar{y})$  is the point of  $\mathfrak{s}(\bar{p}, \bar{y})$  over  $z$  when  $z \in S(p, 2\beta) \cap S(p, y)$ , in particular  $\bar{y}_\beta = y_\beta \Phi^{-1}(\bar{p}, \beta)$  lies over  $y_\beta$ . The mapping  $y \rightarrow \bar{y}$  is continuous. For if  $y_\nu \rightarrow y_0$ ,  $py_0 \geq 2\beta$ , let  $(py_\nu, py_\nu)$  and  $py_\nu = \beta$ . Then  $\bar{y}_\nu, \beta \rightarrow \bar{y}_0, \beta$  and  $(\bar{p}\bar{y}_\nu, \bar{p}\bar{y}_\nu)$ ; hence every accumulation point  $\bar{y}$  of  $\{\bar{y}_\nu\}$  satisfies  $(\bar{p}\bar{y}_0, \bar{p}\bar{y})$ , so that  $\bar{y} = \bar{y}_0$ .

This shows that  $\Phi$  maps the set  $\bar{D}(\bar{p})$  consisting of  $S(\bar{p}, 2\beta)$  and all the points  $\bar{y}$  defined above topologically on the set  $D^+(p)$ .

Let  $\bar{E}$  be any set which contains  $\bar{p}$  and which  $\Phi$  maps topologically on  $D^+(p)$ . If  $y$  is any point of  $\bar{E}$  then  $\bar{E}$  contains a curve from  $\bar{p}$  to  $\bar{y}$  which lies over the (unique) segment  $\mathfrak{s}(p, y)$  in  $D^+(p)$ , hence  $\bar{y}$  coincides by (12.8) with the point of  $\bar{D}(\bar{p})$  over  $y$ , so that  $\bar{E} = \bar{D}(\bar{p})$ .

If there were two different points  $\bar{p}_1$  and  $\bar{p}_2$  over  $p$  such that  $\bar{D}(\bar{p}_1) \cap \bar{D}(\bar{p}_2)$  contains a point  $\bar{y}$ , then (according to the construction of  $\bar{D}(\bar{p}_i)$ )  $\mathfrak{s}(\bar{p}_1, \bar{y})$  and  $\mathfrak{s}(\bar{p}_2, \bar{y})$  would be different segments over  $\mathfrak{s}(p, y)$ , which contradicts (12.8).

$$(12.10) \quad D^+(p) \supset S'(p, \eta_2(p)), \quad D^-(p) \supset S''(p, \eta_2(p))$$

follows from (9.8). Hence  $\Phi(\bar{p}, \eta_2(p))$  is a topological mapping.  $\Phi(\bar{p}, \eta_2(p)/3)$  is an isometry. This follows from (12.7) and the fact (compare (1.15)) that  $a, b \in S(p, \eta_2(p)/3)$  implies  $\mathfrak{s}(a, b) \subset S(p, \eta_2(p))$ . This leads to

(12.11) THEOREM. Let  $R_\Phi$  be a covering space of the  $E$ -space  $R$ . If  $\bar{p}\Phi = p$ , then  $\Phi$  maps  $S(\bar{p}, \eta_2(p))$  topologically on  $S(p, \eta_2(p))$ ; in addition  $\Phi$  maps  $S(\bar{p}, \eta_2(p)/3)$  isometrically on  $S(p, \eta_2(p)/3)$ . If  $\bar{p}_1 \neq \bar{p}_2$  and  $\bar{p}_i\Phi = p$ , then  $S(\bar{p}_1, \eta_2(p)) \cap S(\bar{p}_2, \eta_2(p)) = 0$  and

$$(12.12) \quad \sigma^*(\bar{p}_1, \bar{p}_2) > \eta_2(p) \quad (\text{in } G\text{-spaces } \bar{p}_1\bar{p}_2 \geq 2\eta_2(p)).$$

Regarding (12.12) observe that by (12.9), (12.10),  $S'(\bar{p}_1, \eta_2(p)) \cap S'(\bar{p}_2, \eta_2(p)) = 0$  so that both  $\bar{p}_1\bar{p}_2 > \eta_2(p)$  and  $\bar{p}_2\bar{p}_1 > \eta_2(p)$ , hence  $\sigma^*(\bar{p}_1, \bar{p}_2) > \eta_2(p)$  (see (1.5)). For the remark about  $G$ -space compare (12.20).

The most important point in (12.11) is that  $\eta_2(p)$  depends only on  $p$  and  $R$ , and not on the covering space  $R_\Phi$ . The topological theory of covering spaces requires only the existence of a function  $\zeta(p)$ , such that  $\Phi(\bar{p}, \zeta(p))$  is

topological for all  $\bar{p}$  over  $p$  and a fixed  $\Phi$  (compare Seifert-Threlfall [1, pp. 182, 183]). (12.15) will show that this difference is essential. For that result some well known implications of the weaker property are needed.

(12.13) There is exactly one continuous curve  $\bar{x}(\tau)$  in  $R_\Phi$  over a given continuous curve  $x(\tau)$ ,  $\alpha \leq \tau \leq \beta$ , in  $R$  with a given point  $\bar{x}(\alpha)$  over  $x(\alpha)$ .

**Proof.** Choose  $\alpha = \tau_1 < \tau_2 < \dots < \tau_\mu = \beta$  such that with  $\zeta_\kappa = \eta_2(x(\tau_\kappa))$  the sphere  $S(x(\tau_\kappa), \zeta_\kappa)$  contains the subarc  $c_\kappa: \tau_\kappa \leq \tau \leq \tau_{\kappa+1}$  of  $x(\tau)$ . Put  $\bar{x}(\tau_1) = x(\alpha)$  and by induction  $\bar{x}(\tau_\nu) = x(\tau_\nu)\Phi^{-1}(\bar{x}(\tau_{\nu-1}), \zeta_{\nu-1})$ . Then  $c_1\Phi^{-1}(\bar{x}(\tau_1), \zeta_1) + \dots + c_{\mu-1}\Phi^{-1}(\bar{x}(\tau_{\mu-1}), \zeta_{\mu-1})$  is a continuous curve over  $x(\tau)$  which satisfies (12.13). The uniqueness of  $\bar{x}(\tau)$  follows the same pattern as previous arguments.

Two continuous curves  $\mathfrak{R}$  and  $\mathfrak{L}$  from  $a$  to  $b$  are homotopic,  $\mathfrak{R} \sim \mathfrak{L}$ , if  $\mathfrak{R}$  can be continuously deformed into  $\mathfrak{L}$  such that  $a$  and  $b$  remain fixed during the process. The exact definition is found in Pontrjagin [1, §46]. (13.13) leads to the following result, which is fundamental in any discussion of covering spaces.

(12.14) If  $\mathfrak{R}_\Phi \sim \mathfrak{L}_\Phi$  in  $R_\Phi$ , then  $\mathfrak{R}_\Phi\Phi \sim \mathfrak{L}_\Phi\Phi$  in  $R$ . Conversely, if  $\mathfrak{R} \sim \mathfrak{L}$  in  $R$  and if the curves  $\mathfrak{R}_\Phi$  and  $\mathfrak{L}_\Phi$  over  $\mathfrak{R}$  and  $\mathfrak{L}$  respectively start at the same point  $\bar{a}$ , then they end at the same point  $\bar{b}$  and  $\mathfrak{R}_\Phi \sim \mathfrak{L}_\Phi$  in  $R_\Phi$ .

The first part is obvious; the proof of the second part is found in Seifert-Threlfall [1, pp. 186, 187].

A known consequence of (12.14) is that there are equally many points of  $R_\Phi$  over every point of  $R$ . For if  $p$  and  $q$  are arbitrary in  $R$ ,  $c$  any curve from  $p$  to  $q$ , and  $\bar{p}_1, \bar{p}_2$  different points over  $p$ , then the curves  $c_{\kappa,\Phi}$  which begin at  $\bar{p}_\kappa$  and lie over  $c$ ,  $\kappa = 1, 2$ , end at different points  $\bar{q}_1$  and  $\bar{q}_2$ , because  $\bar{q}_1 = \bar{q}_2$  and  $c^{-1} \sim c^{-1}$  would imply  $c_{1,\Phi}^{-1} \sim c_{2,\Phi}^{-1}$  so that the end points of  $c_{1,\Phi}^{-1}$  and  $c_{2,\Phi}^{-1}$  would by (12.14) coincide.

Now consider a locally isometric mapping  $\Phi$  of the  $E$ -space  $R$  on itself. Then  $R$  may be considered as a covering space of itself. Let  $p_1\Phi = p_2\Phi = p$ ,  $p_1 \neq p_2$ . Since there are equally many points over every point of  $R$ , there are points  $p_{\kappa\lambda}$ ,  $\kappa, \lambda = 1, 2$ , such that  $p_{\kappa\lambda}\Phi = p_\kappa$  and  $p_{\kappa 1} \neq p_{\kappa 2}$ . Moreover  $p_{1\kappa} \neq p_{2\lambda}$  because  $p_1 \neq p_2$ . Similarly there are points  $p_{\kappa\lambda\mu}$ ,  $\kappa, \lambda, \mu = 1, 2$ , such that  $p_{\kappa\lambda\mu}\Phi = p_{\kappa\lambda}$  and  $p_{\kappa\lambda 1} \neq p_{\kappa\lambda 2}$ ; again the 8 points  $p_{\kappa\lambda\mu}$  are all different. The mapping  $\Phi^3$  is still locally isometric and  $p_{\kappa\lambda\mu}\Phi^3 = p$ . Induction shows the existence of  $2^p$  different points  $p^1, \dots, p^{2^p}$  such that  $p^r\Phi^p = p$ . Applying (12.11) to  $\Phi^p$ , it is seen that the neighborhoods  $S(p^r, \eta_2(p))$  are all disjoint. Since  $p$  is arbitrary it follows that  $R$  cannot be compact. Hence, if  $R$  is compact,  $\Phi$  must be one-to-one, so that  $\Phi^{-1}$  is also a locally isometric mapping of  $R$  on itself. (12.6) yields that  $\Phi$  preserves distances. This proves

(12.15) THEOREM. *A locally isometric mapping of a compact E-space on itself is a motion.*

This theorem has no topological counterpart, as the locally topological mapping  $e^{i\phi} \rightarrow e^{2i\phi}$  of the unit circle of the complex plane on itself shows. Nor does (12.15) hold for general unbounded  $E$ -spaces.

A cylinder with a locally hyperbolic metric will furnish an example. Consider a  $(\xi, \eta)$ -plane and define as distance of two points  $\bar{p}_1, \bar{p}_2$  the number

$$(12.16) \quad \bar{p}_1 \bar{p}_2 = \inf_c \int_c (d\xi^2 + \exp(-2\xi)d\eta^2)$$

where  $c$  traverses all curves of class  $D'$  from  $\bar{p}_1$  to  $\bar{p}_2$ . With this distance the  $(\xi, \eta)$ -plane becomes hyperbolic. The curves  $\eta = \text{const.}$  are geodesics and the curves  $\xi = \text{const.}$  are limit circles with the positive  $\xi$ -axis as central ray (compare for instance Liebmann [1, pp. 143, 163]). The transformation

$$(12.17) \quad \xi' = \xi + \alpha, \quad \eta' = \eta e^\alpha + \beta$$

leaves (12.16) obviously invariant and is therefore a motion of the hyperbolic plane.

Now divide the points of the  $(\xi, \eta)$ -plane into classes putting two points  $(\xi_1, \eta_1)$  and  $(\xi_2, \eta_2)$  in the same class if and only if  $\xi_1 = \xi_2$  and  $\eta_1 - \eta_2$  is an integer. The space  $R^9$  is defined as the set of these classes with the distance  $p_1 p_2 = \min \bar{p}_1 \bar{p}_2$ , where  $\bar{p}_1$  and  $\bar{p}_2$  traverse independently the classes  $p_1$  and  $p_2$ . For  $\alpha = \log \nu$ , where  $\nu$  is a positive integer, and  $\beta = 0$  the mapping (12.17) becomes  $\xi' = \xi + \log \nu$ ,  $\eta' = \nu \eta$ . This is a continuous  $\nu$ -to-1 mapping of the cylinder  $R^9$  on itself, which is locally isometric because (12.17) leaves (12.16) invariant.

If  $\Phi$  is a locally isometric mapping of the  $E$ -space  $\bar{R}$  on the  $E$ -space  $R$ , and  $\bar{p}\Phi = p$ , then two curves which begin and end at  $\bar{p}$  are by (12.14) homotopic if and only if their images are homotopic. Hence  $\Phi$  associates with a class of homotopic curves beginning and ending at  $\bar{p}$  a class of homotopic curves which begin and end at  $p$ , and yields thus an isomorphic mapping of the fundamental group of  $\bar{R}$  in the fundamental group of  $R$  (compare Seifert-Threlfall [1, p. 188]). If there are two different points of  $\bar{R}$  over  $p$ , the fundamental group of  $\bar{R}$  will be mapped on a proper sub-group of the fundamental group of  $R$ . This proves

(12.18) *A locally isometric mapping of an  $E$ -space  $\bar{R}$  on an  $E$ -space  $R$  is isometric in the large when the fundamental group of  $\bar{R}$  is not isomorphic to a proper subgroup of the fundamental group of  $R$ .*

The last condition is trivially satisfied when  $R$  is simply connected. The following addition to (12.15) is a corollary of (12.18).

(12.19) *If the fundamental group of  $R$  is not isomorphic to a proper subgroup of itself, then every locally isometric mapping of  $R$  on itself is a motion.*

The theory of covering spaces for  $G$ -spaces is contained in the preceding discussion because (12.5), (12.6), (12.7) yield readily

(12.20) A covering space  $\bar{R}$  of an  $E$ -space  $R$  is symmetric if and only if  $R$  is symmetric.

**13. The universal covering space.** The construction of the universal covering space for connected, locally connected, and locally simply connected topo-

logical spaces is found in Pontrjagin [1, §46] (quoted as P). Only some metric properties will be added here.

The existence of segments obviously makes an  $E$ -space connected and locally connected. The locally simple connectedness is contained in the following more general statement (as the special case  $x(\tau) = a$  for all  $\tau$ ):

(13.1) *If  $\mathfrak{R}: x(\tau)$  and  $\mathfrak{L}: y(\tau)$  are two continuous curves defined for  $\alpha \leq \tau \leq \beta$  with  $x(\alpha) = y(\alpha)$ ,  $x(\beta) = y(\beta)$ , and  $x(\tau)y(\tau) \leq \eta_2(x(\tau))$ , then  $\mathfrak{L} \sim \mathfrak{R}$ . (This implies that an  $E$ -space with  $\eta_\lambda(x) \equiv \infty$  is simply connected.)*

For the segment  $\mathfrak{s}(x(\tau), y(\tau))$  is unique for every  $\tau$  and depends therefore continuously on  $\tau$ . For every  $0 \leq \sigma \leq 1$  define  $y_\sigma(\tau)$  as the point on  $\mathfrak{s}(x(\tau), y(\tau))$  with  $y_\sigma(\tau)y(\tau) = \sigma \cdot x(\tau)y(\tau)$ . Then  $y_0(\tau) = y(\tau)$  and  $y_1(\tau) = x(\tau)$ , hence  $y_\sigma(\tau)$  yields a continuous deformation of  $\mathfrak{L}$  into  $\mathfrak{R}$ .

(13.1) leads to the following important fact.

(13.2) *Among all curves from  $a$  to  $b$  in an  $E$ -space which are homotopic to a given curve  $\mathfrak{R}$  from  $a$  to  $b$ , there is at least one rectifiable shortest one. Each such shortest curve is equivalent to an extremal arc from  $a$  to  $b$ , which degenerates into  $a$  if and only if  $a = b$  and  $\mathfrak{R} \sim 0$ .*

For the definition of equivalence see §1. If  $\mathfrak{R}$  and  $\mathfrak{R}'$  are equivalent then  $\mathfrak{R}' \sim \mathfrak{R}$ , hence  $\mathfrak{R} \sim \mathfrak{L}$  implies  $\mathfrak{R}' \sim \mathfrak{L}$ . To prove (13.2) observe first that (13.1) implies the existence of polygons (that is, curves of the form  $\sum \mathfrak{s}(a_k, a_{k+1})$ ) from  $a$  to  $b$  which are homotopic to  $\mathfrak{R}$ , hence  $\delta = \inf \lambda(\mathfrak{R}')$ ,  $\mathfrak{R}' \sim \mathfrak{R}$ , is finite. If  $\delta = 0$ , then  $a = b$  and  $\mathfrak{R} \sim 0$  because of (13.1).

Let  $\delta > 0$  and choose  $\mathfrak{R}' \sim \mathfrak{R}$  such that  $\lim \lambda(\mathfrak{R}') = \delta$ . By (1.19) there is a curve  $x''(\tau)$ ,  $0 \leq \tau \leq \alpha$ , equivalent to  $\mathfrak{R}'$  and such that  $x''(\tau)$  tends uniformly to a continuous curve  $\mathfrak{R}'': x''(\tau)$  from  $a$  to  $b$ . Since  $\eta = \min_\tau \eta(x''(\tau))$ ,  $0 \leq \tau \leq \alpha$ , is positive (compare (4.9)), the uniformity of the convergence and (14.1) yield  $\mathfrak{R}' \sim \mathfrak{R}''$  for large  $\nu$ , hence  $\mathfrak{R}'' \sim \mathfrak{R}$ . Since  $\lambda(\mathfrak{R}'') \leq \lim \lambda(\mathfrak{R}') = \delta$  (see (1.10)), it follows from the definition of  $\delta$  and  $\mathfrak{R}'' \sim \mathfrak{R}$  that  $\lambda(\mathfrak{R}'') = \delta$ .

(13.1) yields, moreover, that  $x''(\tau)$  must, locally, be a segment, and is therefore equivalent to a subarc of an extremal. If finally  $a = b$  and  $\mathfrak{R}$  is not homotopic to 0, it follows from (13.1) that this arc must contain a point outside of  $S(a, \eta(a))$  and has therefore at least length  $\eta(a)$  (for  $\lambda(x''(\tau)) \geq ay + ya > \sigma(a, y) > \eta(a)$ ). This proves the "only if" part of the last statement in (13.2).

The universal covering space  $\bar{R}$  of a given  $E$ -space  $R$  may now be constructed as follows. Fix a point  $p$  in  $R$ . With every point  $x$  in  $R$  and every class ( $\mathfrak{R}$ ) of curves from  $p$  to  $x$  homotopic to a given curve  $\mathfrak{R}$  we associate a point  $\bar{x}_K$  of  $\bar{R}$ . The distance of two points  $\bar{x}_K$  and  $\bar{y}_L$  is defined as

$$(13.3) \quad \bar{x}_K \bar{y}_L = \inf \lambda(c), \quad \text{where } c \sim \mathfrak{R}^{-1}\mathfrak{L}.$$

(13.2) shows that  $\bar{x}_K \bar{y}_L$  is always finite and that  $\bar{x}_K \bar{y}_L = 0$  if and only if  $x = y$  and  $\mathfrak{R} \sim \mathfrak{L}$ ; moreover that an extremal arc  $c$  exists with

$$(13.4) \quad \lambda(c) = \bar{x}_K \bar{y}_L \quad \text{and} \quad c \sim \mathfrak{R}^{-1}\mathfrak{L}.$$

Choose  $c_1 \sim \mathfrak{R}^{-1}\mathfrak{Q}$  and  $c_2 \sim \mathfrak{Q}^{-1}\mathfrak{M}$  such that  $\lambda(c_1) = \bar{x}_K \bar{y}_L$  and  $\lambda(c_2) = \bar{y}_L \bar{z}_M$ . Then  $c_1 c_2 \sim \mathfrak{R}^{-1}\mathfrak{Q}\mathfrak{Q}^{-1}\mathfrak{M} \sim \mathfrak{R}^{-1}\mathfrak{M}$ , hence  $\bar{x}_K \bar{z}_M \leq \lambda(c_1) + \lambda(c_2) = \bar{x}_K \bar{y}_L + \bar{y}_L \bar{z}_M$ .

(13.4) implies that Axiom C holds in  $\bar{R}$ . Namely, let  $\bar{x}_K \neq \bar{y}_L$  and let, with the notation of (13.4),  $z$  be any point on  $c$  different from  $x$  and  $y$ . Designate by  $c_1$  and  $c_2$  subarcs of  $c$  from  $x$  to  $z$  and from  $z$  to  $y$  respectively with  $c_1 c_2 = c$  and  $\lambda(c_1) + \lambda(c_2) = \lambda(c)$  (compare (1.9)). Put  $\mathfrak{R}c_1 = \mathfrak{M}$ . Then  $c_1 \sim \mathfrak{R}^{-1}\mathfrak{M}$  and  $c_2 \sim c_1^{-1}c \sim \mathfrak{M}^{-1}\mathfrak{R}\mathfrak{R}^{-1}\mathfrak{Q} = \mathfrak{M}^{-1}\mathfrak{Q}$ , hence  $\lambda(c) = \lambda(c_1) + \lambda(c_2) \geq \bar{x}_K \bar{z}_M + \bar{z}_M \bar{y}_L \geq \bar{x}_K \bar{y}_L = \lambda(c)$ , so that  $(\bar{x}_K \bar{z}_M \bar{y}_L)$ .

To see that  $\bar{R}$  is finitely compact, let  $\bar{x}_\nu$  be a point of  $\bar{R}$  which belongs to  $x_\nu$  and the homotopy class  $(c_\nu)$  and for which  $\bar{p}_0 \bar{x}_\nu < \beta$ . Because of (13.4),  $c_\nu$  may be assumed to be an extremal arc with  $\lambda(c_\nu) = \bar{p}_0 \bar{x}_\nu$ . Call  $\mathfrak{r}_\nu^+$  the half extremal with  $\mathfrak{r}_\nu^+(c_\nu) = c_\nu$ . Then  $\{\mathfrak{r}_\nu^+\}$  is bounded because of (9.1) and  $\bar{p}_0 \bar{x}_\nu \leq \lambda(c_\nu) = \bar{p}_0 \bar{x}_\nu < \beta$ . Hence  $\{\mathfrak{r}_\nu^+\}$  contains a subsequence  $\{\mathfrak{r}_\mu^+\}$  which tends to a half extremal  $\mathfrak{r}^+$  with origin  $p$  and for which  $\lim \lambda(c_\mu) = \lambda$  exists. Then  $x_\mu = x_\mu^+(c_\mu) \rightarrow x^+(\lambda) = x$ . Call  $\mathfrak{s}_\mu$  the segment from  $x$  to  $x_\mu$ . By (9.2),  $x_\mu^+(\tau)$  tends uniformly to  $x^+(\tau)$  in every interval  $0 \leq \tau \leq \alpha$ . Hence it follows from (13.1) that  $c_\mu \mathfrak{s}_\mu^{-1} \sim \mathfrak{r}^+(\lambda) = c$ , or  $\mathfrak{s}_\mu \sim c^{-1}c_\mu$ , so that  $\bar{x}_c \bar{x}_\mu \leq \lambda(s_\mu) \rightarrow 0$ .

Finally it has to be shown that the mapping  $\Phi$  of  $\bar{R}$  on  $R$  which associates  $x$  with  $\bar{x}_K$  satisfies (12.1), that is, that for every  $\bar{q}_M$  in  $\bar{R}$  an  $\alpha > 0$  exists such that  $\bar{x}_K \bar{y}_L = xy$  for  $\sigma(\bar{x}_K, \bar{q}_M) < \alpha$  and  $\sigma(\bar{y}_L, \bar{q}_M) < \alpha$ .

There is an  $\alpha > 0$  such that

$$(13.5) \quad S'(q, 3\alpha) \cup S''(q, 3\alpha) \subset S(q, \eta(q)).$$

If  $x, y \in S(q, \alpha)$ , then  $\mathfrak{s}(x, y) \subset S(q, 3\alpha)$  (see (1.15)). Any curve  $c$  from  $x$  to  $y$  in  $S(q, \eta(q))$  is by (13.1) homotopic to  $\mathfrak{s}(x, y)$  since  $\mathfrak{s}(x, y)c^{-1} \sim 0$ . Hence if  $c'$  is not homotopic to  $\mathfrak{s}(x, y)$  it contains a point  $z$  outside of  $S(q, \eta(q))$ . Since  $x, y$  lie in  $S(q, \alpha) = S'(q, \alpha) \cap S''(q, \alpha)$  it follows from (13.5) that

$$(13.6) \quad \lambda(c') \geq xz + zy \geq qz - qx + zq - yq \geq 3\alpha - \alpha + 3\alpha - \alpha = 4\alpha.$$

Now let  $\sigma(\bar{q}_M, \bar{x}_K) < \alpha$  and  $\sigma(\bar{q}_M, \bar{y}_L) < \alpha$ . By (13.4) there is a curve  $c \sim \mathfrak{R}^{-1}\mathfrak{Q}$  such that  $\lambda(c) = \bar{x}_K \bar{y}_L < 2\alpha$ . Because of (13.6),  $c \sim \mathfrak{s}(x, y)$ , hence  $\mathfrak{s}(x, y) \sim \mathfrak{R}^{-1}\mathfrak{Q}$  and  $\bar{x}_K \bar{y}_L \leq \lambda(\mathfrak{s}(x, y)) = xy$ . But (13.3) implies  $\bar{x}_K \bar{y}_L \geq xy$ , which proves  $\bar{x}_K \bar{y}_L = xy$ .

This completes the proof that  $\bar{R}$  is a covering space of  $R$ . The space  $\bar{R}$  is simply connected (see P, Theorem 59). This property is characteristic for  $\bar{R}$ , as are various others, which all go back to

(13.7) *If  $R_\Phi$  is a simply connected and  $R_X$  any covering space of  $R$ , and  $\bar{p}_1, \bar{p}_2$  points of  $R_\Phi$  and  $R_X$  respectively which lie over the same point  $p$  of  $R$ , then a locally isometric mapping  $\Psi$  of  $R_\Phi$  on  $R_X$  exists with  $\bar{p}_1 \Psi = \bar{p}_2$  and  $\Psi X = \Phi$ .*

**Proof.** Consider any point  $\bar{x}_1$  of  $R_\Phi$ . A curve  $\mathfrak{R}_1\Phi$  from  $\bar{p}_1$  to  $\bar{x}_1$  lies over a curve  $\mathfrak{R}$  from  $p$  to  $x$ . By (12.13) there is exactly one continuous curve  $\mathfrak{R}_{2X}$  over  $\mathfrak{R}$  in  $R_X$  which starts at  $\bar{p}_2$ . Define  $\bar{x}_2 = \bar{x}_1 \Psi$  as the end point of  $\mathfrak{R}_{2X}$ . This

definition does not depend on the choice of  $\mathfrak{R}_{1\Phi}$ . For if  $\mathfrak{L}_{1\Phi}$  is any other curve from  $\bar{p}_1$  to  $\bar{x}_1$  then  $\mathfrak{L}_{1\Phi} \sim \mathfrak{R}_{1\Phi}$  because  $R_\Phi$  is simply connected, hence  $\mathfrak{L}_1 \sim \mathfrak{R}_1$  by the first part of (12.14). Since  $\mathfrak{L}_{2x}$  and  $\mathfrak{R}_{2x}$  start at the same point  $p_2$  it follows from the second part of (12.14) that they end at the same point  $\bar{x}_2$ . It is easily seen that  $\Psi$  maps  $R_\Phi$  on all of  $R_X$ . Finally it follows from (12.11) that  $\Psi$  maps  $S(\bar{x}_1, \eta_2(x)/3)$  isometrically on  $S(\bar{x}_2, \eta_2(x)/3)$ . The relation  $\Psi X = \Phi$  expresses merely that  $\bar{x}_1$  and  $\bar{x}_1\Psi$  lie over the same point  $x$  of  $R$ , which follows from the construction.

If  $R_X$  is also simply connected, it follows from (12.18) that  $\Psi$  is an isometry. This yields the following important fact:

(13.8) THEOREM. *A given E-space has a simply connected covering space. This space is unique up to isometries.*

This uniqueness in the metric sense justifies the term "the universal covering space."

The motion  $\Psi_\Phi$  of the covering space  $R_\Phi$  of  $R$  is said to lie over the motion  $\Psi$  of  $R$  if  $\Psi_\Phi\Phi = \Phi\Psi$ . (This means geometrically that  $\bar{x}\Psi_\Phi$  lies always over  $x\Psi$ , for then  $\bar{x}\Psi_\Phi\Phi = x\Psi = \bar{x}\Phi\Psi$  for all  $\bar{x} \in R_\Phi$ .) Two motions  $\Psi_{1\Phi}$ ,  $\Psi_{2\Phi}$  of  $R_\Phi$  lie over the same motion  $\Psi$  of  $R$  if and only if  $\Psi_{1\Phi}^{-1}\Psi_{2\Phi}$  lies over the identity  $E$  of  $R$ , because  $\Psi_{1\Phi}\Phi = \Phi\Psi = \Psi_{2\Phi}\Phi$  implies  $\Phi E = \Phi = \Psi_{1\Phi}^{-1}\Psi_{2\Phi}\Phi$  and conversely.

In general there is neither a motion  $\Psi_\Phi$  over a given motion  $\Psi$  of  $R$ , nor a motion  $\Psi$  of  $R$  over which a given motion  $\Psi_\Phi$  of  $R_\Phi$  lies. However

(13.9) *If  $R_\Phi$  is the universal covering space of  $R$ , then at least one motion of  $R_\Phi$  lies over a given motion  $\Psi$  of  $R$ .*

For a proof we go back to the construction of  $R_\Phi$  and the mapping  $\Phi$ . A point  $\bar{x}_K$  of  $\bar{R}$  corresponded to a point  $x$  of  $R$  and a class  $(\mathfrak{R})$  of curves from  $p$  to  $x$  homotopic to  $\mathfrak{R}$ . Under  $\Psi$  the class  $(\mathfrak{R})$  is transformed into a set of curves from  $p\Psi$  to  $x\Psi$ , which is the homotopy class  $(\mathfrak{R}\Psi)$  because  $\Psi$  is a motion. Define the mapping  $\Psi_\Phi$  of  $R_\Phi$  on itself by  $\bar{x}_K\Psi_\Phi = x^*\mathfrak{K}_\Phi$  where  $x^* = x\Phi$  and (later)  $y^* = y\Phi$ . The definition (13.3) yields  $\bar{x}_K\Psi_\Phi\bar{x}_L\Psi_\Phi = x^*\mathfrak{K}_\Phi y^*\mathfrak{L}_\Phi = \inf \lambda(c)$  with  $c \sim (\mathfrak{R}\Phi)^{-1}\mathfrak{Q}\Phi$ . Since  $\Phi$  is a motion,  $c\Phi^{-1} \sim \mathfrak{R}^{-1}\mathfrak{Q}$  and  $\lambda(c\Phi^{-1}) = \lambda(c)$  which shows that  $\bar{x}_K\Psi_\Phi\bar{y}_L\Psi_\Phi = \bar{x}_K\bar{y}_L$ . Hence  $\Psi$  is a motion of  $\bar{R}$ ; moreover  $\Psi_\Phi$  lies obviously over  $\Psi$ .

For the further discussion the following lemma will prove useful.

(13.10) If the motion  $\Psi$  of the E-space  $R$  leaves all points of a neighborhood  $S(p, \rho)$ ,  $\rho > 0$ , of  $p$  fixed, then  $\Psi$  is the identity.

It is to be shown that  $x\Psi = x$  for every  $x$  with  $\sigma(p, x) \geq \rho$ . A segment  $s(p, x)$  contains (by (1.4)) a point  $y$  with  $(pyx)$  and  $\sigma(p, y) < \rho$ . Then  $(p\Psi y\Psi x\Psi)$  and  $y\Psi x\Psi = yx$  because  $\Psi$  is a motion; and  $p\Psi = p$ ,  $y\Psi = y$  because  $p, y \in S(p, \rho)$ , hence  $(pyx\Psi)$  and  $yx\Psi = yx$ . Now (4.1) yields  $x\Psi = x$ .

It is now easy to prove

(13.11) *The motions of the universal covering space  $R_\Phi$  of the E-space  $R$  which lie over the identity of  $R$  form a group  $H_\Phi$ . The group  $H_\Phi$  is discrete, simply*



*transitive on the points  $\bar{x}$  which lie over a given point  $x$  of  $R$ , and isomorphic to the fundamental group of  $R$ .*

It is obvious that the motions of  $R_\Phi$  over  $E$  form a group  $H_\Phi$ . Let  $\bar{x}_1$  and  $\bar{x}_2$  be two points of  $R_\Phi$  over  $x$ . By (13.7) there is a locally isometric mapping of  $R_\Phi$  on itself which carries  $\bar{p}_1$  into  $\bar{p}_2$ ; this mapping is by (12.19) a motion.

Hence the orbit  $T(H_\Phi, \bar{x})$  (defined in §3) consists of all points of  $R_\Phi$  over  $x$ . If the motions  $\Psi_{1\Phi}$  and  $\Psi_{2\Phi}$  of  $H_\Phi$  both carry  $\bar{x}_1$  into  $\bar{x}_2$ , then  $\Psi_{1\Phi}^{-1}\Psi_{2\Phi}$  leaves, according to (12.11), every point of  $S(\bar{x}_2, \eta_2(x)/3)$  fixed. (13.10) shows that  $\Psi_{1\Phi}^{-1}\Psi_{2\Phi}$  is the identity so that  $H_\Phi$  is simply transitive on  $T(H_\Phi, \bar{x})$ .

The discreteness of  $H_\Phi$  refers to the topology (3.3) for the motions of  $R_\Phi$ , and is an immediate consequence of (12.12). The usual argument (compare Seifert-Threlfall [1, p. 197]) shows that  $H_\Phi$  is isomorphic to the fundamental group of  $R$ .

(13.12) *Let  $\Gamma^*$  be a group of motions of  $R$ . The motions of  $R_\Phi$  which lie over motions of  $\Gamma^*$  form a group  $\Gamma_\Phi^*$  and  $\Gamma_\Phi^*/H_\Phi \simeq \Gamma^*$ . If  $\Gamma^*$  is closed, transitive, or simply transitive, then  $\Gamma_\Phi^*$  has the same property.*

**Proof.** Let  $\Psi_{\kappa\Phi}$  lie over  $\Psi_\kappa$ ,  $\kappa = 1, 2$ . Then  $\Psi_{1\Phi}\Psi_{2\Phi}\Phi = \Psi_{1\Phi}\Phi\Psi_2 = \Phi\Psi_1\Psi_2$  and  $\Psi_{1\Phi}^{-1}\Phi = \Psi_{1\Phi}^{-1}\Phi\Psi_1\Psi_1^{-1} = \Psi_{1\Phi}^{-1}\Psi_{1\Phi}\Phi\Psi_1^{-1} = \Phi\Psi_1^{-1}$  which shows that  $\Gamma_\Phi^*$  is a group and that  $\Psi_{\kappa\Phi} \rightarrow \Psi_\kappa$  is a homomorphism. Its kernel consists of the motions which lie over the identity of  $R$ , hence  $\Gamma_\Phi^*/H_\Phi \simeq \Gamma^*$ .

Let the motion  $\Psi_{\nu\Phi}$ ,  $\nu = 1, 2, \dots$ , of  $R_\Phi$  lie over the motion  $\Psi_\nu$  of  $R$  and let  $\rho_{\bar{p}}(\Psi_{\nu\Phi}, \Psi_\Phi) \rightarrow 0$ . Then (see (3.12))  $\bar{x}\Psi_{\nu\Phi} \rightarrow \bar{x}\Psi_\Phi$  for every  $\bar{x} \in R_\Phi$ , hence  $\bar{x}\Psi_{\nu\Phi}\Phi = \bar{x}\Phi\Psi_\nu = x\Psi_\nu \rightarrow \bar{x}\Psi_\Phi$ . Putting  $\bar{x}\Psi_\Phi\Phi = x\Psi$ , it follows from (3.12) that  $\Psi$  is a motion of  $R$  and that  $\rho_p(\Psi_\nu, \Psi) \rightarrow 0$ . Moreover  $\Psi_\Phi$  lies over  $\Psi$  since  $x\Psi = \bar{x}\Phi\Psi$ . If  $\Gamma^*$  is closed, then  $\Psi \in \Gamma^*$ , hence  $\Psi_\Phi \in \Gamma_\Phi^*$  so that  $\Gamma_\Phi^*$  is closed.

Let  $\Gamma^*$  be transitive and  $\bar{x}$  and  $\bar{y}$  any two points of  $R_\Phi$ . There is a motion  $\Psi$  in  $\Gamma^*$  which carries  $x$  into  $y$ . Any motion  $\Psi_\Phi$  over  $\Psi$  carries  $\bar{x}$  into a point  $\bar{y}'$  over  $y$ . By (13.11) an element  $\Xi$  in  $H_\Phi$  exists which carries  $\bar{y}'$  into  $\bar{y}$ ; then  $\bar{x}\Psi_\Phi\Xi = \bar{y}$  and  $\Psi_\Phi\Xi \in \Gamma_\Phi^*$  because  $H_\Phi \subset \Gamma_\Phi^*$ , so that  $\Gamma_\Phi^*$  is transitive.

If  $\Psi$  is the only motion in  $\Gamma^*$  which carries  $x$  into  $y$  then any two motions  $\Psi_{1\Phi}$ ,  $\Psi_{2\Phi}$  of  $\Gamma_\Phi^*$  which carry  $\bar{x}$  into  $\bar{y}$  must lie over  $\Psi$ , or  $\Psi_{1\Phi}^{-1}\Psi_{2\Phi}$  must lie over the identity. This means  $\Psi_{1\Phi}^{-1}\Psi_{2\Phi} = \Xi \in H_\Phi$  and  $\bar{y} = \bar{y}\Psi_{1\Phi}^{-1}\Psi_{2\Phi} = \bar{y}\Xi$ . Since  $H_\Phi$  is simply transitive on  $T(H_\Phi, y)$  it follows that  $\Xi$  is the identity or  $\Psi_{1\Phi} = \Psi_{2\Phi}$ .

$H_\Phi$  is a normal subgroup of the group  $\Gamma_\Phi$  of all motions of  $R_\Phi$  which lie over motions of  $R$ , or  $\Psi_\Phi H_\Phi = H_\Phi \Psi_\Phi$  for every  $\Psi_\Phi \in \Gamma_\Phi$ . If  $\Xi_1, \Xi_2, \dots$  are the motions of  $H_\Phi$ , this means the existence of relations

$$(13.13) \quad \Xi_\kappa \Psi_\Phi = \Psi_\Phi \Xi_{\lambda(\kappa)} \quad \text{and} \quad \Psi_\Phi \Xi_\kappa = \Xi_{\mu(\kappa)} \Psi_\Phi, \quad \kappa = 1, 2, \dots$$

It will be shown that these relations are also sufficient for a motion  $\Psi_\Phi$  of  $R_\Phi$  to lie over a motion  $\Psi$  of  $R$ . Define a mapping  $\Psi$  of  $R$  on itself as follows: for a given point  $p$  of  $R$  choose any point  $\bar{p}$  over  $p$  (that is,  $\bar{p}\Phi = p$ ) and define the image  $p\Psi$  of  $p$  as the point  $\bar{p}\Psi_\Phi\Phi$ . The point  $p\Psi$  does not depend on the choice of  $\bar{p}$  over  $p$ . For if  $\bar{p}_1$  is another point over  $p$ , then a motion  $\Xi_\kappa \in H_\Phi$

exists with  $\bar{p}\Xi_\kappa = \bar{p}_1$ . Then by (13.13),  $\bar{p}_1\Psi_\Phi\Phi = \bar{p}\Xi_\kappa\Psi_\Phi\Phi = \bar{p}\Psi_\Phi\Xi_{\lambda(\kappa)}\Phi = \bar{p}\Psi_\Phi\Phi = p\Psi$ .

The definition  $\bar{p}\Phi_\Phi\Phi = p\Psi = \bar{p}\Phi\Psi$  implies  $\Psi_\Phi = \Phi\Psi$ , hence it remains only to be shown that  $\Psi$  is a motion, or that  $pq = p\Psi q\Psi$  for all  $p, q$ .

By (12.8),  $\bar{p}$  and  $\bar{q}$  over  $p$  and  $q$  exist such that  $\bar{p}\bar{q} = pq$ . Then (12.6) yields  $pq = \bar{p}\bar{q} = \bar{p}\Psi_\Phi\bar{q}\Psi_\Phi \geq \bar{p}\Psi_\Phi\Phi\bar{q}\Psi_\Phi\Phi = p\Psi q\Psi$ . Next choose by (12.8) a point  $\bar{r}$  over  $q\Psi$  such that  $\bar{p}\Psi_\Phi\bar{r} = p\Psi q\Psi$ . Since  $\bar{r}$  and  $\bar{q}\Psi_\Phi$  lie over  $q\Psi$ , there is a  $\Xi_\kappa \in H_\Phi$  with  $q\Psi_\Phi\Xi_\kappa = \bar{r}$ . Then by (13.13),  $p\Psi q\Psi = \bar{p}\Psi_\Phi\bar{r} = \bar{p}\Psi_\Phi q\Psi_\Phi\Xi_\kappa = \bar{p}\Psi_\Phi\bar{q}\Xi_{\mu(\kappa)}\Psi_\Phi = \bar{p}\bar{q}\Xi_{\mu(\kappa)} \geq pq$ . This proves

(13.14) *The motion  $\Psi_\Phi$  of the universal covering space  $R_\Phi$  of  $R$  lies over a motion of  $R$  if and only if  $\Psi_\Phi H_\Phi = H_\Phi \Psi_\Phi$ .*

Any  $E$ -space  $R$  whose fundamental group is not abelian furnishes examples which show that in general a motion  $\Psi_\Phi$  in  $\Gamma_\Phi$  will not commute with every element of  $H_\Phi$ . Killing observed in the case of Riemann spaces that a motion  $\Psi_\Phi$  of  $\Gamma_\Phi$ , which belongs to a one-parameter subgroup of motions over  $R$ , commutes with every element of  $H_\Phi$  (compare Cartan [1, p. 88]). That this is also true for  $E$ -spaces will follow from the following lemma.

(13.15) Let  $\Psi_\Phi, X_\Phi \in \Gamma_\Phi$ . If motions  $\Psi_{\nu\Phi}$  over  $\Psi_\Phi$ , with  $\Psi_{\nu\Phi}^{\lambda_\nu} = \Psi_\Phi$ ,  $\Psi_{\nu\Phi} \rightarrow E_\Phi$  and  $\Psi_\Phi X = X\Psi_\Phi$ , exist then  $\Psi_\Phi X_\Phi = X_\Phi \Psi_\Phi$ .

Since the mapping of  $\Gamma_\Phi$  on  $\Gamma$  which associates a motion of  $R_\Phi$  with the motion of  $R$  over which it lies is a homomorphism,  $\Psi_{\nu\Phi} X_\Phi$  and  $X_\Phi \Psi_{\nu\Phi}$  lie over the same motion  $\Psi_\Phi X = X\Psi_\Phi$ . By (13.13) there is a motion  $\Xi_\nu$  in  $H_\Phi$  with  $\Xi_\nu \Psi_{\nu\Phi} X_\Phi = X_\Phi \Psi_{\nu\Phi}$ . Then it follows from (3.6), (3.7), (3.8), if  $\rho_{\bar{r}}$  is simply denoted by  $\bar{\rho}$ , that

$$\begin{aligned}\bar{\rho}(E_\Phi, \Xi_\nu) &= \bar{\rho}(\Psi_{\nu\Phi} X_\Phi, \Xi_\nu \Psi_{\nu\Phi} X_\Phi) = \bar{\rho}(\Psi_{\nu\Phi} X_\Phi, X_\Phi \Psi_{\nu\Phi}) \\ &\leq \bar{\rho}(\Psi_{\nu\Phi} X_\Phi, X_\Phi) + \bar{\rho}(X_\Phi, X_\Phi \Psi_{\nu\Phi}) = \bar{\rho}(\Psi_{\nu\Phi}, E_\Phi) + \bar{\rho}_q(E_\Phi, \Psi_{\nu\Phi}) \\ &\leq \bar{\rho}(\Psi_{\nu\Phi}, E_\Phi)(1 + \exp \sigma(\bar{p}, \bar{p}X_\Phi)), \quad \text{where } q = \bar{p}X_\Phi.\end{aligned}$$

The right side tends to 0 because  $\Psi_{\nu\Phi} \rightarrow E_\Phi$ , hence  $\bar{\rho}(E_\Phi, \Xi_\nu) \rightarrow 0$ . Since  $H_\Phi$  is discrete (compare (13.11)),  $\Xi_\nu = E_\Phi$  for large  $\nu$ , hence  $\bar{\rho}(\Psi_{\nu\Phi} X_\Phi, X_\Phi \Psi_{\nu\Phi}) = 0$  for large  $\nu$ , so that also  $\Psi_{\nu\Phi}^{\lambda_\nu} X_\Phi = X_\Phi \Psi_{\nu\Phi}^{\lambda_\nu}$  or  $\Psi_\Phi X_\Phi = X_\Phi \Psi_\Phi$ .

The assumption  $\Psi_\Phi X = X\Psi_\Phi$  is trivially satisfied when  $X = E$  or  $X_\Phi \in H_\Phi$ . When  $\Psi_\Phi$  belongs to a one-parameter subgroup of  $\Gamma_\Phi$ , the other assumptions are satisfied, so that Killing's theorem is contained in the following corollary of (13.15).

(13.16) *If  $\Psi_\Phi \in \Gamma_\Phi$  and if a sequence of motions  $\Psi_{\nu\Phi} \in \Gamma_\Phi$  exists which tends to the identity and such that  $\Psi_{\nu\Phi}^{\lambda_\nu} = \Psi_\Phi$  then  $\Psi_\Phi \Xi_\kappa = \Xi_\kappa \Psi_\Phi$  for every  $\Xi_\kappa \in H_\Phi$ .*

The existence of the universal covering space yields the existence of other covering spaces by means of

(13.17) **THEOREM.** *If  $\Delta$  is a discrete group of motions without fixed points of the  $E$ -space  $R$ , then the orbits  $x^* = T(\Delta, x)$  form with the metric  $x^*y^* = T(\Delta, x)T(\Delta, y)$  an  $E$ -space  $R^*$ , and  $x \rightarrow x^*$  is a locally isometric mapping  $\Psi$  of  $R$  on  $R^*$  (so that  $R$  is the covering space  $R_\Psi^*$  of  $R^*$ ).*

**Proof.** As a discrete group  $\Delta$  is closed. By (3.23),  $R^*$  is finitely compact and convex.

For a given point  $x$  there is a  $\rho > 0$  such that  $\sigma^*(z, z\Phi) > \rho$  for  $\Phi \neq E$  in  $\Delta$  and  $z \in S(x, 1)$ . Otherwise sequences of points  $z_\nu \in S(x, 1)$  and of motions  $\Phi_\nu \neq E$  in  $\Delta$  would exist such that  $\sigma^*(z_\nu, z_\nu\Phi_\nu) \rightarrow 0$ . Then  $\sigma^*(x, x\Phi_\nu) \leq \sigma(x, z_\nu) + \sigma^*(z_\nu, z_\nu\Phi_\nu)$  would show that  $\sigma(x, x\Phi_\nu)$  and  $\rho_\nu(E, \Phi_\nu)$  are bounded (compare (1.17) and (3.9)), hence  $\{\Phi_\nu\}$  would contain a converging subsequence  $\{\Phi_\mu\}$  (see (3.13)), whose limit  $\Phi$  lies in  $\Delta$ . Since  $\Delta$  is discrete,  $E \neq \Phi_\mu = \Phi$  for large  $\mu$ . On the other hand  $\Phi$  would leave any accumulation point of  $\{z_\mu\}$  fixed.

Now let  $\eta = \min(1, \rho/5)$  and  $y, z \in S(x, \eta)$ . The relation  $yz' \geq zz' - yz \geq \sigma^*(z, z') - 2\eta$  holds for any  $z' \in T(\Delta, z)$ . If  $z' \neq z$ , that is,  $z' = z\Phi$ ,  $\Phi \in \Delta$ , then  $\sigma^*(z, z') > \rho$  and  $yz' > 3\rho/5 > yz$ . Hence if  $z'$  is such that  $yz' = y^*z^* \leq yz$  (compare the proof of (3.23)) then  $z' = z$ . This shows that  $x \rightarrow x^*$  maps  $S(x, \eta)$  isometrically in  $S(x^*, \eta)$ . Conversely, if  $y^* \in S(x^*, \eta)$  is given then  $y' \in T(\Delta, y)$  with  $xy' = x^*y^*$  exists and it follows as before that  $y' \in S(x, \eta)$ , so that  $x \rightarrow x^*$  maps  $S(x, \eta)$  on  $S(x^*, \eta)$ .

The proof of (12.2) shows that  $R^*$  satisfies Axiom D at  $x^*$ , hence  $R^*$  is an  $E$ -space and  $R = R_\Psi^*$  a covering space of  $R^*$ .

This leads to the following fact, whose topological analogue is well known:

(13.18) The fundamental group  $\Delta$  of any covering space  $R_X$  of an  $E$ -space  $R$  is isomorphic to a subgroup of the fundamental group  $H$  of  $R$ . For a given subgroup  $\Delta$  of  $H$  there is a covering space  $R_X$  of  $R$  with  $\Delta$  as fundamental group.

The first part is contained in the proof of (12.18).

Any subgroup  $\Delta$  of  $H$  satisfies the assumptions of (13.17) (because of (13.11)). The mapping  $\Psi$  which associates the point  $x^* = T(\Delta, \bar{x})$  of  $R^*$  with the point  $\bar{x}$  in the universal covering space  $R_\Phi$  of  $R$  is locally isometric, so that  $R_\Phi = R_\Psi^*$ , being simply connected, is the universal covering space of  $R^*$  and  $\Delta$  is the fundamental group of  $R^*$ . Moreover  $\Phi$  is the mapping  $T(H, \bar{x}) \rightarrow x$  of  $R_\Phi$  on  $R$ , whence it easily follows that  $T(\Delta, \bar{x}) \rightarrow T(H, \bar{x})$  is a locally isometric mapping  $X$  of  $R^*(=R_X)$  on  $R$  and that  $\Psi X = \Phi$ .

**14. Some applications.** By (8.8) the only one-dimensional  $G$ -spaces are the straight line and the great circle. The former is the universal covering space of the latter.

A two-dimensional  $G$ -space is always a manifold (see B [1, chapter I, §4, Theorem 4]) hence its universal covering space is homeomorphic either to the two-sphere or to the  $E^2$ .

The three simplest types of higher dimensional  $G$ -spaces may be characterized as follows:

(1) All geodesics are straight lines (or  $\eta_2(x) = \infty$ ). Menger [1] called these spaces *straight line spaces* (Geradenräume).

(2) All geodesics are great circles of the same length and two different geodesics intersect at most once. These spaces will be said to be of the *elliptic type*.

(3) *Spherelike spaces*, as defined at the end of §9.

Types (2) and (3) are not essentially different. In the one-dimensional case they are identical. If the dimension is greater than one, the mapping of a spherelike space  $\bar{R}$  on itself which associates with every point  $p$  its antipodal point  $p'$  is a motion  $\Psi$  (compare §9).  $\Psi$  forms together with the identity a group  $\Delta$  which satisfies (13.17). Hence identification of antipodal points in  $\bar{R}$  yields a  $G$ -space which is easily seen to be of the elliptic type (see B [1, Theorem 1]).

Conversely a space of the elliptic type and dimension not less than 2 has a spherelike space as universal covering space. The straight line spaces and the spaces of the elliptic type are the only  $G$ -spaces in which the geodesic through two points is unique. The proof of these facts in B [2] rests on the following lemma:

(14.1) *If in a  $G$ -space  $R$  the geodesic connecting two points is unique, and  $R$  contains a great circle, then  $R$  is not simply connected.*

In the case of a general  $G$ -space the direct construction of a covering space as in B [2] seems to be the simplest way to establish (14.1). We observed after (8.8) that the geodesics of  $R$  are all great circles or straight lines. It is worth noting that this fact makes (14.1) trivial for a  $\nu$ -dimensional Finsler space  $R$  in the usual sense. For  $R$  can then be mapped topologically in the  $\nu$ -dimensional elliptic space  $R^*$  in such a way that each geodesic through a fixed point  $p$  is mapped in an elliptic straight line (great circle) through the image  $p^*$  of  $p$ . A great circle  $\mathcal{G}$  through  $p$  will be mapped on an elliptic line  $\mathcal{G}^*$  through  $p^*$ . Since  $\mathcal{G}$  cannot be contracted to a point in  $R^*$ , it cannot be contracted in the subset of  $R^*$  which is the image of  $R$ , hence  $\mathcal{G}$  cannot be contracted in  $R$ .

If (14.1) is granted the above mentioned results will follow from (14.2) (which is independent of (14.1)).

(14.2) *If all geodesics of a  $G$ -space  $R$  are straight and if  $R$  is not simply connected and has dimension not less than 2, then  $R$  is of the elliptic type and has a spherelike space as two-sheeted universal covering space.*

Consider the universal covering space  $R_\Phi$  of  $R$ . Since the geodesics of  $R$  are straight, their carriers are closed point sets, hence it follows easily by (12.1), (12.3) that geodesics in  $R_\Phi$  are carried by closed sets.

The geodesics of  $R_\Phi$  are straight. For if the geodesic  $\mathcal{G}_\Phi$  of  $R_\Phi$  over  $\mathcal{G}$  in  $R$  were not straight it would contain two points  $\bar{p}$  and  $\bar{q}$  such that  $\mathcal{G}_\Phi$  contains no segment  $t(\bar{p}, \bar{q})$ . Since the carrier of  $\mathcal{G}_\Phi$  is closed, the segments contained in  $\mathcal{G}_\Phi$  form a closed subset among all segments. Therefore  $\bar{r}$  may be chosen on  $\mathcal{G}_\Phi$  close to  $\bar{q}$  such that  $\mathcal{G}_\Phi$  contains no segment  $t(\bar{p}, \bar{r})$  and  $r$  is neither identical with nor conjugate to  $p$  on  $\mathcal{G}$ .

Then a segment  $t(\bar{p}, \bar{r})$  belongs to a geodesic  $\mathcal{G}_{1\Phi}$  which lies over a geodesic  $\mathcal{G}_1 \neq \mathcal{G}$  in  $R$ . The geodesics  $\mathcal{G}_1$  and  $\mathcal{G}$  would be straight and intersect at points  $p$  and  $q$  which are not conjugate on  $\mathcal{G}$ . This contradicts the following fact:

(14.3) Two different straight geodesics  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  of a  $G$ -space have not more than two common points. If they have the two common points  $p$  and  $q$ , then  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  are great circles of the same length and  $p$  and  $q$  are conjugate to each other on both geodesics.

For if  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  have  $p$  and  $q$  in common, then both  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  contain segments  $t(p, q)$  since they are straight, and it follows from (9.6), (9.7) that  $p$  and  $q$  are conjugate to each other on  $\mathfrak{G}_1$  as well as  $\mathfrak{G}_2$ . Hence  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  are great circles of the same length  $2pq$ . Moreover  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  cannot have a third common point  $r$  because  $(prq)$  for  $r \neq p, q$  on  $\mathfrak{G}_\kappa$  and the segment  $t(p, r)$  is therefore unique.

There is only one geodesic of  $R_\Phi$  over a given geodesic  $\mathfrak{G}$  of  $R$ . For if both  $\mathfrak{G}_{1\Phi}$  and  $\mathfrak{G}_{2\Phi}$  lie over  $\mathfrak{G}$ , choose  $p_1$  and  $p_2$  on  $\mathfrak{G}$  such that  $p_2$  is neither identical with nor conjugate to  $p_1$ . Let  $\tilde{p}_\kappa \in \mathfrak{G}_{\kappa\Phi}$  lie over  $p_\kappa$ ,  $\kappa = 1, 2$ , and consider a geodesic  $\mathfrak{F}_\Phi$  in  $R_\Phi$  through  $\tilde{p}_1$  and  $\tilde{p}_2$ . Because both  $\mathfrak{G}$  and  $\mathfrak{F}$  contain  $p_1$  and  $p_2$  and  $p_2$  is not conjugate to  $p_1$  on  $\mathfrak{G}$  it follows from (14.3) that  $\mathfrak{G} = \mathfrak{F}$ . Since  $\mathfrak{F}_\Phi$  and  $\mathfrak{G}_{\kappa\Phi}$  lie both over  $\mathfrak{G}$  and have the common point  $\tilde{p}_\kappa$ , (12.4) yields  $\mathfrak{G}_{\kappa\Phi} = \mathfrak{F}_\Phi$  and  $\mathfrak{G}_{1\Phi} = \mathfrak{G}_{2\Phi}$ .

Since there is only one geodesic  $\mathfrak{G}_\Phi$  over a given geodesic  $\mathfrak{G}$  of  $R$ , it follows from (12.4) that  $\mathfrak{G}_\Phi$  must contain all points which are over a given point  $p$  of  $\mathfrak{G}$ . Therefore two different geodesics through  $\tilde{p}$  have all points over  $p$  in common. There are at least two such points since  $R$  is not simply connected, and at most two by (14.3), which shows also that the geodesics of  $R_\Phi$  are all great circles. Hence  $R_\Phi$  is spherelike.

Finally the geodesic of  $R_\Phi$  through two points  $\tilde{p}$  and  $\tilde{q}$  which do not lie over the same point  $p$  of  $R$  is, by (14.3), unique, since each geodesic through  $\tilde{p}$  and  $\tilde{q}$  contains also the other points over  $p$  and  $q$ . Therefore the geodesic  $\mathfrak{G}$  through  $p$  and  $q \neq p$  in  $R$  is unique. Obviously the geodesics of  $R$  have half the length of those in  $R_\Phi$ .

(14.1) and (14.2) yield

(14.4) THEOREM. *A  $G$ -space is of the elliptic type if and only if its geodesics are straight and the space is not simply connected.*

(14.5) THEOREM. *If the geodesic through two points of a  $G$ -space is unique then the space is either a straight line space or of the elliptic type.*

(14.4) leads to the problem of determining the simply connected  $G$ -spaces with straight geodesics. In two dimensions, (1) and (3) are the only types (see B [2, Theorem 6]).

The following theorem, which characterizes the  $E$ -spaces which admit a transitive abelian group of motions, is of prime importance for the transition from  $E$ -spaces to Finsler spaces in the usual sense (compare B [1, chapter II]).

(14.6) THEOREM. *If the  $E$ -space  $R$  admits a transitive abelian group  $\Delta$  of motions, then  $R$  is homeomorphic to the product of a finite number  $\nu$  of circles and straight lines, and the universal covering space  $R_\Phi$  of  $R$  is Minkowskian.*

( $\Delta$  and its universal covering group  $\Delta_\Phi$  are homeomorphic to  $R$  and  $R_\Phi$  respectively;  $\Delta_\Phi$  is isomorphic to the group of translations of the  $E^r$ .)

The main part of this theorem, that is, the structure of  $\Delta$  and  $\Delta_\Phi$ , is an immediate consequence of Theorem 43 in Pontrjagin [1]. But the importance of (14.6) for metric geometry makes a separate proof desirable. The following proof is a compromise. While using Pontrjagin's result, it gives enough metric details to provide the tools for a direct approach (compare a similar proof in B [1, chapter IV, §8]).

If  $\Delta$  is metrized by (3.3), it follows from (3.24) that for any point  $a$  of  $R$  the correspondence  $\Psi \rightarrow a\Psi$  maps  $\Delta$  isometrically on  $R$  with the metric  $\sigma$ , so that  $\Delta$  and  $R$  are homeomorphic. The extremals of  $R$  are one-parameter subgroups of  $\Delta$ . This follows immediately from

(14.7) If  $ab = bc = ac/2$  and  $a\Psi = b$  then  $a\Psi^2 = c$ .

First let there be only one point  $y$  with  $ay = yc = ac/2$ . Denote by  $X$  the motion of  $\Delta$  which carries  $b$  into  $c$ . Then  $aX\Psi = a\Psi X = c$ ; hence  $aXc = aXa\Psi X = aa\Psi = ac/2$ . Moreover by (3.24),  $aaX = bbX = bc = bc/2$ . Because of the uniqueness of  $y$ , the point  $aX$  must coincide with  $b = a\Psi$ . Then  $X = \Psi$ , because  $\Delta$  is simply transitive (see (3.24)).

If  $b$  is not the only point  $y$  with  $ay = yc = ac/2$ , let  $ad = db = ab/2$ . The point  $d$  is by (4.1) unique. Let  $a\Omega = d$ . The first part of this proof yields  $a\Omega^2 = b$ , hence  $\Omega^2 = \Psi$  and  $a\Psi^2 = a\Omega^4 = c$  because  $(aa\Omega a\Omega^2)$  implies  $(a\Omega a\Omega^2 a\Omega^3)$ .

Consider the universal covering space  $R_\Phi$  of  $R$  and the group  $\Delta_\Phi$  of motions of  $R_\Phi$  which lie over elements of  $\Delta$ . By (3.24) and (13.12),  $\Delta_\Phi$  is closed and simply transitive. It then follows easily from (12.3), (12.4) and (14.5) that the extremals of  $R$  are one-parameter subgroups of  $\Delta_\Phi$ .

$\Delta_\Phi$  is also abelian. For since the extremals are one-parameter subgroups, every motion  $\Psi_\Phi$  of  $\Delta_\Phi$  belongs to such a subgroup, so that motions  $\Psi_{\Phi\Phi}$  in  $\Delta_\Phi$  with  $\Psi_{\Phi\Phi} = \Psi_\Phi$  and  $\Psi_{\Phi\Phi} \rightarrow E_\Phi$  exist. If  $X_\Phi$  is any motion of  $\Delta_\Phi$ , then  $\Psi_\Phi X_\Phi = X_\Phi \Psi_\Phi$ , because  $\Delta$  is abelian, hence it follows from (13.15) that  $\Psi_\Phi X_\Phi = X_\Phi \Psi_\Phi$ . Since  $\Delta_\Phi$  is abelian  $\Psi_\Phi \rightarrow a\Psi_\Phi$  maps  $\Delta_\Phi$  isometrically on  $R_\Phi$  with the metric  $\sigma$  (see (3.24)).

By Pontrjagin [1, Theorem 43] both  $\Delta$  and  $\Delta_\Phi$  are direct products of a finite number of groups isomorphic to the additive group of real numbers (straight lines) and an at most countable number of groups isomorphic with the additive group of real numbers mod 1 (circles). Since  $\Delta_\Phi$  is simply connected, it is isomorphic to the product of some finite number  $\nu$  of groups of the first type, or  $\Delta_\Phi$  is isomorphic to the group of translations of the  $E^r$ .

Therefore  $\Delta$ , being covered by  $\Delta_\Phi$ , is isomorphic to the product of  $\rho$ ,  $0 \leq \rho \leq \nu$ , groups of the first type and  $\nu - \rho$  groups of the second type.

The metric of  $R_\Phi$  is now easily determined. Choose coordinates  $\xi_1, \dots, \xi_r$  in  $E^r$  such that the motions of  $\Delta_\Phi$  are given by the translations  $\xi'_\kappa = \xi_\kappa + \alpha_\kappa$ . Then the distance  $xy$  of two points  $x = (\xi_\kappa)$  and  $y = (\eta_\kappa)$  has the form  $\rho(\xi_\kappa - \eta_\kappa)$ . Let  $(a \rightarrow b)$  denote the motion which carries  $a$  into  $b$ . If  $\zeta_\kappa = (1/2)(\xi_\kappa + \eta_\kappa)$  and

( $\xi_\kappa$ ) =  $z$ , then  $(x \rightarrow z)$  is the only translation whose square equals  $(x \rightarrow y)$ . If  $c$  is a center of  $x$  and  $z$  it follows from (14.7) that  $(x \rightarrow c)^2 = (x \rightarrow y)$ , so that  $c$  is unique and coincides with  $z$ . This implies  $\rho(\kappa\xi) = \kappa\rho(\xi)$  for  $\kappa > 0$ .

To obtain the notation of the example  $R^2$  in §2 put  $\xi/\rho(\xi) = \eta(\xi)$  when  $\xi \neq (0, \dots, 0)$ . Then  $xy$  will be Minkowskian when  $\rho(\xi) = 1$  is a strictly convex surface  $K$  of the  $E^n$ .

It is well known that the triangle inequality (1.3) in conjunction with the uniqueness of the center  $c$  of  $x$  and  $y$  is equivalent to the strict convexity of  $K$  (compare Minkowski [1, §§17, 18] or B [1, pp. 34–36]).

**15. Characterization of the spaces which are locally euclidean, hyperbolic or spherical.** Spaces with constant curvature will be the last subject of the present paper. They will be characterized first by properties of their bisectors. The bisector  $B(a, a')$  of two distinct points  $a, a'$  (the notation  $B(a, a')$  will henceforth imply that  $a \neq a'$ ) in the  $G$ -space  $R$  is the locus of those points  $x$  in  $R$  for which  $xa = xa'$ . Notice that the bisector  $B(a, a')$  has this property:

(15.1)  $aB(a, a') = a'B(a, a') = aa'/2$ ; hence the feet of  $a$  (or  $a'$ ) on  $B(a, a')$  are the centers of  $a$  and  $a'$ .

For, every center  $c$  of  $a$  and  $a'$  belongs to  $B(a, a')$  and any point  $x$  of  $B(a, a')$  satisfies the inequality  $2ax = ax + xb \geq ab = 2ac$ , so that  $c$  is a foot of  $a$  on  $B(a, a')$ . If  $x$  is any foot of  $a$ , then  $2ax = ab = 2ac$  and  $ax = xb$ , so that  $x$  is a center of  $a$  and  $b$ .

$S(p, \rho(p))$  is briefly said to have *linear bisectors*, if the following holds.

(\*) *For any two distinct points  $a, a'$  in  $S(p, \rho(p))$  the bisector  $B(a, a')$  contains with any two points  $x, y$  at least one segment  $t(x, y)$ .*

The main result, from which the various characterizations of spaces with constant curvature will follow, is this:

(15.2) *If  $S(w, \rho)$ ,  $0 < \rho \leq \eta(w)$ , has linear bisectors, then  $S(w, \rho)$  is congruent to the interior of a sphere of a finite dimensional euclidean, spherical, or hyperbolic space.*

The special case of a  $G$ -space with  $\eta_2(p) \equiv \infty$  (that is, all geodesics are straight lines) and  $\rho = \infty$  (that is, the bisectors are linear in the large) is the subject of B [1, chapter IV, §8]. The central point of the proof is that a certain topological mapping  $\Omega(B)$  (compare (15.16)) is a motion. In the special case, the existence of the mapping  $\Omega(B)$  offers no difficulty, but in the present case it requires some involved local considerations. After the existence and certain topological facts have been established, the procedure of B [1] may be followed.

(15.2) is trivial for one-dimensional  $G$ -spaces, because they are by (8.8) great circles or straight lines. Therefore it may be assumed that the space, and then by (4.12) every  $S(x, \beta)$ ,  $\beta > 0$ , has at least dimension 2.

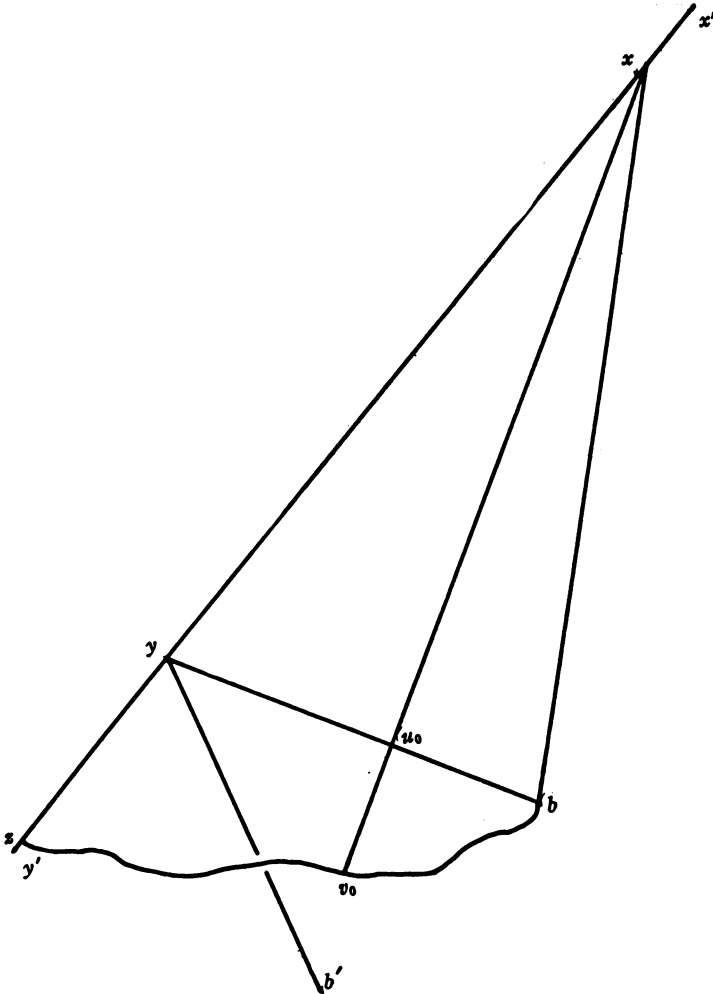
Observe first that with  $S(w, \rho)$  any sphere  $S(p, \sigma)$  which is contained in  $S(w, \rho)$  has linear bisectors, in particular every  $S(p, \bar{\rho}(p))$  with  $p \in S(w, \rho)$  and

$\bar{\rho}(p) = \rho - wp$ . The relations (4.9) and  $\bar{\rho}(w) = \rho \leq \eta(w)$  imply  $\bar{\rho}(p) \leq \eta(p)$ ; moreover  $|\bar{\rho}(p) - \bar{\rho}(q)| = |wq - wp| \leq pq$ . For convenience put

$$(15.3) \quad 6\rho(p) = \rho'(p) = \bar{\rho}(p)/7, \quad \text{so that} \quad |\rho(p) - \rho(q)| < pq/40.$$

The whole discussion rests on two lemmas which will be proved next:

(15.4) *If  $x, y \in B(b, b')$ , where  $x, y, b, b'$  lie in  $S(p, \rho'(p))$ , then a segment  $t(x', y')$  (with  $(xyy')$ ) of length  $6\rho'(p)$  exists which contains  $t(x, y)$  as co-central subsegment and  $t(x', y') \subset B(b, b')$ .*



**Proof.**  $t(x', y')$  exists because  $\rho'(p) < \eta(p) \leq \eta_6(p)$ . Assume for an indirect proof that some point  $z$  of  $t(x', y')$  does not belong to  $B(b, b')$ . Because



$S(p, \bar{p}(p))$  has linear bisectors  $z$  cannot belong to  $t(x, y)$ . Without loss of generality let  $(xyz)$  and  $b'z < bz$ . The segment  $t(b, y)$  lies in  $S(p, 2\rho'(p))$  (compare (1.15)). Therefore any point  $u$  of  $t(b, y)$  satisfies the inequality

$$(15.5) \quad xu + yz \cdot ub/yb \leq xu + yz \leq xp + pu + yy' < 6\rho'(p).$$

There is a segment  $t(x'', u')$  (with  $(xuu')$ ) which contains  $t(x, u)$  as co-central subsegment and has length  $6\eta_6(p)$ . Then  $u'x > x''u/2 = 3\eta_6(p) > 6\rho'(p)$ , so that  $t(x'', u')$  contains by (15.5) a point  $v$  with  $xv = xu + yz \cdot ub/yb$ .

The point  $v$  depends continuously on  $u$ . Now  $b'v < bv$  for  $u=b$  because then  $v=z$ ; and  $b'v > bv$  for  $u=b$  because then  $v=b$ . Therefore  $t(b, y)$  contains a point  $u_0$  such that for the corresponding point  $v_0$  the relation  $b'v_0 = bv_0$  or  $v_0 \in B(b, b')$  holds.  $xv_0 < 6\rho'(p)$  and  $xp < \rho'(p)$  show that  $pv_0 < 7\rho'(p) = \bar{p}(p) \leq \eta(p)$ . Hence  $t(x, v_0)$  is unique and lies on  $B(b, b')$  because  $S(p, \bar{p}(p))$  has linear bisectors. In particular  $u_0 \in B(b, b')$  so that  $b'u_0 = bu_0 = by - yu_0 = b'y - yu_0$ . But then (4.1) would yield  $b' = b$ .

The second lemma is a consequence of the first:

(15.6) *Let  $c$  be the center of  $b$  and  $b' \neq b$ , where  $b, b' \in S(p, \rho'(p)/2)$ . If  $c \neq x \in B(b, b') \cap S(p, \rho'(p))$  and  $x'c = cx = xx'/2$ , then  $t(x, x') \subset B(b, b)$ ,  $t(b, b') \subset B(x, x')$ , and  $xb = x'b = xb' = x'b'$ .*

**Proof.** The end point  $c'$  of the segment  $t(x'', c')$  with  $(xcc')$  which contains  $t(x, c)$  as co-central subsegment and has length  $6\rho'(p)$  has from  $b$  distance  $c'b \geq c'x - xb \geq 3\rho'(p) - 3\rho'(p)/2 > xb$ , and, by (15.1),  $xb > cb$ . Hence  $t(x'', c')$  contains a point  $x^*$  with  $(xcx^*)$  and  $bx^* = bx$ . By (15.4),  $t(x'', c')$  lies on  $B(b, b')$ , in particular  $x^*b' = x^*b = xb = xb'$ . Since  $S(p, \bar{p}(p))$  has linear bisectors,  $t(x, x^*) \subset B(b, b')$  and  $t(b, b') \subset B(x, x^*)$ . The second relation implies  $cx = cx^* = xx^*/2$ . Therefore  $x^* = x'$  which proves (15.6).

This result yields

(15.7) A point  $p$  of  $S(w, \rho)$  has exactly one foot on any segment  $t$  that contains points of  $S(p, \rho'(p)/2)$ .

For, any foot of  $p$  on  $t$  lies in  $S(p, \rho'(p)/2)$ . If  $p$  had two different feet  $f_1$  and  $f_2$  on  $t$ , then  $pf_1 = pf_2$ , hence  $p \in B(f_1, f_2)$ . If  $c$  is the center of  $f_1$  and  $f_2$  and  $p'c = cp = p'p/2$  it follows from (15.6) that  $t(f_1, f_2) \subset B(p, p')$ . By (15.1),  $c$  would be the only foot of  $p$  on  $B(p, p')$ , hence also on  $t(f_1, f_2)$ .

A corollary of (15.7) and (7.11) is

(15.8) *The spheres  $K(p, \sigma)$  with  $p \in S(w, \rho)$  and  $0 < \sigma < 3\rho(p)$  are strictly convex.*

Now consider a definite point  $p$  in  $S(w, \rho)$ ; let  $0 < \alpha \leq \rho(p)$  and put  $S = S(p, \alpha)$ .

(15.9) Let  $B(q, q')$ , where  $q, q' \in S(p, 40\rho(p))$ , contain points of  $S$ . Then any point  $x \in S$  has exactly one foot on  $B(q, q')$ .

For if  $y \in B(q, q') \cap S$  then  $xy \leq xp + py < 2\rho(p) < 3(\rho(p) - xp/40) < 3\rho(x)$ . If  $x$  had two different feet  $f_1, f_2$  on  $B(q, q')$  they would lie in  $S(x, 3\rho(x))$  and

would be feet of  $x$  on  $t(f_1, f_2)$  because  $t(f_1, f_2)$  lies in  $B(q, q')$  (this follows from  $40\rho(p) < \bar{\rho}(p)$ ).

Any segment  $t(a', b')$  whose end points lie on  $K(p, \alpha)$  lies by (15.8), except for  $a'$  and  $b'$ , in  $S$ . The open segment  $t(a', b') \cap S$  will be denoted as a line  $g$  in  $S$ , and the center of  $t(a', b')$  as the center of  $g$ . Any two distinct points  $a, b$  of  $S$  lie on exactly one line  $g(a, b)$ . The symbol  $B_p(q, q')$  will be used for the intersection  $B(q, q') \cap S$  when it is not empty and at least one of the points  $q, q'$  lies in  $S$ . It follows from (15.1) that  $qq' = 2qB(q, q') < 4\alpha$ .

The line  $g$  is *perpendicular* to the set  $D$  at the point  $f$ , if  $f \in g \cap D$  and every point of  $g$  has  $f$  as foot on  $D$ . Obviously  $f = g \cap D$  and, by (7.9),  $f$  is the unique foot of any point  $x \in g$  on  $D$ .

By (15.4),  $B_p(q, q')$  contains with any two points  $a, b$  the line  $g(a, b)$  and by (15.9), every point  $x$  of  $S$  has exactly one foot on  $B(q, q')$ . If  $f \in S$  then  $g(x, f)$  is perpendicular to  $B(q, q')$  and  $B_p(q, q')$ , and to all lines in  $B_p(q, q')$  through  $f$ .

(15.10) Let  $f$  and  $q_\kappa \neq f$ ,  $\kappa = 1, 2$ , be points of  $g$ , and  $q'_\kappa f = fq_\kappa = q'_\kappa q_\kappa / 2$ . Then  $B_p(q_1, q'_1) = B_p(q_2, q'_2)$ .

For if  $f \neq x \in B_p(q_1, q'_1)$  and  $x'f = fx = x'x/2$ , then by (15.6),  $t(q_1, q'_1) \subset B(x, x')$ , by (15.4),  $t(q_2, q'_2) \subset B(x, x')$ , and again by (15.6),  $t(x, x') \subset B(q_2, q'_2)$ , so that  $x \in B_p(q_2, q'_2)$  and  $B_p(q_1, q'_1) \subset B_p(q_2, q'_2)$ , q.e.d.

If  $B_p(q, q')$  contains the center  $f$  of  $q$  and  $q'$ , it depends by the last result only on  $f$  and the line  $g$  through  $q$  and  $q'$ . It may therefore be denoted by  $B_p(f, g)$ . Now the following two important facts may be established.

(15.11) The line  $g$  through the point  $f$  of  $B = B_p(a, a')$  is a perpendicular to  $B$  if  $g$  contains any point  $q^* \neq f$  whose foot on  $B$  is  $f$ .

(15.12) The line  $g$  is perpendicular to  $B_p(a, a')$  at  $f$  if and only if there is a pair of points  $(q, q')$  on  $g$  with center  $f$  such that  $B_p(q, q') = B_p(a, a')$ .

First let the assumption of (15.11) be satisfied (which is trivially the case when  $g$  is already perpendicular to  $B$ ) and choose  $q, q'$  on  $g$  with center  $f$  such that  $(q^*qf)$ . By (7.9),  $f$  is the unique foot of  $q$  on  $B$ . Let  $f \neq y \in B$ . Since  $f$  is also the unique foot of  $q$  on  $g(f, y)$ , this line contains points  $d, d'$  with  $(dfd')$  and  $qd = qd'$ . Let  $c$  be the center of  $d$  and  $d'$  and  $q''c = cq = q''q/2$ . Then  $t(q, q'') \subset B(d, d'')$  and  $t(d, d') \subset g(f, y) = g(d, d') \subset B_p(q, q'')$ . By (15.3),  $c$  is the foot of  $q$  on  $g(d, d')$ , hence  $f = c$  and  $q' = q''$ , so that  $y \in B_p(q, q')$  and  $B \subset B_p(q, q')$ . Similarly  $B_p(q, q') \subset B$ .

It has been shown that under the assumption of (15.11) points  $q, q'$  on  $g$  with center  $f$  exist such that  $B_p(q, q') = B$ . This establishes the necessity part of (15.12). Moreover, if  $y$  is any point of  $g$  and  $yf = fy' = yy'/2$  then, by (15.10),  $B_p(y, y') = B_p(q, q')$  so that  $f$  is the foot of  $y$  on  $B$  (see (15.1)); this proves (15.11) and the sufficiency part of (15.12).

Notice the following corollary of (15.12) and (15.6).

(15.13) If  $g$  is perpendicular to  $h$ , then  $h$  is perpendicular to  $g$ .  $B_p(f, g)$  is the locus of all lines through  $f$  and perpendicular to  $g$ .

(15.14) *There is exactly one perpendicular  $g(f)$  to  $B=B_p(a, a')$  at a given point  $f$  of  $B$ , and  $g(f)$  depends continuously on  $f$ .*

**Proof.** For an indirect proof of the uniqueness let there be two perpendiculars  $g$  and  $h$  through  $f$ . Let  $q$  and  $q'$  be points of  $g$  with center  $f$ . If  $x$  tends on  $h$  to  $f$ , then the line  $g(q, x)$  tends to  $g$ , and contains therefore a point  $y$  of  $B$ . Let  $xf=fx'=xx'/2$  and  $yf=fy'=yy'/2$ . By (15.12),  $B_p(q, q')=B_p(x, x')=B$ , hence by (15.6) both  $q$  and  $x$  would lie in  $B(y, y')$ , but then  $y \in g(q, x) \subset B(y, y')$  which is impossible.

It will be proved next: If  $f, \in B$  and  $f, \rightarrow f \in B$  and the perpendiculars  $g(f,)$  exist, then  $g(f)$  exists and  $g(f,)\rightarrow g(f)$ . The relation  $S(f, \alpha - pf) \subset S$  shows that  $g(f,)$  contains for large  $\nu$  a point  $q_\nu$  with  $q_\nu f, = (\alpha - pf)/2$  and  $q_\nu a > q_\nu a'$ . Since  $f,$  is a foot of  $q,$  on  $B$ , every accumulation point  $q$  of  $\{q_\nu\}$  has  $f$  as foot on  $B$ . By (15.11),  $g(q, f)$  is perpendicular to  $B$  at  $f$ , by the first part of this proof there is only one perpendicular  $g(f)$  hence  $q_\nu \rightarrow q$  and  $g(f,)\rightarrow g(f)$ .

It remains to be shown that there is at least one perpendicular  $g(f)$ . Let  $q_\nu a > q_\nu a'$  and  $q_\nu \rightarrow f$ . Then the foot  $f,$  of  $q_\nu$  on  $B$  tends to  $f$ . By (15.11),  $g(q_\nu, f, )$  is perpendicular to  $B$ , and tends by the second part of this proof to  $g(f)$ .

(15.15) *If  $B=B_p(a, a')$  contains  $p$ , then  $S=S(p, \alpha)$  is the union of all perpendiculars to  $B$ . The center of a perpendicular is its intersection with  $B$ .*

**Proof.** Let  $f \in B$  and  $f \neq x \in g(f)$ . If  $xf=fx'=xx'/2$ , it follows from  $p \subset B=B_p(x, x')$  that  $\alpha > px=px'$ , hence  $x' \in g$ . This shows that  $f$  is the center of  $g$ .

To see that every point  $y$  of  $S$  lies on a perpendicular to  $B$ , let  $f^*$  be the foot of  $y$  on  $B(a, a')$  and  $y'f^*=f^*y=y'y/2$ . Then  $yf^* < \alpha$ ,  $pf^* < 2\alpha$  and  $py' < 2\alpha < \rho'(p)/2$ . Since  $f^*$  is the foot of  $y$ , points  $d, d'$  with  $(pdf^*), (df^*d')$  and  $pd' < \rho'(p)/2$  exist such that  $yd=yd'$ . As in the proof of (15.12) it follows that  $f^*$  is the center of  $d$  and  $d'$  and that  $p \in B(y, y')$ , which shows that  $y'p=y'p < \alpha$ . Since  $K(p, \alpha)$  is convex,  $f \in t(y, y')$  must lie in  $S$ , so that  $y \in g(f^*)$ .

With every set  $B=B_p(a, a')$  that contains  $p$ , an involutoric topological mapping  $\Omega(B)$  of  $S(p, \alpha)=S$  on itself may now be associated as follows:  $x\Omega(B)=x$  for  $x \in B$ . If  $x \in S-B$ , the point  $x'=x\Omega(B)$  is determined by  $x'f=fx=x'x/2$ , where  $f$  is the foot of  $x$  on  $B$ . Because of (15.14), (15.15),  $\Omega(B)$  is a topological mapping of  $S$  on itself. The main point is that according to (15.10), (15.12),  $x'$  may also be defined as the point for which  $B_p(x, x')=B$ ; for it yields

(15.16)  $\Omega(B)$  maps a line  $g$  that intersects  $B$  at some point  $s$  isometrically on a line  $g'$  through  $s$ .

If  $a, b \in g$  and  $(asb)$ , then  $a$  and  $b$  are on different sides of  $B$ , hence their images are on different sides of  $B$ , so that  $t(a', b')$  intersects  $B$  at some point  $s'$ . The relation  $B=B_p(a, a')=B_p(b, b')$  yields  $ab \leq as'+s'b=a's'+s'b'=a'b' \leq a's+sb'=as+sb=ab$ , whence  $ab=a'b'$  and  $s=s'$ . If  $c$  is any point of  $g$  different from  $a, b, s$ , then either  $(asc)$  or  $(csb)$ . In the first case it is seen as before that  $(a's'c')$  and  $ac=a'c'$ , hence also  $bc=b'c'$  which proves (15.16).

This fact makes it likely that under  $\Omega(B)$  not only distances on lines which intersect  $B$  are preserved, but that  $\Omega(B)$  is a motion. A proof is obtained by repeated application of (15.16) and from the following topological fact:

(15.17)  $S(p, \alpha)$  is homeomorphic to a euclidean space of some finite dimension  $\nu$ .

For a proof the following notation will be useful: a set  $A$  in  $S(p, \alpha) = S$  is said to be linear if, with any two distinct points  $a, b$ , it also contains the whole line  $g(a, b)$ . The sets  $B_p(a, a')$  are linear. The intersection of any number of linear sets is empty or linear.

Consider a line  $g_1$  through  $p$  and put  $B_1 = B_p(p, g_1)$ . If  $B_1$  does not consist of  $p$  alone it contains a line  $g_2$  through  $p$ . Put  $B_2 = B_1 \cap B_p(p, g_2)$ . If  $B_2 \neq p$  it contains, as a linear set, a line  $g_3$  through  $p$ ; put  $B_3 = B_2 \cap B_p(p, g_3)$  and continue. There is a finite subscript  $\nu$  such that  $B_\nu = p$ . For let  $q_\kappa, q'_\kappa$  be points of  $g_\kappa$  with center  $p$  and  $q_\kappa q'_\kappa = \alpha$ . By construction  $q_\mu \in B_p(q_\kappa, q'_\kappa) = B_p(p, g_\kappa)$  for  $\mu > \kappa$  hence, by (15.1),  $q_\kappa q_\mu > q_\kappa q'_\kappa / 2 = \alpha / 2 = q_\kappa p$ . This relation and the finite compactness of the space imply that the  $q_\kappa$  are finite in number. If  $q_{\nu-1}$  is the last  $q_\kappa$  then  $B_\nu = p$ .

By (15.15),  $S(p, \alpha)$  is the topological product of  $B_1$  and an open segment.  $B_2$  is the locus of those points in  $B_1$  which have equal distance from  $q_2$  and  $q'_2$ . Then (15.14) applied to  $B_1$  as space shows that  $B_1$  is the product of  $B_2$  and an open segment, generally  $B_\kappa$  is the product of  $B_{\kappa-1}$  and an open segment. Now  $B_{\nu-1}$  is the line  $g_{\nu-1}$ , hence  $B_{\nu-2}$  is homeomorphic to  $E^2$ , generally  $B_{\nu-\rho}$  is homeomorphic to  $E^\rho$ , and finally  $S(p, \alpha)$  to  $E^\nu$ .

Let, generally,  $L^\mu$  denote a linear set in  $S = S(p, \alpha)$  of dimension  $\mu$ . The argument of the last proof yields that any  $L^\mu$  through  $p$  is homeomorphic to  $E^\mu$ . A set  $B_p(p, g)$  either contains an  $L^\mu$ ,  $\mu \leq \nu - 1$ , through  $p$  or decomposes it, so that  $\dim B_p(p, g) \cap L^\mu \geq \mu - 1$  (compare Hurewicz-Wallman [1, p. 48, Theorem IV, 4]).

Consider  $\nu$  distinct points  $p, q_1, \dots, q_{\nu-1}$  in  $S$  and put  $B_\kappa = B_p(p, g(p, q_\kappa))$ ,  $\kappa = 1, 2, \dots, \nu - 1$ . The preceding remark yields generally  $\dim B_1 \cap \dots \cap B_\rho \geq \nu - \rho$ , in particular  $\dim B_1 \cap \dots \cap B_{\nu-1} \geq 1$ , so that this intersection contains a line  $g$ . By (15.13),  $g$  is perpendicular to all lines  $g(p, q_\kappa)$ , and  $B_p(p, g)$  contains these lines, hence also the points  $p, q_1, \dots, q_{\nu-1}$ .

Next let  $\nu - 1$  distinct points  $p, q_1, \dots, q_{\nu-2}$  be given in  $S$ . Choose any point  $q'$  different from these points. Then a set  $B_p(p, g')$  exists which contains the given points and  $q'$ . Let  $q'_{\nu-1}$  be any point of  $S$  not in  $B_p(p, g')$ . Then a set  $B_p(p, g'')$  through the given  $\nu - 1$  points and  $q''$  exists and intersects  $B_p(p, g')$  in an  $L^{\nu-2}$ . This procedure yields generally that  $\rho$  points  $p, q_1, \dots, q_{\rho-1}$  lie in an  $L^{\rho-1}$ .

Let  $\alpha = \rho(p)/3$  and  $x \in S(p, \alpha)$ ; then  $S(x, \rho(x)) \supset S(p, \rho(x) - xp) \supset S(p, \rho(p) - xp/40 - xp) \supset S(p, \alpha)$ . Hence, if  $p', q_1, \dots, q_{\rho-1}$  are given points in  $S(p, \alpha)$ , they lie also in  $S(p', \rho(p'))$ . By the last result the given points lie in a  $(\rho - 1)$ -dimensional linear set  $L$  of  $S(p', \rho(p'))$ . Then  $L \cap S(p, \alpha)$  is an  $L^{\rho-1}$  in  $S(p, \alpha)$ .

which contains the given points. Thus any  $\rho \leq \nu$  points of  $S(p, \rho(p)/3)$  lie in an  $L^{\nu-1}$  of  $S(p, \rho(p)/3)$ .

This *reduces the problem* (to prove that  $\Omega(B)$  is a motion) *to two dimensions*, and the procedure of B [1, pp. 171–173] may be carried over to the neighborhood  $S(p, \rho(p)/12)$  because generally

$$(15.18) \quad S(p, \lambda) \supset S(q, \lambda/2) \supset S(p, \lambda/4) \quad \text{for } q \in S(p, \lambda/4).$$

At this point it becomes important that the mapping  $\Omega(B_q(q, g))$  corresponding to a set  $B_q(q, g)$  in  $S(q, \alpha)$ ,  $\alpha \leq \rho(q)$ , maps by (15.15) *all* of  $S(q, \alpha)$  on itself.

Thus the following result is obtained: Let  $L$  be an  $L^2$  through  $p$  in  $S = S(p, \rho(p)/12)$  and  $g$  any line in  $L$  through  $p$ . The mapping  $\Omega(B_p(p, g))$  of  $S$  on itself maps  $L$  isometrically on itself.

$B_p(p, g) \cap L$  is a line  $h$  in  $L$ , and the mapping  $\Omega$  of  $L$  on itself is, in the usual terminology, a reflection of  $L$  in  $h$ . Successive reflections of  $L$  in two lines through  $p$  yield a rotation of  $L$  about  $p$ . Hence the full group rotations of  $L$  about  $p$  exists.

Consider two points  $p_1$  and  $p_2$  on  $h$  in  $S(p, \rho(p)/48)$  and the two-dimensional sets  $L_\kappa = L \cap S(p_\kappa, \rho(p)/24)$ ,  $\kappa = 1, 2$ . Let  $h_\kappa$  be the perpendicular to  $h \cap L_\kappa$  in  $L_\kappa$ . The estimate (15.18) shows that the reflection of  $L_1$  in  $h_1$  followed by the reflection of  $L_2$  in  $h_2$  yields a translation of the subset  $L_1$  of  $L$  along  $h$ . Hence all three elementary types of motions exist in  $L$  in the small.

Classical arguments show now that the metric of  $L$  is euclidean, hyperbolic, or spherical<sup>(10)</sup>. It follows that the neighborhood  $S(p, \rho(p)/48)$  of a given point  $p$  is congruent to the interior of a sphere with radius  $\rho(p)/48$  of a  $\nu$ -dimensional euclidean, hyperbolic, or spherical space (the same for different points  $p$ ).

This yields finally (15.2).

If every point of a  $G$ -space  $R$  has a neighborhood  $S(p, \rho)$  with linear bisectors, it follows from (15.2) that the space is locally euclidean, hyperbolic, or elliptic.

Spaces with this property are studied extensively in Cartan [1]. His results will be used. Notice first that by the definition (15.3) of  $\rho(p)$  and (12.11) the neighborhood  $S(\bar{p}, \rho(\bar{p})/48)$  of any point  $\bar{p}$  over  $p$  of the universal covering space  $R_\Phi$  of  $R$  is congruent to  $S(p, \rho(p)/48)$ , so that  $R_\Phi$  is also locally euclidean, hyperbolic, or spherical. By Cartan [1, pp. 71 and 169, 170] the euclidean, hyperbolic, or spherical spaces (the spherical space of dimension 1 is an exception) are the only simply connected spaces with this property, hence  $R_\Phi$  must be one of these spaces. This yields:

**FIRST CHARACTERIZATION OF THE SPACES WITH CONSTANT CURVATURE.**  
*The universal covering space of a  $G$ -space  $R$  is a finite dimensional euclidean, hyperbolic, or spherical space, if and only if, locally, the bisectors are linear (more*

<sup>(10)</sup> The quickest way seems to be the method of De Tilly [1, 2] which leads directly to certain functional equations yielding the three types of trigonometries.

precisely if and only if every point  $p$  of  $R$  has a neighborhood  $S(p, \rho(p))$ , which satisfies (\*).

**16. The Helmholtz-Lie problem.** The first characterization of the spaces with constant curvature leads to a second one by the "*axiome de libre mobilité*" of Cartan; or by the property that every point  $p$  of the space has a neighborhood  $S(p, \rho)$  ( $\rho$  is an arbitrary positive function of  $p$ ) which *permits rotation* in the following sense:

(16.1) If  $a, b, a', b'$  are points in  $S(p, \rho)$  with  $pa = pa', pb = pb'$  and  $ab = a'b'$ , then a motion of  $S(p, \rho)$  exists which carries  $p, a, b$  into  $p, a', b'$  respectively.

The hypothesis that *every* point of the space has a neighborhood which satisfies (16.1) can be considerably reduced. While this question will not be discussed here, the following statements will give some indications about it.

(16.2) If  $S(p, \rho(p))$  and  $S(q, \rho(q))$  permit rotation and  $pq = pq < \rho(p)$ , then  $S(q_1, \min(\rho(q), \rho(p) - pq))$  permits rotation.

Let  $q_1a_1 = q_1a'_1 < \sigma = \min(\rho(q), \rho(p) - pq)$ ,  $q_1b_1 = q_1b'_1 < \sigma$  and  $a_1b_1 = a'_1b'_1$ . There is a motion  $\Phi$  of  $S(p, \rho(p))$  which carries  $q_1$  into  $q$ . The points  $q_1, a_1, a'_1, b_1, b'_1$  lie in  $S(p, \rho(p))$ . Therefore  $\Phi$  carries them into points  $q, a, a', b, b'$  with  $qa = qa' < \sigma$ ,  $qb = qb' < \sigma$ , and  $ab = a'b'$ . Since  $S(q, \rho(q))$  permits rotation,  $S(q, \sigma)$  does too. Hence a motion  $\Psi$  of  $S(q, \sigma)$  exists which carries  $q, a, b$  into  $q, a', b'$ . Then  $\Phi\Psi\Phi^{-1}$  is an isometric mapping of  $S(q, \sigma)$  on itself which carries  $q_1, a_1, b_1$  into  $q_1, a'_1, b'_1$ .

(16.3) Under the assumptions of (16.2) let  $pq < \eta(p)$ . If  $\lambda = \min(\rho(q), \rho(p) - pq, pq)$  and  $x \in S(q, \lambda)$  then  $S(x, \lambda - qx)$  permits rotation.

**Proof.** Since  $pq < \eta(p) \leq \eta_6(p)$ , there is a point  $p'$  with  $p'p = pq = p'q/2$ . Denote by  $t_1$  and  $t_2$  the centers of  $t(p', p)$  and  $t(p, q)$  respectively, and let  $t$  be any point distinct from  $t_1$  and  $t_2$  with distance  $pt = pq/2$  from  $p$ . ( $t$  exists only if the space has at least dimension 2. This may be assumed because the one-dimensional case is trivial.) The segments  $t(t_1, t)$  lie by (1.15) in  $S(p, pq)$ . For any point  $u$  of  $t(t_1, t) + t(t, t_2)$  there exists therefore a point  $v$  with  $(puv)$  and  $p'v = pq$ . As  $u$  traverses  $t(t_1, t) + t(t, t_2)$  the point  $v$  traverses a Jordan arc  $\mathcal{C}$  from  $p'$  to  $q$  which lies entirely on the sphere  $K(p, q)$ .

This arc  $\mathcal{C}$  intersects every sphere  $K(q, \tau)$  with  $\tau \leq \lambda$  at some point  $y$ , because  $p'q = 2pq > \lambda$ . By (16.2),  $S(q, \min(\rho(q), \rho(p) - pq))$  permits rotation, hence  $S(q, \lambda)$  does too. If  $x \in S(q, \lambda)$ , then  $x \in K(q, \tau)$  for some  $\tau < \lambda$ . Now (16.2) applied to  $q, y$ , and  $x$  instead of  $p, q$  and  $q_1$  shows that  $S(x, \lambda - qx)$  permits rotation.

The following fact yields the reduction of the present problem to the results of the last section.

(16.4) If  $S(p, \rho(p))$  and  $S(q, \rho(q))$  permit rotation,  $pq < \min(\eta(p), \rho(p))$ , then  $S(q, \lambda/3)$  has linear bisectors ( $\lambda$  is defined as in (16.3)).

Observe that the assumptions of (16.3) hold. When  $x \in S(q, \lambda/3)$ , then  $2\lambda/3 < \lambda - qx$ , hence  $S(x, 2\lambda/3)$  permits, by (16.3), rotation. Let  $a, a', x, y$

be points of  $S(q, \lambda/3)$  such that  $a \neq a'$  and  $x$  and  $y$  lie on  $B(a, a')$ . Then  $xa = xa' < 2\lambda/3$ ,  $xy = xy < 2\lambda/3$ , and  $ay = a'y$ . Hence a motion  $\Phi$  of  $S(x, 2\lambda/3)$  exists which carries  $x, a, y$  into  $x, a', y$  respectively. If  $(xuy)$  then  $(xu\Phi y)$  and  $xu = xu\Phi$ . Since  $t(x, y)$  is unique,  $u\Phi = u$ , and  $au = a'u\Phi = a'u$ . This shows that  $u \subset B(a, a')$ , hence  $t(x, y) \subset B(a, a')$ , q.e.d.

Now it is easy to obtain the basic result of the present discussion.

(16.5) *Let  $S(p, \rho)$ ,  $0 < \rho < \eta(p)$ , permit rotation; and let every sphere  $K(p, \tau)$ ,  $0 < \tau < \rho$ , contain a point  $q_\tau$ , such that  $S(q_\tau, \rho(q_\tau))$  permits rotation for some  $\rho(q_\tau) > 0$ . Then  $S(p, \rho)$  is congruent to the interior of a sphere of a finite dimensional euclidean, hyperbolic or spherical space.*

*Note.* It evidently suffices to require that  $S(p, \rho)$  permits rotation and that points  $q_\kappa$ ,  $\kappa = 1, 2, \dots$ , and numbers  $\rho(q_\kappa) > 0$  exist, such that  $K(q_\kappa, \rho(q_\kappa))$  permits rotation and the intervals  $pq_\kappa - \rho(q_\kappa) < \tau < pq_\kappa + \rho(q_\kappa)$  of the real axis cover the interval  $0 < \tau < \rho$ . This remark contains a well known theorem of F. Schur, compare B [1, pp. 191, 192].

To see (16.5) observe that (16.4) and (15.2) imply that every point  $x \neq p$  of  $S(p, \rho)$  has a neighborhood  $S(x, \sigma)$  which is congruent to a sphere of a  $\nu$ -dimensional euclidean, hyperbolic, or spherical space (the same for different  $x$ ). By (4.12),  $S(p, \rho)$  is homeomorphic to  $E^\nu$ , and the metric  $xy$  has in  $S(p, \eta(p))$  the convexity and continuity properties of a  $G$ -space. A familiar argument on removable singularities shows that the metric is also euclidean, hyperbolic, or spherical at  $p$ . This leads to the

#### SECOND CHARACTERIZATION OF THE SPACES WITH CONSTANT CURVATURE.

*The universal covering space of a  $G$ -space  $R$  is a finite dimensional euclidean, hyperbolic, or spherical space if and only if every point  $p$  of  $R$  has a neighborhood  $S(p, \rho(p))$ ,  $\rho(p) > 0$ , with the following property: When  $pa = pa' < \rho(p)$ ,  $pb = pb' < \rho(p)$ , and  $ab = a'b'$  then a motion of  $S(p, \rho(p))$  exists which carries  $p, a, b$  into  $p, a', b'$  respectively.*

The two characterizations of the spaces with constant curvature suggest the problems to determine the spaces in which either every sphere  $S(p, \rho)$  has linear bisectors or every sphere  $S(p, \rho)$  permits rotation. Both problems are well known in slightly different formulations and the pertaining literature is prodigious. The first question is answered in

(16.6) THEOREM. *If (and only if) each bisector  $B(a, a')$  of a  $g$ -space  $R$  contains with any two points  $x, y$  at least one segment  $t(x, y)$ , then  $R$  is either a straight line, or a euclidean, hyperbolic, or spherical space of a finite dimension  $\nu \geq 2$ .*

*Note.* The hypothesis of (16.6) holds when every bisector  $B(a, a')$  of  $R$  is a  $G$ -space<sup>(11)</sup>.

<sup>(11)</sup> In the terminology of Hadamard and Cartan,  $B(a, a')$  is required to be totally geodesic (totalement géodésique). For the literature on (16.6) compare Zacharias [1] and Enriques [1].

**Proof.** By a procedure similar to the proof of (15.4) it will first be shown that  $B(a, a')$  is a  $G$ -space.  $B(a, a')$  is closed, therefore finitely compact, moreover convex, because it contains with  $x$  and  $y$  a segment  $t(x, y)$ ; that is,  $B$  satisfies Axioms A, B, C. To see Axiom D, it suffices to show that  $B(a, a')$  contains with any segment  $t(x, y)$  of length at most  $\eta(y)$  also the segment  $t(x, y')$  where  $xy = yy' = xy'/2$ .

If the point  $z$  of  $t(x, y')$  did not lie on  $B(a, a')$  then  $az < a'z$ , say. Choose  $b$  with  $(yba')$  and  $yb < \eta(y)$ . Then  $ba > ba'$ . For  $ba \geq ay - yb = a'y - yb = ba'$ . The equality sign would imply  $(yba)$ , which, together with  $(yba')$ ,  $ya = ya'$ , and (4.1), would yield  $a' = a$ .

For  $u \in t(y, b)$  let  $v$  be defined by  $(vux)$  and  $vu:zy = bu:by$  ( $v$  exists because  $vu \leq zy < \eta(y)$ ,  $xu < 2\eta(y)$  and  $\eta(y) \leq \eta_0(y)$ ). Now  $v = z$  and  $av < a'v$  when  $u = y$ ; but  $v = b$  and  $av > a'v$  when  $u = b$ . Hence a point  $u_0$  with  $(yu_0b)$  exists such that the corresponding point  $v_0$  lies on  $B(a, a')$ . Since  $t(v_0, x)$  is unique, it lies by hypothesis on  $B(a, a')$ . But  $u_0$  does not lie on  $B(a, a')$ , by the argument that proved  $ba > ba'$ .

It is now not hard to establish (16.6) with the help of (15.2) by direct geometric arguments, similar to those used in §15. But (16.6) may be reduced to two rather simple facts on two-dimensional spaces with constant curvature. By §15 the universal covering space  $R_\Phi$  of  $R$  is a  $\nu$ -dimensional ( $\nu \geq 2$  is assumed) euclidean, hyperbolic, or spherical space. Consider the neighborhood  $S(p, \eta(p))$  of an arbitrary point  $p$  of  $R$  and any  $L^2$  in  $S(p, \eta(p))$  through  $p$ . The geodesics which contain lines (compare §15) in the  $L^2$  through  $p$  form a two-dimensional  $G$ -space  $M^2$ <sup>(12)</sup>. Let  $g$  be any geodesic through a line  $l_1$  in the  $L^2$  through  $p$ , and  $l_2$  the line in the  $L^2$  perpendicular to  $l_1$  at  $p$ . If  $a, a'$  are points of  $l_2$  with center  $p$ , then  $B_2 = B(a, a') \cap M^2$  contains  $g$ , but may conceivably contain other geodesics. (Bisectors in the elliptic plane consist, for instance, of two geodesics.)  $B_2$  cannot contain two line elements  $l'$  and  $l''$  with the same origin  $x$ , because both  $M^2$  and  $B(a, a')$  would then contain the whole  $L^2$  in  $S(x, \eta(x))$  which is determined by  $l'$  and  $l''$ ; and it would follow that  $B(a, a') \supset M^2 \supset a$ . This yields

- (1) Every geodesic in  $R$  is simple (see §8).
- (2) Any two geodesics in  $B(a, a') \cap M^2$  are disjoint.

First let  $R_\Phi$  be a spherical space. If  $M^2$  is any two-dimensional  $G$ -space in  $R$ , its images in  $R_\Phi$  are disjoint two-dimensional spheres, each of which is a covering space of  $M^2$ . Unless  $M^2$  is a sphere, it is the elliptic plane (see Cartan [1, p. 174]). If  $B(a, a')$  is defined similarly as before,  $B(a, a') \cap M^2$  is the bisector of  $a$  and  $a'$  in  $M^2$ , and would consist of two geodesics. These would intersect, which contradicts (2). Hence every two-dimensional  $G$ -space in  $R$  is spherical, whence it easily follows that  $R$  itself is spherical.

Now let  $R_\Phi$  be euclidean or hyperbolic, and  $M^2$  a given two-dimensional

<sup>(12)</sup> This and some of the following facts are simple properties of totally geodesic manifolds, compare Cartan [1, pp. 119 ff].



$G$ -space in  $R$ . If  $M^2$  were a surface of constant curvature not greater than 0 but not a euclidean or hyperbolic plane, it would contain points  $p, q$  with  $0 < pq < \eta(p)$  such that the geodesic  $\mathfrak{h}$  which contains  $t(p, q)$  is not closed<sup>(13)</sup>. Let  $\bar{M}^2$  be the plane over  $M^2$  through the point  $\bar{p}$  over  $p$  in  $R_\Phi$ . If  $M^2$  were not simply connected  $\bar{M}^2$  would contain a point  $\bar{q}_1$  over  $q$  different from the point  $\bar{q}$  over  $q$  in  $S(\bar{p}, \eta(p))$ . Then  $\Phi$  maps the geodesic  $\mathfrak{G}(t(\bar{p}, \bar{q}_1))$  on a geodesic  $\mathfrak{h}'$  which contains an arc  $c$  from  $p$  to  $q$  different from  $t(p, q)$ , and  $\mathfrak{h} \neq \mathfrak{h}'$ . For, the line elements of  $\mathfrak{h}$  and  $\mathfrak{h}'$  are different at either  $p$  or  $q$  because  $\mathfrak{h}$  is not closed, and  $\mathfrak{h}$  has by (1) no multiple point.

Choose  $\bar{b}$  and  $\bar{b}'$  in  $S(\bar{p}, \eta(p))$  with center  $\bar{p}$  such that  $\mathfrak{G}(t(\bar{p}, \bar{q}_1))$  is the bisector of  $\bar{b}$  and  $\bar{b}'$  in  $\bar{M}^2$ . If  $b = \bar{b}\Phi$ ,  $b' = \bar{b}'\Phi$ , then  $h' \subset B(b, b') \cap M^2$ . On the other hand  $t(p, q)$  is unique and lies therefore in  $B(b, b') \cap M^2$ . Therefore  $\mathfrak{h}$  and  $\mathfrak{h}'$  would lie in  $B(b, b') \cap M^2$ , which contradicts (2).

This completes the proof of (16.6).

Call two ordered triples of points  $a, b, c$  and  $a', b', c'$  congruent if  $ab = a'b'$ ,  $bc = b'c'$ ,  $ca = c'a'$ . With this notation the second problem mentioned above is answered by

(16.7) THEOREM. *If (and only if) for any two congruent triples  $a, b, c$  and  $a', b', c'$  of the  $G$ -space  $R$  a motion of  $R$  exists which carries  $a, b, c$  into  $a', b', c'$ , then  $R$  is a finite dimensional euclidean, hyperbolic, or spherical space<sup>(14)</sup>.*

**Proof.** If  $p, a, b$  and  $p, a', b'$  are congruent triples, then a motion  $\Phi$  of  $R$  exists which carries  $p, p, a$  into  $p, p, a'$ . Let  $b\Phi = b^*$ . Then  $pb^* = pb = pb'$  and  $a'b^* = ab = a'b'$ . Therefore a motion  $\Psi$  of  $R$  exists which carries  $p, a', b^*$  into  $p, a', b'$ . The motion  $\Phi\Psi$  carries  $p, a, b$  into  $p, a', b'$ . This shows that every sphere  $S(p, \rho)$  permits rotation. Hence  $R$  is a space of constant curvature.

Let  $q, a, b$  and  $q', a', b'$  be congruent triples and  $p$  a center of  $q$  and  $q'$ . A motion  $\Omega_1$  of  $R$  exists which carries  $p, p, q$  into  $p, p, q'$ . If  $a\Omega_1 = a^*$ ,  $b\Omega_1 = b^*$ , then  $q, a, b$  is congruent to  $q', a^*, b^*$ , hence  $q', a', b'$  is also congruent to  $q', a^*, b^*$ . There is a motion  $\Omega_2$  that carries  $q', a^*, b^*$  into  $q', a', b'$ . Hence  $\Omega = \Omega_1\Omega_2$  carries  $q, a, b$  into  $q', a', b'$ . If the segment  $t(q, a)$  is unique then  $t(q', a')$  is unique and  $\Omega$  maps  $t(q, a)$  on  $t(q', a')$  and  $\mathfrak{G}(t(q, a))$  on  $\mathfrak{G}(t(q', a'))$ . Hence any two geodesics of  $R$  are congruent.

On no surface of a constant curvature  $\kappa \leq 0$  are all geodesics congruent, except in the euclidean or hyperbolic planes. Therefore every geodesic of  $R$  is a straight line, when  $R$  is locally euclidean or hyperbolic. This shows that  $R$  is either a euclidean or a hyperbolic space.

Besides the sphere, the elliptic plane is the only  $G$ -space of constant posi-

<sup>(13)</sup> The surfaces with constant curvature 0 are enumerated in Cartan [1, chapter III, §VII]. For surfaces of negative curvature the reader may be referred to Hadamard [1]. However, the statement admits of a very simple direct proof.

<sup>(14)</sup> The one-dimensional case is no exception. The theorem is stated without proof in B [1, p. 134]. A related result is found in Birkhoff [1].

tive curvature (Cartan [1, p. 174]), and all its geodesics are congruent. Therefore another argument is necessary to show that every two-dimensional  $G$ -space in  $R$  is a sphere.

On a sphere  $A$  with longitude and latitude as coordinates consider the points  $a = (0, 0)$ ,  $b = (\pi/3, 0)$ ,  $c = (2\pi/3, 0)$ , and  $a' = (\pi/2, \alpha)$ , where  $0 < \alpha < \pi/2$  and  $\cos \alpha = 3^{-1/2}$ . If  $d = (\pi/2, 0)$ , then  $bd = cd = \delta \cdot \pi/6$  ( $\delta$  is a factor of proportionality),  $\cos(a'b/\delta) = \cos(a'c/\delta) = \cos(bd/\delta) \cos \alpha = 2^{-1}$ , hence  $a'b = a'c = \delta \cdot \pi/3$ . In the elliptic space  $A'$  obtained from  $A$  by identification of antipodal points,  $a'b = a'c = \delta \cdot \pi/3$ . Therefore  $a, b, c$  and  $a', b, c$  are in  $A'$  congruent triples. But there is evidently no motion of  $R$  which carries  $a, b, c$  into  $a', b, c$  because  $t(a, b) + t(b, c) + t(c, a)$  is a geodesic and  $t(a', b) + t(b, c) + t(c, a')$  is not.

This shows that  $R$  cannot contain an elliptic plane, and completes the proof of (16.7).

This theorem answers the well known *Helmholtz-Lie problem*<sup>(15)</sup> in the large, after the second characterization of the spaces with constant curvature had answered it in the small. As far as I am aware, no theorem to this effect is known in classical differential geometry. The assumptions of Helmholtz, Lie, and Weyl [1], which contains the most complete treatment along these lines, amount in the present notation to requiring that in a  $\nu$ -dimensional Finsler space congruent ordered  $(\nu+1)$ -tuples of points can be carried into each other by a motion of the space. It seems therefore worthwhile to indicate how the present result may be obtained by the classical methods, when the necessary differentiability conditions are granted.

(16.8) *If in a symmetric Finsler space  $R$  a motion of  $R$  exists which carries a given direction  $\mathfrak{d}_1$  with origin  $p$  into a given direction  $\mathfrak{d}_2$  with origin  $p$ , then  $R$  is Riemannian at  $p$ .*

Let the metric of  $R$  be determined by the integrand  $\phi(\xi_1, \dots, \xi_\nu; \eta_1, \dots, \eta_\nu)$  where  $\phi(\xi; \eta)$  satisfies the usual conditions (positive, positively homogeneous, regular). A motion of  $R$  which leaves  $p = (\xi_1, \dots, \xi_\nu)$  fixed induces an affine transformation of the  $\eta$ -space (tangent space at  $p$ ), which leaves the origin  $\eta_i = 0$  fixed, and carries the indicatrix  $K_p: \phi(p; \eta_1, \dots, \eta_\nu) = 1$  into itself. If every direction with origin  $p$  can be transformed into any other by a motion of  $R$ , then  $K_p$  admits a transitive group of affine transformations in itself. Since  $K_p$  is convex and has 0 as center it must be an ellipsoid with center 0<sup>(16)</sup>. This shows that  $R$  is Riemannian at 0.

Notice the corollary: *A Finsler space  $R$  is a Riemann space if every point  $p$  of  $R$  has a neighborhood  $S(p, \rho(p))$  with this property: when  $pa = pa' < \rho(p)$ , then a motion of  $S(p, \rho(p))$  exists which leaves  $p$  fixed and carries  $a$  into  $a'$ .*

The hypothesis of this corollary holds when every point  $p$  of  $R$  has a neighborhood  $S(p, \rho(p))$  that permits rotation. But a Riemann space with this property is known to have constant curvature.

<sup>(15)</sup> For the literature on the problem see Weyl [1] and Enriques [1, pp. 107 ff].

<sup>(16)</sup> See Blaschke [1, §89].

## REFERENCES

G. BIRKHOFF

1. *Metric foundations of geometry*, Proc. Nat. Acad. Sci. U.S.A. vol. 27 (1941) pp. 402-406.

W. BLASCHKE

1. *Vorlesungen über Differentialgeometrie*. II, Berlin, 1923.

H. BUSEMANN

1. (B [1]) *Metric methods in Finsler spaces and in the foundations of geometry*, Annals of Mathematics Studies No. 8, Princeton, 1942.
2. (B [2]) *On spaces in which two points determine a geodesic*, Trans. Amer. Math. Soc. vol. 54 (1943) pp. 171-184.

E. CARTAN

1. *Leçons sur la géométrie des espaces de Riemann*, Paris, 1928.

M. DE TILLY

1. *Études de mécanique abstraite, Mémoires couronnés et autres*, Mémoires de l'Académie Royale de Belgique vol. 21 (1870).
2. *Essai sur les principes fondamentaux de la géométrie et de la mécanique*, Mémoires de la Société des Sciences de Bordeaux (3) vol. 1 (1878).

F. ENRIQUES

1. *Prinzipien der Geometrie*, Encyklopädie der mathematischen wissenschaften vol. III (1907), pp. 1-129.

J. HADAMARD

1. *Les surfaces à courbures opposées*, J. Math. Pures Appl. (5) vol. 4 (1898) pp. 27-73.

F. HAUSDORFF

1. *Mengenlehre*, 2nd ed., Berlin, 1927.

W. HUREWICZ and H. WALLMAN

1. *Dimension theory*, Princeton Mathematical Series, vol. 4, 1941.

S. LEFSCHETZ

1. *Algebraic topology*, Amer. Math. Soc. Colloquium Publications, vol. 27, 1942.

H. LIEBMANN

1. *Nichteuklidische Geometrie*, 3rd ed., Berlin, 1923.

K. Menger

1. *Metrische Geometrie und Variationsrechnung*, Fund. Math. vol. 25 (1935) pp. 441-458.
2. *Die Metrische Methode in der Variationsrechnung*, Ergebnisse eines mathematischen Kolloquiums vol. 8 (1937) pp. 1-32.
3. *Untersuchungen über allgemeine Metrik*. I, II, III, Math. Ann. vol. 100 (1928) pp. 75-163.

H. MINKOWSKI

1. *Geometrie der Zahlen*, Leipzig, 1910.

L. PONTRJAGIN

1. *Topological groups*, Princeton Mathematical Series, vol. 2, 1939.

H. SEIFERT and W. THRELFAH

1. *Lehrbuch der Topologie*, Leipzig, 1934.

H. WEYL

1. *Mathematische Analyse des Raumproblems*, Berlin, 1923.

M. ZACHARIAS

1. *Elementargeometrie und elementare nicht-euklidische Geometrie in synthetischer Behandlung*, Encyklopädie der mathematischen wissenschaften vol. I.

ILLINOIS INSTITUTE OF TECHNOLOGY,  
CHICAGO, ILL.