Metric Methods in Finsler Spaces and in the Foundations of Geometry

HERBERT BUSEMANN

Metric Methods in Finsler Spaces and in the Foundations of Geometry

BY

HERBERT BUSEMANN

PRINCETON
PRINCETON UNIVERSITY PRESS
LONDON: HUMPHREY MILFORD
OXFORD UNIVERSITY PRESS
1942

Lithoprinted in U.S.A.
EDWARDS BROTHERS, INC.
ANN ARBOR, MICHIGAN
1942

PREFACE

Among the earliest speculations on the foundations of geometry we find many attempts to introduce the straight lines as geodesics. But no abstract concept of a metric being known, let alone metrics other than the euclidean and perhaps the spherical, these attempts were futile.

Today it is not difficult to formulate axioms for a space in which geodesics exist. This book treats some of the many problems which arise when one attempts to develop geometry with the geodesic as basic concept. The problems studied here fall essentially under four topics which may be listed roughly as Finsler spaces, parallels, convexity of spheres, and motions.

The choice of these topics is, of course, due partly to personal preference but partly also to the desire to impress the reader with the large variety of questions which fall under the scope of the metric methods. It goes without saying that there are many unsolved problems, very different in character and difficulty. A number of these will be formulated in the text.

The book is divided into five chapters, each of which is preceded by a rather detailed introduction. A reader who wishes to get information beforehand concerning the whole content is asked to turn to these introductions.

The idea of completing Frechet's axioms for a metric space, so as to ensure the existence of geodesics, is due to Menger. His results as far as we shall need them, will be proved in the text. Some familiarity with the topology of metric spaces is assumed, and theorems on convex bodies are used. Results of Riemannian geometry will be frequently referred to for comparison, but not actually applied (except in Chapter V §3). All facts from other theories, if not proved in the text, will be stated in exact form and reference to literature will be made.

Although Menger was the first to study geodesics in metric spaces, both his and his students' contributions to the foundations of geometry and the calculus of variations have so different a trend that the existence of geodesics plays hardly any role in their work. For that reason the material presented here is almost entirely different from the theories found in Blumenthal's Distance Geometries.

Because the topics of this book are interrelated with several different fields, even a moderately complete list of the pertaining literature was impracticable. To avoid inconsistencies the bibliography contains (with the exception of Finsler's dissertation) nothing but references for results actually quoted in the text.

Herbert Busemann
Illinois Institute of Technology

TABLE OF CONTENTS

Chapter I.

METRIC SPACES WITH GEODESICS

<u>Introduction</u>. The geodesics of the Riemann Spaces and Finsler Spaces which one usually considers have the following properties:

1) Any two distinct points P,Q can be connected by a shortest arc.

2) If the point Q is sufficiently close to P, this shortest connection is unique.

3) Any shortest connection between two points is contained in one and only one geodesic.

In the present chapter we give a set of axioms for a general metric space which guarantee that properties 1), 2), 3) hold. We first compile the definitions and theorems on metric spaces which we shall need later. In Section 2 we formulate the basic axioms A, B, C, D and show that 1) and 2) hold. In Section 3 we define geodesics and prove that they have property 3). Finally (Section 4) we discuss the topological structure of the spaces which satisfy axioms A - D and have dimensions 1 or 2. In both cases we shall find that the space is a manifold. The corresponding question for higher dimensional spaces is open.

§1. METRIC SPACES; NOTATIONS

Points, unless expressed by their coordinates, will be designated by Latin Capitals. A point set Σ is a metric space if a real number XY, the <u>distance</u> from X to Y (or of X and Y), is defined for every pair X, Y of points in Σ and satisfies the conditions:

A_1 $XX = 0$

A_2 $XY = YX > 0$ for $X \neq Y$ (symmetry)

A_3 $XY + YZ \geq XZ$ (triangle inequality)

Any subset σ of a metric space Σ becomes itself a metric space, if we define as distance for points X,Y in σ the distance XY of these points in Σ. Whenever we speak of a subset σ of a metric space Σ without defining its metric, we imply that σ is metrized in this way.

We say the point Y lies between the points X and Z, and write (XYZ), when Y is different from X and Z and $XY + YZ = XZ$. The relation (XYZ) has the following obvious but very useful properties:

THEOREM 1. If (XYZ) then (ZYX). If (WXY) and (WYZ), then (XYZ) and (WXZ).

The use of the words limit point, closed set, open set being uniform we do not re-define these concepts. (Definitions and proofs, which are omitted here, can be found in standard works as HAUSDORFF [2] or KURATOWSKI [1].) We call the set σ in a metric space Σ bounded, if a point P and a number α exist so that $PX < \alpha$ for $X \subset \sigma$.

A metric space Σ will be called compact, if every infinite sequence of points in Σ contains a converging subsequence; finitely compact, if every bounded, infinite sequence contains a converging subsequence.

We remind the reader of the following facts:

THEOREM 2. A closed subset of a compact metric space is compact. A bounded closed set in a finitely compact metric space is compact. A closed set in a finitely compact space is finitely compact.

THEOREM 3. A finitely compact metric space Σ is separable, i.e. there is a sequence of points $P_1, P_2, \ldots$ in Σ, so that every point of Σ is limit point of a suitable subsequence of $\{P_\nu\}$.

Consider two metric spaces Σ and Σ' and a mapping of the subset σ of Σ onto the subset σ' of Σ', i.e. for every point X of σ the image $X' = F(X)$ of X in σ' is uniquely determined and F(X) traverses all of σ' when X traverses σ. The mapping $X \to F(X)$ is called

a) continuous, if $\lim P_\nu = P$, $P_\nu, P \subset \sigma$, implies $\lim F(P_\nu) = \lim F(P)$.

b) topological, if it is one-to-one and continuous both ways. The sets σ and σ' are homeomorphic if a topological mapping of σ onto σ' exists.

c) a congruence, if

$$PQ = F(P)\,F(Q)$$

for any points P,Q in σ. A congruence is a topological mapping. The sets σ and σ' are congruent, if σ can be mapped onto σ' by a congruence.

In case Σ or Σ' is the real axis, special terms are used. Namely, if Σ is the real axis $-\infty < t < \infty$ with the absolute value $|t_1 - t_2|$ as metric, and σ the subset $a \leq t \leq b$, $a < b$, of Σ, then a continuous image of Σ is called a continuous curve in Σ', a topological image of Σ is an open Jordan curve in Σ'. A set congruent to Σ will be called an open straight line. Menger uses the simpler term straight line. We shall use this word in a wider sense.

Furthermore, a continuous image of σ is a continuous arc (in Σ'); a topological image of σ is a Jordan arc. We follow Menger in denoting a set congruent to σ as segment.

We see from the definition that a segment admits a representation of the form $P(t)$, $a \leq t \leq b$ with $P(t_1)\,P(t_2) = |t_1 - t_2|$. Such a representation of a segment will be called isometric. If $P(t)$, $a \leq t \leq b$, is an isometric representation of a segment σ we call $A = P(a)$

and B = P(b) the endpoints of σ, and say that σ connects A and B. This definition does not depend on the choice of the isometric representation of σ because A and B can be characterized by the property that (AXB) for every point $X \neq A, B$ of σ. The notation $\overline{AB}$ will be used for any segment with A and B as endpoints. It is convenient to put $\overline{AA} = A$. Of any three points on a segment one is between the two others.

If Σ' is the real axis, and $X \rightarrow F(X) = X'$ maps the subset σ of the metric space Σ onto the set σ' in Σ', then F(X) becomes a real valued function of the point X, with σ as domain of definition and σ' as range of F(X).

Many known theorems on functions of real variables can be extended to mappings of metric spaces. We shall need only:

THEOREM 4. If $X \rightarrow X' = F(X)$ is a continuous mapping of σ onto σ', and if σ is compact, then F(X) is uniformly continuous; i.e. for a given $\epsilon > 0$ a $\delta > 0$ can be found such that $F(P)\,F(Q) < \epsilon$ as soon as $PQ < \delta$.

and

THEOREM 5. A continuous real-valued function F(X), defined on a compact subset σ of a metric space reaches its minimum and its maximum.

As distance $\sigma\tau$ (or $d(\sigma, \tau)$ when there is a possibility of taking $\sigma\tau$ for the intersection of σ and τ) of two subsets σ and τ of a metric space Σ we define the greatest lower bound of the distances XY, where X traverses σ and Y traverses τ. We would express this definition in a formula as

$$\sigma\tau = \inf_{X\subset\sigma,\, Y\subset\tau} XY$$

In particular $P\sigma = \inf_{X\subset\sigma} PX$ will be the distance of the point P from the set σ.

THEOREM 6. For an arbitrary set σ and any two points P,Q we have

(1) $$|P\sigma - Q\sigma| \leqq PQ$$

Consequently the distance $X\sigma$ is a continuous function of X.

PROOF. Let F_ν and G_ν, $\nu = 1, 2, \ldots$ be (not necessarily different) points in σ with $PF_\nu \to P\sigma$ and $QG_\nu \to Q\sigma$.
Then

$$P\sigma - Q\sigma \leqq \underline{\lim}\, PG_\nu - \lim QG_\nu \leqq PQ$$

and

$$Q\sigma - P\sigma \leqq \underline{\lim}\, QF_\nu - \lim PF_\nu \leqq PQ$$

In case σ contains a point F with $PF = P\sigma$ we call F a foot of P on σ. We have

THEOREM 7. If σ is a (non-empty) finitely compact subset of Σ, then every point P has a foot on σ.

PROOF. Let Q be any point of σ, then

(2) $P\sigma \leqq PQ$

There is either a foot of P or a sequence of points F_ν in σ with

$$PF_\nu \longrightarrow P\sigma$$

The sequence $\{PF_\nu\}$ is bounded because of (2), therefore $\{F_\nu\}$ is bounded and has an accumulation point F in σ, because σ is finitely compact. It follows from the last theorem that F is a foot of P.

THEOREM 8. If every point P of the set

τ has exactly one foot F on the finitely compact set σ then F depends continuously on P, when P varies in τ.

PROOF. Let $P_\nu \longrightarrow P_o$, $P_\nu \subset \tau$, $\nu = o, 1, 2, \ldots$ and let F_ν be the foot of P_ν on σ. The sequence $F_1, F_2, \ldots$ is bounded because

$$P_oF_\nu \leqq P_\nu F_\nu + P_\nu P_o = P_\nu\sigma + P_\nu P_o$$

$$\leqq P_o\sigma + 2P_\nu P_o .$$

Every accumulation point of $\{F_\nu\}$ belongs to σ because, being finitely compact, σ is closed. There are accumulation points F because σ is finitely compact. Each F is a foot of P_o on account of Theorem 6, hence $F = F_o$ and $\lim F_\nu = F_o$.

The same pointset Σ may carry different metrics AB and $d(A,B)$. These are called <u>topologically equivalent</u>, when they lead to the same limit concepts, i.e. when $AA_\nu \longrightarrow o$ if, and only if, $d(A,A_\nu) \longrightarrow o$. If a limit is already defined in Σ (for instance in terms of AB) then the introduction of a metric $d(A,B)$ which leads to the same limit concept, is called a <u>metrization</u> of Σ. Metrizations are most frequently obtained in this way: Let $X \longrightarrow X' = f(X)$ be a topological mapping of the metric space Σ onto the metric space Σ'. Then $d(A,B) = f(A)f(B)$ is a metrization of Σ.

Next we discuss the arclength of continuous curves. (The definition and properties I.a,b,c,d,e) of the arclength are due to Menger [1,3,4] who studies much more general cases than ours.) Let $c : P(t)$, $a \leqq t \leqq b$, be a continuous arc. Take any subdivision

$$\Delta: \quad a = t_o < t_1 < t_2 < \ldots < t_n = b$$

of the interval (a,b) and form

$$L(\Delta) = \sum_{i=0}^{n-1} P(t_i)P(t_{i+1}).$$

The least upper bound of $L(\Delta)$ as Δ traverses all subdivisions of (a,b) is called the <u>arclength</u> or simply the <u>length</u> $L(c)$ of c (∞ admitted).

For every sequence $\Delta_i = (t_1^i, \dots, t_{n_i}^i)$ of subdivisions of (a,b) with

$$\lim_{i\to\infty} \max_{1\leq j\leq n_i} (t_j^i - t_{j-1}^i) = 0$$

we have

$$\text{PROPERTY L a)} \qquad L(\Delta_i) \longrightarrow L(c)$$

For, obviously

$$L(\Delta_i) \leq L(c).$$

For a given $\epsilon > 0$ let $\Delta = (t_o, \dots, t_n)$ be a subdivision of (a,b) with

$$L(\Delta) > L(c) - \frac{\epsilon}{2} \quad (L(\Delta) > N+\frac{\epsilon}{2} \text{ when } L(c) = \infty)$$

The interval (a,b) being compact it follows from Theorem 4 that a $\delta > 0$ exists such that

(3) $\qquad P(t)P(t') \leq \frac{\epsilon}{4n} \qquad$ for $|t-t'| < \delta$.

We now choose i so large that

$$\max_{1\leq j\leq n_i} (t_j^i - t_{j-1}^i) < \frac{1}{3}\min(3\delta, t_1-t_o, \dots, t_n-t_{n-1})$$

For a given t_k let $t_{j_k}^i$ be one (of the possibly 2) t_ν^i which has the minimal distance from t_k.

We have

$$|t_k - t_{j_k}^i| < \delta \text{ and } t_{j_k}^i < t_{j_{k+1}}^i, \; k = 0, \dots, n-1.$$

Hence

$$L(\Delta_i) \geq \sum P(t_{j_k}^i)P(t_{j_{k+1}}^i) \geq \sum P(t_k)P(t_{k+1}) - 2\sum P(t_k)P(t_{j_k}^i)$$

$$\geq L(\Delta) - 2n\frac{\epsilon}{4n} \geq L(c) - \epsilon .$$

This proves L a).

From L a) we conclude that the arc length is additive:

PROPERTY L b). If $a < d < b$ and c_1 and c_2 designate the subarcs $a \leq t \leq d$ and $d \leq t \leq b$ of $c = P(t)$, $a \leq t \leq b$, then

$$L(c_1) + L(c_2) = L(c)$$

The same point set may carry different continuous curves, since it may appear in different ways as a continuous image of an interval. (These curves will in general have different lengths. The usual discussion as to which changes of the parametrization do not change the length, can be applied without change to metric spaces.) When we say the points $P(t)$, $a \leq t \leq b$ of a continuous curve form a segment, we mean that the set of all points $P(t)$ can be mapped congruently onto a closed interval of the real axis, but we do not imply that $P(t) \to t$ is such a mapping (isometric representation).

PROPERTY L c). For every continuous curve $c = P(t)$, $a \leq t \leq b$ we have

(4) $P(a)P(b) \leq L(c)$

and if the equality sign holds the points $P(t)$ form a segment. If the points $P(t)$ form a segment and if $P(t_1) = P(t_2)$ for $t_1 < t_2$ implies $P(t) = P(t_1)$ for $t_1 \leq t \leq t_2$ the equality sign holds in (4).

PROOF. We have for every subdivision $\Delta = (t_0, \ldots, t_n)$:

$$P(a)P(b) \leq \sum P(t_i)P(t_{i+1}) \leq L(c).$$

If $P(a)P(b) = L(c)$ we have $P(a)P(b) = \sum P(t_i)P(t_{i+1})$ for every Δ therefore

(5) $P(t')P(t'') + P(t'')P(t''') = P(t')P(t''')$ for any $t' < t'' < t'''$. We now map $P(t)$ on the point $\tau = P(a)P(t)$ of the interval $0 \leq \tau \leq L(c)$. Let $0 \leq t_1 < t_2$ and $P(0)P(t_i) = \tau_i$, $i = 1,2$. It follows from (5) that

$$P(t_1)P(t_2) = P(a)P(t_2) - P(a)P(t_1) = \tau_2 - \tau_1$$

which proves that the $P(t)$ form a segment. $P(t_1) = P(t_2)$ for $t_1 < t_2$ may occur; for (4) implies in this case only that $P(t) = P(t_1)$ for $t_1 \leqq t \leqq t_2$.

This justifies the last part of L c). To prove it let $t' < t'' < t'''$. If two of the corresponding points $P(t^{(i)})$ coincide, (4) follows from our assumption. If the points are different we have either:
$(P(t')P(t'')P(t'''))$ or $(P(t')P(t''')P(t''))$ or $(P(t'')P(t')P(t'''))$ since the points belong to a segment. The last two relations are impossible. For in the case of $(P(t')P(t''')P(t''))$ we would vary t from t' to t'' and thereby pass a value t^* for which $P(a)P(t^*) = P(a)P(t''')$. Since the $P(t)$ form a segment we must have $P(t^*) = P(t''')$. Hence we should have $P(t) = P(t''')$ for $t^* \leqq t \leqq t'''$ according to our second assumption; but $P(t'') \neq P(t''')$. We see that (5) holds and herewith $L(\Delta) = P(a)P(b)$ for every Δ.

It follows from c) that a segment is a shortest connection of its end points, and that the points of any shortest connection of two points form a segment, if a segment joining the two points exists.

PROPERTY L d). <u>Lower semicontinuity of the arc length</u>. Let $c_\nu = P_\nu(t)$, $a_\nu \leqq t \leqq b_\nu$, $\nu = 0,1,2,\ldots$ be continuous arcs, with $a_\nu \longrightarrow a_0$, $b_\nu \longrightarrow b_0$, $P_\nu(a_\nu) \longrightarrow P_0(a_0)$ $P_\nu(b_\nu) \longrightarrow P_0(b_0)$ and $P_\nu(t) \longrightarrow P_0(t)$ for $a_0 < t < b_0$, then

$$(6) \quad \underline{\lim}\, L(c_\nu) \geqq L(c_0).$$

PROOF. Choose for a given $\epsilon > 0$ or N a subdivision $a_0 = t_0 < t_1 < \ldots < t_n = b_0$ of (a_0, b_0) so that

$$\sum_{i=0}^{n-1} P_0(t_i)P_0(t_{i+1}) > \begin{cases} L(c_0) - \epsilon & \text{if } L(c_0) < \infty \\ N & \text{if } L(c_0) = \infty \end{cases}$$

For large ν we have $t_1 > a_\nu$ and $t_{n-1} < b_\nu$ hence putting $t_0^\nu = a_\nu$, $t_n^\nu = b_\nu$, $t_i^\nu = t_i$ for $i \neq 0,n$

$$L(c_\nu) \geq \sum_{i=0}^{n-1} P_\nu(t_i^\nu)P_\nu(t_{i+1}^\nu) \rightarrow \sum_{i=0}^{n-1} P_0(t_i)P_0(t_{i+1})$$

Now (6) follows from the arbitrariness of ϵ or N.

PROPERTY L e). <u>Existence of a minimum.</u> In the finitely compact metric space Σ let a continuous arc of finite length from A to B exist. Then a continuous arc c from A to B exists whose length is smaller than or equal to the length of any continuous arc from A to B in Σ.

PROOF. Call b the greatest lower bound of the lengths of all continuous arcs from A to B. There is a sequence of (not necessarily different) arcs c_ν from A to B, so that $b_\nu = L(c_\nu) \rightarrow b$. On c_ν we may introduce the arc length s as parameter because the length is thereby not increased. We may thus get the representation $P_\nu(s)$, $0 \leq s \leq b_\nu$, $P_\nu(0) = A$, $P_\nu(b_\nu) = B$ for c_ν. For every s_ν with $0 \leq s_\nu \leq b_\nu$ the sequence of points $P_\nu(s_\nu)$ is bounded on account of L c).

Let $r_1, r_2, \ldots$ be the sequence of rational numbers between 0 and 1. We choose the subsequence $\{1_\nu\}$ of $\{\nu\}$ so that the points $P_{1_\nu}(r_1 b_{1_\nu})$ converge, then in $\{1_\nu\}$ a subsequence $\{2_\nu\}$ so that the points $P_{2_\nu}(r_2 b_{2_\nu})$ converge and so forth. We form the diagonal sequence $P_{\nu_\nu}(s)$. We conclude from the triangle inequality that $P_{\nu_\nu}(s)$ converges for every s between 0 and b, to a limit point P(s) and that P(s), $0 \leq s \leq b$ is a continuous arc c from A to B. We see from the definition of b that $L(c) \geq b$ and from L d) that

$$b = \lim L(c_{\nu_\nu}) \geq L(c),$$

so that $L(c) = b$, q.e.d.