METRIC FOUNDATIONS OF GEOMETRY. I

BY

GARRETT BIRKHOFF

1. **Introduction.** It is shown below by elementary methods that n-dimensional Euclidean, spherical, and hyperbolic geometry can be characterized by the following postulates.

Postulate I. Space is metric (in the sense of Fréchet).

Postulate II. Through any two points a line segment can be drawn, *locally* uniquely.

Postulate III. Any isometry between subsets of space can be extended to a self-isometry of all space.

The outline of the proof has been presented elsewhere (G. Birkhoff [2](1)). Earlier, H. Busemann [1] had shown that if, further, exactly one straight line connected *any* two points, and if bounded sets were compact, then the space was Euclidean or hyperbolic; moreover for this, Postulate III need only be assumed for triples of points. Later, Busemann extended the result to elliptic space, assuming that exactly one *geodesic* connected any two points, that bounded sets were compact, and that Postulate III held *locally* for triples of points (H. Busemann [2, p. 186]).

We recall Lie's celebrated proof that any Riemannian variety having local free mobility is locally Euclidean, spherical, or hyperbolic. Since our Postulate II, unlike Busemann's straight line postulates, holds in any Riemannian variety in which bounded sets are compact, our result may be regarded as freeing Lie's theorem for the first time from its analytical hypotheses and its local character.

What is more important, we use direct geometric arguments, where Lie used sophisticated analytical and algebraic methods(2). Our methods are also far more direct and elementary than those of Busemann, who relies on a deep theorem of Hurewicz. Our deepest assumption is that(3) a continuous function $f(x)$, which satisfies $f(a)=A$, $f(b)=B$, must take on all values between A and B in the interval $a \leqq x \leqq b$.

The direct geometric arguments used below are based, above all, on full use for the first time of the hypothesis of free mobility. This hypothesis appears, like Proteus, in a dozen different shapes to yield as many different

Presented to the Society, September 5, 1941; received by the editors November 24, 1942.

(1) Numbers in brackets refer to the Bibliography at the end of the paper.

(2) For a modern exposition of Lie's proof, cf. H. Weyl, *Das Raumproblem*, Springer, 1922. The undue reliance of Lie on analyticity is also criticized in Coolidge [1, p. 286].

(3) This is inevitable; thus it is needed to prove that the circles involved in Euclid's Prop. 1, Book I, actually do intersect. Hilbert's equivalent is Dedekind's axiom of "linear completeness." Cf. Heath [1, vol. I, p. 237 ff.].

theorems. Its use is thus analogous to Leibniz's use of the Principle of Sufficient Reason. It is seen as the "method of superposition," but freed from logical taint and flaw(4)—a fact which makes it possible to use it with a freedom and daring which was never possible with the method of superposition.

In Euclidean geometry, a similar role is played by the (apparently new) condition of *free expansibility*, Postulate III'. Any similitude between portions of space can be extended to a self-similitude of all space. The author hopes to demonstrate this, and discuss the two-dimensional case in detail, in a later paper.

I. Postulates and examples

2. **Metric spaces.** We shall begin by asserting that space is a *metric space* in the usual sense of Fréchet(5). This assertion is contained in the following two postulates, which the author regards as a slight improvement on the known postulates of Lindenbaum(6).

Postulate I. Space is a collection of undefined entities called *points*. Any two points x and y determine a *real number* called the *distance* from x to y, and written $|x-y|$. This distance satisfies identically,

$$|x-x| = 0, \quad \text{while} \quad |x-y| \neq 0 \quad \text{if} \quad x \neq y, \tag{1}$$

$$|x-y| + |y-z| \geqq |z-x| \qquad \text{(Circularity)}. \tag{2}$$

Setting $z=y$ in identity (2), and using (1), we get $|x-y| \geqq |y-x|$; interchanging x and y, we get the reverse inequality. Combining the two inequalities, we get

$$|x-y| = |y-x| \qquad \text{(Symmetry)}. \tag{3}$$

Again, substituting x for z in (2), and using (3), (1), we get $|x-y|+|x-y| \geqq 0$, whence, dividing by two,

$$|x-y| \geqq 0 \qquad \text{(Positiveness)}. \tag{4}$$

Finally, combining (2) and (3), we get immediately

$$|x-y| + |y-z| \geqq |x-z| \quad \text{(Triangle inequality)}.. \tag{5}$$

Incidentally, the second half of (1) is not required in the proofs of (3)–(5), so that the first half of (1) and (2) suffice to define quasi-metric spaces.

A system satisfying Postulate I is called a *metric space*, and all the notions

(4) Cf. Heath [1, vol. I, p. 249, p. 226, and related discussion]. This discussion also brings out the close relation between free mobility and the method of superposition.

(5) Called "classe (E)" in M. Fréchet, *Sur quelques points du calcul fonctionnel*, Rend. Circ. Mat. Palermo vol. 22 (1906) pp. 1–74. For the theory of metric spaces, cf. L. M. Blumenthal [1].

(6) A. Lindenbaum, *Contributions a l'etude de l'espace metrique*, Fund. Math. vol. 8 (1926), pp. 209–222. The author feels that the "circularity postulate" has a clear intuitive content: if one journeys from p to q and then from q to r, the minimum energy required to get back to p is not more than that already expended. We thus assume Euclid's Prop. I. 20.

of general topology apply to metric spaces(7). To indicate this, we recall the following common notions, applying to general metric spaces.

Definition. The *sphere* of (positive) radius ρ and center c is the set of all points x satisfying $|x-c| \leqq \rho$. A set S is *bounded* if it is contained in some sphere —or equivalently, if $\sup_{x,y \in S} |x-y|$ is finite. This supremum is called the *diameter* of S. A point a is called a *cluster point* of S if every sphere with center a contains a point of S not a. A *sequence* $\{x_i\}$ is said to *converge* to the *limit* a (in symbols, $x_i \to a$) if and only if $|x_i - a| \to 0$. A sequence $\{x_i\}$ such that $|x_i - x_j| \to 0$ as $i, j \to \infty$ is called a *Cauchy sequence.* Every convergent sequence is a Cauchy sequence; if every Cauchy sequence is convergent, the space is called *complete.* A metric space is called *compact* if every sequence contains a convergent subsequence; *locally compact,* if every point is the center of some compact sphere. A subset S of a metric space M is called *dense* if every point of M is a point or cluster point of S. A metric space is called *separable* if it contains a finite or countable dense subset.

An excellent list of examples of metric spaces may be found in Blumenthal [1, chap. 1, §5]; others are described in Busemann [2].

3. **Digression: scales of distance.** Metric spaces go into metric spaces under any *convex change of scale.* That is, let $f(\rho)$ be any real function such that $f(0)=0$, such that $f(\rho) < f(\sigma)$ whenever $\rho < \sigma$, and such that $f(\rho+\sigma) \leqq f(\rho)+f(\sigma)$. Then identical substitution of $f(|p-q|)$ for $|p-q|$ in any metric space E gives a new metric space E_f. Such an E_f is called a *metric transform* of E.

For much of what follows, it is not necessary that the distance between two points be a real number. It can be an element of an ordered field(8), of an ordered Abelian group, or even a partially ordered Abelian group(9). In all these cases, the metric postulates have meaning, as do the straight line postulates and isometry postulates to be discussed below. The study of the consequences of these postulates under such more general notions of distance might yield results in metric geometry as fruitful as those which have been obtained in the foundations of projective geometry by constructing abstract projective geometries over general fields and skew-fields.

4. **Straight line postulates.** It is well known that one can define straight lines, and so on, in general metric spaces, as follows(10).

Definition. A set whose points can be put into one-one correspondence with a segment of the real line, so that distance is preserved, is called a *straigh*

(7) This observation is mainly due to Fréchet, op. cit.; for a systematic treatment, c Hausdorff, *Mengenlehre.*

(8) Cf. K. Reidemeister, *Grundlagen der Geometrie,* Springer, 1930, p. 40. This fact is signi cant in non-Archimedean geometry.

(9) Cf. G. B. Price, *A generalization of a metric space, with applications to spaces whose ments are sets,* Amer. J. Math. vol. 63 (1941) pp. 46–56.

(10) Cf. K. Menger [1]; also Blumenthal [1] and Busemann [2].

segment. If the segment is the whole real line, the given set is called a *straight line.* Three points on the same straight segment are called *collinear.* The term *ray* (half-line) is self-explanatory.

If we replace the hypothesis that distance is preserved by the weaker hypothesis that distance is preserved *locally*, we obtain metric definitions of a *geodesic segment* and of a *geodesic line* without multiple points.

Now it is well known that Euclid's postulate that "through two given points passes one and only one straight line" does not hold on the sphere, or on numerous other important manifolds; it is too strong in the large. Hence we shall replace it by the following weaker postulates.

Postulate II. Any two points can be joined by a straight segment.

Postulate II'. Given a, there exists a $\delta>0$ so small that two straight segments issuing from a of equal length less than δ are either identical or have only a in common.

Postulate II' is equivalent to Axiom D of Busemann [2]; it holds in any Riemannian variety (hence differentiable manifold). Postulate II holds provided the variety is metrically complete([11]) (that is, in *closed* differentiable manifolds).

It is clear that Postulate II describes behavior in the large, while II' describes behavior in the small. If we wished to reverse this, we could postulate that (IIa) any two sufficiently near points can be joined by a straight segment, and (II'a) no two points are connected by more than one geodesic. Postulate IIa would hold, for example, in a disconnected space. Postulates II and II'a hold in the *straight line spaces* (Geradenräume) of Menger (cf. Busemann [2, chap. 3]); they hold, for example, in any simply connected Riemann space of negative curvature. They also hold in Euclidean, hyperbolic, and elliptic n-space; but not in spherical n-space.

5. **Digression: relation to betweenness and convexity.** Menger [1] has given a very interesting analysis of straight line postulates in terms of the relation of metric betweenness, as defined below.

Definition. The point x is *between* a and b (in symbols, axb) if and only if $|a-x|+|x-b|=|a-b|$.

He has pointed out that various of the usual([12]) properties of betweenness are valid. Moreover (i) in a *complete* metric space, Postulate II holds if and only if a point can be found "between" any two points. Again (ii) the segment ab is unique if and only if axb and ayb imply axy or ayx ("inner con-

([11]) For the general situation regarding Postulate II, cf. M. Morse, *The calculus of variations in the large*, New York, 1934, chap. 5. Regarding Postulate II', cf. Busemann [2, p. 55]; the proof is easy if bounded sets are compact.

([12]) E. V. Huntington and J. R. Kline, *Sets of independent postulates for betweenness*, Trans. Amer. Math. Soc. vol. 18 (1917) pp. 301–325. Cf. also ibid. vol. 26 (1924) pp. 257–282, and M. Pasch, *Neuere geometrie*, Leipzig, 1882, pp. 64–72. Cf. Blumenthal [1].

vexity"), while (iii) the segment ab can be extended uniquely beyond b if and only if abx and aby imply axy or ayx ("outer convexity").

Postulate II′ is equivalent to assuming that (ii) and (iii) hold locally. We note that (iii), which excludes the possibility of a geodesic branching in two, holds in the large if it holds locally; whereas (ii) holds locally, but not in the large, on the sphere and many other Riemannian spaces. Also, that Euclid distinguished inner from outer convexity in his Postulates 1, 2.

It is natural to ask why the author preferred to assume Postulates II–II′ to conditions (i), (iii), and (ii) locally. The answer is that he would then have had to *assume* that space was not only metric but complete; and this assumption seemed to him far more recondite than the ones made.

However, he feels that Postulate II′ is still in a rather unsatisfactory form.

6. **Isometry postulate.** An *isometry* between metric spaces M and N is a one-one correspondence between the points of M and the points of N which preserves distance. A *similitude* is one which multiplies all distances by the same nonzero factor. A *homeomorphism* is one which preserves convergence.

If there exists an isometry between M and N, they are called *isometric* (in symbols, $M \cong N$); if there is a similitude between them, they are called *similar* (in symbols, $M \simeq N$); if there is a homeomorphism between them, they are called *homeomorphic* (in symbols, $M \sim N$). Evidently isometry is a special case of similitude, and similitude is a special case of homeomorphism.

Evidently also, any similitude carries straight lines into straight lines, and spheres into spheres with corresponding centers. In fact, the entire theory developed below is invariant under similitudes.

It is well known, and evident, that the set of all isometries of any metric space with itself is a *group*, called the group of *self-isometries* (autometries) of the space. For example, the group of autometries of the real line consists of the *identity* $x \to x$, the *translations* $x \to x+a$ $(a \neq 0)$, and the *reflections* $x \to a-x$. There are also the self-similitudes $x \to ax+b$ $(a \neq 0)$; these again form a group.

Since the time of Helmholtz, it has been realized that Euclidean geometry depended largely on a property of "free mobility" of bodies in space; and that this property held equally in spherical and hyperbolic space. The precise formulation of this property was accomplished by Pasch([13]), and we shall assume it as our last postulate.

Postulate III. Any isometry between subsets of space can be "extended" to an autometry of all space.

Explanation. By an *extension* of a correspondence σ between sets M and N,

---

([13]) M. Pasch, *Neuere Geometrie*, Leipzig, 1882, p. 109. For Helmholtz, cf. *Über die Tatsachen, die der Geometrie zu Grunde liegen*, Nachr. Ges. Wiss. Göttingen (1868). For refinements of the condition see Blumenthal, chap. 3, §2 (or §7 below); Postulate III asserts, in Blumenthal's language, that space as a whole is "freely movable." Cf. footnote 4 above.

is meant any correspondence between sets $S \supset M$ and $T \supset N$ which coincides on M with the original correspondence σ.

Postulate III is not to be confused with the frequently stated weaker condition: "If M and N are isometric sets, then there exists a self-isometry of space which carries M into N." It should also be distinguished from the *n-point homogeneity* condition: "Any isometry between two sets of n or fewer points can be extended to a self-isometry of space." Thus Hilbert space has n-point homogeneity for every finite order n, yet does not satisfy the free mobility postulate; the same is true of Urysohn space (P. Urysohn, *Sur un espace métrique universel*, Bull. Sci. Math. vol. 51 (1927) pp. 43–64, 74–90).

Historically, 3-point homogeneity has been the most used; it is equivalent to Euclid's Prop. I.8. Thus it is just what is needed for the theory of Riemann-Helmholtz-Lie (cf. Coolidge [1, p. 279]); it is also what is used by Busemann [1, 2]. The example of Hilbert space is ruled out traditionally by a hypothesis implying finite-dimensionality. We mention in passing that 1-point homogeneity amounts to the condition that the group of isometries is transitive.

We suggest calling Postulate III the *hyperhomogeneity* condition, and Postulate III′ of §1, the *hypersimilitude* condition.

It is well known, and not too hard to show, that Postulate III is valid in Euclidean n-space. From this its validity in spherical n-space follows from the following obvious principle.

If a space satisfies the free mobility postulate, then so do all its spheres, and so do all its metric transforms.

For any isometry between sets M and N on a sphere S with center a can be extended to an isometry $(a, M) \cong (a, N)$, and then extended to an autometry of all space leaving a invariant; this will induce an autometry on S. So much for the first assertion; the second is true simply because the meaning of isometry is invariant.

But now, spherical n-space is a metric transform of a sphere in Euclidean n-space, completing the proof. As regards the free mobility of hyperbolic n-space, it seems generally known(14), although the author has not seen any explicit proof of the fact.

On the other hand, elliptic 2-space does not satisfy the free mobility postulate. Thus the triangles whose vertices (in latitude and longitude) are (60°, 0°), (−60°, 0°), (0°, 30°) and (60°, 0°), (60°, 180°), (30°, 90°) are isosceles triangles with corresponding sides equal, and in fact having identical bases; yet they are not congruent (corresponding angles are not equal). Neither (cf. Theorem 8 infra) does elliptic n-space ($n > 2$) satisfy the free mobility postulate.

II. Elementary theorems

7. **Straight lines.** Hyperhomogeneity implies directly strong limitations

(14) Cf. Busemann [1, p. 101]; or Coolidge [1, p. 29, Axiom XIX].

on the behavior of straight lines. This does not appear to have been remarked before (cf. footnote 16).

THEOREM 1. *Through any two points passes a geodesic cycle*([15]). *The self-isometries of this include reflection in any point and translations through any distance.*

Proof. The translation-isometry between one end of a given straight segment and the opposite end can be extended, locally uniquely in the presence of Postulate II′. This construction, iterated indefinitely in both directions, gives a geodesic cycle. Further, any translation or reflection of a small segment of a geodesic is, by definition, an isometry. By Postulate III, it can therefore be extended to all space—but it clearly carries geodesics into geodesics. Since geodesics can be *uniquely* extended (outer convexity), it therefore carries the geodesic into itself. We conclude that the group of autometries of any geodesic includes reflections in all points, and (taking powers of small translations) translations through any amount. Similarly, the obvious isometry between any two equilong segments of two geodesics can be (uniquely) extended to an isometry between the geodesics as a whole.

THEOREM 2. *If a and b can be joined by more than one straight segment, then $|a-b|$ is the diameter of space.*

Proof. Suppose $|x-y| \geq |a-b|$ for some x, y. Mark off on xy a point z such that $xz \cong ab$. By double homogeneity (III), there will be *two* straight segments joining x and z (cf. Fig. 1). Hence yz could be extended in two ways, which contradicts outer convexity, and so Postulate II′.

FIG. 1

COROLLARY 1. *If space is unbounded, then one and only one straight line passes through two given points*([16]).

COROLLARY 2. *Postulates* I, II, II′ *and* III′ *imply Postulate* II′a.

In any metric space, we can introduce a parameter s of length along any geodesic, and speak of the "length" of a geodesic segment. What we have done is to show that all geodesic segments of equal length are isometric. Hence if space is unbounded, they are all straight segments. Even if space has finite diameter d, all segments of length less than d (hence, by continuity,

([15]) We mean this word in the topological sense, that the geodesic has no boundary (is non-terminating). Actually, this is implied by homogeneity.

([16]) Thus all Busemann's postulates hold. Since only double homogeneity was used, it follows that Busemann could have replaced his strong uniqueness Postulate II′a by the weaker Postulate II′ and the hypothesis of unboundedness.

all segments of length d) are straight segments. To summarize, we state the following theorems.

THEOREM 3. *If space is unbounded, all geodesics are infinite straight lines; if space has finite diameter d, all geodesic segments of length d are straight segments.*

THEOREM 4. *If space has finite diameter d, then every geodesic is periodic, of period 2d.*

Proof. Denote by s' the point with geodesic parameter s; by translation-homogeneity (Theorem 1), we need only show that O' and $(2d)'$ coincide. To show this, given $\delta>0$, let $y=d+\delta$ and $z=|y'-O'|$. By hypothesis, $z\leqq d$; and by Theorem 3 and the triangle inequality, $|z-O|\geqq|d'-O'|-|y'-d'|$. $=d-\delta$; hence $|y-z|\leqq 2\delta$. Moreover, by construction, the correspondence $O'y'z'\rightarrow O'z'y'$ is an isometry. Its extension to space will clearly leave the geodesic $y'z'$ setwise invariant, reflecting it in a point within δ of d'; hence it will carry O' into a point within 2δ of $2d'$. But, by construction, it leaves O' fixed; hence $|2d'-O'|\leqq 2\delta$ for all $\delta>0$, and $2d'=O'$.

COROLLARY. *In the bounded case, more than one straight segment joins at least one pair of distinct points.*

That is, inner convexity implies that space is unbounded. It is worth noting that, in the proofs of Theorems 1–3, only double homogeneity was used; while Theorem 4 requires only the 3-point form of Postulate III. Full hyperhomogeneity has not been used up to now.

8. **Digression: partial generalizations.** Even if Postulates II–II′ are replaced by much weaker conditions, hyperhomogeneity implies limitations on the behavior of straight lines. Let us first see what can be proved if we combine Postulate III with the first of the following two weak conditions:

(IIb) There exist two distinct points which can be joined by a straight segment.

(II′b) There exist two points joined by only one straight segment.

Clearly any pair of nearer points can then be joined by a straight segment, using double homogeneity. Let m be the least upper bound to the lengths of straight segments. We can distinguish two cases.

Case I: $m=+\infty$. This is necessarily the case if Postulate III′ is assumed—a similitude exists extending part of a line segment, and hence all space, in the ratio 1:2. In this case, by iterated extension, one can construct an infinite straight line. But by (III) again, since this must contain points arbitrarily far apart, we have

(IIc) Through any two points passes an infinite straight line.

Case II: $0<m<+\infty$. As in Theorem 1, we can construct an infinite geodesic through any two points. But as this need not be unique, it is hard to

see why all its geodesic segments of length between $m/2$ and m need be straight. However, any two points whose distance apart is less than m can be joined by a straight segment; no two points further apart than m can.

Now note that if two points a, b can be joined by more than one straight segment, then so can any two points x, y with $|x-y| \geqq |a-b|$, as in Theorem 2 (Fig. 1) above. Hence (II'b) and (III) imply *local inner convexity.*

If we assume (II), then we can limit the possibility of inner convexity unaccompanied by outer convexity.

THEOREM. *If a metric space satisfies* (II), (III), *and* (II'b), *but not* (II'), *then it is topologically one-dimensional and is not locally compact*[17].

Proof. Without loss of generality, by change of scale, we can assume inner convexity for segments of length less than 2. Now let o be any point, and S resp. T the spheres $|x-o| = 1/2$ and $|x-o| = 1$. Let θ be the map $T \to S$ such that $t\theta = s$ means

$$|s-t| = |s-o| = 1/2; \tag{A}$$

that is, shrinking T into S along radii to o. By inner convexity, θ is single-valued; it maps a closed set onto a given s, by substitution in the definition (A). By hyperhomogeneity, each ot ($t \in T$) has a branch point (it is here that the lack of outer convexity comes in) at its midpoint s, the other branch extending to some t' on T. By inner convexity, $t' \neq t$; and clearly o, s, t and o, s, t' are collinear sets. Moreover if $x \in T$ and $|x-t| = |t'-t|$, then the correspondence o, t, $t' \to o$, t, x is isometric, and so (by Postulate III) can be extended to a self-isometry of space. The latter will leave ot, including s, pointwise invariant, and will map ot' onto ox; hence s is also the midpoint of ox. We conclude that θ maps all points of T on a circle of radius $|t'-t|$ about t onto s. The same will be true a fortiori for points of T inside the circle (the rays to o can never escape).

We infer at this point that if space were locally compact, then, since the surface of the sphere would be covered by a finite number of such circles, S would have only a finite number of points, leading quickly to an absurdity. In any case, the circles are closed (being the antecedents of closed sets under a continuous map), open (contain with t' all x with $|x-t'| \leqq |t-t'|$), disjoint, and may (replacing $1/2$ by $1-e$, where e is small) be made arbitrarily small. Hence each sphere is *totally disconnected* (zero-dimensional), and the space is one-dimensional topologically.

9. **Angles; reflection in a point.** We define an *m-cross* with *vertex* a as an ordered set $L_1, \cdots, L_m$ of m geodesics through a; and define an *m-hedral*

---

[17] This is interesting, since Busemann postulates local compactness. Again, we need only the 3-point form of Postulate III; this shows that Busemann need have assumed only inner convexity.

with vertex a similarly as an ordered set of rays([18]) (half-geodesics) issuing from a. A 2-hedral is called an angle, and is written L_1aL_2.

We can also define *straight angles* (trivially), and *equal angles*: $\angle LaM = \angle L'a'M'$ means that there is an isometry which carries L into L', a into a', and M into M'. But we cannot yet speak of the *sum* of two angles, or even of one angle being *greater* than another. We shall be able to compare angles in §13, but we shall not be able to add them until we construct 2-dimensional subspaces (planes).

It is clear that any two straight angles are equal; moreover we have the following theorem.

THEOREM 5. *Any angle is equal to its transpose*([19]); *and the opposite angles of any 2-cross are equal.*

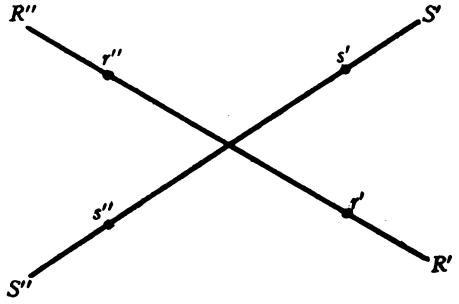

FIG. 2

Proof. Let the lines be R and S meeting at a, cut by a into rays R', R'', S', S''. Mark off on R' and S', r' and s' at equal distances (less than the diameter of space) from a. Then the correspondence $a, r', s' \to a, s', r'$ is an isometry, which can be extended to space by Postulate III. The extension will carry $\angle R'aS'$ into $\angle S'aR'$ and interchange $R'aS''$ with the opposite angle $\angle S'aR''$, proving the two assertions of the theorem.

If we mark off r'' and s'' at the same distance on R'' and S'', we get a similar symmetry of space inducing the permutation $a, r', s', r'', s'' \to a, s'', r'', s', r'$. The product of this with the preceding isometry will yield an isometry permuting $a, r', s', r'', s'' \to a, r'', s'', r', s'$. Thus this isometry will *reflect* R and S in the point a, from which follows the lemma:

LEMMA 1. *If we reflect any two geodesics through a in a, we get an isometry of the 2-cross formed by the lines.*

([18]) In the bounded case, a ray covers an entire geodesic; to avoid ambiguity, one must think of it as sensed with the origin a. In the unbounded case, one can identify m-hedrals with the m-dimensional sectors which they enclose.

([19]) The transpose of $\angle LaM$ is $\angle MaL$. Note that thus Hilbert's Postulate III.5 [1, p. 14] is demonstrable in our system. Also, Euclid's Prop. I. 15 is demonstrable without appeal to dubious univalent "differences" of angles.

Definition. By the *reflection* of space in a point a is meant the transformation which reflects each geodesic through a in the point a.

We remark that this is a single-valued transformation of points. For unless b is antipodal([20]) to a, there is only one geodesic through a and b; while if b is antipodal to a, reflection in a of any geodesic through a and b leaves b fixed. By Lemma 1, it is isometric; hence we have the following theorem.

THEOREM 6. *Reflection in any point is an isometry of all space*([21]).

10. **Comparison of angles; antipodal points.** Let $\theta = \angle LaL'$ be any angle, and let $g(\theta; \rho)$ denote the distance $|p-q|$ for $p \in L$, $q \in L'$, and $|p-a| = |q-a| = \rho$. Then $g(\theta; \rho)$ is a single-valued function; in the Euclidean case, $g(\theta; \rho) = 2\rho \sin(\theta/2)$.

Definition. By $\theta > \theta'$, we mean that $g(\theta; \rho) > g(\theta'; \rho)$ for all $\rho[0 < \rho < d]$.

LEMMA 2. *If* $g(\theta; \rho) = g(\theta'; \rho)$ *for one* ρ $[0 < \rho < d]$, *then* $\theta = \theta'$.

Proof. The correspondence $a, p, q \to a', p', q'$ between the vertices of the triangles with angles θ, θ' and sides ρ, ρ' (cf. Fig. 3) will then be an isometry; its extension to space will map θ on θ'.

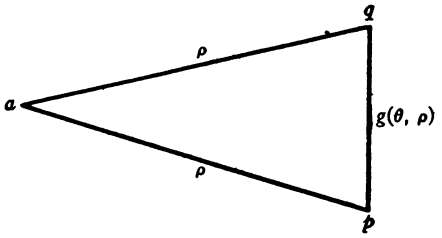

FIG. 3

LEMMA 3. *For any* θ, $|g(\theta; \rho) - g(\theta; \rho')| \leqq 2 \cdot |\rho - \rho'|$.

The proof is by the triangle inequality, and is valid in any metric space.

LEMMA 4. *If* $g(\theta; \rho) > g(\theta'; \rho)$ *for one* ρ $[0 < \rho < d]$, *then* $\theta > \theta'$.

Proof. At the g.l.b. σ_0 of the $\sigma > \epsilon$ $[\epsilon > 0]$ such that $g(\theta; \sigma) \leqq g(\theta'; \sigma)$ we would have by continuity (Lemma 2) $g(\theta; \sigma_0) = g(\theta'; \sigma_0)$; hence (by Lemma 1) $\theta = \theta'$ or $\sigma_0 \geqq d$. But $\theta = \theta'$ implies $g(\theta; \rho) = g(\theta'; \rho)$, and thus is contrary to hypothesis.

([20]) Two points are called *antipodal* when their distance apart is the diameter of space. We shall prove shortly that antipodal points occur in pairs, like the north and south poles.

([21]) The assumption that Theorem 6 holds is the basic assumption of E. Cartan's theory of symmetric Riemann spaces (*La theorie des groupes finis et continus et l'analysis situs*, Gauthiers-Villars, Paris, 1930; *Les espaces riemanniens symmetriques*, Zurich Congress, 1932, vol. I, pp. 152–161). Note also Busemann [2, p. 183, Theorem 8].

THEOREM 7. *Under the relation $\theta > \theta'$, angles form a simply ordered set, order-isomorphic to a subset of reals, $0 \leqq \lambda \leqq 1$. (Unless space is a point or line.)*

Proof. Fix any ρ_0 $[0 < \rho_0 < d]$, and make the definition $\lambda = g(\theta; \rho_0)/2\rho_0$. We shall show later (by homogeneity arguments) that all reals are included.

THEOREM 8. *In the bounded case, every point p has a unique antipode p'; every ray of length d issuing from p terminates in p'; $g(\theta; d) = 0$ for all θ.*

Remark. This is the first (and principal) consequence of Postulates I–III which does not apply to elliptic n-space, but only to spherical n-space.

Proof. First, if π denotes a straight angle, then by Theorem 3, $g(\pi; \rho) = 2\rho$ for all $\rho < d/2$. Hence, by the triangle inequality and Lemmas 2, 4, $\pi \geqq \theta$ for all θ, and so (again by Theorem 3) $2\rho = g(\pi; d-\rho) \geqq g(\theta; d-\rho) \geqq 0$ for all ρ, $0 < \rho < d/2$. Hence, letting $\rho \to 0$, by continuity (Lemma 2) $g(\theta; d) = 0$ for all θ. This gives both the second and third statements of Theorem 8. As regards the first statement, if p' and p'' are both antipodal to p, then by Postulate II we can find rays of length d from p to both p' and p'', whence by the second statement, $p' = p''$.

COROLLARY 3. *If space is bounded, any two geodesics which intersect have two antipodal points in common, and meet at the same angles at these two intersection points.*

COROLLARY 4. *If space is bounded, the antipodal involution $p \to p'$ is an isometry.*

Proof. Given p, q, form $pq = pqp'q'$. The antipodal involution acts like a translation through a distance d.

We note that the antipodal involution is easily shown to be permutable with every isometry.

11. **Right angles.** We have seen above that a 2-cross C always has three symmetries besides the identity. These may be thought of as reflections ϕ', ϕ'' in the bisectors of the angles, and reflection ϕ in the vertex. If C has any other isometry θ, then θ and the products $\theta\phi', \theta\phi'', \theta\phi$ give four more isometries, or eight in all([22]). But there are only eight ways to permute four half-lines so that opposite (collinear) half-lines stay opposite; hence no more symmetries are possible. We conclude with the following theorem:

THEOREM 9. *A 2-cross admits either exactly four or exactly eight symmetries (autometries).*

Definition. An angle is a *right angle* if the 2-cross L, M formed by extend-

([22]) We omit the proof that these are distinct. The reader familiar with Lagrange's theorem on finite groups will recognize the type of argument.

ing its sides admits eight symmetries; the lines L, M will then be called *perpendicular* (in symbols, $L \perp M$).

The preceding argument shows that perpendicularity of a 2-cross is equivalent to each of the following conditions: (i) the existence of a symmetry permuting the four rays cyclically, (ii) the existence of a symmetry holding one line pointwise invariant and reflecting the other in the vertex, (iii) the equality of each angle with its "supplement"([23]). (The last is Euclid's definition.)

Consider the *point-line* configuration consisting of a line L and a point p not on L. By Theorem 8, p cannot be antipodal to any point $q \in L$; hence by Theorem 2, p and q can be joined by exactly one straight segment. Moreover clearly $pq \perp L$ if and only if reflection of L in q is an isometry relative to p; such a point we shall call a *symmetry point* of L for p.

THEOREM 10. *Let m be the midpoint of rr' on L. Then $pm \perp L$ if and only if $|p-r| = |p-r'|$.*

Proof. The necessity is trivial. As regards the sufficiency, the isometry $p, m, r, r' \to p, m, r', r$ can be extended to all space([24]). The extension leaves p fixed, and reflects $L = rmr'$ in m.

COROLLARY. *In intervals of L containing no symmetry points for p, $f(s) = |s-p|$ is monotone.*

Explanation. It is understood, as in the proof of Theorem 4, that s measures arc-length on L.

Proof. Unless $f(s)$ is monotone, we shall have([25]) $f(s_1) = f(s_2)$ for some $s_1 \neq s_2$; hence $(s_1+s_2)/2$ is a symmetry point.

THEOREM 11. *If space is unbounded, then exactly one perpendicular pq can be dropped from p to L; moreover $f(s) = |s-p|$ increases monotonely as s recedes from q.*

Proof. The product of two reflection symmetries of L for p gives a translation symmetry. If space is unbounded, this leads to a contradiction: iteration of such a symmetry would carry any $s_1 \in L$ arbitrarily far from its original position, hence could not leave its distance from p constant. The bitone nature of $f(s)$ is now a consequence of the corollary of Theorem 10.

([23]) It is easy to define the "supplement" of an angle, and to show that supplements of equal angles are equal. We have not yet proved, however, that all right angles are equal.

([24]) This uses the 4-point form of Postulate III. If we used the 3-point form (omitting m), the case $|r-r'| = d$ would be exceptional. Theorem 10 implies that the base angles of an isosceles triangle are equal.

([25]) We omit the proof of this result of function theory. It is a corollary of the fact that a continuous function assumes a continuous range of values. This theorem can be stated geometrically as a postulate on the behavior of lines and spheres. Thus $f(s)$ is monotone on an interval when no sphere with center p cuts the interval more than once, and so on.

COROLLARY. *Every sphere with center p and radius exceeding $|p-L|$ cuts L exactly twice.*

LEMMA. *If space has finite diameter d, if $|q-q'|<d$, and if $|q-p|=|q'-p|=d/2$, then every point x on qq' satisfies $|x-p|=d/2$.*

Proof. Consider the isometry engendered by reflection in p followed by the antipodal involution. It leaves fixed those points x such that p bisects xx' and no other points—hence, by Theorem 8, those such that $|x-p|=d/2$. But clearly any isometry which leaves q, q' fixed ($|q-q'|<d$) leaves qq' pointwise fixed—hence the lemma.

THEOREM 12. *If space is bounded, then either* (i) *for all $q\in L$, $|p-q|=d/2$ and $pq\perp L$, or* (ii) *exactly one geodesic through p is perpendicular to L; it cuts L at antipodal points q, q', with $|p-q|<d/2<|p-q'|$.*

Proof. If every $x\in L$ satisfies the equation

$$|x-p|=d/2, \tag{B}$$

then by Theorem 10 every pm ($m\in L$) is perpendicular to L.

Otherwise, since for any antipodal points s, $s'\in L$ we have $|s-p|+|s'-p|=d$ (cf. Theorems 4, 8), for some such pair we have $|s-p|<d/2<|s'-p|$. Hence, since a continuous function assumes a continuous range of values, (B) must have at least two solutions x_1, x_2, one in each of the intervals of L separating s, s'. But it cannot have more than two solutions, as otherwise $|x-y|<d$ for two solutions, and by the lemma $|x-p|=d/2$ for all $x\in L$, contrary to hypothesis. Hence it has exactly two antipodal solutions x_1, x_2. The antipodal midpoints q, q' of the intervals of L between x_1, x_2 will be, by Theorem 10, feet of perpendiculars from p to L. But now, reflection of L in the foot of *any* perpendicular from p to L will permute the (two) solutions of (B) among themselves, and q, q' are the only points which do this([26]). Hence the geodesic qpq' cuts L twice orthogonally, and no other line through p cuts L orthogonally, q.e.d.

Arguments almost identical with those used in Theorem 11 will prove Euclid's Prop. I.16 in case space is unbounded. In the bounded case, an argument like that used in proving Theorem 12 will show that the angles between pq and L oscillate monotonely between extreme values at x_1 and x_2. For (cf. Fig. 4) if the same exterior angle is assumed twice, translation of L so as to move q' into q will be an isometry—hence interchange x_1 and x_2, hence be a rotation of L through d (180°).

12. **Triangles.** A "triangle" means a configuration which consists of three

([26]) The argument just presented resolves an old difficulty noted by Proclus (Heath [1, vol. 1, p. 272]), but never before resolved. We also show that "spheres are convex" (cf. Busemann [2, chap. 4]).

points, called "vertices," and three straight segments, called "sides," joining the vertices in pairs. Thus any triangle has three angles, one at each vertex. The vertices clearly determine the sides, unless space is bounded and two of the vertices are antipodal.

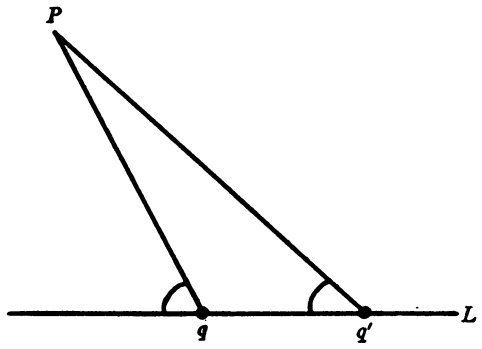

FIG. 4

THEOREM 13. *Two triangles are isometric if two sides and the included angle are equal, or—unless one side has length d—if their sides are equal in pairs. Either no triangle, one right triangle, or two triangles have two sides and a given angle adjacent to one side—unless the sides are d/2 and the angle a right angle ("Ambiguous Case").*

Proof. Under the first hypothesis, there exists an isometry carrying the angle of one into the equal angle of the other. This will then have to carry the other two vertices of the first triangle into the corresponding vertices of the other; hence the triangles are isometric. The second assertion is a corollary of Postulate III applied to the vertices, and the exceptional case is genuine (cf. lunar triangles on spheres).

As regards the last assertion, let $\angle LaM$ be the given angle, and $ap=L$ the given adjacent side. The number of possibilities is equal to the number of lines of the given length which can be dropped from p to M, the opposite side. Now use Theorems 11 and 12, and their proof.

THEOREM 14. *Two right triangles which have equal sides opposite right angles ("hypotenuses"), and one pair of sides of equal length λ adjacent to these angles, are isometric unless $2\lambda=d$.*

Proof. Let the triangles be qpr and $q'p'r'$, with right angles at p and p', $|p-r|=|p'-r'|$, and $|q-r|=|q'-r'|$. Extend rp and $r'p'$ geodesically to s and s', so that $|p-s|=|p-r|$ and $|p'-s'|=|p'-r'|$. Then, by symmetry about pq and $p'q'$, $|q-s|=|q-r|=|q'-r'|=|q'-s'|$. Hence, by Theorem 13 there is an isometry between Δqrs and $\Delta q'r's'$ unless $|r-s|=d$

(the exceptional case $2\lambda = d$); this will take p into p' and hence map $\triangle qpr$ isometrically on $\triangle q'p'r'$, as asserted.

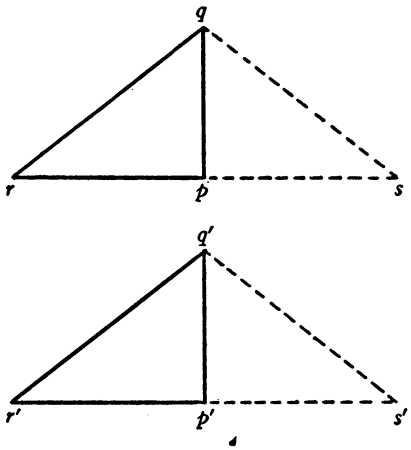

FIG. 5

THEOREM 15. *All right angles are equal*([27]).

Proof. Let $\angle LpM$ and $\angle L'p'M'$ be right angles. Lay off short segments pq and $p'q'$ on L resp. L' of equal length λ. Consider the function $f(s) = |r-q|$, where r is s units along M from p. Since $f(0) = \lambda$, $f(3\lambda) \geqq 2\lambda$, and $f(\lambda)$ is continuous, we must have $f(s) = |r-q| = 2\lambda$ for some r. Similarly, for some r' on M', $|r'-q'| = 2\lambda = |r-q|$. Now we can apply Theorem 14.

COROLLARY. *Two right triangles are isometric if their adjacent sides are equal, respectively.*

13. **Dimension.** It is natural to characterize the "dimension" of a space as the maximum number of concurrent mutually perpendicular straight lines which can be found in it. The theorem that all right angles are equal, combined with hyperhomogeneity, makes this characterization a workable definition.

Definition. By an *orthocross*, we mean a cross whose axes are perpendicular. An orthocross which is not a part of any larger orthocross will be called a *rectangular basis.*

THEOREM 16. *Any space has a rectangular basis.*

Proof. Take any center and any straight line through the center. This is

([27]) This is Postulate 4 in Euclid's system, but modern writers prove it from other assumptions. Cf. Hilbert, p. 23, footnote. The essential point in all proofs is that the triangular distance functions increase monotonely with the angle.

an orthocross. Unless a hypercross is maximal, it can be enlarged. But the process of adding on new axes, if continued without limit, will ultimately (transfinitely) exhaust all points of space. When it cannot be further continued, it will give us a rectangular basis.

THEOREM 17. *Any one-one correspondence between the axes of two orthocrosses can be induced by a suitable isometry.*

For since all right angles are equal (Theorem 15), as soon as a positive and negative direction on each axis has been prescribed, the correspondence between points at an equal distance from the center on corresponding axes *is* an isometry.

THEOREM 18. *Any two rectangular bases of the same space contain the same finite number of axes.*

Proof. Unless two abstract classes have the same finite cardinal number, there is known (Dedekind-Peirce) to exist a one-one correspondence between one class and a proper subclass of the other. By Theorem 17, there will thus be an isometry between a rectangular basis and an orthocross not a basis, and by Postulate III this can be extended to all space. But any isometry of space leaves invariant the definition of a rectangular basis, and so carries rectangular bases into rectangular bases, giving a contradiction.

Definition. The number of axes in each rectangular basis of a space I is called the *dimension* of the space, and denoted by $d(I)$.

III. SUBSPACES AND INDUCTION

14. **Hyperplanes.** By analogy with the terminology of modern algebra, we shall define a *subspace* of any "space" S as a subset of S which itself satisfies our postulates under the same metric. Since Postulates I–II′ hold automatically, only Postulates II–III are needed for this. A *flat* will be defined as a subset which contains, with any two non-antipodal points, the unique geodesic through them. A *convex* subset is one which contains, with any two non-antipodal points, the unique straight segment between them.

The proofs of the following lemmas are immediate.

LEMMA 5. *Any geodesic is a subspace.*

LEMMA 6. *Any subspace is a flat.*

We shall show later (Theorem 23) that, conversely, any flat is a subspace or consists of two antipodal points.

LEMMA 7. *A subset U of a subspace T of a space S is a subspace of S if and only if it is a subspace of T.*

LEMMA 8. *Any intersection of flats is a flat.*

COROLLARY. *The flats of any space are a complete lattice.*

LEMMA 9. *If space is bounded, then any flat consists of just one point, or is antipodally closed.*

For it contains, with any point p, a straight line pq through p. We shall avoid exceptions below by requiring flats to be antipodally closed.

LEMMA 10. *The set of all points left fixed by any isometry θ is a topologically closed flat.*

Proof. It is topologically closed in any metric space, provided only θ is continuous. Now suppose $a\theta = a$ and $b\theta = b$, while a and b are non-antipodal. Then θ will carry the geodesic through a and b into itself setwise, and leave a and b fixed; hence it must act like the identity on ab. The proof that it is antipodally closed is trivial.

It will be shown later (Theorem 24) that, conversely, any flat is the set of fixpoints of a suitable isometric involution (namely, reflection in that flat).

COROLLARY. *The set of points transformed alike by any two isometries θ and θ' is a topologically closed flat.*

For it is the transform under θ^{-1} of the set of points invariant under $\theta^{-1}\theta'$.

Definition. The set of points equidistant from two given points p and p' is called a *hyperplane*, and denoted by $H(p, p')$.

THEOREM 19. *Any hyperplane $H(p, p')$ separates space in two, and is a maximal flat.*

Proof. In *any* metric space, $H = H(p, p')$ separates the set of x such that $|x-p| < |x-p'|$ (the "side of H nearer p") from the set of y such that $|y-p'| < |y-p|$. Next([28]), suppose $a, b \in H$ non-antipodal. Then $a, b, p \to a, b, p'$ is an isometry, and so can be extended to space. Its extension will leave ab pointwise invariant, whence $|x-p| = |x-p'|$ for all $x \in ab$, and $ab \subset H$.

Finally, suppose F is a flat containing H and a point $x \notin H$ as well. Let y be any point on the opposite side of H. Draw a straight segment xy; it must cut H at some point h, with $|x-h| < |x-y| \leqq d$. Hence F, being a flat, contains $xh = xhy$ and so y. Repeating the argument with respect to y, F contains also every point on the same side of H as x; hence it contains all space, and H is a *maximal* proper flat.

([28]) This proof, due to Busemann [1, p. 111], uses only the 3-point form of Postulate III. Another proof can be based on Lemma 6 and the fact that H is the set of fixpoints under reflection in H (Theorem 20). It should be mentioned that if bounded sets are compact and space is unbounded, Theorem 19 alone is sufficient for a Euclidean or hyperbolic metric (Busemann).

THEOREM 20. *There is one and only one proper isometry of space which leaves a given hyperplane pointwise invariant.*

Proof. The correspondence $p, p', H \to p', p, H$ (pointwise on $H = H(p, p')$) is an isometry, and so can be extended to a proper isometry of all space. Incidentally, θ will interchange the two sides of H.

Conversely, let θ be any isometry not the identity, which leaves H pointwise invariant. For any q not on H, $q\theta = q$ would imply that θ was the identity by the maximality of H and Lemma 6; hence $q\theta \neq q$. Moreover for all x in H, $|x-q| = |x\theta - q\theta| = |x - q\theta|$; hence $H \leqq H(q, q\theta)$. But H is maximal and $H(q, q\theta) \subset S - q < S$; hence $H < H(q, q\theta)$ is impossible, and $H = H(q, q\theta)$ *for any q not in H.*

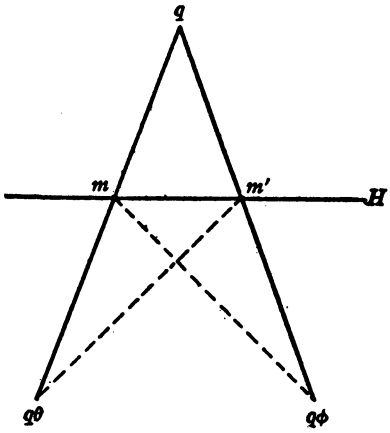

FIG. 6

Now if ϕ is any second such isometry, we shall have $H = H(q, q\theta) = H(q, q\phi)$. Draw straight segments $qq\theta$ and $qq\phi$, with midpoints m and m' (cf. Fig. 6). Then $m, m' \in H$, and so

$$|q - q\phi| \leqq |q - m| + |m - q\phi| = 2|q - m| = |q - q\theta|.$$

Similarly, $|q - q\theta| \leqq |q - q\phi|$, and so $|q - q\theta| = |q - q\phi|$, and

$$|q - m| = |m - q\theta| = 1/2|q - q\theta| = 1/2|q - q\phi| = |q - m'| = |m' - q\theta|.$$

Hence $|q - m| + |m - q\phi| = |q - q\phi|$, and $q\phi$, like $q\theta$, is on qm extended beyond m to a distance $|q\theta - m|$. We conclude that $q\phi = q\theta$ for any $q \notin H$; but $q\phi = q\theta = q$ for any $q \in H$; hence $\phi = \theta$, and the latter is unique, q.e.d.

Definition. If H is any hyperplane, the unique proper isometry θ which leaves H pointwise invariant will be called *reflection* in H.

It is a corollary of Theorem 20 that reflection in H is *involutory* (its own inverse).

15. **Hyperplanes are subspaces.** We shall now prove that every hyperplane is a subspace—that is (cf. §14, paragraph one), satisfies Postulates II–III.

LEMMA 11. *Any flat either satisfies Postulate* II *or consists of two antipodal points.*

Proof. The existence of pq for $|p-q|<d$ is trivial. The existence for $|p-q|=d$, when F contains some r not p or q, follows since pr must go through $q=p'$.

LEMMA 12. *Reflection ϕ in any point p of any flat F leaves the flat setwise invariant.*

Proof. If $q\in F$ and $|p-q|=d$, then $q\phi=q\in F$ as in Theorem 6; if $q\in F$ and $|p-q|<d$, then there is just one geodesic through p and q; it will contain $q\phi$ and lie in F, so $q\phi\in F$.

THEOREM 21. *Any flat F is homogeneous.*

This means that the group of its isometries is transitive.

Proof. If $p, q\in F$ and $|p-q|<d$, then $pq\in F$, and reflection in the midpoint of pq is an isometry of F which interchanges p and q. If $|p-q|=d$, then F is antipodally closed and the antipodal involution leaves F setwise invariant and interchanges p and q.

LEMMA 13. *Given any point s on a hyperplane H, there exist points p, p' such that psp' and $H=H(p, p')$.*

Proof. By definition of hyperplane, $H=H(r, r')$ for some r, r'; the midpoint m of rr' will be in H. By Theorem 21, there exists an isometry ϕ leaving H setwise invariant and carrying m into s; set $p=r\phi$, $p'=r'\phi$. Then $H=H\phi=H(r\phi, r'\phi)=H(p, p')$ and $p=r\phi$, $s=m\phi$, $p'=r'\phi$ are collinear, as asserted.

THEOREM 22. *Any hyperplane H is hyperhomogeneous.*

Proof. Let $S\subset H$, $T\subset H$ be given, and let θ be any isometry mapping S setwise on T. Then by Lemma 13, there exist for any $s\in S$ and $s\theta=t\in T$ points p, p', q, q' such that $H=H(p, p')=H(q, q')$, $|p-s|=|q-t|$, and psp', qtq'. It follows by the corollary to Theorem 15 that for any $x\in S$, since xp is the hypotenuse of the right triangle with sides ps and sx, while $x\theta q$ is that of the right triangle with equal sides qt and $tx\theta=x(sx)\theta$, $|x-p|=|x\theta-q|$, and similarly $|x-p'|=|x\theta-q'|$. Consequently the extension $\theta^*: S\to S\theta$, $p\to q$, $p'\to q'$ of θ is still an isometry, and so can be extended to all space. But any exten-

sion of θ^* will carry $H(p, p')$ into $H(q, q')$, and so give an isometry of H which is an extension of θ, as desired[29].

Combining Theorem 22 with Lemma 11, we see that any hyperplane $H=H(p, p')$ in n-dimensional space is a subspace, or consists of two antipodal points. It remains to ascertain the dimension of H.

We recall that reflection θ in H reflects $pp'=pp\theta$ in its midpoint o. But it leaves H, and hence every goedesic in H, pointwise invariant. We infer that pp' is perpendicular to every geodesic in H through o; in this case, it is *perpendicular to H at o*. But conversely, if a line L and pp' are perpendicular at o, then $|p-x|=|p'-x|$ for all $x\in L$, and so $L\in H(p, p')$. We conclude with the following lemma.

LEMMA 14. *The perpendiculars to any straight segment pp' at its midpoint constitute the hyperplane $H(p, p')$.*

COROLLARY. *Any two hyperplanes of the same space are isometric.*

But now, we can clearly extend the 1-orthocross through o consisting of $pp'=L_1$ into a rectangular basis $L_1, \cdots, L_n$. Moreover the orthocross $L_2, \cdots, L_n$ is maximal in $H(p, p')$, since if it could be enlarged to $L_2, \cdots, L_n, M$, then by Lemma 14, $L_1, L_2, \cdots, L_n, M$ would be an orthocross, contradicting the choice of $L_1, \cdots, L_n$. The dimension of H is therefore $n-1$.

THEOREM 23. *Any hyperplane in n-dimensional space is an $(n-1)$-dimensional subspace, or else consists of two antipodal points.*

The exceptional case is clearly genuine; it is clearly the case that space consists of a single circle through the "poles" of the hyperplane.

16. **Reflections; involutions in general.** Let F be any flat, p any point not in F, q any point in F. We shall say that q is the foot of a perpendicular from p to F (in symbols, $pq\perp F$) if $pq\perp qr$ for every line qr in F through q.

Definition. By a *reflection* in a flat F is meant any transformation which leaves F pointwise invariant and reflects each $p\notin F$ in the foot of a perpendicular from p to F.

Remark. Suppose q, r are both feet of perpendiculars from p to F. In case $|q-r|=d$, then reflection of space in q amounts to reflection of space in r (§9); hence both affect p alike. In case $|q-r|<d$, then $qr\leqq F$ and two perpendiculars can be dropped from p to qr; hence (Theorems 11–12) $|p-q|=|p-r|=d/2$, and reflection of p in either q or r carries p into its antipode p'; hence both affect p alike. Therefore reflection in F is *uniquely* determined

[29] The preceding proof is analogous to that (given in §6) of the hyperhomogeneity of spheres.

by F. Furthermore, repetition of the transformation yields the identity; in this sense, the transformation is of period two.

Definition. By an *involution* is meant any transformation of period two.

Conversely, let θ be any *isometric* involution. Unless space is bounded and θ is the antipodal involution, $|p-p\theta| < d$ for some p. In this case, the midpoint o of $pp\theta$ is a fixpoint of θ. The set of all fixpoints of θ is thus a non-void flat (Lemma 10), which we shall denote by F_θ. But now, $pp\theta$ is reflected in o by θ, while if $or \leqq F_\theta$, then or is pointwise invariant under θ; hence $po \perp F_\theta$, and θ is identical with reflection in F_θ. We conclude the following theorem.

THEOREM 24. *There is only one transformation which reflects space in a given flat F; this is an involution. Conversely, any isometric involution θ amounts to reflection in the flat F_θ of its fixpoints, unless space is bounded and θ is the antipodal involution.*

Furthermore, unless F_θ is all space (and θ is the identity), there exists a point p such that $p \neq p\theta$. Clearly, for all $x \in F_\theta$, $|x-p| = |x\theta-p\theta| = |x-p\theta|$; hence $F_\theta \leqq H(p, p\theta)$. But now, since θ is of period two, the hyperplane $H=H(p, p\theta)=H(p\theta, p)$ is invariant under θ—which thus is an isometric involution on H, which still has F_θ for the set of its fixpoints. By induction on dimension, we can infer that *F_θ is a subspace.*

17. **Flats and subspaces.** Our next object is to prove that the concepts of flat and subspace are effectively equivalent—that any flat is a subspace or consists of two antipodal points.

LEMMA 15. *The only flat F which contains a rectangular basis $L_1, \cdots, L_n$ is all space.*

Proof. Choose p, p' on L_1, equidistant from the vertex o of the basis. Then the hyperplane $H=H(p, p')$ contains $L_2, \cdots, L_n$ by orthogonality; hence so does $F \cap H$, which is a flat in H. By induction on n, we infer $F \cap H = H$ and so $F \geqq H$. But H is a maximal flat (Theorem 19); hence $F \geqq H \cup L_1 > H$ must include all space.

Next, we define a *k-orthohedral* to consist of a vertex o and a sequence of k mutually perpendicular rays (that is, half-lines) issuing from o. Thus a k-orthohedral is a k-orthocross whose axes have been ordered and sensed.

LEMMA 16. *In n-dimensional space, the group of isometries of space is simply transitive on its n-orthohedrals. Hence the only isometry which leaves an n-orthohedral invariant is the identity.*

Proof. Clearly any isometry carries an n-orthohedral into another n-orthohedral; clearly also, all n-orthohedrals are isometric; therefore (by free mobility) the isometries of space are transitive on n-orthohedrals. To prove that

they are simply transitive, suppose θ, θ' transform the n-orthohedral of rays $R_1, \cdots, R_n$ alike. The set of points transformed alike by θ and θ' is a flat (Lemma 9); it includes the rectangular basis $L_1, \cdots, L_n$ formed by extending each R_i into a line L_i; hence (Lemma 15) it includes all space.

LEMMA 17. *The intersection of all flats containing any orthocross $L_1, \cdots, L_m$ is a subspace S, and isometric reflection in S is possible.*

Proof. We can extend $L_1, \cdots, L_m$ to a rectangular basis $L_1, \cdots, L_n$ of space, with the same vertex o. Let θ denote the transformation which leaves $L_1, \cdots, L_m$ pointwise invariant and reflects $L_{m+1}, \cdots, L_n$ in o; and let θ' denote the dual transformation reflecting $L_1, \cdots, L_m$ in o and leaving $L_{m+1}, \cdots, L_n$ pointwise invariant. By Lemma 16, both transformations can be uniquely extended to space, and both are isometric *involutions*. Let S, S' denote the subspaces (Theorem 24) of their fixpoints.

Then $L_1, \cdots, L_m$ is a maximal orthocross in S, since $L_{m+1}, \cdots, L_n$ are perpendicular to every oq in S and space is n-dimensional. Therefore, if F is any flat containing $L_1, \cdots, L_m$, then $F \cap S$ is a flat in S containing a rectangular basis of S—hence $F \cap S = S$ by Lemma 15, and $F \supseteq S$. But S itself is such a flat; hence S is the intersection of all flats containing $L_1, \cdots, L_m$, q.e.d.

THEOREM 25. *Any flat F is a subspace; and reflection in any flat is possible and isometric.*

Proof. Let $L_1, \cdots, L_m$ be an orthocross in F with m maximal. Then F contains the subspace S generated by $L_1, \cdots, L_m$ as in Lemma 17. If F contained another point p, then F would contain the line $pp\theta$ obtained by reflecting p in S (§16); the midpoint q of $pp\theta$ could be used as the vertex of an m-orthocross $M_1, \cdots, M_m$, $pp\theta$ would be an $(m+1)$-orthocross in F, with vertex q, contrary to hypothesis. Hence $F = S$, from which the conclusion follows immediately.

18. **Lattice of subspaces.** The intersection of any family of flats is itself a flat; hence the flats of any space I which satisfies our postulates form a *complete lattice*([30]). Combining this with Theorem 25, we get the following lemma.

LEMMA 18. *The subspaces of any space I form a complete lattice $L(I)$.*

It is this lattice which now interests us. First, we note that its minimal elements are single points—or, in the bounded case, pairs of antipodal points. Since flats are antipodally closed, it is clear that in either case *any subspace is the join of the minimal elements contained in it.*

([30]) For the terminology of lattice theory, cf. G. Birkhoff [1], esp. pp. 17–18 and 63–65. Our results in §§18–20 below will not be needed for later developments.

LEMMA 19. *If p is a minimal element, and S is an m-dimensional subspace not containing p, then the join $S \cup p$ is $(m+1)$-dimensional, and minimal among subspaces properly containing S.*

Proof. Clearly $S \cup p$ contains a line $pp\theta$, where $p\theta$ is the reflection of p in S; moreover the midpoint q of $pp\theta$ must be in S. We can form a rectangular basis $L_1, \cdots, L_m$ for S with vertex q, and this will be an m-orthocross in I. Also, $L_1, \cdots, L_m, qp$ is an $(m+1)$-orthocross in $S \cup p$; it can be extended to a rectangular basis $L_1, \cdots, L_m, qp, L_{m+2}, \cdots, L_n$ for I. As in Lemma 17, $L_1, \cdots, L_m, qp$ is a maximal orthocross (rectangular basis) in the intersection of all flats containing $L_1, \cdots, L_m, qp$; and this is clearly $S \cup p$; hence $S \cup p$ is an $(m+1)$-dimensional subspace.

Further, if T were any subspace such that $S \cup p > T > S$, then points t, r would exist in T but not in S, resp. in $S \cup p$ but not in T. By the preceding paragraph, $S \cup t$ would be at least $(m+1)$-dimensional, and $S \cup p \geqq (S \cup t) \cup r$ at least $(m+2)$-dimensional, which gives a contradiction.

This completes the proof of MacLane's "exchange axiom." As a corollary of this and Theorem 4.10 of the author's *Lattice theory* we have the following theorem.

THEOREM 26. *The subspaces of any n-dimensional space form an n-dimensional* ***matroid lattice*** *(or "exchange lattice")*[31].

19. **Duality and polarity.** By the *dual* of a lattice is meant the abstract system obtained if the inclusion relation is replaced by its converse. As is well known to experts in lattice theory, the notion of a matroid lattice is not self-dual; in fact, provided there are at least three points on every line, any matroid lattice whose dual is a matroid lattice is an *abstract projective geometry* (*Lattice Theory*, §§70, 78).

Case I. *Space is bounded.* We define $p \perp q$ to mean $|p-q| = d/2$. If R is any set, we shall let R^* (in words, the *polar* of R) denote the set of all points q such that $q \perp r$ for all $r \in R$. If $R = (R^*)^*$, we shall say that R is *orthoclosed*.

It is clear that antipodal points have the same polars. Furthermore, if we identify antipodal points (as in forming elliptic space), the relation $q \perp r$ becomes anti-reflexive and symmetric. From this fact alone there follows, by the general theory of polarity with respect to a relation (cf. *Lattice theory*, §32):

(1) The correspondence $R \to (R^*)^* = \overline{R}$ is a closure operation in the sense that it is isotone, idempotent, and enlarging.

(2) Any R^* is orthoclosed: $((R^*)^*)^* = R^*$ for all R.

(3) The correspondence $R \to R^*$ is antitone, and one-one on orthoclosed sets.

---

[31] This result goes back to ideas of K. Menger [2], amplified in his Rice Institute Lectures *On algebra of geometry* (1940); technically, the present discussion is entirely new.

(4) The orthoclosed sets are a (complete) complemented lattice, in which R and R^* are complementary.

In the present case, since the polar of a point p is the hyperplane $H(p, p')$ of points equidistant from p and its antipode p', since the polar of a set of points p_i is the intersection $\wedge_i H(p_i, p_i')$ of the polars of its members, and since any intersection of hyperplanes is a subspace, we have

(5) Any orthoclosed set is a subspace.

Again, let S be any subspace, and let θ be reflection in S. If $P \perp S$, then $p\theta$ is obtained from p by reflecting in a point $m \in S$ (the midpoint of $pp\theta$) $d/2$ from p; hence p and $p\theta$ are antipodal. Conversely, if p and $p\theta$ are antipodal, then for all $x \in S$, $|p-x| = |p\theta - x|$; hence $|p-x| = d/2$ for all $x \in S$. It follows that S^* is the subspace of fixpoints of $\theta\phi = \phi\theta$, where ϕ is the antipodal involution. But $(\theta\phi)^2 = \theta\phi\theta\phi = \theta\theta\phi\phi$ is the identity; hence (Theorem 24) $\theta\phi$ is reflection in S^*. Repeating the argument, $\theta = (\theta\phi)\phi$ is reflection in $(S^*)^*$; hence $(S^*)^* = S$, and

(5′) Any subspace is an orthoclosed set.

Combining (3) and (5)–(5′), we see that $L(I)$ is self-dual. Combining this with paragraph one above, we get the theorem:

THEOREM 27. *If space is bounded, then its subspaces coincide with its orthoclosed sets, and form an abstract projective geometry.*

Case II. *Space is unbounded.* Here we have no absolute polarity like that of Theorem 27. But relative to any designated center o, we can develop a similar theory. We define $p \perp q \pmod{o}$ to mean that poq is a right angle; we then define R^* from this relation as in the bounded case. If $R = (R^*)^*$, we say that R is orthoclosed at o. Since perpendicularity is symmetric and anti-reflexive for points not o, we can again read off (1)–(4). Further, it is evident that

(4′) $o^* = I$ and $I^* = o$. However, conditions (5)–(5′) must be modified.

By Lemma 14, the polar p^* of any point p is the hyperplane $H(p, p')$, where p' denotes the reflection of p in o. By general lattice theory, the polar of any set X is the intersection of the polars of the points in X. But (Theorem 25) any intersection of hyperplanes is a subspace. Consequently

(5_0) Any orthoclosed set is a subspace through o. Conversely, let S be a subspace through o. By definition, $p \perp S \pmod{o}$ if and only if o is the foot of the perpendicular from p to S—hence if and only if reflection θ in S and reflection ϕ in o transform p alike. In symbols, $p \in S^*$ means $p\theta = p\phi$, or $p(\theta\phi) = p$. Hence S^* is the set of fixpoints of $\theta\phi$. Since $o\theta = o$, $\theta\phi = \phi\theta$, and so $\theta\phi$ is not only an isometry but also of period two—that is, it is an isometric involution whose fixpoints constitute S^*. Repeating the argument, $\theta = (\theta\phi)\phi$ is reflection in $(S^*)^*$. Hence (we rely implicitly on Theorem 24) $(S^*)^* = S$, and so

($5_0'$) Any subspace through o is orthoclosed mod o. From these results we conclude, as in Theorem 27, the following result.

THEOREM 28. *If space is unbounded, then its subspaces are the subsets orthoclosed at all their points, and the subspaces which contain any selected point form an abstract projective geometry.*

20. **Dependence and rank.** From Theorem 26 and the postulate theory of linear dependence of Whitney([32]), it follows directly that we can define notions of "dependence," "rank," and "basis" in any n-dimensional space I, with all the usual properties.

More precisely, let the *rank* of a subset X of I be defined as $d[\overline{X}]-1$, where $\overline{X}$ denotes the subspace spanned by X and $d[\overline{X}]$ denotes the dimension of $\overline{X}$. According to Whitney's theory, to say that p is *linearly dependent* on X means that $p\in\overline{X}$. A subset B of I is a *basis* if B spans I, whereas no proper subset of B spans I. The interdeducibility of these notions and their fundamental properties has been shown by Whitney (loc. cit.), and has no special interest in the present context. What seems interesting is the relation between these linear notions and certain metric notions.

Definition. A point p of a metric space I is *metrically dependent* on a subset X of I if and only if every isometry of I which leaves X pointwise invariant leaves p fixed. More generally, if G is any group of transformations of a domain I, a point p in I will be called *dependent* on a subset X of I (mod G) if and only if every transformation of I which leaves X pointwise invariant leaves p fixed.

It is easy to show that this implies that if we know every $x\theta$ $[x\in X]$ and that $\theta\in G$, then we know $p\theta$. In the special case that X is a point, the points dependent on X (mod G) form Lie's "systatic variety for X"([33]); thus we generalize the notion of a systatic group. In case I is a separable normal field, and G is the group of automorphisms of X, the elements dependent on X form the subfield generated by X. For metric spaces which satisfy the free mobility postulate, it may also be shown that Menger's notion of a "metric basis" (cf. Blumenthal [1, p. 60]) is equivalent to the notion of metric dependence just defined([34]).

THEOREM 29. *In a metric space satisfying our postulates, linear dependence and metric dependence are equivalent concepts.*

([32]) H. Whitney, *On the abstract properties of linear dependence*, Amer. J. Math. vol. 57 (1935) pp. 509–533. Direct connection can be made between Lemma 19 and Whitney's postulates on rank.

([33]) Cf. L. P. Eisenhart, *Continuous groups of transformations*, Princeton, 1933, p. 84.

([34]) The latter is presumably also closely related to the notion of "independence" defined in Blumenthal [1, p. 83].

There also seems to be a connection, going back to first principles, between a subset's having free mobility and being metrically closed. The author believes that these ideas merit further study for their own sake.

Proof. By Theorem 24, any subspace is metrically closed, in the sense of containing all the points metrically dependent on it. Conversely, by Lemma 10, the metric closure of any subset is a flat—and so a subspace.

21. **Relation to a note of Kolmogoroff.** An ingenious but sketchy note of Kolmogoroff [1] purports to give sufficient conditions that a topological space R with a group G be "equivalent" (in the sense of Lie) to n-dimensional hyperbolic, Euclidean, spherical, or elliptic space. His conditions are topological, but directly implied by the following metric postulates: (KI) R is metric, (KII) R is connected, (KII′) R is locally compact, (KIII) G consists of isometries of R (and so is uniformly continuous), (KIV) G is transitive on R, (KV) if $S(x_1, \cdots, x_n/y)$ denotes the set of all transforms of $y \in S(x_1, \cdots, x_{n-1}/x_n)$ under isometries leaving $x_1, \cdots, x_n$ fixed, then given distinct sets $S(x_1, \cdots, x_n/y)$ and $S(x_1, \cdots, x_n/z)$, one separates the other (topologically) from x_n on $S(x_1, \cdots, x_{n-1}/x_n)$.

Kolmogoroff asserts without further details that finite-dimensionality can be proved from KI–KV, and that then his conclusion can be proved by induction on $r-n$. This may be true, but in the author's opinion it was not shown by Kolmogoroff.

Specifically, it is not clear why "spheres" of different radii about a fixed center must have the same dimension. More generally, how can one reconstruct r-space from $(r-1)$-spheres without something like straight lines to correlate concentric spheres? Another vital omission is a proof that spheres of all dimensions are *connected*—a fact needed for induction. Finally, it is not even clear why local compactness implies finite-dimensionality, although this is less important.

However, it may be worth pointing out that KI–KV are almost immediate consequences of I, II, II′, III of the present paper, except as regards local compactness. Most follow trivially: I implies KI, II implies KII, KIII is a definition, and III implies KIV. Less trivially, III implies KV.

Proof of KV. We first show that in any metric space satisfying the $(n+1)$-point form of Postulate III, $S(x_1, \cdots, x_n/y)$ is a sphere with center x_n on $S(x_1, \cdots, x_{n-1}/x_n)$, in the sense that it consists of all points p of $S(x_1, \cdots, x_{n-1}/x_n)$ which satisfy $|p-x_n| = |y-x_n|$. For *all* points p on $S(x_1, \cdots, x_{n-1}/x_n)$ satisfy $|p-x_i| = |x_{i+1}-x_i| = \rho_i$ $[i=1, \cdots, n-1]$. Hence if also $|p-x_n| = |y-x_n|$, then we can extend the isometry $x_1, \cdots, x_n, y \to x_1, \cdots, x_n, p$ to a self-isometry of R. But now, in any metric space, of two concentric spheres, one separates the other from the center. Hence KV holds if R is any hyperhomogeneous metric space and G is the group of all isometries of R, regardless of Postulates II, II′.

As regards local compactness, this can be proved fairly directly from the results of the preceding paper. But it seems more elegant to prove the isometry with Euclidean, hyperbolic, or spherical n-space first, and this the author proposes to do trigonometrically in a later paper.

BIBLIOGRAPHY

G. BIRKHOFF

1. *Lattice theory*, Amer. Math. Soc. Colloquium Publications, vol. 25, New York, 1940.
2. *Metric foundations of geometry*, Proc. Nat. Acad. Sci. U. S. A. vol. 27 (1941) pp. 402–406.

L. BLUMENTHAL

1. *Distance geometries*, University of Missouri Press, 1938.

H. BUSEMANN

1. *On Leibniz's definition of planes*, Amer. J. Math. vol. 63 (1941) pp. 101–111.
2. *Metric methods in Finsler spaces and in the foundations of geometry*, Annals of Mathematics Studies, no. 8, Princeton, 1942, with bibliography.

J. L. COOLIDGE

1. *The elements of non-Euclidean geometry*, Oxford, 1909.

T. L. HEATH

1. *The thirteen books of Euclid's elements*, Cambridge University Press, 1908, 1926.

D. HILBERT

1. *Grundlagen der Geometrie*, 7th ed., Teubner, 1930.

A. KOLMOGOROFF

1. *Zur topologisch-gruppentheoretischen Begrundung der Geometrie*, Nachr. Ges. Wiss. Göttingen (1930) pp. 208–210.

K. MENGER

1. *Untersuchungen über allgemeine Metrik*, Math. Ann. vol. 100 (1928) pp. 73–163.
2. *New foundations of projective and affine geometry*, Ann. of Math. (2) vol. 37 (1936) pp. 456–482.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.