Synthetic Geometry

Synthetic geometry is that kind of geometry which deals purely with geometric objects directly endowed with geometrical properties by abstract axioms. This is in contrast with a procedure which constructs "geometric objects" from other things; as, for example, analytic geometry which, with the artifice of a coordinate system, models points by n-tuples of numbers. Synthetic geometry is the kind of geometry for which Euclid is famous and that we all learned in high school.

Modern synthetic geometry, however, has a more logically complete and consistent foundation. In this chapter the pattern of this foundation will be adapted, informed by the previous physical considerations, to develop a synthetic system of axioms which do not entail such things as uniformity or isotropy. This geometry is a global one but it is hoped that its elaboration, like that of Euclidean geometry, will be instructive for the development of a similar, but more general, local theory.

2.1 Incidence Geometry

The most basic of geometrical notions are introduced by the concept of an *incidence geometry*. Assuming basic ideas from set theory, it is abstractly defined in terms of a set $\mathcal{P}$ whose elements will be called *points* and a collection of subsets $\mathcal{L}$ of $\mathcal{P}$ called *lines* satisfying three axioms.

Axiom 1: Every line contains at least two points.

Axiom 2: There are three points which are not all contained in one line.

Axiom 3: There is a unique line containing any two points.

The points referred to in these axioms are, of course, all distinct. Generally, points will be denoted by capital letters from beginning or middle of the alphabet and lines will be denoted by small letters from the middle of the alphabet.

The first axiom just rules out vacuity. The second axiom guarantees "at least two dimensions"; otherwise it would hardly be worth calling a geometry. The third incidence axiom has more substantial content. It both* requires that there be a line connecting any pair of points and prohibits lines from converging to intersect twice.

These simple ideas are sufficient to establish some of the usual terminology and little else. A point contained in a line is said to *lie on the line*. Points are *collinear* if they are contained in the same line. Two lines with a point in common are said to *intersect*. Two lines which do not intersect are called *parallels*. The theorems that are possible are little more than restatements of the axioms such as the following.

Theorem 2.1.1. If two lines have two points in common then they are identical.

Incidence geometry is just the beginning of, not the object of, this study. There is no point in dwelling on it or belaboring the obvious.

2.2 Metric Geometry

There are two different approaches to synthetic geometry that can be followed from this point. One is the pure synthetic geometry† of Euclid and Hilbert which continues with additional abstract axioms. And the other is metric geometry‡, due to George David Birkhoff, which introduces the concept of a numerical measure of distance etc. in clear departure from the spirit of the Greek approach to geometry. The course taken herein is closer to this last. This would have been regarded with great suspicion by the ancient Greeks and may not seem genuinely synthetic to some.

---

* In a local version of this theory both contraries should be accommodated globally. Therefore this axiom will have to be modified in a more general theory.

† See the book on non-Euclidean geometry by Greenberg.

‡ See the book on metric geometry by Millman and Parker. People whose background is in topology and analysis rather than geometry should be made aware that this established geometric term is not a synonym for metric space. Nor is the word "metric" being used in the sense with which they are accustomed; rather, following Birkhoff, it just refers to the existence of numeric measures.

However, the significant differentiation between analytic and synthetic geometry does not turn on the use of analysis per se. Modern treatments of pure synthetic geometry essentially contain the axioms of the real number system* which entail analysis anyway. But they are not based on coordinate systems (that is, modeling the geometry by n-tuples) like manifold theory, do not hypothesize a Riemannian metric, etc. It is precisely the assumptions about space that the use of such methods entail which this paper aims to avoid because, by generalizing from an algebraic model, they may be inadvertently specialized in a physically inappropriate way. By this criterion, referring to the approach of this paper as synthetic is justified.

For there to be a genuine distinction between the pure and metric approaches, some of the properties which allow the construction of the number system would have to be dropped from the pure axiom system. For instance, a large part of geometry can be obtained without continuity. Perhaps there is another, purer, way to obtain essentially the same results. From the point of view of physics this possibility is especially interesting since the concept of distance poses considerable epistemological problems.

In this regard, the question arises: is the distance measure unique? Or: can it be transformed in some way so as to preserve the interweaving structure† of the lines? If not, then the measure must, in some sense, be derived from the structure and is not fundamental. Being able to construct a distance from physical structure would be a considerable conceptual advance for physics. However, such possibilities and thoughts will not be pursued further here. They might be worth more consideration if the results obtained in this work turn out to be of physical use.

---

* In Greenberg's Axiom system: the first 3 betweenness axioms, aspects of the first 3 congruence axioms and the continuity axioms.

† For example, one possibility is to apply a scale factor, smoothly varying with direction, along each line.

Directly introducing a distance measure is simpler and clearer than pure synthetic geometry as long as no questionable physical significance is attributed to it in the process. The result is considerable consolidation. Furthermore, this facilitates disentangling those essential aspects (which concern things like continuity) of the usual axioms of pure synthetic geometry from "physical" aspects (which concern things like isotropy) which are too special. Therefore, such "impure" means will be used without further apology.

Definition[*]. For a line, $m \in \mathfrak{L}$, a function $r_m : m \times m \rightarrow \mathrm{R}$ which has the following properties will be called a *ruler* for *m*.

(i) It is a bijection on either variable if the other is held constant.

(ii) It is antisymmetic: $r_m(A,B) = -r_m(B,A)\ \forall A,B \in m$.

(iii) It is additive: $r_m(A,B) + r_m(B,C) = r_m(A,C)\ \forall A,B,C \in m$.

A line which has a ruler will be called a *ruled line*.

The points on a ruled line inherit properties by being in a one-to-one correspondence with real numbers. Roughly speaking, a ruling defines continuity and imposes an order on a line. That is, it orders the points in the line so that there are points around and between other points without any holes. It provides such properties at once and in a unified manner in place of a collection of pure axioms which do so piecemeal.

Because of antisymmetry: $r_m(A,A) = 0\ \forall\ A \in m$ for any ruler. Therefore $r_m(A,B) = 0$ iff $A = B$ since a ruler is a bijection. Furthermore, given any fixed origin point, *O*, on a ruled line, *m*, then for all other points, *A*, on that line the quantities $r_m(O,A)$ are a bijection between those points and the reals from which the

---

[*] The definition of a ruler used here differs from the usual concept (in metric geometry) which is essentially a synonym for a coordinate system. It is more physically intuitive and is more adaptable to a localized theory. Line coordinatization will be deduced. This is actually somewhere between the pure synthetic approach and that of Birkhoff.

ruler can be obtained,

$$r_m(A,B) = r_m(O,B) - r_m(O,A),$$

because of addition and antisymmetry. This is called a *coordinate system* for the line and it exhaustively characterizes the nature of lines.

Note that any one-to-one function composed with such a coordinate system is also a coordinate system corresponding to a ruler produced from it as above. Since such a transformation may rearrange all the points, just the existence of rulers and coordinate systems are rather weak properties. Even composition by a continuous, monotonic function, tho it will yield the same topological properties, may produce a quite different coordinate system and rulers.

However, by fixing a particular ruler to be associated with each line it is possible to proceed to attribute more geometric significance to these special rulers.

> **Axiom 4**. There is a particular collection of rulers which contains exactly one for every line. The rulers in this collection will be called *geometric rulers*.

These geometric rulers order the lines internally, measure the distances between their points and establish congruences in a particular and fixed way. They exhaustively characterize the internal nature of lines but they do not, by themselves, entail <u>any</u> relationships between different lines. Nevertheless, by means of these rulers, the pervasive intersecting lines are, so to speak, an ordered and continuous framework which knits space together in some way. By this means, characteristically in synthetic geometry, the lines and their geometric rulers serve the function which the artifice of coordinates do in analytic geometry or charts do for a manifold.

Note that multiplying any one of these rulers by any real number except zero yields another ruler which endows the line with exactly the same ordering and congruences. It is possible, therefore, to change the "scale factor" of a geometric ruler without altering the internal structure of its line. Since this scale factor cannot be continuously passed thru zero, however, there are two disconnected collections of these equivalent geometric rulers with "opposite signs". These distinguish the two different directions along the line and will be said to have opposite *orientations*. The absolute scale of the rulers will become geometrically significant only when making comparisons between lines. The orientation will be crucial in relating "nearby" lines to each other is a consistent way.

Since between any two points, A and B, there is a line, say m, (and therefore a geometric ruling which can be used) a *distance function*[*] can be defined by

$$d(A,B) = |r_m(A,B)| = |r_m(O,B) - r_m(O,A)| \qquad 2.2.1$$

between any two points. An incidence geometry in which all the lines have a coordinate system related to a distance in this way is called[†] a *metric geometry*. In distinction to this usual distance function, a signed distance measure, based on orientation, will be introduced later. Its use will be found preferable in this paper.

2.3 Elementary Concepts[‡].

In terms of the geometric rulers some needed concepts can now be introduced. For any two *endpoints*, $A \neq C$, on a line m define[§] a *segment* as follows.

$$AC = \{B \in m | 0 \le r_m(A,B)/r_m(A,C) \le 1\} \qquad 2.3.1$$

---

[*] The standard geometrical definition is given on p.28 of Millman and Parker. It is distinguished from the topological concept of a metric (Roman, v.1, p.189) by not requiring the triangle inequality.
[†] Millman and Parker, p.30.
[‡] These concepts are all familiar and entailed by metric geometry per se. For example, see Chapter 3 of Millman and Parker. The presentation is merely adapted to use the present axiom system and illustrates its application. But the important thing to understand is that, altho the setting may be unfamiliar, these ideas are exactly like those of high school geometry.
[§] Note that the notation, AB, used for segments (and also that, $\overline{AB}$, which will be introduced later for segment measure) is not that used in many books. This convention follows that used in Greenberg; which is more consistent than the more common notation. For example, if $\overline{AB}$ were used to denote a segment then logically a point should be denoted by $\overline{A}$ and $\overline{ABC}$ would be a triangle!

Obviously this is independent of scale and so, in particular, gives the same result for either possible orientation of the geometric ruler. Now, from the addition and symmetry properties of rulers

$$\frac{r_m(A,B)}{r_m(A,C)} = \frac{r_m(A,C) - r_m(B,C)}{r_m(A,C)} = 1 - \frac{r_m(C,B)}{r_m(C,A)}$$

and so

$$0 \le \frac{r_m(A,B)}{r_m(A,C)} \le 1 \Leftrightarrow 0 \le \frac{r_m(C,B)}{r_m(C,A)} \le 1 .$$

Therefore $AC = CA$; that is, the segments are the same, as a set, irrespective of order even tho the values of a ruler change sign with order. The distance function provides a notion of congruence between segments in the usual way.

If $B \in AC$ and $B \neq A,C$ then B is said to be *between A* and *C*. Clearly, from the above, *B* is also between *C* and *A*. If *A*, *B* and *C* are any three distinct points on a line *m* then there are only three possibilities.

(i) $$\frac{r_m(A,B)}{r_m(A,C)} < 0 \Rightarrow 0 < \frac{r_m(C,A)}{r_m(C,B)} = \frac{1}{1 - r_m(A,B)/r_m(A,C)} < 1$$

(ii) $$0 < \frac{r_m(A,B)}{r_m(A,C)} < 1$$

(iii) $$1 < \frac{r_m(A,B)}{r_m(A,C)} \Rightarrow 0 < \frac{r_m(B,C)}{r_m(B,A)} = 1 - \frac{r_m(A,C)}{r_m(A,B)} < 1 .$$

Thus one, and only one, of any three distinct points on a line must be between the other two.

If *B* is between *A* and *C* then

$$AC = \left\{ D \in m | 0 \le \frac{r_m(A,D)}{r_m(A,C)} \le \frac{r_m(A,B)}{r_m(A,C)} \right\} \cup \left\{ D \in m | \frac{r_m(A,B)}{r_m(A,C)} \le \frac{r_m(A,D)}{r_m(A,C)} \le 1 \right\}.$$

But since $0 < r_m(A,B)/r_m(A,C)$, dividing by this factor gives:

$$\left\{ D \in m | 0 \le \frac{r_m(A,D)}{r_m(A,C)} \le \frac{r_m(A,B)}{r_m(A,C)} \right\} = \left\{ D \in m | 0 \le \frac{r_m(A,D)}{r_m(A,B)} \le 1 \right\} = AB .$$

Considerably more, but similar, algebra gives:

$$\left\{ D \in m | \frac{r_m(A,B)}{r_m(A,C)} \le \frac{r_m(A,D)}{r_m(A,C)} \le 1 \right\} = \left\{ D \in m | 1 - \frac{r_m(C,B)}{r_m(C,A)} \le \frac{r_m(A,D)}{r_m(A,C)} \le 1 \right\}$$
$$= \left\{ D \in m | 0 \le 1 - \frac{r_m(A,D)}{r_m(A,C)} \le \frac{r_m(C,B)}{r_m(C,A)} \right\} = \left\{ D \in m | 0 \le \frac{r_m(C,D)}{r_m(C,B)} \le 1 \right\} = BC .$$

Therefore if B is between A and C then $AC = AB \cup BC$.

Other common results can be obtained also. For example, by combining several cases of the above, then

$$AC \subseteq AB \cup BC$$

for A, B and C any three distinct points on a line.

Any point, A, on a line, m, divides it into two parts. These two parts are the sets

$$\{ B \in m \,|\, r_m(A,B) \ge 0 \} \quad \text{and} \quad \{ B \in m \,|\, -r_m(A,B) \ge 0 \},$$

called *rays*, which have only the point A, called their *vertex*, in common. These will be called *opposite* rays. They result from either the line's geometric ruler or its negative (a geometric ruler of opposite orientation) and so correspond to the two orientations of their line. Note that if two rays of a line have the same orientation, that is if they are both derived from either the line's ruler or its negative, then one is contained in the other. Their mutual intersection is one of the original rays but their intersection with any ray of the opposite orientation is not. This is a way to tell whether or not two rays of a line have the same orientation without reference to the rulers.

This makes it possible to make a general definition that any two rays whose intersection is one of themselves have the *same direction* and to further define the equivalence class of such rays to be a *direction* in space. Directions will be a principal element of the geometric theory being developed. A direction entails a line and an orientation of that line. Every line has two directions corresponding to its two orientations which will be called *opposite* or *inverse* directions. It is most convenient to deal with directions and it will usually not be necessary to explicitly specify their lines. A direction and a point on its line determine a ray; this is the most convenient way for rays to be identified. When the vertex point is clear from the context a ray may be referred to by the direction alone. Generally, both directions and rays will be denoted by capital letters from the end of the alphabet. In figures they will be indicated by arrows pointing out the extension of the (associated) rays. Also, if T is a direction or ray then the notation T^{-} will be used to denote the direction or ray opposite to it. These notations are illustrated in the following figure.

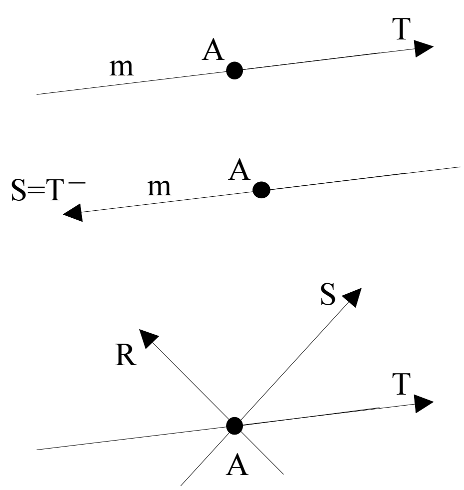


Figure 1: Rays and Directions

2.4 Oriented Measure

It is necessary to think carefully about and settle on the appropriate way to measure the "distance" between points. Such a measure has two different functions. One is: given a measure to locate a point on a line. For this purpose a signed measure, like a ruler, always smoothly distinguishes between nearby points but a distance function does not. The other function is: given points to determine the "distance" between them; as, for example, for establishing congruences. For this a distance function or the absolute value is usually used. The first is like an independent variable or the argument of a function. The second is like a dependent variable or the value of a function. It is preferable to always measure "distance" in the same way. Otherwise the theory will be confused by apparent special cases and discontinuities that arise only from the inconsistency. Since it is natural to use a signed (continuous) measure for function arguments these considerations urge that it be used in all cases. Only in this way can functions involving "distance" be inverted, expressed in implicit form, or composed with each other.

The problem with doing this consistently in geometry arises from the fact that there is more than one path connecting two points. One path may traverse the line, *m*, connecting, say, the points *A* to *C*, to get between them. The other may "go around" (via *A*-*B*-*C*) as in the figure below. Consider the distance between the intermediate

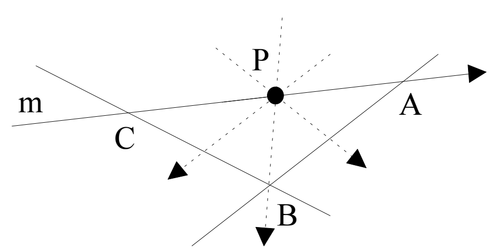


Figure 2: Alternate Paths

points on these trajectories and a fixed point, *P*, located between the endpoints. The direct line distances, which go thru zero along that path, must have opposite signs at the endpoints. The roundabout distances can never vanish on their path since it does not go thru *P*. So, if they vary continuously the endpoint distances must have the same sign. This seems to be an inconsistency: the distance measure for distinct points ought not to have two different signs! The conventional practice of just taking the absolute value amounts to ignoring this problem. Resolving it leads to a useful concept.

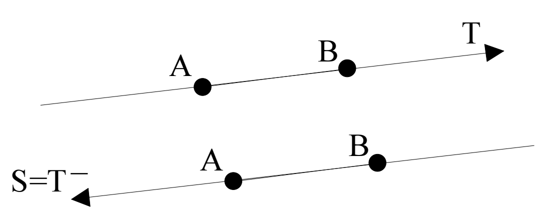
This resolution results from realizing that two different things are actually being measured. The "distances" on the direct path are all measured on a single line as it is traversed by a point. The "distances" to points on the path *A-B-C* are measured on a variety of different lines which rotate away from the line *m* and back to it. The change that transpires is not so much in the "location" of the point, whose "distance" to P on its line may well be nearly constant, as in the direction of its ray. The initial and final measures are actually being taken according to measures of opposite orientation. These observations are the basis of how segments will be measured in this paper.

Definition. Given two points, *A* and *B*, on a line, *m*, and an orientation or direction, *T*, for that line then

$$\overline{AB} = \begin{cases} r_m(A,B) & \text{if } T \text{ has the same orientation as the ruler} \\ -r_m(A,B) & \text{otherwise} \end{cases}$$

is the *signed distance* or *oriented segment measure*. This is not a function of the points alone but also of a specified orientation. This is all and the only thing that the notation $\overline{AB}$ will ever denote.

Since two points determine a line, explicit reference to the line of a measure is redundant and will hereafter be omitted. The line is implied by the orientation or direction anyway. And as long as context makes clear which of the two possible orientations is intended the orientation might also be omitted. In any case, for simplicity and compactness the intended orientation will not be indicated on the measure. It will be indicated separately; usually in diagrams by an arrow as in the



$\overline{AB} > 0$

$\overline{AB} < 0$

Figure 3: Segment Measure and Orientation

figure above. In the first case, *T* denotes the orientation being applied. The second case demonstrates the application of the opposite orientation.

This way of measuring segments is more complicated than the usual distance function and it may seem inconvenient at first. But it becomes clear as the theory is developed that it is natural. It turns out to be the key to handling various essential ideas, such as defining trigonometric functions, consistently. (Note that it includes the usual distance as a special case: just select the orientation so as to always make the measure positive. It is sometimes essential, however, not to have to do this.) The important thing to remember is that, in measuring a segment, a direction needs to be specified also. In this way, direction becomes a fundamental element of the theory.

Certain direction specifications may become ambiguous in special cases and produce confusing complications. (For example, in the case of the distance function mentioned above the direction specification is ambiguous at zero.) Use of the signed segment measure provides an understanding of the source of such problems and terms with which to deal with them.

2.5 Continuity

It is desirable that, roughly speaking, the segment measures on "neighboring" lines fit together in a smooth and continuous way. This will be ensured by the following.

> **Axiom 5**. In any geometric construction, those segment measures which are mutually dependent are mutually thrice differentiable (except, possibly, in situations where a direction becomes ambiguous). When there is more than one independent variable in the construction this refers to partial derivatives and includes the mixed partials as well.

Consider an example to illustrate the application of this axiom and the use of segment measure. In the figure shown below, there is a fixed point P on a line m. For an orientation T of m, a variable point Q on the line will be located with respect to P by the signed segment measure $\overline{PQ}$. Then, a fixed point A not on m will determine a (non-zero) distance $\overline{QA}$ which, as a function of $\overline{PQ}$, will be called the *point-to-line function*. Note that this is a real valued function of a real variable. One

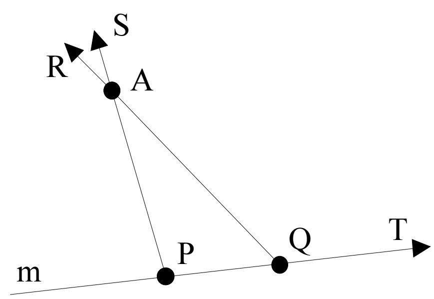


Figure 4: Point-to-Line Function

possible choice for the orientation, R, that determines the value of this function might be that for the usual distance: that is, so that $\overline{QA}$ is always positive. Then the continuity axiom just requires that this function always be thrice differentiable (except when $\overline{PA}=0$ and R becomes ambiguous for vanishing argument). In the Euclidean case, for example, this function is:

$$\sqrt{\overline{PA}^2 - 2\overline{PA}\,\overline{PQ}\cos\theta + \overline{PQ}^2}\ ,$$

where θ is the angle (properly defined) between T and S, an arbitrarily chosen direction for the line between P and A. It is apparent that this quantity is actually a function of both directions and both distances.

This function is an important concept. But, it will turn out that the more natural form is that given by the following definition. This is only easy to express by use of the signed measure.

Definition. For the orientation, *R*, chosen such that $\overline{QA}$ has the same sign as $\overline{PA}$ the family of functions with the *value*

$$\overline{QA} = S_T(\overline{PA}; \overline{PQ})$$

and the *argument* $\overline{PQ}$ will be called *line-to-line functions*. A specific function in this family is identified by the two directions *S* and *T* (which determine the line, *m*, and the intersection point *P*) and the segment measure $\overline{PA}$ (called the *parameter*, which locates the point *A* on *S*). The direction *S* will be called the *base* ray while *T* will be called the *solitary* ray.

These functions have two variables. There is an ambiguity of direction when $\overline{PA} = 0$, as before. Otherwise, according to Axiom 4, the third derivatives of both variables, including the mixed partials, exist.

For an example consider the Euclidean plane. Let the directions *T* and *S* be determined by the angles θ_T and θ_S, respectively, counterclockwise from some arbitrary reference direction. If $\overline{PA} = L$ and $\overline{PQ} = x$ then

$$S_T(L;x) = L\sqrt{1 - 2\frac{x}{L}\cos(\theta_T - \theta_S) + \left(\frac{x}{L}\right)^2} \qquad 2.5.1$$

is the line-to-line function. The positive square root is always used. In particular, note that when *L* is negative this expression gives the right answer with no need to consider another case for the complementary angle and that the function varies smoothly with the directions.

2.6 The Geodesic Hypothesis

In the conventional developments of synthetic geometry substantial axioms concerning the congruence of angles and triangles are introduced. In Birkhoff's approach this takes the form of axioms for angle measure. These produce the familiar uniform, isotropic and twist free space of neutral geometry. In accordance with the ideas of the introduction this will not be done. Instead the *geodesic hypothesi*s will be introduced.

> **Axiom 6**. For any three distinct points, $A, B, C \in \mathcal{P}$,
>
> $$\left|\overline{AB}\right| + \left|\overline{BC}\right| \geq \left|\overline{AC}\right|,$$
>
> the triangle inequality is satisfied. The equality obtains only if the points are collinear and B is between A and C.

The absolute values of the measures are distance-like *on their lines* and can be used to establish "congruences" between segments on different lines. Note, in passing, that the distance function mentioned before, in view of the geodesic hypothesis, satisfies all the axioms needed to make the space of this paper a metric space. The geodesic hypothesis means that all lines are, roughly speaking, "straight" or "geodesics". It relates measures on different lines to each other in a geometric way that goes beyond mere continuity and gives some meaning to the congruence relations they entail. This hypothesis makes the way that the geometric rulings knit the space together of a special kind and is the source of the richness of the resulting theory.

While the internal structure of lines has been given by hypothesis, it remains to give as complete a characterization as possible of the nature of the rest of the space. It is surprising that, on the basis of the axioms given, it will eventually be possible to derive nearly definitive properties of angles and triangles in place of the abandoned congruence axioms. Furthermore, the synthetic construction of an angle-like natural measure of direction will be possible. All these results flow from the geodesic hypothesis. The Euclidean axioms by which these sorts of things are usually obtained are, therefore, not necessary for them. Rather the usual axioms merely give them in a particular form. For instance, they give the natural angle measure in the form of the familiar Euclidean angle which turns out to be a special case of a more general quantity.

2.7 Euclidean vs. Physical Geometry

The differential synthetic geometry entailed by this axiom system will be called a *physical geometry* for convenience. In many ways the two dimensional version of physical geometry is similar to ordinary synthetic Euclidean plane geometry. Axioms 1-3 are the same. Axiom 4 has been shown to be equivalent to the existence of coordinate systems on the lines which is the corresponding axiom in Birkhoff's ruler and protractor formulation. It is known that Euclidean geometry satisfies the continuity axiom and that its lines satisfy[*] the geodesic hypothesis. Therefore, Euclidean geometry is a special case of physical geometry and so it must be possible to specialize any result obtained to the Euclidean case. Considering this example will be instructive at every point of the investigation.

The differences are that all the assumptions which lead to uniformity or isotropy have been excised and replaced with other, physically motivated, axioms which make no such assumptions. Accordingly, it is to be expected that the local geometrical characteristics of the space specified by these axioms may vary with position and direction. It is reasonable to call such a geometry a *differential geometry* altho it arises from a point of view which may be somewhat alien to conventional differential geometers.

2.8 Busemann's Examples

Noted mathematician Herbert Busemann has done some work tangential to that of this paper. In the course of it he gave a type of example which in some cases are

---

[*] See the Appendix.

physical geometries. They show that rather singular behavior is possible in a physical geometry: For instance, either all parallel lines can converge at infinity or all parallel lines can diverge at infinity. The demonstration of such possibilities is interesting but not physically remarkable.

A more interesting point* is that since Bussman's examples are not Riemannian this shows that physical geometry is not necessarily Riemannian.

These examples also suggest a more telling physical conclusion because their distance function is constructed by a physically bizarre process. That is, it is a mathematical contrivance, not based on any physically intelligible principles, solely for the purpose of artificially satisfying the axioms. This is not to disparage Busemann's ingenuity or the usefulness of such contrivances for the purpose of supplying mathematical counterexamples. However, that it is possible for such a distance function to satisfy the axioms is a clear indication of their physical inadequacy. The present axiom system must be incomplete. A physically satisfactory geometry ought to be restricted to more natural structures. This suspicion will ultimately be reinforced by the general characterization which can be derived for physical geometry's tangent space. So far, however, suitable additional, physically based, axioms have not been discovered to complete the theory.

The appendix contains a discussion of Busemann's construction where anyone can read it but it will not be a pointless diversion from the subject at hand.

2.9 A Natural Example

Ultimately it will be possible to deduce the most general scaling, isotropic physical geometry (tho this demonstration is beyond the scope of the introductory material in

* Due to Dr. Plaut.

this paper). Scaling means that any figure may be constructed at an arbitrary scale and all other properties (e.g. the angle derivatives, ratios of segments etc.) remain invariant. Isotropy means that the properties of the space do not vary with direction: "rotation" of figures is an isomorphism. This special geometry is natural in the sense that it is a consequence of combining these physically meaningful characteristics with the axioms of physical geometry.

This geometry is a Cartesian plane (that is, the points are ordered pairs of numbers and the lines are the usual lines) upon which the following distance function is imposed.

$$d(x,y) = \sqrt{t(x_1 - y_1)^2 + m(x_1 - y_1)(x_2 - y_2) + g(x_2 - y_2)^2}\ e^{-\frac{m}{\Delta}\phi} \qquad 2.9.1$$

where $\phi = \tan^{-1}\left(\frac{2g}{\Delta}\frac{x_2 - y_2}{x_1 - y_1} + \frac{m}{\Delta}\right)$. The various constants are related by

$$\Delta = \sqrt{4tg - m^2}\ ,\ \sqrt{t} = e^{\frac{m}{\Delta}\tan^{-1}\frac{m}{\Delta}} \text{ and } \sqrt{g} = e^{\frac{m\pi}{2\Delta}}$$

which leaves only one free parameter in the geometry.

Obviously, when m vanishes this reduces to Euclidean geometry as a special case. Here it is tacitly assumed that the discriminant in the square root is positive. This entails a transcendental inequality

$$4e^{\frac{m}{\Delta}\left(2\tan^{-1}\frac{m}{\Delta}+\pi\right)} \geq m^2$$

which it is clearly possible to satisfy for small m (that is, in the interesting vicinity of the Euclidean case) of either sign. (For the other sign of the discriminant the nature of the geometry changes in character.) This is an example of a physical geometry with a naturally derived structure which nevertheless has both non-Euclidean and non-Riemannian characteristics (it does not have a Euclidean tangent space).

2.10 Global and Local Differential Geometry

Like ordinary synthetic geometry, this geometric system is a global one. That is, the axioms apply between any points and over all distances. Because such a global geometry cannot include many geometries of great mathematical or physical interest, such as those of the sphere or of planetary orbits, it might be considered of minor or specialized interest.

To include the more interesting phenomenon a geometry must be local. That is, the axioms, rather than applying to the whole plane, must apply only to any sufficiently small, but finite, neighborhood of any point. In this way the possibility of topological complications are not inadvertently and unnecessarily excluded while specifying the nature of the geometry.

There are also conceptual physical considerations which argue toward a similar attitude. Making distant comparisons, or even comparisons between substantially distinct directions, are epistemologically questionable procedures which would be best avoided. Accordingly, for instance, it would be physically preferable to regard the triangle inequality as pertaining only to triangles all of whose sides "nearly" coincide with a common line. No attempt will be made here to give this a precise meaning; that is a task to be undertaken when a local theory is developed.

Nevertheless, the asymptotically flat geometries which global physical geometry can be expected to include ought to be of considerable intrinsic interest to both physics and mathematics. Physically, if these axioms hold then space must at least have their logically consequent properties. If, on the other hand, some of the consequences are not physical then space must violate these axioms in some way. Similarly, everything deduced from the axioms must hold for the geometry of Riemann or Riemann's geometry must violate the axioms in some way. Understanding any of these possibilities promises insight.

Furthermore, this global geometry is the essential foundation for the desired local theories. Working out its consequences can be expected to be instructive as to how to construct a similar local geometry. There are two ways it can then be used to construct a local theory. The first is to modify the axiom system to make it local based on the understanding gained from the global theory. The second is to construct a "manifold" theory based on the tangent space of this theory in a way similar to that by which conventional manifolds are based on a Euclidean tangent space.

2.11 Next

The next chapter will begin the exploration of the consequences of the axiomatic system that has now been set forth. Then, for simplicity, Chapter 4 will add an axiom that will specialize to the case of two dimensions. Building on these foundations, in Chapter 5 the concept of direction will be developed setting the stage for the investigation of tangent space (not included in this thesis).