MATHEMATICAL ANALYSIS OF BINOCULAR VISION

R. K. Luneburg

Published on demand by
UNIVERSITY MICROFILMS
University Microfilms Limited, High Wycomb, England
A Xerox Company, Ann Arbor, Michigan, U.S.A.

MATHEMATICAL ANALYSIS OF BINOCULAR VISION

by

RUDOLF K. LUNEBURG

Published for the
HANOVER INSTITUTE
BY PRINCETON UNIVERSITY PRESS
Princeton, New Jesrey
1947

Lithoprinted in U.S.A.
EDWARDS BROTHERS, INC.
ANN ARBOR, MICHIGAN
1948

FOREWORD

I wish to express my gratitude to the Dartmouth Eye Institute especially to Professor Adelbert Ames, Jr., and to Mr. John Pearson. The following pages would not have been written without their hospitality which I had the privilege of enjoying at the Institute and without the interest they accorded this mathematical theory.

In particular I have to thank Professor Ames for many stimulating discussions and demonstrations. His thesis that our sensations are related to the outside stimulus patterns but cannot be derived from them has been a guiding line in the following considerations.

To Dr. Anna Stein (Bureau of Visual Science, American Optical Company), assistant to Professor Ames, I am greatly indebted on the mathematical side. Her critical help has been invaluable in the discussion and solution of the following mathematical problems, and last but not least, in the preparation of the manuscript.

I also wish to express my thanks to Mrs. Alice Weymouth for her work of typing the various drafts of the following text with its many mathematical formulae.

Rudolf Luneburg
Dartmouth Eye Institute
Hanover, N. H.

December 26, 1946

CONTENTS

MATHEMATICAL ANALYSIS OF BINOCULAR VISION

INTRODUCTION

Our aim in the following investigation is to develop a mathematical theory of visual perception. In particular we are concerned with binocular vision, i.e., with the perception provided by the concerted action of two eyes. We hope to demonstrate that certain observations analyzed from a general geometrical point of view lead to a theory of binocular vision, which has some rather interesting consequences and which gives a natural explanation to certain well-known phenomena of visual optics. Before developing this theory in detail, we shall outline the general premises upon which our solution of the problem is based.

1. We recognize, by binocular vision, that we are surrounded by a three-dimensional manifold of objects. These objects have, besides characteristic qualities of color and brightness, form and localization. In a visual sensation we thus are not only immediately aware of a distribution of colors and brightnesses, but also of the fact that certain of these qualities are combined to unities, namely, objects, which have a definite geometrical form and a definite localization in a three-dimensional space. We shall call this space the visual space. Our problem is to investigate its geometrical character, i.e., the qualities of form and localization in visual sensations.

The concept of the visual space becomes clearer from the following consideration. We can coordinate the "sensed" points in a particular visual sensation to the points of a three-dimensional geometrical manifold. This, of course, can be done in many different ways. We shall call the result of such a coordination a geometrical map of the visual sensation. Consider, for example, the coordination which is the basis of the projection theory of binocular vision. A sensed point is represented by the intersection point of two projection lines which are drawn from two fixed points of the Euclidean space. The base points are the centers of rotation of the eyes and the projection lines the optical axes. We obtain by this construction a Euclidean map of the visual sensation. However, we cannot be sure that the map represents truly the sensed qualities of form and localization of the objects. This would be the case if the apparent distance of any two sensed points were always proportional to the geometrical distance of the associated points of the Euclidean map.

Clearly, this is not true. Astronomical objects like the sun or the moon are seen at finite distances; their sensed size is also finite and in no way proportional to astronomical dimensions. Even the sky itself gives the impression of a dome of finite radius. It certainly does not introduce any special size sensation comparable to Euclidean infinity in its relation to finite Euclidean size.

These considerations indicate our actual problem: To find a coordination of the sensed points of a visual sensation to the points of a geometrical manifold such that the apparent distance of any two sensed points is always proportional to the geometrical distance of the correlated points. A coordination of this type is called a psychometric coordination. Whether or not such a coordination is possible is a psychological problem which requires a special consideration. Certain basic psychological facts which we shall discuss in §1 indicate, however, that in the case of visual sensations, a psychometric coordination is possible. Moreover, we shall prove from these facts that the geometric manifold in which a psychometric map of visual sensations can be obtained is uniquely determined. The geometry in this manifold then represents the visual qualities of form and localization in mathematical formulation. It establishes the possibility of measuring in the visual space.

Our above discussion of the Euclidean map obtained by projection from two centers does not prove that the visual space is a non-Euclidean manifold. It only shows that this method of coordination yields a map which is not psychometric. There could still exist other coordinations of sensed points to points of the Euclidean space which lead to psychometric maps. However, such a coordination will be impossible, if the visual space should be non-Euclidean.

Let us illustrate this situation by the two-dimensional non-Euclidean manifold of points on the sphere. We can coordinate the points of the sphere to the points of the Euclidean plane and thus construct a plane map of the sphere. We may study the spherical geometry by the plane map and its principle of construction. However, we must not try to judge the actual size or shape of objects on the sphere by the Euclidean size and shape of their images on the map. A map on which this is allowed is called isometric and the coordination an isometric transformation. In our example such a transformation is impossible: A sphere cannot be mapped isometrically to a plane.

Instead of the spherical geometry let us consider the geometry on a cylinder or on a cone. In these cases it is possible to construct isometric maps. We also may say that a cylinder or cone can be applied to a plane without "stretching" the material from which it is made.

The problem of isometric mapping occupies a significant position in mathematics. In fact, we may consider this originally practical problem as the beginning of one of the roads which have led to the establishment of geometries different from the Euclidean geometry. In Gauss' theory of curved surfaces conditions were given for isometric transformation of surfaces onto each other. Riemann, after Gauss, generalized these results to manifolds of three and more dimensions and formulated their significance for the general space problem. The general result is as follows: The geometry in a manifold can be derived from its metric, i.e., from a rule for measuring the size of small line elements. Two such manifolds can be coordinated isometrically to each other only under certain conditions which the metric of the first manifold must satisfy in relation to the metric of the other. If the second manifold is Euclidean, then these conditions give the answer to the question of whether an isometric Euclidean map of the first manifold can be constructed. Manifolds where the answer is negative are called non-Euclidean; in this case no Euclidean map can be considered as true in all respects.

Suppose now that we study visual sensations by the Euclidean map obtained by projection from two centers, or, in fact, by any other Euclidean map. This means that we try to interpret visual observations by applying indiscriminately the relations of Euclidean geometry. If the visual space should be non-Euclidean, then any conclusion we draw from our results must be questioned and we must expect eventually to find contradictions with observations. Such contradictions then can be eliminated only by reinterpreting our observations in a non-Euclidean visual space.

2. We can explore our environment in an entirely different way, namely, by physical measurements. With the aid of certain general principles the results of these measurements are combined mathematically and the environment is recognized

as a manifold of physical objects. Their qualities are principally different from the sensed qualities of visual perception. Instead of colors and brightness we obtain optical qualities referring to reflection or transmission of light waves. Instead of sensed form and localization we have measured physical form and physical localization in a three-dimensional space. We shall call this space the physical space and have to distinguish it carefully from the visual space. We assume the physical space to be Euclidean in what is to follow. This is certainly justified in the environment where sensory depth perception by binocular vision is effective.

Of course there is a certain relationship between the two spaces. This relation is established by the stimuli provided by the light which is emitted or reflected by physical objects. A small part of this radiated energy is picked up by the dioptric system of our eyes and, by certain electrical and chemical disturbances, transmitted to the brain. The immediate and definite character of the associated visual sensation may tempt us to the belief that it is determined in all its qualities by these light stimuli. Indeed, if we subscribe to the projection theory of binocular vision, we tacitly make this assumption, since we identify physical and visual space. But even by considering these spaces as metrically different we can still believe in a necessary one-to-one correspondence of physical and visual space. A configuration of physical objects seems to create, by necessity, one and the same visual sensation for a given observer.

However, this belief does not stand a critical test. Actually, a visual sensation is the response of a living organism to physical stimuli. Thus we can scarcely hope to find the explanation of visual sensations and their sensed qualities in the complicated chain of physical events by which the organism is stimulated. We must take account of other factors which are given by the organism itself and not by the stimuli. These are psychological factors determined by the purposes, expectations, and the experiential background of the observer.

By adopting this point of view we have to consider the following possibility. Objects can be identical in certain aspects of physical form and localization but are seen as objects which differ in these aspects. Vice versa, two sensations can be identical in all their qualities though related to different physical objects. That this is true even in the realm of binocular vision is clearly shown by some experiments carried out at the Dartmouth Eye Institute. A set of rooms with curved walls has been constructed; the walls are provided with curved window patterns. Every one of these distorted rooms gives the appearance of the same rectangular room, i.e., the same sensation is related to an infinite set of physically different rooms. In a second experiment, perspective patterns are drawn upon a vertical board. The apparent localization of the board changes strikingly if the pattern is varied though physically the board is not moved. An infinite set of apparently different localizations thus can be related to the same physical localization. We stress the point that in both demonstrations the observation is binocular.*

We conclude from the above experiments that it would be futile to attempt to express the relation of visual and physical space in the form of a necessary one-to-one correspondence. The qualities of visual sensations are not uniquely determined by the physical stimuli. Since, on the other hand, we cannot consider sensations and stimuli as entirely unrelated, we are forced to the conclusion that only certain special elements of visual sensations are determined by the stimuli.

*These demonstrations have been designed by A. Ames, Jr. The mathematical analysis of the experiments has been given by A. Stein.

The essence of the following theory will be that, in fact, there exist immutable relationships but that they are confined to the assignments of apparent size to physical line elements. Consider two infinitely close luminous points in the physical space. Our hypothesis is that their apparent relative distance is determined by the physical coordinates of the two points. For these differential or primitive size sensations and only for these we shall assume a necessary functional dependence upon the cooresponding differentials of the stimuli.

We stress the point that this relation of apparent size to physical qualities of localization is purely a mathematical relation. It does not introduce the concept of physical causality and thus does not express sensory size qualities by physical units. In fact, it is a relation of two geometrical manifolds to each other: The visual space obtained by psychometric coordination and the geometrical manifold which represents the physical space.

Our postulate is clearly compatible with the fact that visual sensations are not uniquely determined by the stimuli. Indeed, an actual sensation requires in addition to assigning size to its differential elements the integration of these elements to a unity. Thus arbitrary parameters of integration are available. These parameters are chosen by the observer and, in the choice, he depends on his psychological condition.

The mathematical expression for the apparent size of a line element in terms of its physical coordinates can be found by analyzing certain observations, i.e., by an inductive empirical investigation. On the other hand, this expression establishes a Riemannian metric for the visual space, namely, a rule for measuring the size of inifinitely small line elements. Though referring to the infinitely small, it nevertheless already determines the general character of the visual space. It thus must give us the answer to the question whether or not the visual space is Euclidean. Our answer will be that, in fact, the metric of visual sensations is non-Euclidean. In particular we hope to demonstrate that the geometry of the visual space is the hyperbolic geometry of Lobachevski.

Section 1

PSYCHOMETRIC COORDINATION

1.1. It seems to be paradox, at first sight, to introduce the concept of a metric in a manifold of sensations, i.e., the concept of measuring psychological qualities by coordination of numbers. Indeed, psychological manifolds, like heat sensations, or sensations of brightness, do not seem to be provided with a metric. We may say that sensation S_1 is greater or smaller than sensation S_2, but not how much greater or smaller. How then can we speak about a manifold of sensations of space which has a non-Euclidean metric?

It is, however, problematic whether greater or smaller is the only property of psychological manifolds which we may recognize. Consider, for example, the sensation of pitch in sound perception. If trained, we are well able to compare three pitch sensations

$$S_1, S_2, S_3$$

i.e., to judge whether the contrast (S_2S_3) is greater or smaller than the contrast (S_1S_2).

In space sensations a similar phenomenon may be observed. Let us consider, for example, the sensation of height. The contrast (S_1S_2), between two such sensations, is interpreted as the size of the object between S_1 and S_2. Obviously we are in the position to judge whether (S_1S_2) is greater or smaller than (S_2S_3). We also may say that, by our sensations of height, we assign vertical size independent of vertical localization.

We shall show next that recognition of greater and smaller and recognition of greater and smaller contrast implies the existence of a metric and that this metric is in essence uniquely determined.

1.2. Let us first, for simplicity's sake, consider a one-dimensional manifold of sensations. Our problem is to coordinate numbers, x, to the sensations, S, of this manifold. This, of course, can be done in a great variety of ways; let

$$x = x(S) \tag{1.21}$$

be such a coordination, and let us also assume that the manifold of coordinated numbers, x, forms a continuous manifold. However, we require that this coordination shall be such that

$$x(S_2) > x(S_1) \quad \text{if} \quad S_2 > S_1$$

and that (1.22)

$$x(S_3) - x(S_2) > x(S_2) - x(S_1)$$

if the contrast $(S_3S_2) > (S_2S_1)$

Only if these conditions are satisfied can we consider the coordination as representing the characteristics of the sensations in question.

We show next that the "function" $x(S)$ is in its essentials uniquely determined. Indeed, let $X = X(S)$ be another coordination of the type (1.22). Then we may consider X as a mathematical function of x. This follows from the fact that to every sensation S there belongs one and only one number x and also one and only one number X. Consequently to a given number x there belongs one and only one number X, i.e., $X = f(x)$. Furthermore, from $x_2 > x_1$ it follows that $S_2 > S_1$ and thus $X_2 > X_1$. Finally, from $x_3 - x_2 > x_2 - x_1$ that $(S_3S_2) > (S_2S_1)$ and thus that $X_3 - X_2 > X_2 - X_1$. This means that the function $X = f(x)$ must satisfy the conditions

$$f(x_2) > f(x_1) \quad \text{if} \quad x_2 > x_1$$

$$f(x_3) - f(x_2) > f(x_2) - f(x_1) \quad \text{if} \quad x_3 - x_2 > x_2 - x_1 \tag{1.23}$$

whatever number x_1 and x_2 may be.

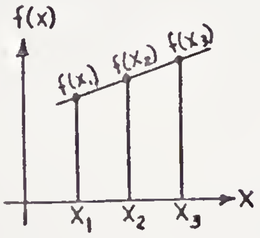
From the last condition (1.23) and from the continuity of the function $f(x)$, it follows that

$$f(x_3) - f(x_2) = f(x_2) - f(x_1) \tag{1.24}$$

if $x_3 - x_2 = x_2 - x_1$.

This last result we may formulate as follows:

If x_2 is the arithmetic mean



$$x_2 = \tfrac{1}{2}(x_1 + x_3) \tag{1.25}$$

of two numbers x_1, x_3, then the value $f(x_2)$ of the function $f(x)$ at x_2 is also the arithmetic mean of the values $f(x_1)$ and $f(x_3)$, i.e.,

$$f(x_2) = \tfrac{1}{2}\,[f(x_1) + f(x_3)] \tag{1.26}$$

The only continuous functions $f(x)$ satisfying this condition for any two values x_1, x_3 are the linear functions

$$X = f(x) = ax + b \tag{1.27}$$

where a and b are arbitrary constants.

The arbitrariness of the constant a means that no absolute size is given but only relative size. (Change of unit of size.) The arbitrariness of b means that the origin of the scale is undetermined.

In these limits, however, we see that the psychometric coordination of numbers to sensations is uniquely determined, if the sensations allow recognition of greater and smaller and of greater and smaller contrast.

We mention that this result can be obtained under very much weaker conditions. We need only to require that contrasts (S_1S_2) (S_2S_3) can be compared if S_2 and S_3 lie in the immediate neighborhood of S_1.

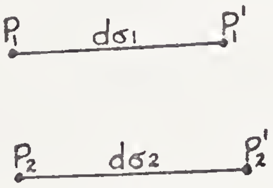
1.3. In the case of sensation manifolds of more than one dimension, we have to proceed a little differently. Let us consider a two-dimensional manifold of sensations and coordinate these sensations S to the points P of a two-dimensional manifold m of points. The geometric relations in the point manifold shall be determined by a quadratic differential.

$$d\sigma^2 = e dx^2 + 2f dx dy + g dy^2 \tag{1.31}$$

where x, y are the coordinates of a point P and e, f, g functions of x, y.

The above differential is called the metric* of the point manifold; it determines the distance of two neighboring points $P = (x,y)$ and $P' = (x+dx,\ y+dy)$.

We assume again that it is possible to compare contrasts of sensations, i.e., to recognize whether the contrast (S_1S_1') of two sensations S_1 and S_1' is greater, equal to, or smaller than the contrast (S_2S_2') of two other sensations S_2 and S_2'.



We now require the coordination of sensations S and points P to be such that the distances $d\sigma_1$ and $d\sigma_2$ of the points P_1, P_1' and P_2, P_2' give a true measure of the contrasts (S_1S_1') and $S_2S_2')$, of two pairs of coordinated "neighboring" sensations S_1, S_1' and S_2, S_2'. In other words, we require that

$$d\sigma_1^2 = e_1dx_1^2 + 2f_1dx_1dy_1 + g_1dy_1^2 = e_2dx_2^2 + 2f_2dx_2dy_2 + g_2dy_2^2 = d\sigma_2^2 \tag{1.32}$$

if

$$(S_1S_1') = (S_2S_2')$$

and vice versa.

Only with this condition satisfied can the coordination of sensations and points be considered as representing truly the characteristics of the sensations in question. We shall call it a psychometric coordination.

Let us now assume that, for a given manifold of sensations, such a psychometric coordination is possible. Then we can show that, in essence, the point manifold m and its metric are uniquely determined.

Indeed, let M be another manifold of points and

$$d\Sigma^2 = E dX^2 + 2F dX dY + G dY^2 \tag{1.33}$$

its metric expressed in certain coordinates X, Y.

Let us assume that our given manifold of sensations can be coordinated psychometrically to M so that always

$$d\Sigma_1^2 \gtreqless d\Sigma_2^2 \quad \text{if} \quad (S_1S_1') \gtreqless (S_2S_2') \tag{1.34}$$

and vice versa. Since the sensations S are coordinated in one-to-one correspondence to the points (x, y) of m as well as to the points (X, Y) of M, it follows that X, Y must be functions of x, y:

$$X = X(x,y)$$
$$Y = Y(x,y) \tag{1.35}$$

We conclude furthermore by (1.32) and (1.34):

---

*In §3 the concept of a metric is explained in greater detail.

The inequalities

$$e_1dx_1^2 + 2f_1dx_1dy_1 + g_1dy_1^2 \gtreqless e_2dx_2^2 + 2f_2dx_2dy_2 + g_2dy_2^2$$

imply the corresponding inequalities

$$E_1dX_1^2 + 2F_1dX_1dY_1 + GdY_1^2 \gtreqless E_2dX_2^2 + 2F_2dX_2dY_2 + G_2dY_2^2$$

whatever dx_1, dy_1; dx_2, dy_2; and x_1, y_1; x_2, y_2 may be.

It is not difficult to see that this is possible only if the quadratic differentials $d\sigma^2$ and $d\Sigma^2$ are related by the identity $d\Sigma^2 \equiv a d\sigma^2$ or

$$EdX^2 + 2FdXdY + GdY^2 \equiv a(edx^2 + 2fdxdy + gdy^2) \tag{1.36}$$

where a is a constant.

In other words: <u>By submitting the differential $d\Sigma^2$ of M to the transformation (1.35) the differential $d\sigma^2$ is obtained multiplied with a constant a.</u> As before, we may interpret this appearance of an arbitrary constant a as indicating the <u>arbitrariness of the unit of size</u> in psychometric evaluation.

In general it is not possible to transform a given quadratic differential $edx^2 + 2fdxdy + gdz^2$ into another one $EdX^2 + 2FdXdY + GdY^2$ with arbitrarily chosen coefficients E, F. G. The result that, in the case of the above differentials $d\sigma^2$ and $d\Sigma^2$, such a transformation is possible, points to the fact that, <u>geometrically, the point manifolds m and M are identical</u>. The points of M are identical with the points of m but characterized by different numbers X, Y instead of x, y, i.e., by different coordinates. It is clear that the geometrical characteristics of a point manifold must be independent of the choice of the coordinate system.

It may be mentioned again that the above result already follows if all the three sensations S_1', S_2, S_2' lie in an infinitesimal neighborhood of S_1; recognition of greater or smaller contrast thus is required only if two pairs of sensations S_1, S_1'; S_2, S_2' are sufficiently near to each other.

1.4. A similar consideration for sensational manifolds of three dimensions leads to an analogous result. If a psychometric coordination of sensations to a point manifold is at all possible, then there exists only one such manifold. The geometrical distance of two neighboring points P and P' given by a quadratic differential

$$d\sigma^2 = g_{11}dx^2 + g_{22}dy^2 + g_{33}dz^2 + 2g_{12}dxdy + 2g_{13}dxdz + 2g_{23}dydz \tag{1.41}$$

measures the contrast of the two associated neighboring sensations S and S'. The unit of the contrast size is the only indeterminacy in this coordination.

The question whether or not a geometrical manifold actually fits a given manifold of sensations according to the above contrast requirements can only be answered by an empirical investigation. For example, as to the space sensations of binocular vision, we have no right to assume a priori that the Euclidean space, i.e., a manifold with the metric

$$d\sigma^2 = dx^2 + dy^2 + dz^2$$

truly represents its characteristics.

In the following, we shall assume the possibility of psychometric coordination of visual space sensations to a geometrical manifold. We shall also assume that contrasts of space sensations can be compared. This means we may compare the sizes of two arbitrary line elements in space even if these line elements are not attached to the same base point. We then know from the above result that there exists only one geometrical manifold which represents the characteristics of binocular vision psychometrically. Our aim is to determine this manifold.

Section 2

BIPOLAR COORDINATES

In order to facilitate the mathematical investigation of our problem, we introduce first a suitable bipolar coordinate system. This system allows us to characterize a point of the physical space by three angles α, β, θ instead of by three Cartesian coordinates x, y, z. We shall discuss in this section the relation of these two coordinate systems.

2.1. The Cartesian coordinate system is oriented relative to the observer as follows: His eyes are at the points $y = \pm 1$ of the y-axis, the x,y-plane is his <u>horizontal plane</u>, the x,z-plane his <u>median plane</u>. We assume first that the observer views his environment without head movements so that--to be more precise--the centers of rotation of his two eyes remain at the points $y = \pm 1$ of the y-axis. We now construct a plane through the y-axis and through a point P of coordinates x, y, z. This plane is called the <u>plane of elevation</u> of the observed point. Let θ be the <u>angle of elevation</u>, i.e., the angle of the plane of elevation with the horizontal plane. We draw next, in the plane of elevation, two lines from the eyes to the point P. α and β are the angles of these lines with the y-axis. As indicated in Fig. 1, we measure the angle α from the positive direction of the y-axis, but β from the negative direction.

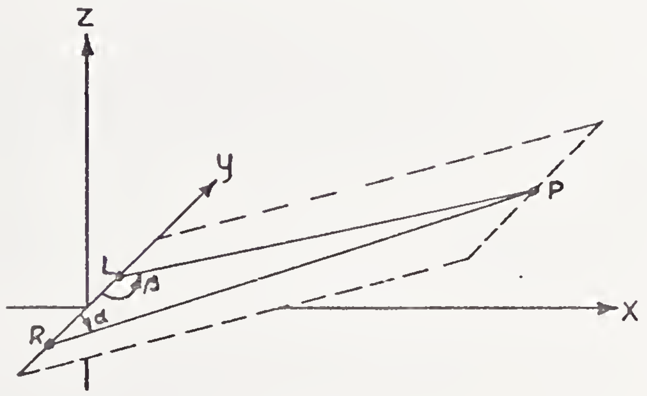


Fig. 1

One verifies easily that the relation between the linear coordinates x, y, z and the angular coordinates α, β, θ is given by the formulae

$$x = \frac{2\cos\theta}{\cot\alpha + \cot\beta}, \quad \cot\alpha = \frac{y+1}{\sqrt{x^2+z^2}}$$

$$y = \frac{\cot\alpha - \cot\beta}{\cot\alpha + \cot\beta}, \quad \cot\beta = \frac{1-y}{\sqrt{x^2+z^2}} \qquad (2.11)$$

$$z = \frac{2\sin\theta}{\cot\alpha + \cot\beta}, \quad \cot\theta = \frac{x}{z}$$

The transformation of the x, y, z space into the angular α, β, θ space is everywhere regular except on the y-axis. Since (2.11) is a transformation which treats each plane of elevation alike, we need only to discuss the transformation of the horizontal plane. In this case $z = 0$; $\theta = 0$, and thus

$$x = \frac{2}{\cot\alpha + \cot\beta}, \quad \cot\alpha = \frac{1+y}{x} \qquad (2.12)$$

$$y = \frac{\cot\alpha - \cot\beta}{\cot\alpha + \cot\beta}, \quad \cot\beta = \frac{1-y}{x}$$

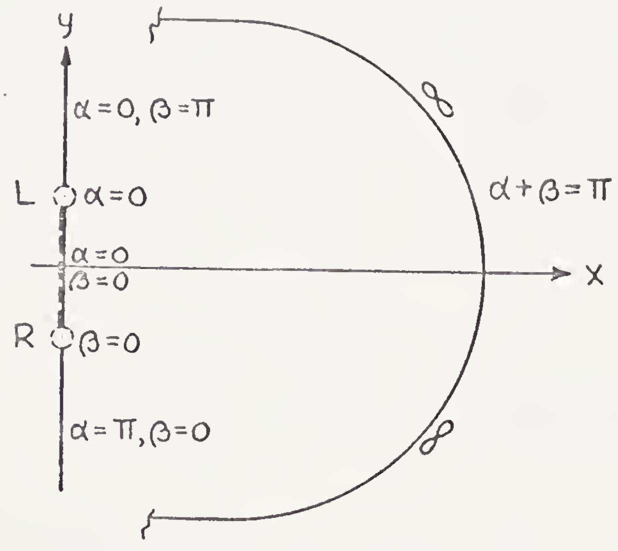


Fig. 2

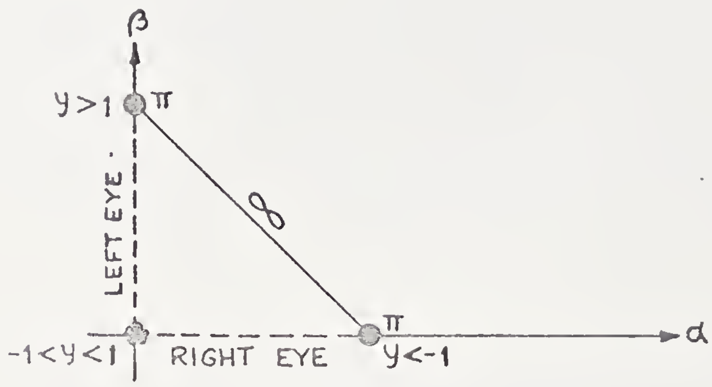


Fig. 3

We illustrate this transformation of the x,y-plane into the angular α, β-plane by determining the region of the α, β-plane which corresponds to the right half-plane $x \geqq 0$.

This region is bounded by the y-axis and an ideal curve infinitely far away. The values of the angular coordinates α, β on the boundary are given in Fig. 2. They determine immediately the boundary of the domain of the α, β-plane which corresponds to the half-plane $x \geqq 0$ (Fig. 3).

We conclude that the half-plane $x \geqq 0$ is transformed into the interior of the triangle shown in Fig. 3. To every point $x > 0$ there belongs one and only one point α, β in the interior of this triangle. The transformation thus is regular in all interior points. As to the transformation of the boundary elements, we notice, however, a striking irregularity. The two eyes are transformed into two lines of the α, β-plane. The sections $-1 < y < 1$, $y < -1$, $y > 1$ of the y-axis, on the other hand, are compressed into 3 separate points (0, 0); (π, 0); (0, π) respectively.

The significance of the bipolar coordinates for the physiological aspects of binocular vision is easily understood. Let us assume that the two eyes are in the "primary position," i.e., the optical axes are parallel to the x-axis. Then a point P with coordinates α, β, θ is projected onto the retina of the right eye with spherical coordinates (α, θ) and onto the retina of the left eye with spherical coordinates (β, θ). Indeed, the planes of elevation θ = const. intersect the two retinae in longitudinal sections, i.e., great circles through the retinal points on the y-axis. The cones α = const. intersect the retina of the right eye in lateral sections, i.e., circles of latitude around the y-axis. The cones β = const. give the lateral sections of the left eye. (Fig. 4)

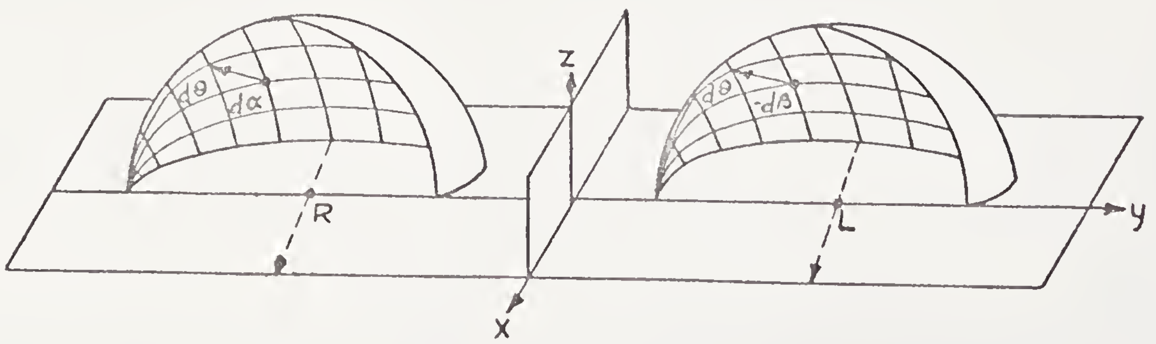


Fig. 4

We also notice that a line element (dx, dy, dz) attached to a point P will be projected as a line element ($d\alpha$, $d\theta$) onto the right retina and as ($-d\beta$, $d\theta$) onto the left retina. $d\alpha$, $d\beta$, $d\theta$ thereby may be found from dx, dy, dz by differentiating the equations (2.11).

The quantity $d\theta$ determines the vertical extension of the line element, the quantity $d\alpha + d\beta$, the horizontal disparity, and finally $\frac{1}{2}(d\beta - d\alpha)$, the horizontal extension.

As is well known in binocular vision, the sensation of depth is directly related to the horizontal disparity. Clearly, it would vanish if we let the distance of the two eyes converge to zero. By the mechanism of our vision the two retinal images are seen as one fused image, provided that the horizontal disparity is not too great. The disparity in the horizontal direction, i.e., the difference of extension of the two images parallel to the y-axis, is perceived as a new space dimension of the line element, namely, depth, in addition to vertical and horizontal extension.

It is characteristic for our bipolar coordinates that the two images of a line element in space (dx, dy, dz) must have the same vertical extension $d\theta$ on the retina. It is, however, easily possible by artificial means, for example in a stereoscope or haploscope, to provide the eyes with individual images which have different vertical extensions $d\theta_1$ and $d\theta_2$. Mathematically this would mean offering to the observer line elements from a four-dimensional manifold $d\alpha$, $d\beta$, $d\theta_1$ and $d\theta_2$. However, even if this vertical disparity $d\theta_1 - d\theta_2$ is small enough that a fused image is seen, there is no sensation of a new space dimension. In other words, our mind refuses to digest the well-meant offer of four-dimensional line elements--either the two images are not fused or, if fused, the vertical disparity is completely ignored.

The bipolar coordinates α, β, θ preserve their good physiological meaning if the eyes do not remain in the primary position but view objects directly by convergence. The angles α, β, θ then determine the position of the two optical axes of the eyes. A point P then is projected into the center of the retina (fovea), the region of clearest vision. Also a line element (dx, dy, dz) attached to P is observed in the center of the retina. Its images, however, are still characterized by the bipolar differentials ($d\alpha$, $d\theta$) and ($d\beta$, $d\theta$). $d\theta$ determines the vertical extension; $\frac{1}{2}(d\beta - d\alpha)$, the horizontal extension; $d\alpha + d\beta$, the horizontal disparity. Again $d\alpha + d\beta$ is sensed as depth.

It is true that eye movements are of a more complicated nature than assumed above. If the optical axes of the eyes are moved to converge at a point P which is not in the horizontal plane, then this movement is accompanied by well-determined rotational movements of the retina around the optic axis. This torsional movement of the eyes has the effect that retinal points on the horizontal section ($\theta = 0$) in Fig. 4, do not lie, after the movement towards P, in the plane of elevation through P. In other words, our differentials ($d\alpha$, $d\theta$); ($d\beta$, $d\theta$) do not determine directly the location of the images of a line element in a coordinate system solidly engraved on the retina. In such a system we would obtain two sets of differentials--($d\alpha^*$, $d\theta_1^*$) and ($-d\beta^*$, $d\theta_2^*$) for the right and left eye respectively--which, of course, can be found from ($d\alpha$, $d\beta$, $d\theta$) if the mechanism of torsion is known. It is a problem which of these different coordinates are significant for the interpretation and localization of the original line element: the coordinates $d\alpha$, $d\beta$, $d\theta$ in our bipolar system oriented on the base line between both eyes, or the coordinates ($d\alpha^*$, $d\theta_1^*$), ($-d\beta^*$, $d\theta_2^*$) in systems solidly attached to the retinae.

The following consideration seems to favor the assumption that the bipolar differentials give the significant clues for interpretation and localization, if the line element is viewed with both eyes. Consider two line elements with $d\theta = 0$, i.e., line elements which lie in a plane of elevation. It is quite inconceivable that the fact of their lying in a plane of elevation will not be recognized directly in spite of the fact that none of the quantities $d\theta_1^*$, $d\theta_2^*$ in the solidly attached coordinate systems vanishes. If the same combination of line elements, however, is viewed only with one eye, we can judge its orientation probably only by referring to fixed coordinate lines on the retina. Binocular vision thus gives us, in addition to depth perception, a greater certainty in directional localization, namely, orientation relative to the base line of the two eyes. Exactly this fact, however, is expressed in our bipolar coordinates.

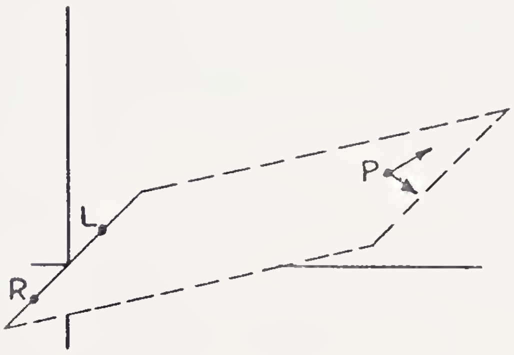


Fig. 5

For this reason we shall base our considerations in the following upon the bipolar coordinates α, β, θ and disregard torsional movements of the eyes.

2.2. Modified bipolar coordinates. For many purposes it is advantageous to use a modification of the bipolar coordinates α, β, θ which expresses more directly their physiological meaning.

We introduce the bipolar latitude

$$\varphi = \tfrac{1}{2}(\beta - \alpha) \tag{2.21}$$

and the bipolar parallax

$$\gamma = \pi - \alpha - \beta \tag{2.22}$$

For a discussion of these modified coordinates γ, φ, θ we may confine ourselves to the horizontal plane $\theta = 0$.

The meaning of γ is clear; the angle subtended by the lines of sight at the point of convergence, P. Obviously, γ may vary from 0 to π. For the interpretation of φ construct the circle through P and the eyes R and L. (Vieth-Muller circle.) The x-axis is intersected at P_0 and Q_0 by this circle. One easily proves that the angle under which the arc P_0P appears from either R, L, or Q_0 is equal to $\varphi = \frac{1}{2}(\beta - \alpha)$. This shows that it is justified to interpret φ as determining the lateral position of a point P.

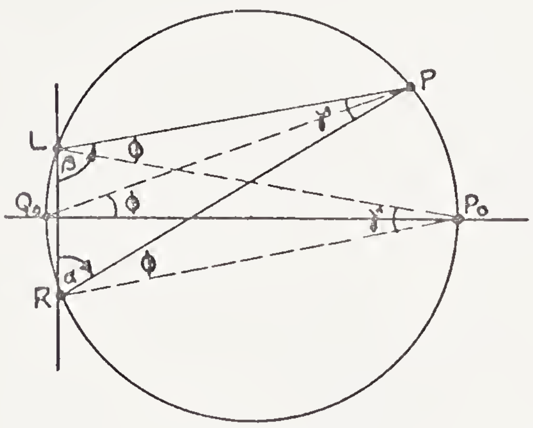


Fig. 6

A line element (dx, dy, dz) attached to a point P can be characterized by the differentials $d\gamma$, $d\varphi$, $d\theta$. From our former considerations it follows that $d\varphi$ and $d\theta$ determine the lateral and vertical extension of the retinal images and that $d\gamma$ measures the horizontal disparity. The latter is sensed as depth extension of the line element.

One proves readily that the relation between Cartesian coordinates x, y, z and bipolar coordinates γ, φ, θ is given by the formulae:

$$x = \frac{\cos 2\varphi + \cos\gamma}{\sin\gamma}\cos\theta, \qquad \tan\gamma = \frac{2\sqrt{x^2+z^2}}{x^2+y^2+z^2-1}$$

$$y = \frac{\sin 2\varphi}{\sin\gamma}, \qquad \tan 2\varphi = \frac{2y\sqrt{x^2+z^2}}{x^2+z^2-y^2+1} \tag{2.23}$$

$$z = \frac{\cos 2\varphi + \cos\gamma}{\sin\gamma}\sin\theta, \qquad \tan\theta = \frac{z}{x}$$

In the horizontal plane we have

$$x = \frac{\cos 2\varphi + \cos\gamma}{\sin\gamma} \qquad \tan\gamma = \frac{2x}{x^2+y^2-1}$$

$$y = \frac{\sin 2\varphi}{\sin\gamma} \qquad \tan 2\varphi = \frac{2xy}{x^2-y^2+1} \tag{2.24}$$

It follows that the curves γ = const. are given by the Vieth-Muller circles through the eyes:

$$x^2 + y^2 - 2x\cot\gamma = 1 \tag{2.25}$$

and the curves φ = const. by the hyperbolae of Hillebrand

$$-x^2 + y^2 + 2xy\cot 2\varphi = 1 \tag{2.26}$$

through the eyes. These hyperbolae have the asymptotes

$$y = x\tan\varphi \tag{2.27}$$

i.e., lines through the origin of direction φ. At any practical distance the hyperbolae φ = const. coincide with these lines, as can be seen in Fig. 7. This demonstrates again the justification of interpreting φ as characteristic for the lateral position of a point P and $d\varphi$ as lateral extension of a line element.

In order to investigate the regularity of the transformation (2.23), let us determine the domain of the γ, φ-plane into which the half-plane $x \geq 0$ is transformed. Fig. 8 shows the values of φ and γ on the boundary of the half-plane $x \geq 0$. At infinity we have $\gamma = 0$.

At the right eye: $\gamma - 2\varphi = \pi$

At the left eye: $\gamma + 2\varphi = \pi$

This gives as domain in the γ, φ-plane a triangle bounded by sections of the lines $\gamma = 0$, $\gamma - 2\varphi = \pi$, $\gamma + 2\varphi = \pi$. The eyes are stretched into the lines

$$\gamma - 2\varphi = \pi$$

and

$$\gamma + 2\varphi = \pi$$

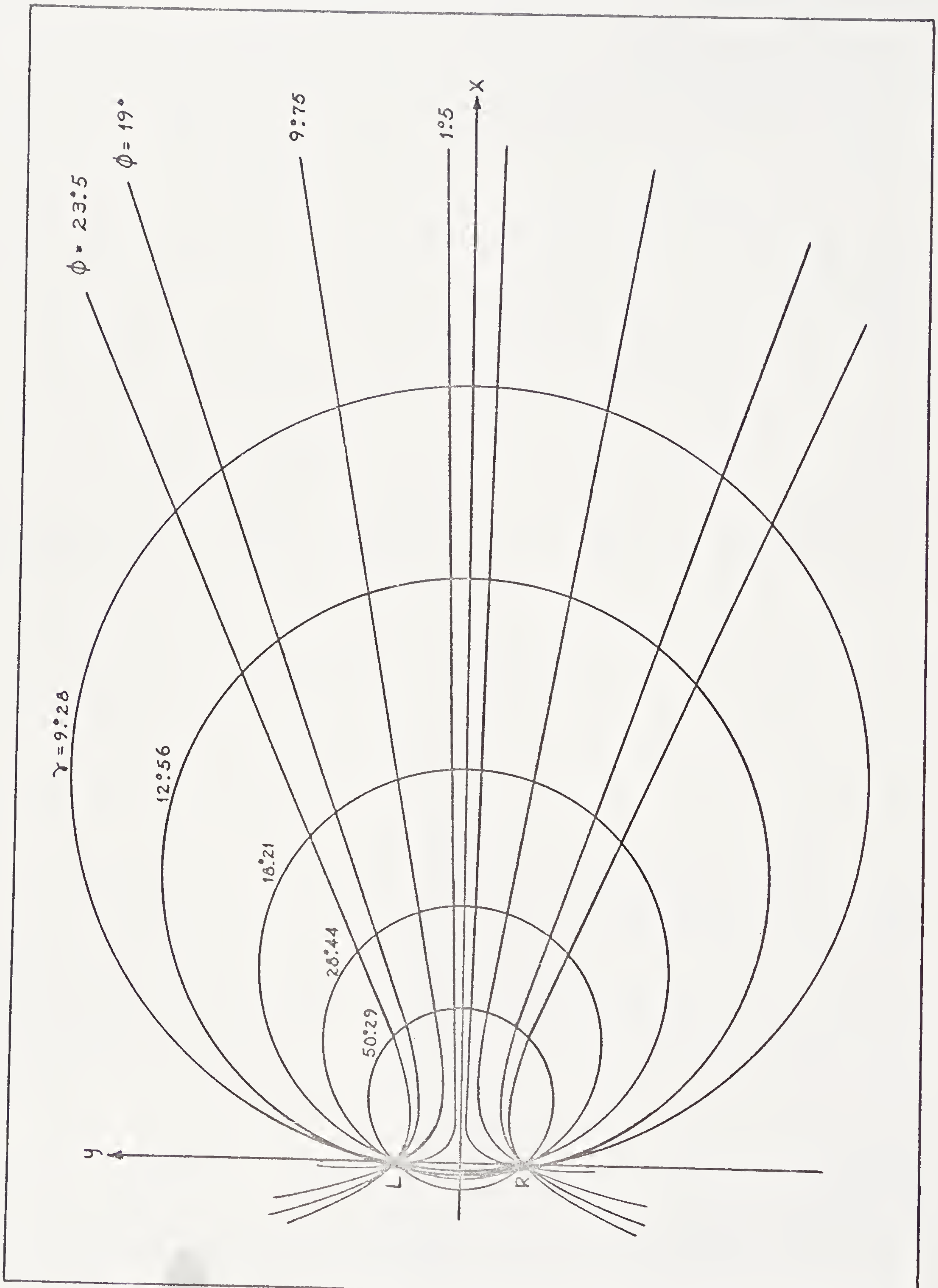


Fig. 7

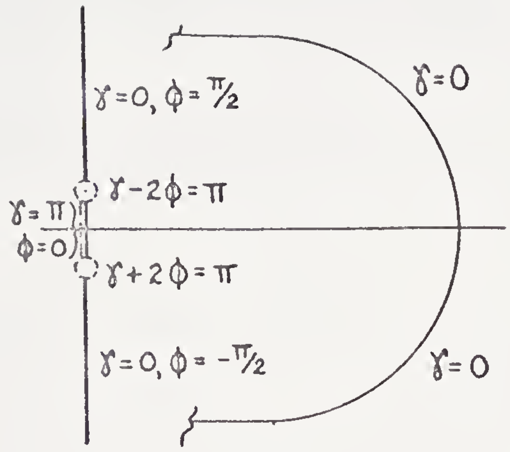


Fig. 8

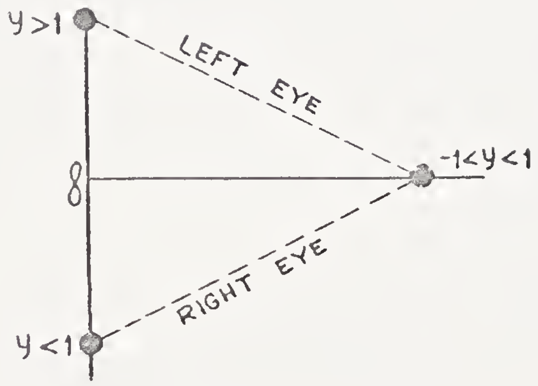


Fig. 9

The remaining sections of the y-axis are compressed into three single points $(0, \pi/2)$; $(0, -\pi/2)$; $(\pi, 0)$. The transformation is regular at interior points but highly irregular on the boundary.

2.3. Simplified relations between Cartesian and bipolar coordinates. In many practical applications we can replace the relations (2.23) by simplified approximate formulae by considering the distance of the eyes and thus the bipolar parallax γ as small. We may replace in (2.23) $\sin\gamma$ by γ and $\cos\gamma$ by 1, and obtain

$$x = \frac{2\cos^2\varphi\ \cos\varphi}{\gamma}, \qquad \gamma = \frac{2\sqrt{x^2+z^2}}{x^2+y^2+z^2}$$
$$y = \frac{2\sin\varphi\ \cos\varphi}{\gamma}, \qquad \tan\varphi = \frac{y}{\sqrt{x^2+z^2}} \tag{2.31}$$
$$z = \frac{2\cos^2\varphi\ \sin\theta}{\gamma}, \qquad \tan\theta = \frac{z}{x}$$

These relations may be used safely for objects which are far enough away from the eyes ($x > 30$, for example).

In the horizontal plane we have

$$x = \frac{2\cos^2\varphi}{\gamma}, \qquad \gamma = \frac{2x}{x^2+y^2}$$
$$y = \frac{2\sin\varphi\ \cos\varphi}{\gamma}, \qquad \tan\varphi = \frac{y}{x} \tag{2.32}$$

The curves γ = const. are now circles

$$x^2 + y^2 - 2\gamma x = 0 \tag{2.33}$$

through the point $x = y = 0$ with centers on the x-axis.

The curves φ = const. are the straight lines

$$y = x\tan\varphi \tag{2.34}$$

through the origin (Fig. 10).

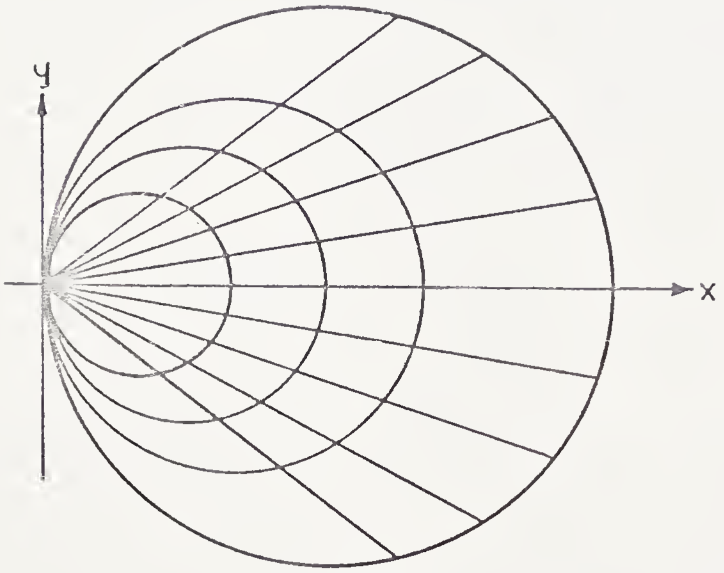


Fig. 10

This result allows us easily to determine the domain in the γ, φ-plane into which the half plane $x \geqq 0$ is transformed. γ can assume all values between ∞ and 0, and φ all values between $-\pi/2$ and $+\pi/2$. The transformation is regular at all interior points $x > 0$.

The irregularity of the boundary coordination is illustrated in Fig. 11 and Fig. 12.

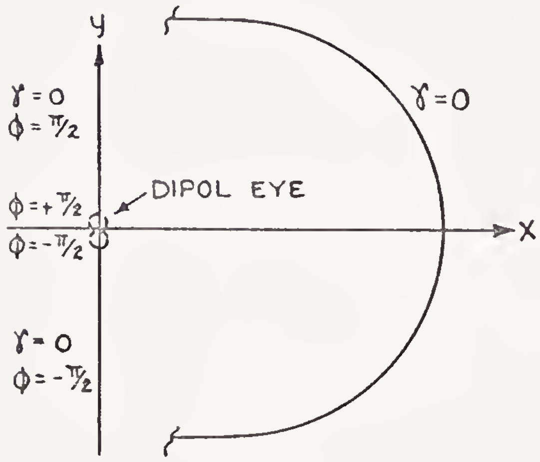


Fig. 11

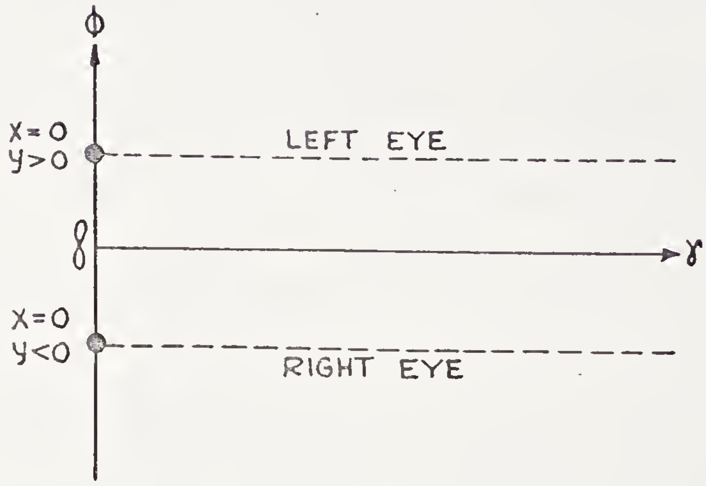


Fig. 12

2.4. <u>Equivalent configurations; iseikonic transformations</u>. Two line elements in space, (dx_1, dy_1, dz_1) and (dx_2, dy_2, dz_2) can have identical binocular line elements so that

$$\begin{aligned} d\alpha_1 &= d\alpha_2 & \text{or} \qquad d\gamma_1 &= d\gamma_2 \\ d\beta_1 &= d\beta_2 & d\varphi_1 &= d\varphi_2 \\ d\theta_1 &= d\theta_2 & d\theta_1 &= d\theta_2 \end{aligned} \tag{2.41}$$

Such line elements give to an observer the same binocular clues on the retinae, i.e., the same horizontal disparity $d\gamma$, the same lateral extension $d\varphi$, and the same vertical extension $d\theta$. For this reason we call such line elements <u>binocularly equivalent</u>.

Since most external objects may be considered as configurations of line elements--a curve in space is a one-parameter manifold of line elements, a surface in space a two-parameter manifold of line elements--we may extend the concept of equivalence to such configurations.

Consider, for example, a curve in space. We may characterize it by three functions

$$\alpha = \alpha(t)$$
$$\beta = \beta(t) \tag{2.42}$$
$$\theta = \theta(t)$$

where t is a parameter. Another curve in space must be considered equivalent to this curve if its line elements $d\alpha'$, $d\beta'$, $d\theta'$ are equal to the line elements of the original curve. This obviously is the case then and only then if

$$\alpha' = \alpha(t) + \delta$$
$$\beta' = \beta(t) + \varepsilon \tag{2.43}$$
$$\theta' = \theta(t) + \lambda$$

δ, ε, λ being arbitrary constants.

In a similar manner we obtain equivalent surfaces. A surface is obtained in parametric form by three functions.

$$\alpha = \alpha(s, t)$$
$$\beta = \beta(s, t) \tag{2.44}$$
$$\theta = \theta(s, t)$$

which determine a two-parameter manifold of line elements

$$d\alpha = \alpha_s\, ds + \alpha_t\, dt$$
$$d\beta = \beta_s\, ds + \beta_t\, dt \tag{2.45}$$
$$d\theta = \theta_s\, ds + \theta_t\, dt$$

Another surface α', β', θ' is equivalent to this surface if its line elements $d\alpha'$, $d\beta'$, $d\theta'$ can be coordinated to the original line elements such that

$$d\alpha' = d\alpha$$
$$d\beta' = d\beta$$
$$d\theta' = d\theta$$

This leads immediately to the result that an equivalent surface must have the form

$$\alpha' = \alpha(s, t) + \delta$$
$$\beta' = \beta(s, t) + \varepsilon \tag{2.46}$$
$$\theta' = \theta(s, t) + \lambda$$

where δ, ε, λ are arbitrary constants.

We may integrate the above results from a more general point of view by interpreting the relations

$$\alpha' = \alpha + \delta$$

$$\beta' = \beta + \varepsilon \qquad (2.47)$$

$$\theta' = \theta + \lambda$$

as a group of transformations of the angular α, β, θ space. In fact, they represent the simplest type of transformations of this space, namely, translatory shifts. If these transformations, however, are formulated in the Cartesian x, y, z coordinates, an interesting and in no way trivial group of transformations of the physical x, y, z space is obtained. A general investigation of these transformations should give us many general results interesting for binocular vision. Any configuration submitted to such transformation will be seen by the same sequence of retinal images before and after the transformation. For this reason we shall call the transformations (2.47) iseikonic transformations. Mathematically we recognize immediately a characteristic feature of these transformations. A cone α = const. through the right eye is transformed into another cone through this eye. Similarly a cone β = const. through the left eye into another such cone. These two basic sets of cones thus are transformed without distorting but only interchanging the individual cones.

If we prefer the use of the modified coordinates γ, φ, θ we can express iseikonic transformations by the relations

$$\gamma' = \gamma + \tau$$

$$\varphi' = \varphi + \sigma \qquad (2.48)$$

$$\theta' = \theta + \lambda$$

where τ, σ, λ are arbitrary constants.

Two configurations of objects such that the one can be transformed into the other by an iseikonic transformation are called equivalent configurations.

2.5. Significance of iseikonic transformations. We ask the question: Are two equivalent configurations of objects indistinguishable in binocular vision? Indeed, both can be observed by identical sequences of images on the retinae.

With regard to the above question, we can have two extremely opposite points of view. If we subscribe to the "projection theory" that our eyes are a kind of measuring device for the angular coordinates γ, φ, θ and that the results of the measurements are directly transformed by our mind into space sensations, then equivalent configurations are sensed as different. However, if we believe the actual values of γ, φ, θ (i.e., the convergence of the eyes, the relation of the optical axes to the median and horizontal plane) are of no consequence and the sequence of retinal images provides the only stimulus for sensations, then equivalent configurations are absolutely indistinguishable, even in binocular vision. The observer depends upon intellectual clues such as perspective to choose the one or the other physical realization of equivalent configurations.

The second hypothesis can be supported by actual experiments. Ames (Dartmouth Eye Institute) has shown experimentally that to a given surface a whole family of surfaces belongs which may be interpreted as the original surface if suitable perspective patterns are drawn on the "wrong" surface. For example, to a rectangular room there belongs a family of non-rectangular rooms which are binocularly indistinguishable from the original room. If the walls are provided with certain distorted windows which correspond to rectangular windows in the original room, then the visual sensation of the distorted room is that of a rectangular room. Indeed, the observer knows that windows are rectangular and thus in his sensation he chooses <u>the</u> physical realization of the impinging pattern which fits this notion.

It is possible to construct mathematically such a family of distorted rooms with the aid of iseikonic transformations. The result is a set of distorted rooms <u>equivalent</u> to the original rectangular room. Rooms which have been constructed on this basis, indeed, give, at least approximately, the above-described effect. For this reason we shall derive in the two following sections the mathematical equations for distorted rooms equivalent to a given rectangular room.

The experiments with distorted rooms seem thus to be fully explained by the above hypothesis that the actual values of φ, θ and especially the value of the convergence γ are insignificant for the visual sensations. However, this theory leads us into difficulties when we try to understand the psychological fact of judging size independent of localization. Indeed, two line elements (dx_1, dy_1, dz_1) and (dx_2, dy_2, dz_2) can have the <u>same</u> bipolar characteristics $d\gamma$, $d\varphi$, $d\theta$, and still be of entirely different physical size. If these "local signs" $d\gamma$, $d\varphi$, $d\theta$ are the only basis for visual sensations, then it is hard to understand how we can judge the difference in actual size with such remarkable accuracy. Is this judgment obtained purely by former experience, or is it at least partly an element of direct visual sensation? In other words, is judgment of size only the result of training, or can we assume that it has developed from a seed which is an immutable part of primitive sensation of space?

The fact that two line elements can have the same impinging characteristics $d\gamma$, $d\varphi$, $d\theta$, but different apparent linear size, forces us to reconsider the significance of the absolute values γ, φ, θ, especially of the convergence γ. We shall, in §4, <u>relate the apparent size ds of a line element</u> $d\gamma$, $d\varphi$, $d\theta$ <u>to the bipolar parallax</u>, γ, <u>i.e., to the convergence of the eyes</u>. We shall not attempt to explain this relation of size estimation to convergence physiologically, but shall consider it as a hypothesis necessary for the solution of our problem: To establish a metric for the manifold of visual sensations.

By the introduction of the convergence γ as a significant element of binocular vision, we have to conclude that <u>equivalent configurations can not be truly indistinguishable</u>. However, we shall see that our postulated relation of apparent size and convergence does not necessarily mean absolute localization in space. On the contrary, the special functional relation of both which we shall establish in §6 allows an even greater group of configurations metrically equivalent to a given configuration. This means that a group of transformations of the space exists which transforms a given configuration into other ones with identical binocular characteristics. These transformations we shall call <u>rigid transformations</u>, and two configurations of this type, <u>congruent configurations</u>. Instead of a three-parameter group as the iseikonic transformations, we shall find a six-parameter group of rigid transformations. Ames's postulate of the existence of a group of surfaces indistinguishable from a given surface thus is even more guaranteed, <u>if we introduce into binocular vision the convergence</u> γ <u>as a significant factor</u>.

In § 8 we shall derive a set of distorted rooms with walls congruent to the plane walls of a rectangular room. The result will be, that these congruent rooms are nearly identical with the equivalent rooms to be derived in § 2.6 and § 2.7, though obtained from entirely different mathematical principles. We thus may consider equivalent rooms as a good approximation of congruent rooms. However, we shall see that the differences between both types are great enough to be easily observable. To compare the impressions of both types of rooms as to the conviction of seeing an ordinary rectangular room can be considered a direct test of the two theories.

2.6. The distorted room equivalent to a rectangular room. We assume that the walls of an originally rectangular rooms are given by the planes

$$x = x_o$$

$$y = \pm y_o \qquad (2.61)$$

$$z = \pm z_o$$

We consider a special iseikonic transformations of the space represented by the relations

$$\gamma' = \gamma + \tau$$

$$\varphi' = \varphi \qquad (2.62)$$

$$\theta' = \theta$$

where τ is an arbitrary constant. The plane walls of the original room are transformed into equivalent curved surfaces. Any pattern drawn on these plane surfaces is transformed into an equivalent pattern on the curved surfaces and will be seen by the same binocular characteristics $d\gamma' = d\gamma$; $d\varphi' = d\varphi$; $d\theta' = d\theta$ as the original pattern. The only difference is the absolute value of the convergence of the eyes. If a special part of the plane walls is observed with an angle of convergence γ, then the corresponding part on the equivalent curved wall is seen with an angle $\gamma' = \gamma + \tau$. The assumption that the convergence of the lines of sight is immaterial for binocular space sensation then leads to the consequence that the observer sees the equivalent curved walls as plane if a suitable pattern on the walls induces him to this interpretation.

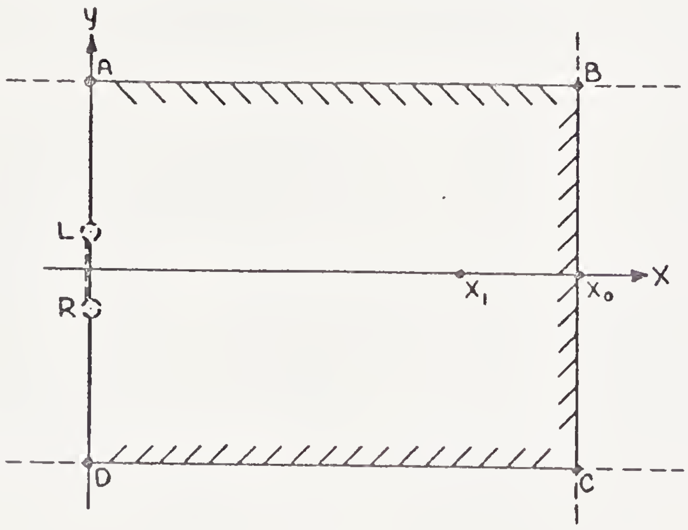


Fig. 13

Before we derive analytic expressions for the curved walls, let us consider a simple topological method which allows us easily to determine the general shape of these walls. We determine for this purpose the curves in the γ, φ-plane which correspond to the rectangular cross section of the original room with the horizontal plane. The front wall $x = x_o$ becomes a curve symmetrical to the γ-axis which reaches infinity ($\gamma = 0$) at the points $\varphi = \pm \pi/2$. The side wall $y = y_o$ may be considered as a line connecting a point $y_o > 1$ of the y-axis with a point at ∞ for which $\varphi = 0$. Consequently, its

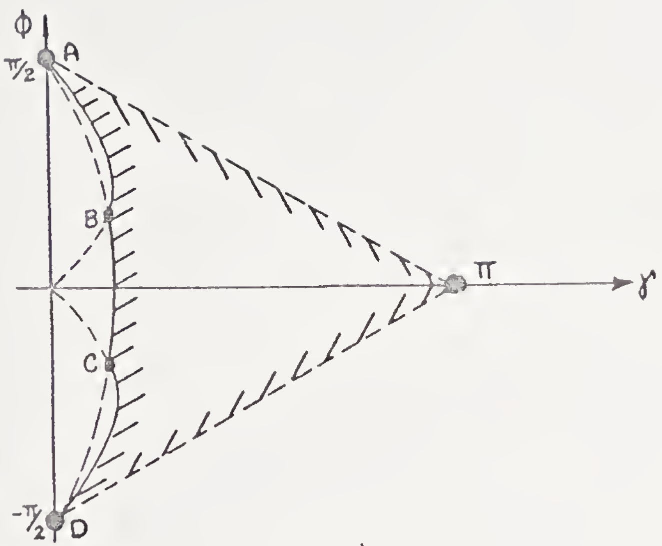


Fig. 14

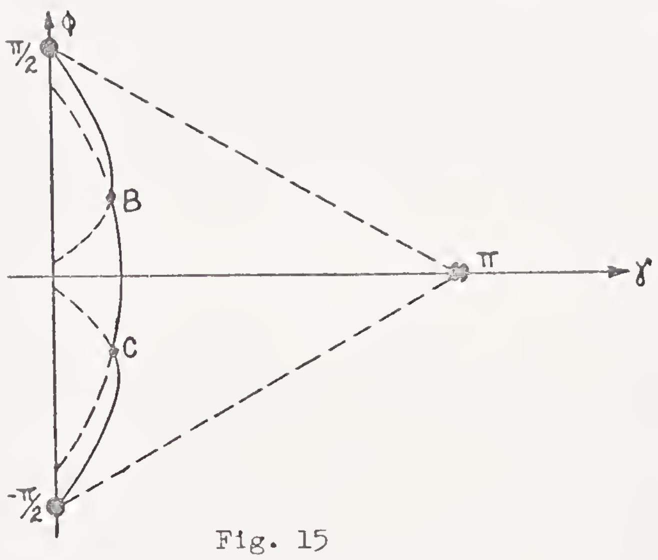


Fig. 15

image in the γ, φ-plane is a curve from the point (0, π/2) to the point (0, 0). This consideration leads to Fig. 14, where the shaded region corresponds to the interior of the rectangle of Fig. 13.

We now submit the γ, φ-plane to the transformation $\gamma' = \gamma + \tau$. Let us first assume that $\tau < 0$. The curve ABCD is shifted to the left by this transformation and thus is located in the basic triangle as shown in Fig. 15. The front wall extends from B to $\infty(\gamma = 0)$ and reaches this line at a value of $\varphi_\infty < \pi/2$. Consequently it must be a curve of hyperbolic shape in the x,y-plane, a curve which is symmetric to the x-axis and approaches ∞ asymptotically with the angles $\pm\varphi_\infty$ where $\varphi_\infty < \pi/2$. The side walls go from B (or C) directly to $\gamma = 0$, and reach it at φ-values smaller than π/2 but greater than the values φ_∞ of the front wall. This means that the side walls also give hyperbola-shaped curves which

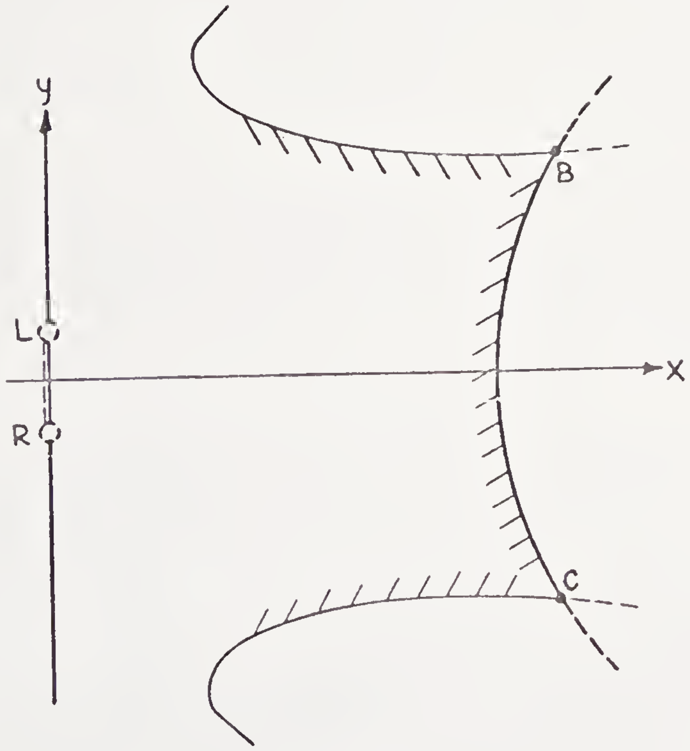


Fig. 16

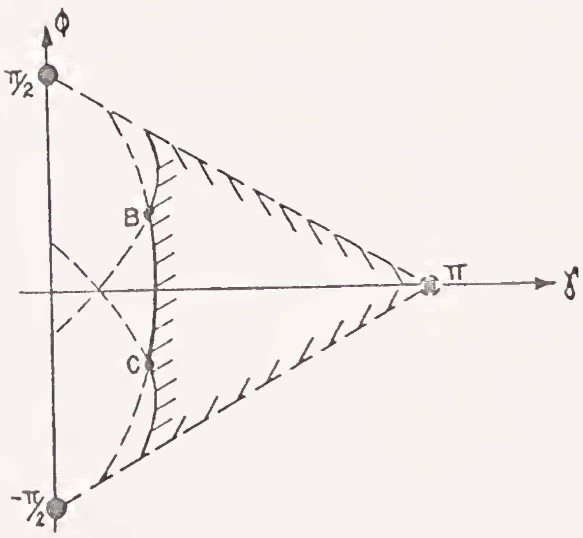


Fig. 17

approach ∞ with an asymptote steeper than the asymptotes of the front wall. From these considerations it follows that the equivalent room must have a cross section with the horizontal plane as illustrated in Fig. 16. The analytic investigation shows that the curves are in fact true hyperbolae.

We consider next the case $\tau > 0$. Now the curve ABCD in Fig. 14 is shifted towards the ri[illegible] and located in the basic triangle as shown in Fig. 17. The extension of the front wall beyond B goes directly to the upper or lower boundary line and thus in the x,y-plane to one of the eyes. It goes through the other eye if extended beyond C. Thus it must be an elliptically shaped curve symmetric to the x-axis and passing through the eyes. The upper side wall goes to the upper boundary line and thus to the left eye in the x,y-plane. Its extension beyond B reaches $\gamma = 0$ at a negative value of $\varphi_\infty > -\pi/2$. This leads to the equivalent room whose cross section with the horizontal plane is shown in Fig. 18

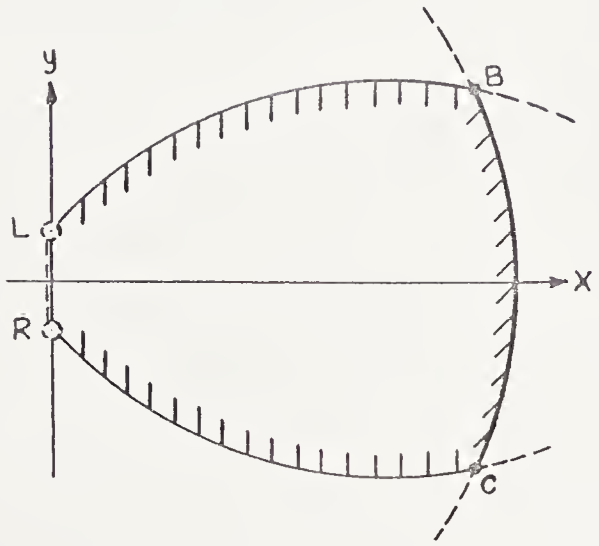


Fig. 18

The analytical treatment will show that the curved walls really intersect the horizontal plane in an ellipse and two hyperbolae.

2.7. The distorted room equivalent to a rectangular room. (Analytical derivation) Since the dimensions of the room are quite large compared with the distances of the eyes, we shall use the simplified relations (2.31) for our purpose.

From these formulae it follows that the transformation

$$\gamma' = \gamma + \tau$$
$$\varphi' = \varphi \tag{2.71}$$
$$\theta' = \theta$$

in the γ, φ, θ space may be written in Cartesian coordinates as follows:

$$\frac{\sqrt{x'^2 + z'^2}}{x'^2 + y'^2 + z'^2} = \frac{\sqrt{x^2 + z^2}}{x^2 + y^2 + z^2} + \frac{\tau}{2}$$
$$\frac{y'}{\sqrt{x'^2 + x'^2}} = \frac{y}{\sqrt{x^2 + z^2}} \tag{2.72}$$
$$\frac{z'}{x'} = \frac{z}{x}$$

The last two equations may be replaced by

$$\frac{y'}{x'} = \frac{y}{x}$$
$$\frac{z'}{x'} = \frac{z}{x} \tag{2.73}$$

We now introduce $y' = \frac{y}{x} x'$ and $z' = \frac{z}{x} x'$ in the first equation, and obtain

$$\frac{x}{x'} = 1 + \frac{\tau}{2} \frac{x^2 + y^2 + z^2}{\sqrt{x^2 + z^2}} \tag{2.74}$$

Combined with (2.73) this leads to the formulae

$$x' = \frac{x}{1 + \frac{\tau}{2} \frac{x^2 + y^2 + z^2}{\sqrt{x^2 + z^2}}}$$

$$y' = \frac{y}{1 + \frac{\tau}{2} \frac{x^2 + y^2 + z^2}{\sqrt{x^2 + z^2}}} \tag{2.75}$$

$$z' = \frac{z}{1 + \frac{\tau}{2} \frac{x^2 + y^2 + z^2}{\sqrt{x^2 + z^2}}}$$

between Cartesian coordinates.

The inversion of these formulae is simply obtained by replacing τ by $-\tau$, i.e., we have

$$x = \frac{x'}{1 - \frac{\tau}{2} \frac{x'^2 + y'^2 + z'^2}{\sqrt{x'^2 + z'^2}}}$$

$$y = \frac{y'}{1 - \frac{\tau}{2} \frac{x'^2 + y'^2 + z'^2}{\sqrt{x'^2 + z'^2}}} \tag{2.76}$$

$$z = \frac{z'}{1 - \frac{\tau}{2} \frac{x'^2 + y'^2 + z'^2}{\sqrt{x'^2 + z'^2}}}$$

The last equations give us immediately the equations for the curved walls of our distorted room. Indeed, since the plane walls of the original room are given by the planes $x = x_o$, $y = y_o$, $z = z_o$, where x_o, y_o, z_o are constants, it follows by (2.76) that (we use x, y, z instead of x', y', z') the equivalent walls are determined by the surfaces

$$x = x_0 \left(1 - \frac{\tau}{2} \frac{x^2 + y^2 + z^2}{\sqrt{x^2 + z^2}}\right)$$

$$y = y_0 \left(1 - \frac{\tau}{2} \frac{x^2 + y^2 + z^2}{\sqrt{x^2 + z^2}}\right) \qquad (2.77)$$

$$z = z_0 \left(1 - \frac{\tau}{2} \frac{x^2 + y^2 + z^2}{\sqrt{x^2 + z^2}}\right)$$

with τ as arbitrary parameter.

The cross sections with the horizontal plane ($z = 0$) are given by the curves:

$$x = x_0 \left(1 - \frac{\tau}{2} \frac{x^2 + y^2}{x}\right)$$

$$y = y_0 \left(1 - \frac{\tau}{2} \frac{x^2 + y^2}{x}\right) \qquad (2.78)$$

One recognizes immediately that these curves are conic sections, namely, hyperbolae if $\tau < 0$, and ellipses (front wall) and hyperbolae (side walls) if $\tau > 0$ in agreement with our former results.

We remark in general that each plane of elevation $z = x \tan \theta$ is intersected in conic sections by our surfaces. Furthermore: The front wall intersects the median plane $y = 0$ in a conic section, namely

$$x = x_0 \left(1 - \frac{\tau}{2} \sqrt{x^2 + z^2}\right) \qquad (2.79)$$

Finally we remark that the side walls are surfaces of revolution

$$y = y_0 \left(1 - \frac{\tau}{2} \frac{\rho^2 + y^2}{\rho}\right) ; \qquad \rho = \sqrt{x^2 + z^2} \qquad (2.791)$$

with the y-axis as axis of revolution.

2.8. Angular coordinates for observation with head movements. By movements of the head we are in a position to view the neighborhood of any point, not only in direct vision but also in symmetrical convergence of the eyes. We assume that the head rotates about a center of rotation so that the eyes are moved on a sphere around this center. The base line of the eyes loses its significance for directional orientation. It is replaced by a reference line given by the position of our shoulders, and we can assume that this reference line remains in our consciousness if we move our head. Similarly we are conscious of the position of the horizontal plane normal to the direction of the gravitational force.

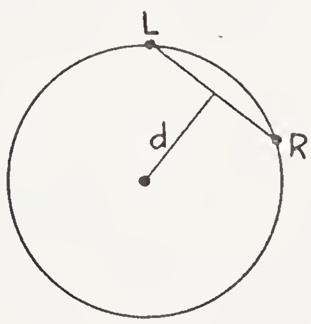


Fig. 19

For this manner of observation we introduce suitable systems of Cartesian and angular coordinates.

The x,y-plane shall be the horizontal plane and the y-axis shall coincide with the direction of the shoulders.

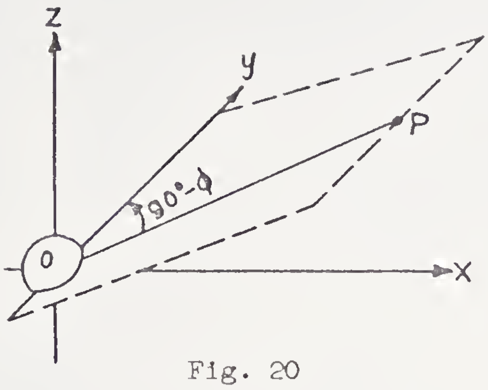


Fig. 20

We place the origin of the coordinate system at the center of rotation of our head.

We replace this Cartesian system by an angular system as follows. We determine the plane of elevation of a point by the plane through P and the y-axis. Let θ^* be its angle with the horizontal plane. We draw then in this plane the line which connects the origin O with P. We characterize this line by its angle $(90 - \varphi^*)$ with the y-axis. The angles of elevation θ^* and the lateral angle φ^* then correspond to longitude and latitude on a sphere with two poles on the y-axis. The distance OP of a point from the origin finally is characterized by the parallax γ^* at P with the eyes as basis. According to our assumption, we consider the lines of sight RP and LP as always symmetrical to the radius vector OP.

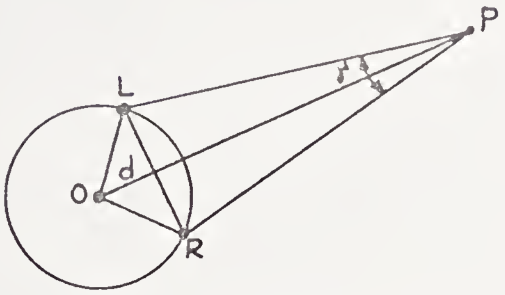


Fig. 20a

The relations of the Cartesian coordinates x, y, z to the angular coordinates are simple. Let d be the distance of the base line RL from the center of rotation O. Then

$$OP = d + \cot \gamma^*/2$$

and hence

$$x = (d + \cot \gamma^*/2) \cos \varphi^* \cos \theta^*$$

$$y = (d + \cot \gamma^*/2) \sin \varphi^* \qquad (2.81)$$

$$z = (d + \cot \gamma^*/2) \cos \varphi^* \sin \theta^*$$

A line element (dx, dy. dz) attached to a point P, can be expressed in terms of our angular differentials $d\gamma^*$, $d\varphi^*$, $d\theta^*$. These differentials determine at the same time the significant characteristics of the retinal images of the line element, namely, the disparity $d\gamma^*$ interpreted as depth and the lateral and vertical extensions $d\varphi^*$ and $d\theta^*$.

We may assume that the base line RL of our eyes remains approximately horizontal if we move the head according to a habit established in the past. Thus we may call the disparity $d\gamma^*$ also in this case the horizontal disparity.

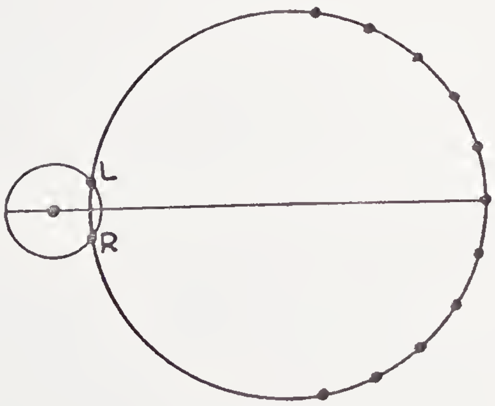


Fig. 21

The assumption of a base line which remains horizontal is, however, not essential for our theory. The existence of a well-established habit we can certainly expect, since otherwise repeated fixation of the same line element would result in erratic judgment of depth in accordance with different disparities $d\gamma^*$.

If a given configuration of objects is observed with the head in fixed position and then with moving head, then it is not self-evident that the two interpretations are identical. Indeed, one can easily demonstrate that this is not the case. For this purpose we construct a number of marks (for example, pins) arranged at equal distances on a Vieth-Muller circle through the eyes. If these pins are observed with the head in fixed position, they give the impression of being arranged on a circle with the observer at its center. This impression does not remain if the pins are viewed with moving head.

On the other hand, let us consider two configurations of points which are physically entirely different, namely

$$x_1^* = (d + \cot \gamma_1/2) \cos \varphi_1 \cos \theta_1$$
$$y_1^* = (d + \cot \gamma_1/2) \sin \varphi_1 \qquad (2.82)$$
$$z_1^* = (d + \cot \gamma_1/2) \cos \varphi_1 \sin \theta_1$$

and

$$x_1 = \frac{\cos 2\varphi_1 + \cos \gamma_1}{\sin \gamma_1} \cos \theta_1$$
$$y_1 = \frac{\sin 2\varphi_1}{\sin \gamma_1} \qquad (2.83)$$
$$z_1 = \frac{\cos 2\varphi_1 + \cos \gamma_1}{\sin \gamma_1} \sin \theta_1$$

where the quantities γ_1, φ_1, θ_1 are in both configurations the same.

We observe the configuration (2.82) with moving head and symmetrical convergence of the eyes, but the configuration (2.83) with the head in fixed position and asymmetrical convergence of the eyes.

We shall adhere in the following to the hypothesis that the observer is led to an identical sensation in both configurations and that he interprets both configurations as identical.

This hypothesis may be tested easily enough by experiment. For example, we can construct a network of wires obtained by rotating the Vieth-Muller circle in Fig. 21 around the base line of the eyes. The result is a network on a torus surface as illustrated in Fig. 22. According to our hypothesis this network would appear, if observed with fixed head, as a spherical network with the observer as center. The wires indicate the circles of longitude and latitude on this sphere. The same sensation would be obtained if an actual spherical network (Fig. 23) is observed with moving head so that the individual parts of the network are seen in symmetrical convergence.

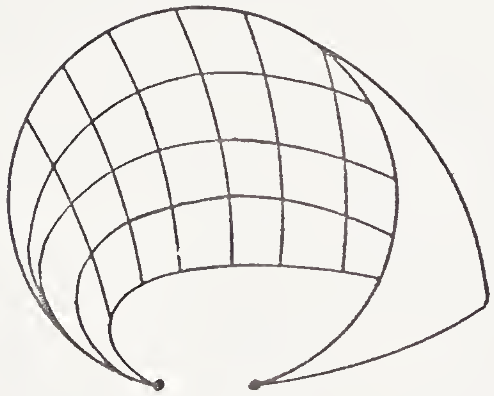


Fig. 22

The above-formulated principle seems to contradict, at first sight, our common experience. It certainly implies that from the same physical configuration of objects different sensations may be obtained, depending upon the manner of observation: with head in fixed position or with moving head. Moreover, since head and eye movements may be combined in an infinity of variations, we must expect that the same physical configuration can lead the observer to an infinite variety of sensations. Still,

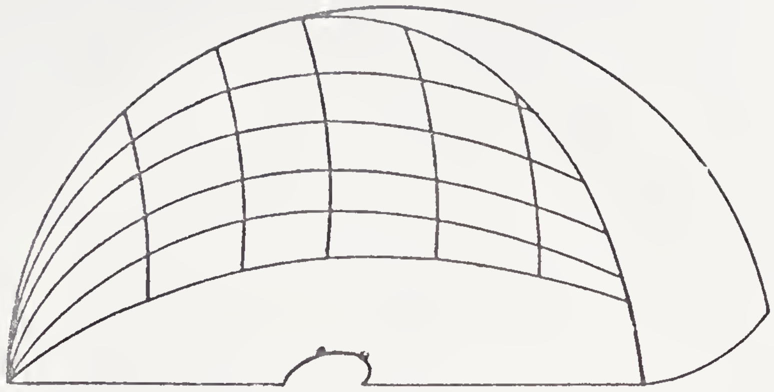


Fig. 23

we are convinced that all these different views are associated with the same and unchanged environment.

Is this conviction the result only of experience, i.e., correlated to a judgment of probability based on repitition of similar observations in the past? This may be the explanation, but we are by no means forced to consider this the only possible explanation. It is possible that all these different interpretations have an invariant element in common, and that this invariant element is the criterion which gives us the conviction of seeing the same external configuration. We may compare this with constructing two different maps of the same configuration by using two different principles of mapping. In fact, we may describe an interpretation of a configuration of objects as a coordination or a mapping of sensations to the Euclidean screen of our intuitive mathematical thinking. To formulate mathematically the invariant element which both Euclidean maps have in common--namely, the same non-Euclidean metric relations--will be one of the subjects of the next sections.

Section 3

CHARACTERIZATION OF A METRIC BY QUADRATIC DIFFERENTIALS

3.1. The <u>metric of a manifold of points</u> is defined as a rule for determining the size of objects in this manifold. It is a significant mathematical fact that it is sufficient to formulate this rule for small objects, indeed, "infinitesimally" small objects. We illustrate this by the example of the Euclidean plane. We introduce Cartesian coordinates x, y and consider two points $P_1 = (x, y)$ and $P_2 = (x + dx, y + dy)$ where dx and dy are the differential increments of the coordinates x, y. The line element ds connecting the two points then is characterized by the differentials (dx, dy), and its size ds is given by Pythagoras' theorem:

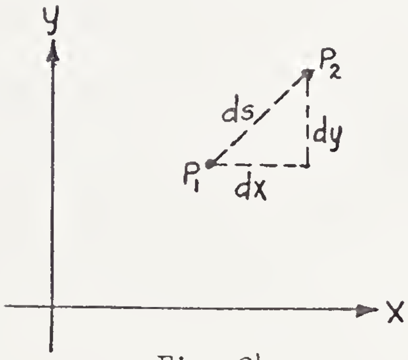


Fig. 24

$$ds = \sqrt{dx^2 + dy^2} \tag{3.11}$$

We also may say that the size is determined by the quadratic differential

$$ds^2 = dx^2 + dy^2 \tag{3.12}$$

which again is an analytical formulation of Pythagoras' theorem. This quadratic differential (3.12) is the mathematical expression for the metric of the Euclidean plane. By purely analytical methods one is in the position to derive the theorems of Euclidean geometry from the fundamental differential (3.12). It is clear that this fact is by no means self evident, since it means that a metric (3.12) referring to points which are infinitely close also determines metric relations of points which are far apart. We may consider the differential (3.12) as an inconspicuous seed out of which we may develop the whole organism of Euclidean geometry with all its variety of relations.

To illustrate this, we show how intimately the measurement of angles is related to the measurement of size through the quadratic differential (3.12). Let us consider two line elements P_1P_2 and P_1P_3 attached to the point P_1. We denote their differential coordinates by (dx, dy) and (δx, δy). Then it follows that the angle included by the two line elements is given by the formula:

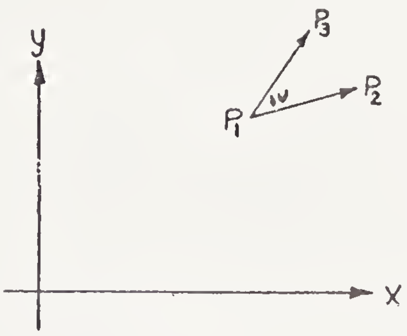


Fig. 25

$$\cos \omega = \frac{dx \enspace \delta x + dy \enspace \delta y}{\sqrt{dx^2 + dy^2} \cdot \sqrt{\delta x^2 + \delta y^2}} \tag{3.13}$$

The expression $dx \cdot \delta x + dy \cdot \delta y$ is called the mixed quadratic differential associated with (3.12). Hence the angular metric follows directly from the original metric of size.

The step from the infinitely small to the finite is represented by the problem of <u>geodesic lines</u>, i.e., by the problem of finding among all curves connecting two given points P_0 and P_1 the special curve which has the shortest length. The length of such a connecting curve is obtained by summation of the size of its line element, i.e., by the integral

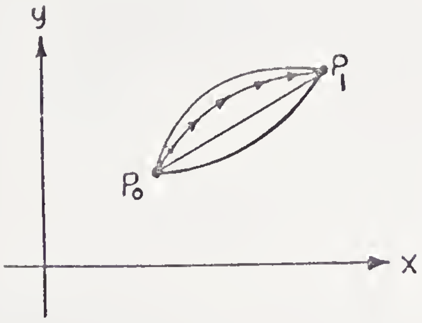


Fig. 26

$$S = \int \sqrt{dx^2 + dy^2} \tag{3.14}$$

and our problem means to determine among all curves connecting P_0 and P_1 one for which this integral assumes a minimum value. This problem

$$\int \sqrt{dx^2 + dy^2} = \text{Minimum} \tag{3.15}$$

can be solved by purely analytical methods, and leads, of course, in our case, to the Euclidean straight line between P_0 and P_1.

3.2. Let us next consider a metric manifold of two dimensions in general. We characterize the points P of the manifold by coordinates x, y in a suitably chosen coordinate system. The size ds of a line element connecting two neighboring points (x, y) and (x + dx, y + dy) is given by the general quadratic differential

$$ds^2 = Edx^2 + 2Fdxdy + Gdy^2 \tag{3.21}$$

where E(x, y), F(x, y), G(x, y) are given functions of x, y.

In the above example of the Euclidean plane we have E = G = 1, F = 0. If the manifold is a sphere of radius one, we have E = 1, F = 0, G = $\sin^2 x$ so that

$$ds^2 = dx^2 + \sin^2 x \cdot dy^2 \tag{3.22}$$

The coordinates x, y in this case are related to the latitude φ and the longitude θ on the sphere, namely:

$$\begin{aligned} x &= \pi/2 - \varphi \\ y &= \theta \end{aligned} \tag{3.23}$$

The geometry in our manifold can be developed by purely analytical methods from the differential (3.21). For example, the angle ω included by two line elements (dx, dy) and (δx, δy) attached to the same point (Fig. 25) is now given by the expression

$$\cos \omega = \frac{Edx\delta x + F(dx\delta y + \delta x dy) + Gdy\delta y}{\sqrt{Edx^2 + 2Fdxdy + Gdy^2} \cdot \sqrt{E\delta x^2 + 2F\delta x\delta y + G\delta y^2}} \tag{3.24}$$

and thus directly related to the basic quadratic differential (3.21). From the infinitely small we go to the finite by the problem of geodesics: To find, among all curves connecting two points P_0, P_1, the one for which the length

$$S = \int \sqrt{Edx^2 + 2Fdxdy + Gdy^2} \tag{3.25}$$

assumes a minimum value.

These curves are called Geodesics; they have a similar significance in our general metric manifold as the straight lines in the Euclidean plane.

We mention another property of Geodesics which is significant for a problem in space perception treated later on. It is possible to relate the concept of parallelism of line elements which are not attached to the same point to the quadratic differential (3.21). Consider a line element attached to a point P_0. We wish to transfer this line element from P_0 to another point P_1 along a given path C, but in such a way that it does not change its direction. The analytic conditions for this parallelism of two line elements (Parallelism of Levi-Civita) can be shown to be directly related to the basic quadratic differential (3.21), i.e., to the metric of size. Without formulating this condition here, we only mention the fact that the line elements of geodesic lines are always parallel. This means the result of constructing a curve by attaching line elements to each other without change of direction is again a geodesic line, i.e., the shortest connection between two of its points (Fig. 28). This is, of course, quite clear in the Euclidean plane, but it is true also in non-Euclidean geometries characterized by other metric differentials.

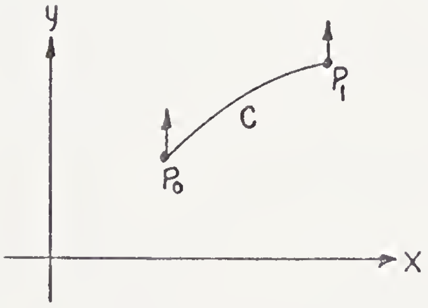


Fig. 27

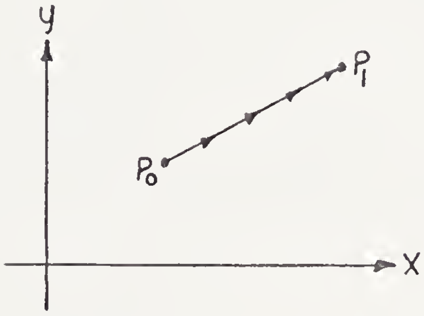


Fig. 28

3.3. The above considerations illustrate the basic significance of the metric differential (3.21) for the geometry of a manifold of points. We may call it the "code script" of this geometry. It allows us to replace geometrical relations by relations of numbers and geometrical methods by analytical methods. And in the end this is based upon the Cartesian idea of coordinating geometrical elements, namely, points, to numbers by introducing a coordinate system. Obviously, a certain arbitrariness is revealed here. The coordination of numbers to points can be done in a great variety of ways. Let us illustrate this situation again by the Euclidean plane. If, instead of Cartesian coordinates, we choose, for example, polar coordinates, then the same point of the plane receives an entirely different pair of numbers. There seems to be an almost unlimited range of different coordinate systems which we may introduce in our Euclidean plane. Mathematically this is expressed by the fact that we may introduce new coordinates ξ, η by a transformation

$$x = f(\xi, \eta) \tag{3.31}$$

$$y = g(\xi, \eta)$$

of the original Cartesian coordinates x, y into new coordinates ξ, η. (The functions f and g are completely arbitrary.) A point P of the Euclidean plane then is located in a system of curved coordinated lines instead of a system of rectilinear lines (Fig. 29). By submitting the quadratic differential (3.12) to the transformation (3.31) we obtain again a quadratic differential, but of a more general type

$$ds^2 = Ed\xi^2 + 2Fd\xi d\eta + Gd\eta^2 \tag{3.32}$$

where the functions $E(\xi, \eta)$, $F(\xi, \eta)$, $G(\xi, \eta)$ are given by

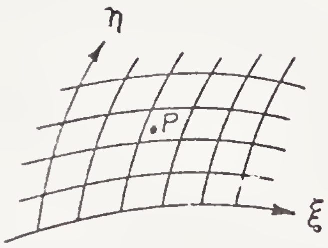


Fig. 29

$$E = f_\xi^{\;2} + g_\xi^{\;2} \qquad (3.33)$$

$$F = f_\xi f_\eta + g_\xi g_\eta$$

$$G = f_\eta^{\;2} + g_\eta$$

This shows that the differential (3.12) may appear in an infinite number of mathematical variations (3.32), depending upon the coordinate system chosen for assigning numbers to the points of the Euclidean plane. In polar coordinates $R = \xi$, $\varphi = \eta$, for example, we have

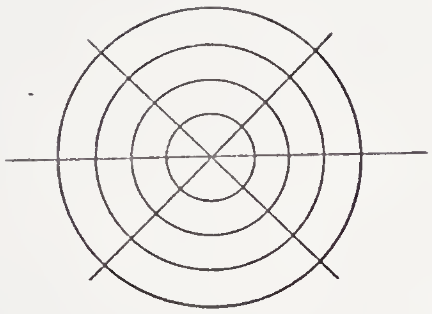


Fig. 30

$$ds^2 = d\xi^2 + \xi^2\, d\eta^2$$

By choosing the bipolar coordinates explained in 2.3, we have, letting

$$\gamma = \xi,\ \varphi = \eta:$$

$$ds^2 = 4\left(\frac{\cos^2\eta}{\xi^4}\, d\xi^2 + \frac{2\sin\eta\cos\eta}{\xi^3}\ .\ d\xi . d\eta + \frac{1}{\xi^2}\, d\eta^2\right)$$

Obviously the relations of the Euclidean geometry itself cannot depend upon the special choice of the coordinate system and thus upon the special analytic form of the differential (3.12). Hence we conclude that all these variations (3.32) must have an element in common which expresses the fact that the geometry developed from any of them is Euclidean. This means, for example, that the angle ω between two line elements characterized by its differentials in a ξ, η system: $d\xi$, $d\eta$; $\delta\xi$, $\delta\eta$, and computed by the formula:

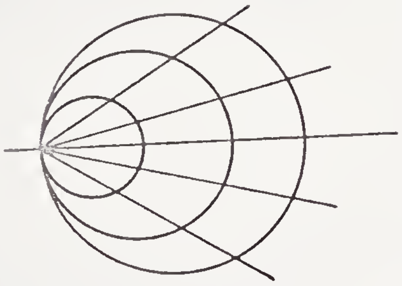


Fig. 31

$$\cos\omega = \frac{Ed\xi\delta\xi + F(d\xi\delta\eta + d\eta\delta\xi) + Gd\eta\delta\eta}{\sqrt{Ed\xi^2 + 2Fd\xi d\eta + Gd\eta^2}\ \sqrt{E\delta\xi^2 + 2F\delta\xi\delta\eta + G\delta\eta^2}} \qquad (3.34)$$

based upon the quadratic differential (3.32) must be the same as the angle found by (3.13).

Furthermore, the geodesic lines of the differential (3.32), i.e., the curves for which the integral

$$\int\sqrt{Ed\xi^2 + 2Fd\xi d\eta + Gd\eta^2}$$

assumes a minimum, must be given by functions which in the curved coordinate system lead to the straight lines of the Euclidean plane.

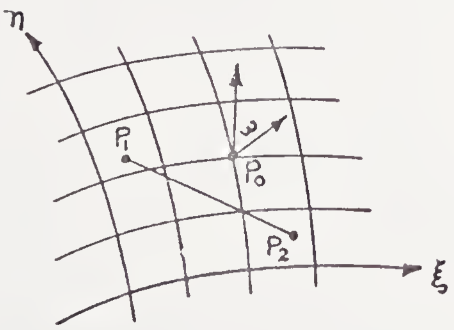


Fig. 32

The property which all the quadratic differentials (3.32) have in common obviously is that they are obtained from (3.12) by a transformation of the coordinates. Vice versa, we may say that, by a suitable transformation of the coordinates, the line element (3.32) can be transformed into the normal form (3.12).

3.4. The above consideration leads us to the question: Suppose a geometrical manifold is given, and, after choosing a coordinate system in it, we obtain a metric differential

$$ds^2 = Ed\xi^2 + 2Fd\xi d\eta + Gd\eta^2 \tag{3.41}$$

where E, F, G, are known functions of ξ and η. Is it then always possible to find a transformation

$$x = f(\xi, \eta) \tag{3.42}$$

$$y = g(\xi, \eta)$$

of the coordinates such that in the new coordinate system the line element (3.41) assumes the Euclidean form $ds^2 = dx^2 + dy^2$? Since the transformation (3.42) is the analytical expression for drawing in the Euclidean x,y-plane a map of the given metric manifold (3.41), it then would be possible to obtain a plane isometric map of the given two-dimensional manifold. In other words: Euclidean measurements of size on the plane map would give the size of objects in the original manifold (3.41).

The answer to our question, however, is negative: Unless the functions E, F, G satisfy a certain mathematical condition, the desired transformation is impossible. Indeed, if the functions E, F, G are given, then the equations (3.33) represent a system of three partial differential equations of first order for two unknown functions $f(\xi, \eta)$ and $g(\xi, \eta)$. The system, obviously, is overdetermined and in general will have no solution. A solution can be expected only if the functions E, F, G satisfy a certain mathematical condition. If we interpret the geometry associated with a quadratic differential

$$ds^2 = Edx^2 + 2Fdxdy + Gdy^2 \tag{3.43}$$

as the geometry on a curved surface in the three-dimensional Euclidean space--as we always can--then the formulation of the above condition is given by Gauss' _theorema egregium_: _The Gaussian Curvature K of a surface can be derived from the coefficients E, F, G of the metric differential (3.43)_. The curvature K of the Euclidean plane is zero, and, as a geometrical quantity, must be zero for any choice of the coordinate system. This means, for a differential (3.43) which has been obtained from the Euclidean differential $ds^2 = dx^2 + dy^2$ by transformation of the coordinates, that the result of introducing the functions E, F, G into the Gaussian expression for K must be identically zero. Also the reverse is true: If the functions E, F, G satisfy the condition $K \equiv 0$, then a transformation (3.42) can be found which transforms the differential (3.43) into the normal Euclidean form $ds^2 = dx^2 + dy^2$. A line element (3.43) of this special type is called a Euclidean line element.

It is not difficult to give examples of non-Euclidean line elements. The geometry on a sphere of radius one is characterized by the metric differential

$$ds^2 = dx^2 + \sin^2 x\, dy^2 \tag{3.44}$$

where $\varphi = \pi/2 - x$ is the latitude and $\theta = y$ the longitude on the sphere. The Gaussian curvature K has the constant value $K = +1$, and it is impossible to transform this line element into the Euclidean form (3.12) for which $K = 0$. The non-Euclidean geometry developed from (3.44) is called the _elliptic geometry_.

Another example is given by the quadratic differential

$$ds^2 = dx^2 + \sinh^2 x\, dy^2 \tag{3.45}$$

and is distinguished by the fact that its Gaussian curvature has the constant value $K = -1$. The corresponding special non-Euclidean geometry is the hyperbolic geometry.

The three geometries based on the line elements (3.12), (3.44) and (3.45) are called geometries of constant curvature, since the Gaussian curvature K is constant. We may expect that these geometries are of special interest as compared with the great variety of other non-Euclidean geometries in which K is variable from point to point.

By introducing other coordinates in the manifolds of constant curvature we obtain a great variety of different forms of the associated quadratic differentials. We mention Riemann's normal form

$$ds^2 = \frac{d\xi^2 + d\eta^2}{\left[1 + \frac{1}{4}K(\xi^2 + \eta^2)\right]^2} \tag{3.46}$$

by which these geometries can be represented simultaneously, K being the constant Gaussian curvature. For $K = 1, 0, -1$ we obtain the elliptic, Euclidean, and hyperbolic geometry respectively.

3.5. Geometrical manifolds of three dimensions may be treated in a similar manner. The solid Euclidean geometry, for example, can be derived from the quadratic differential

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{3.51}$$

The Euclidean angle ω between two line elements (dx, dy, dz) and (δx, δy, δz) is given by the expression

$$\cos\omega = \frac{dx\delta x + dy\delta y + dz\delta z}{\sqrt{dx^2 + dy^2 + dz^2}\cdot\sqrt{\delta x^2 + \delta y^2 + \delta z^2}} \tag{3.52}$$

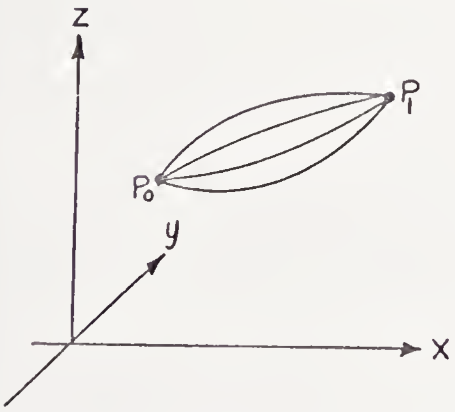


Fig. 33

The problem of finding the shortest connection of two points P_0 and P_1 introduces finite elements into the geometry, namely, the geodesic lines. On these geodesics the integral

$$\int \sqrt{dx^2 + dy^2 + dz^2}$$ assumes a minimum value; the geodesics of the Euclidean space are, of course, the straight lines in space.

We may characterize in general the points of a three-dimensional manifold by three suitable coordinates x, y, z and determine the distance ds of two infinitesimally distant points (x, y, z) and (x + dx, y + dy, z + dz) by the general quadratic differential

$$ds^2 = g_{11}\,dx^2 + g_{22}dy^2 + g_{33}dz^2 + 2g_{12}dxdy + 2g_{13}dxdz + 2g_{23}dydz \tag{3.53}$$

where the six coefficients g_{ik} are given functions of x, y, z. The angle of two line elements (dx, dy, dz) and (δx, δy, δz) attached to the same point P can then be found with the aid of (3.53), namely, by the expression

$$\cos\omega = \frac{(ds, \delta s)}{\sqrt{(ds, ds)}\cdot\sqrt{(\delta s, \delta s)}} \tag{3.54}$$

where (ds, ds) and (δ s, δ s) are the quadratic differentials

$$(ds, ds) = g_{11}dx^2 + g_{22}dy^2 + g_{33}dz^2 + 2g_{12}dxdy + 2g_{13}dxdz + 2g_{33}dydz$$

$$(\delta s, \delta s) = g_{11}\delta x^2 + g_{22}\delta y^2 + g_{33}\delta z^2 + 2g_{12}\delta x\delta y + 2g_{13}\delta x\delta z + 2g_{23}\delta y\delta z$$

and (ds, δs) the mixed quadratic differential

$$(ds, \delta s) = g_{11}dx\delta x + g_{22}dy\delta y + g_{33}dz\delta z + g_{12}(dx\delta y + \delta xdy) + g_{13}(dx\delta z + \delta xdz) + g_{23}(dy\delta z + \delta ydz)$$

The geodesics finally are determined by the solution of the problem of variation

$$S = \int\sqrt{g_{11}dx^2 + g_{22}dy^2 + g_{33}dz^2 + 2g_{12}dxdy + 2g_{13}dxdz + 2g_{23}dydz} = \text{Minimum} \quad (3.55)$$

3.6. The special analytic form of the differential (3.51) of the Euclidean geometry depends again on the special choice of a rectangular Cartesian coordinate system. By introducing other coordinates by a transformation

$$x = f(\xi, \eta, \zeta)$$
$$y = g(\xi, \eta, \zeta) \quad (3.61)$$
$$z = h(\xi, \eta, \zeta)$$

a great variety of analytic representations of (3.51) may be obtained, namely, quadratic differentials of the general form

$$ds^2 = g_{11}d\xi^2 + g_{22}d\eta^2 + g_{33}d\zeta^2 + 2g_{12}d\xi d\eta + 2g_{13}d\xi d\zeta + 2g_{23}d\eta d\zeta \quad (3.62)$$

The coefficients g_{ik} are certain functions of ξ, η, ζ, namely, quadratic combinations of the first derivatives of the functions f, g, h:

$$g_{11} = f_\xi^2 + g_\xi^2 + h_\xi^2 \qquad g_{12} = f_\xi f_\eta + g_\xi g_\eta + h_\xi h_\eta$$
$$g_{22} = f_\eta^2 + g_\eta^2 + h_\eta^2 \qquad g_{13} = f_\xi f_\zeta + g_\xi g_\zeta + h_\xi h_\zeta \quad (3.63)$$
$$g_{33} = f_\zeta^2 + g_\zeta^2 + h_\zeta^2 \qquad g_{23} = f_\eta f_\zeta + g_\eta g_\zeta + h_\eta h_\zeta$$

All these different forms have in common the fact that a transformation can be found which transforms (3.62) into the Euclidean normal form (3.51). The geometries derived from these differentials thus have the character of the solid Euclidean geometry.

This leads us to the following problem: Consider a general quadratic differential (3. 62). Which conditions must be satisfied by the six coefficients g_{ik} so that it be possible to transform (3.62) into the Euclidean form (3.51)? In general such a transformation is impossible. In fact, with g_{ik} being given, we obtain for the three functions f, g, h, of a transformation (3.61), a system (3.63) of six quadratic differential equations. This system is overdetermined and has thus, in general, no solution. In order to insure the existence of a solution, three conditions must be satisfied by the coefficients g_{ik}. These conditions can be derived from Riemann's Tensor of Curvature, a generalization of Gauss' Curvature to the case of manifolds of three dimensions. The three essential components of Riemann's tensor can be expressed in terms of the functions g_{ik} and their derivatives. If the differential (3.62) represents a Euclidean line element, then all components of Riemann's tensor are identically zero. Vice versa, if the components of this tensor are identically zero, then a transformation (3.61) may be found which transforms (3.62) into the Euclidean normal form (3.51), so that the manifold in question is isometric to the Euclidean space. This result implies the remarkable fact that the question whether or not a given metric manifold of three dimensions is Euclidean can be answered by measurements in the manifold itself. It means the curvature of such a manifold can be recognized even if observation from a viewpoint in an additional fourth dimension is impossible.

3.7. Among the general metric differentials

$$ds^2 = g_{11}dx^2 + g_{22}dy^2 + g_{33}dz^2 + 2g_{12}dxdy + 2g_{13}dxdz + 2g_{23}dydz \tag{3.71}$$

there exists a special group distinguished by the property that Riemann's tensor of curvature is constant, i.e., its three components are independent of the localization. Riemann has shown that all these line elements of constant curvature can be transformed into a normal form similar to (3.46) in case of two dimensions:

$$ds^2 = \frac{d\xi^2 + d\eta^2 + d\zeta^2}{\left[1 + \frac{1}{4} K (\xi^2 + \eta^2 + \zeta^2)\right]^2} \tag{3.72}$$

where K is a constant, namely, the constant Riemannian curvature of the three-dimensional manifold.

For $K = 0$ we obtain the Euclidean geometry:

$$ds^2 = d\xi^2 + d\eta^2 + d\zeta^2 \tag{3.73}$$

For $K < 0$, for example $K = +1$, the elliptic geometry:

$$ds^2 = \frac{d\xi^2 + d\eta^2 + d\zeta^2}{\left[1 + \frac{1}{4} (\xi^2 + \eta^2 + \zeta^2)\right]^2} \tag{3.74}$$

For $K > 0$, for example $K = -1$, the hyperbolic geometry:

$$ds^2 = \frac{d\xi^2 + d\eta^2 + d\zeta^2}{\left[1 - \frac{1}{4} (\xi^2 + \eta^2 + \zeta^2)\right]^2} \tag{3.75}$$

Instead of Riemann's normal form, we shall use later on another form which is found by (3.72) by introducing polar coordinates

$$\begin{aligned} \xi &= R \cos\varphi \cos\theta \\ \eta &= R \sin\varphi \\ \zeta &= R \cos\varphi \sin\theta \end{aligned} \tag{3.76}$$

This gives

$$ds^2 = \frac{dR^2 + R^2 (d\varphi^2 + \cos^2\varphi \; d\theta^2)}{(1 + \frac{K}{4} R^2)^2} \qquad (3.77)$$

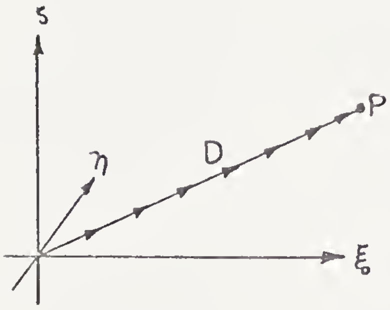


Fig. 34

The geodesic distance D of a point P from the origin obviously is obtained by summing up line elements ds which have no lateral or vertical extension d φ and d θ. Hence

$$D = \int_0^R \frac{dR}{1 + \frac{K}{4} R^2} = \frac{2}{\sqrt{K}} \text{ arc tan } (\tfrac{1}{2}\sqrt{K}\, R) \qquad (3.78)$$

By introducing this geodesic distance D instead of R in (3.77), we find the line element

$$ds^2 = dD^2 + M^2 (d\varphi^2 + \cos^2\varphi \; d\theta^2) \qquad (3.781)$$

where M is the function

$$M = \frac{1}{\sqrt{K}} \sin \sqrt{K}\, D \qquad (3.782)$$

This function may be interpreted as determining the linear size of a line element with no depth extension (dD = 0) but only lateral extension given by the angular coordinates dφ, dθ. The dependence of this size factor M on the distance D is given by (3.782), i.e., it is

$$\begin{aligned} M &= \sin D \quad && \text{in the elliptic case} \quad K = 1 \\ M &= D && \text{in the Euclidean case} \quad K = 0 \qquad (3.783) \\ M &= \sinh D && \text{in the hyperbolic case} \quad K = -1 \end{aligned}$$

For a direct application to our binocular problem another form of Riemann's line element is advantageous. We submit (3.781) to the following transformations

$$\begin{aligned} K = 1: &\quad -\tan \tfrac{1}{2} D = \sigma(\gamma + \mu) \\ K = 0: &\quad -\log \tfrac{1}{2} D = \sigma(\gamma + \mu) \qquad (3.79) \\ K = -1: &\quad -\tanh \tfrac{1}{2} D = \sigma(\gamma + \mu) \end{aligned}$$

where σ and μ are constants and γ a variable replacing the distance D. It follows quite easily that

$$\begin{aligned} K = 1: &\quad ds^2 = \frac{1}{\cosh^2 \sigma(\gamma + \mu)} \left\{\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi \; d\theta^2\right\} \\ K = 0: &\quad ds^2 = 2\, e^{-2\sigma(\gamma+\mu)} \left\{\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi \; d\theta^2\right\} \qquad (3.791) \\ K = -1: &\quad ds^2 = \frac{1}{\sinh^2 \sigma(\gamma + \mu)} \left\{\sigma^2 d\gamma^2 + d\theta^2 + \cos^2\varphi \; d\theta^2\right\} \end{aligned}$$

All these line elements have the same general form

$$ds^2 = M^2 (\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi \, d\theta^2) \qquad (3.792)$$

where the size factor $M(\gamma)$ is given by the functions

$M = \dfrac{1}{\cosh \sigma(\gamma + \mu)}$ in case of the elliptic geometry

$M = 2e^{-\sigma(\gamma+\mu)}$ in case of the the Euclidean geometry

$M = \dfrac{1}{\sinh \sigma(\gamma + \mu)}$ in case of the hyperbolic geometry.

3.8. Constancy of size. Rigid transformations. The Euclidean geometry is distinguished by the fact that objects can be moved without changing their size or their other metric characteristics. A triangle in the Euclidean plane can be moved freely in this plane to any other position without distortion of its characteristics. The result is a triangle congruent to the original one.

A movement of objects in a metric manifold can be described mathematically by a point transformation of the manifold in itself. This means, for example, in a Euclidean plane, that to any point (x, y) another point (x', y') is coordinated

$$x' = f(x, y) \qquad (3.81)$$
$$y' = g(x, y)$$

namely, the point (x', y') to which an object located at x, y is moved. Such a point transformation is principally different from a transformation (3.31) of a coordinate system, although both are mathematically represented in a similar way. In (3.31) the same point of a plane is associated with different numbers, i.e., the points are considered fixed but the coordinate system is changed. In (3.81), however, we keep the same coordinate system but interchange the points of the plane. Suppose now that a configuration of objects x, y is transformed into a configuration x', y', by (3.81). Since the coordinate system is unchanged, we have the same metric differential $ds^2 = dx^2 + dy^2$ before and after the transformation (3.81), and thus we must expect in general that the new configuration x', y' is metrically different from the original configuration. This distortion is a consequence of the fact that

$dx'^2 + dy'^2 = ds'^2 = (f_x^2 + g_x^2)\, dx^2 + 2(f_x f_y + g_x g_y)\, dxdy + (f_y^2 + g_y^2)\, dy^2$ is in general not identical with $dx^2 + dy^2 = ds^2$.

There exist, however, in the Euclidean plane, special transformations (3.81) which preserve the metric of a configuration, i.e., transformations for which

$$dx'^2 + dy'^2 = dx^2 + dy^2 \qquad (3.82)$$

These special transformation are called rigid transformations of the plane, and are given by the three-parameter group of transformations

$$x' = x \cos \omega - y \sin \omega + a \qquad (3.83)$$
$$y' = x \sin \omega + y \cos \omega + b$$

where ω, a, b are arbitrary constants. The transformations (3.83) can be described as rotations around the origin by an angle ω plus an additional translatory shift. If any configuration of objects is submitted to such a transformation, then the resulting configuration is congruent to the original one.

In a similar manner we may determine the rigid transformations of the three-dimensional Euclidean space. It is a six-parameter group of transformations characterized by rotations around the origin (3 parameters) plus an additional translatory shift (3 parameters).

By these groups of transformations we thus may move any line element dx, dy, dz to any other position and direction without changing its size. We can say that the existence of such transformations is a necessary and sufficient condition for the existence of objects independent of localization, i.e., objects may be moved freely without distorting their metric characteristics.

The existence of a group of rigid transformations of sufficiently great number of parameters (6 in three dimensions; 3 in two dimensions) so that complete movability of objects is insured, is by no means self-evident. Indeed, in the case of a general quadratic differential ds^2 we must expect that no transformation of this type exists; objects in such a manifold are frozen to their positions; any movement will result in a distortion of their metric characteristics.

The remarkable feature of the geometries of constant curvature is now that in all three types, elliptical, Euclidean, hyperbolic, the same complete movability of objects is found. Indeed, this is Riemann's result: These geometries are the only ones which have this character. Other metric differentials may allow partial movability, i.e., possess a group of rigid transformations of lower number of parameters with the result that objects may be moved into certain restricted positions, but only the three quadratic differentials of constant curvature allow objects to be moved into any position without distortion of their metric characteristics. These three geometries thus are the only ones with true size constancy. We shall, later on, discuss the rigid transformations of the hyperbolic geometry in mathematical detail, and recognize their significance for binocular vision.

Section 4

THE PSYCHOMETRIC OF VISUAL SENSATIONS

4.1. We have recognized in §2 the physiological significance of the angular differentials $d\gamma$, $d\varphi$, $d\theta$ for the binocular observation of a physical line element (dx, dy dz). These angular differentials determine directly the essential characteristics of the images of the line element on the retinae of the observer, provided his head remains in a fixed position. If, on the other hand, the line element is observed with moving head, then the eyes are directed in symmetrical convergence towards the line element, and its retinal images are given by the differentials $d\gamma^*$, $d\varphi^*$, $d\theta^*$ as explained in §2.8.

Our problem in this section is the sensation of size which we have when observing a given physical line element. Our basic assumption is that assignment of size to physical line elements is a primitive sensation, i.e., a sensation completely determined by the physical characteristics of the line element. Though the retinal images of a line element are determined by the above angular differentials, we do not sense these differentials as angles but as linear quantities. Even objects as far away as the moon or clouds appear to have a certain linear size which, of course, is in no way identical with their actual physical size. This transformation of angular into linear distances is quite characteristic for our visual sensations, and, in the case of uniocular vision, can be demonstrated by a simple experiment. An angular wedge made from cardboard is placed in front of one eye of the observer so that the vertex of the angle coincides with the center of rotation of the eye. The wedge may be vertical or horizontal or in intermediate position. With the other eye closed, the legs of the wedge do not appear converging but parallel to each other, i.e., as two lines which have a constant linear distance from each other. Physically they include, however, a constant angle. We may express this peculiar situation by saying that the angular coordinates of a polar system are interpreted as linear coordinates of a Cartesian system.

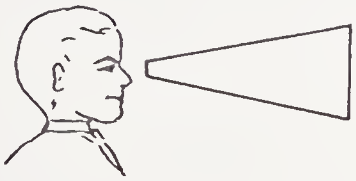


Fig. 35

The assignment of linear size to angular impingements should not be understood as assigning physical length measurable in cm. to the objects. We may do so by intellectual association of sensation and physical experience, but the primitive act of transforming angles into size is based on physiological units quite different from units of physical length. However, the character of these physiological units is not the problem we are here concerned with. From a geometrical point of view we need not be interested in the unit employed in size assignments. The metric characteristics of the visual apparent space are entirely independent of the physiological dimensions of this unit.

4.2. Let us now consider a physical line element (dx, dy, dz) attached to a point P = (x, y, z) of the physical space. By using bipolar coordinates we may characterize the base point P by the three angles γ, φ, θ and the line element by the differentials $d\gamma$, $d\varphi$, $d\theta$. These differentials can be found from dx, dy, dz by differentiating the formulae

$$x = \frac{\cos 2\varphi + \cos\gamma}{\sin\gamma}\cos\theta$$

$$y = \frac{\sin 2\varphi}{\sin\gamma} \tag{4.21}$$

$$z = \frac{\cos 2\varphi + \cos\gamma}{\sin\gamma}\sin\theta$$

which relate Cartesian and bipolar coordinates.

We observe this line element with fixed head so that the optical axes of the eyes converge at P. Our assumption is that the observer is led to a definite sensation of the size and that this apparent size ds is a function of the differentials $d\gamma$, $d\varphi$, $d\theta$, namely, given by a quadratic differential

$$ds^2 = C^2 (g_{11}\, d\gamma^2 + g_{22}\, d\varphi^2 + g_{33}\, d\theta^2 + 2g_{12}\, d\gamma d\varphi + 2g_{13}\, d\gamma d\theta + 2g_{23}\, d\varphi d\theta) \tag{4.22}$$

The coefficients g_{ik} are functions of the coordinates γ, φ, θ of the base point P. The constant C determines the unit of the apparent size, which, as we point out again, is not a unit of physical length. Since this constant is immaterial for the geometry based upon the differential (4.22), we shall omit it in the following and thus assume $C = 1$. The size ds of the line element $d\gamma$, $d\varphi$, $d\theta$ thus will be a pure number of our number system.

4.3. Let us next observe the above line element with moving head so that the eyes are directed in symmetrical convergence towards the point P. We now characterize the point P by the angular coordinates γ^*, φ^*, θ^* of §2.8 and the line element by the differentials $d\gamma^*$, $d\varphi^*$, $d\theta^*$, which are found from dx, dy, dz by differentiating the equations

$$x = (d + \cot \tfrac{1}{2}\gamma^*) \cos\varphi^* \cos\theta^*$$

$$y = (d + \cot \tfrac{1}{2}\gamma^*) \sin\varphi^* \tag{4.31}$$

$$z = (d + \cot \tfrac{1}{2}\gamma^*) \cos\varphi^* \sin\theta^*$$

relating the coordinates x, y, z and γ^*, φ^*, θ^*.

The apparent size ds^* which we assign to our line element by this manner of observation is based upon the differentials $d\gamma^*$, $d\varphi^*$, $d\theta^*$ and upon the coordinates φ^*, γ^*, θ^* of the point P. Indeed, the differentials $d\gamma^*$, $d\varphi^*$, $d\theta^*$ determine the psychologically significant characteristics of the retinal images of the line element and γ^*, φ^*, θ^*, the position of the eyes and their optical axes. The principle formulated in §2.8 about the relationship of observations with fixed and moving head enforces the implication that <u>ds^* can be found from γ^*, φ^* θ^* by the same formula as ds from γ, φ, θ</u>. This means that ds^* is given by the quadratic differential

$$ds^{*2} = g_{11}\, d\gamma^{*2} + g_{22}\, d\varphi^{*2} + g_{33}\, d\theta^{*2} + 2g_{12}\, d\gamma^* d\varphi^* + 2g_{13}\, d\gamma^*\, d\theta^* + 2g_{23}\, d\varphi^*\, d\theta^* \tag{4.32}$$

where <u>the coefficients g_{ik} $(\gamma^*, \varphi^*, \theta^*)$ are the same functions as the coefficients g_{ik} $(\gamma, \varphi, \theta)$ in (4.22)</u>.

Indeed, consider the following pair of two different physical line elements:

$$(dx,\ dy,\ dz) \text{ at a point } P = (x,\ y,\ z) \text{ and}$$

$$(dx^*,\ dy^*,\ dz^*) \text{ at a point } P^* = (x^*,\ y^*,\ z^*)$$

Let the first one be chosen arbitrarily and be given by the bipolar differentials $d\gamma$, $d\varphi$, $d\theta$ and the base point $P = (\gamma, \varphi, \theta)$. We determine the second base point P^* by its coordinates $(\gamma^*, \varphi^*, \theta^*)$ for head movements requiring that

$$\gamma^* = \gamma$$

$$\varphi^* = \varphi$$

$$\theta^* = \theta$$

We next construct the line element (dx^*, dy^*, dz^*) at P^* by the condition that, with regard to the coordinates $(\gamma^*, \varphi^*, \theta^*)$ it shall have the differentials

$$d\gamma^* = d\gamma$$

$$d\varphi^* = d\varphi$$

$$d\theta^* = d\theta$$

The second line element thus has the same coordinates in the γ^*, φ^*, θ^* system for head movements as the first one in the bipolar γ, φ, θ system for observation with fixed head. Obviously, the second line element is uniquely determined by the first one. From the general principle of §2.8, it now follows that the same size sensation must be obtained if the first line element is observed with fixed head and the second with moving head. If this be true for any such pair of line elements, then the quadratic differentials (4.22) and (4.32) must be formally identical.

4.4. The above result has the consequence that to the same physical line element (dx, dy, dz) a different size will be assigned if it is first viewed with fixed and second with moving head. Let us first consider an observation with fixed head. With the aid of the relations (4.21) we express the quadratic differential (4.22) by the Cartesian coordinates x, y, z and their differentials dx, dy, dz. The result of this transformation is a quadratic differential of the form

$$ds^2 = A_{11}dx^2 + A_{22}dy^2 + A_{33}dz^2 + 2A_{12}dxdy + 2A_{13}dxdz + 2A_{23}dydz \qquad (4.41)$$

where the coefficients A_{ik} are certain functions of x, y, z. It determines directly the apparent size of the physical line element dx, dy, dz observed at the point P = (x, y, z).

On the other hand, let us observe the same line element with moving head and thus direct the eyes in symmetrical convergence towards the point P. We now transform the differential (4.32) which is formally identical with (4.22) with the aid of the relations (4.31), and obtain another quadratic differential

$$ds^{*2} = B_{11}dx^2 + B_{22}dy^2 + B_{33}dz^2 + 2B_{12}dxdy + 2B_{13}dxdz + 2B_{23}dydz \qquad (4.42)$$

The coefficients $B_{ik}(x, y, z)$ are of course not the same as the coefficients $A_{ik}(x, y, z)$ in (4.41), and consequently the apparent size ds^* of our line element differs in general from its apparent size ds.

However, we know that the two quadratic differentials (4.41) and (4.42) have been obtained by transformation from two differentials (4.22) and (4.32) which are formally identical. In other words, they are the result of two different transformations of <u>the same basic quadratic differential</u>

$$ds^2 = g_{11}d\gamma^2 + g_{22}d\varphi^2 + g_{33}d\theta^2 + 2g_{12}d\gamma d\theta + 2g_{13}d\gamma d\theta + 2g_{23}d\varphi d\theta \qquad (4.43)$$

We may also say that both differentials (4.41) and (4.42) can be transformed into each other. We have seen in §3 that this property of two metric differentials establishes an intimate relationship between them, namely, that both differentials characterize the same invariant geometry but in different coordinate systems. They are different analytic expressions for the same metric, and the geometries developed from them are identical.

By observing the same physical configuration, whether with fixed or moving head, we find in ourselves the conviction that the resulting sensations, though actually different, belong to the same unchanged environment of objects. In our theory, this conviction of observing the same objects but by different methods of observation finds its mathematical expression in the fact that the two quadratic differentials associated with these sensations represent the same invariant geometrical relations.

4.5. Our problem is to determine the coefficients $g_{ik}(\gamma, \varphi, \theta)$ of the basic quadratic differential

$$ds^2 = g_{11}d\gamma^2 + g_{22}d\varphi^2 + g_{33}d\theta^2 + 2g_{12}d\gamma d\varphi + 2g_{13}d\gamma d\theta + 2g_{23}d\varphi d\theta \qquad (4.51)$$

In general we shall interpret φ, φ, θ as bipolar coordinates referring to observation with fixed head. However, in this section, we shall base our discussion upon observation with moving head. We know the corresponding differential

$$ds^2 = g_{11}d\gamma^{*2} + g_{22}d\varphi^{*2} + g_{33}d\theta^{*2} + 2g_{12}d\gamma^*d\theta^* + 2g_{13}d\gamma^*d\theta^* + 2g_{23}d\varphi^*d\theta^* \qquad (4.52)$$

must be formally identical with (4.51). Hence we are sure that any information obtained about the functions g_{ik} by observation with moving head can be used directly for the metric (4.51) of the sensations associated with observation with fixed head.

Observation with moving head as described in §2.8 must be considered as spherically symmetrical to the origin. Indeed, we can safely assume that two line elements (dx, dy, dz) and (dx', dy', dz') of equal length, located at the same distance R from the observer and including the same angle with the corresponding lines of view are judged to be of equal size (provided that we turn our head towards them). Mathematically we may express this as follows: Rotations of the space around the origin must not change the metric differential (4.52). In other words, these rotations must be rigid transformations of the differential (4.52).

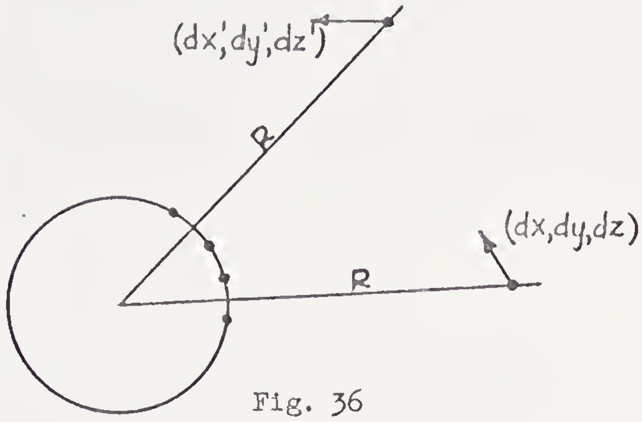


Fig. 36

It is not difficult to prove that this condition is satisfied then and only then if (4.52) has the form

$$ds^{*2} = A^2(\gamma^*)\ d\gamma^{*2} + M^2(\gamma^*)\ (d\varphi^{*2} + \cos^2\varphi^*\ d\theta^{*2}) \qquad (4.53)$$

where $A(\gamma^*)$ and $M(\gamma^*)$ are arbitrary functions of γ^* alone.*)

Since the differentials ds^* and ds are formally identical, we conclude that the metric differential (4.51) for observation with fixed head must also have the general form

$$ds^2 = A^2(\gamma)\ d\gamma^2 + M^2(\gamma)\ (d\varphi^2 + \cos^2\varphi\ d\theta^2) \qquad (4.54)$$

where γ, φ, θ are the ordinary bipolar coordinates. Instead of six unknown functions g_{ik} of three variables γ, φ, θ there are only two functions $A(\gamma)$ and $M(\gamma)$ unknown in (4.54). With this simplification we abandon the discussion of observations with moving head and confine ourselves in the following to the investigation of visual sensations associated with observation with fixed head.

4.6. The function $M(\gamma)$ in (4.54) determines the apparent size of a line element $d\varphi$, $d\theta$ which has no disparity $d\gamma$. The function $A(\gamma)$ gives the depth extension of a line element with disparity $d\gamma$ but with no lateral or vertical extension $d\varphi$, $d\theta$. By summing up line elements of the latter type, we obtain the apparent distance D of two points P_o and P from each other

$$D = \int_{\gamma_o}^{\gamma} A(\gamma)\ d\gamma \qquad (4.61)$$

The arbitrariness of the functions $M(\gamma)$ and $A(\gamma)$ illustrates the fact that there is no general relation between the concepts of size and distance, i.e., of size and localization.

We show next that in the case of binocular vision there exists such a relation. Let us consider two line elements (dx, dy, dz) and (dx', dy', dz') located at two different points of the same radius vector. Without loss of generality, we may assume that the line elements are lying in the horizontal plane on the x-axis. Let us assume that both line elements have the same bipolar differentials, i.e.,

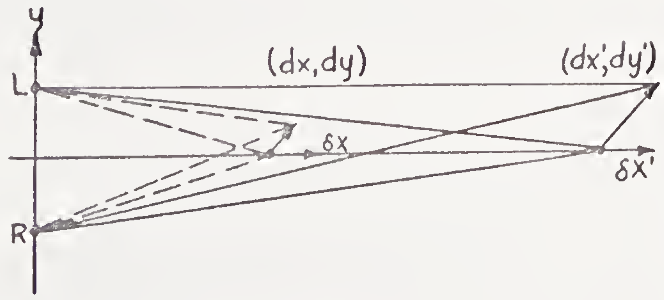


Fig. 37

$$d\gamma' = d\gamma \qquad d\varphi' = d\varphi \qquad (4.62)$$

so that we observe the same impinging angular characteristics in the two cases. We determine the apparent angle ω which these line elements include with the x-axis, i.e., with two line elements

$$\delta\gamma' = \delta\gamma \qquad \delta\varphi' = \delta\varphi = 0 \qquad (4.63)$$

*A proof can be found in: Levi-Civita, Der absolute Differential Calcul, Berlin: 1928, pp. 278-283.

The apparent angle ω is given by the formula (refer to 3.54):

$$\cos\omega = \frac{A^2(\gamma)d\gamma\delta\gamma + M^2(\gamma)d\varphi\delta\varphi}{\sqrt{A^2(\gamma)d\gamma^2 + M^2(\gamma)d\varphi^2}\cdot\sqrt{A^2(\gamma)\delta\gamma^2 + M^2(\gamma)\delta\varphi^2}}$$

i.e., on account of (4.63) by

$$\cos\omega = \frac{A(\gamma)d\gamma}{\sqrt{A^2(\gamma)d\gamma^2 + M^2(\gamma)d\varphi^2}} = \frac{1}{\sqrt{1 + \frac{M^2(\gamma)}{A^2(\gamma)}\left(\frac{d\varphi}{d\gamma}\right)^2}} \qquad (4.64$$

Similarly we find ω' by

$$\cos\omega' = \frac{1}{\sqrt{1 + \frac{M^2(\gamma')}{A^2(\gamma')}\left(\frac{d\varphi}{d\gamma}\right)^2}}. \qquad (4.65)$$

We notice that the two angles ω and ω' are different unless the ratio $A(\gamma)/M(\gamma)$ is independent of γ, i.e., equal to a constant σ.

Actually, two such line elements on the same line of view which give the same angular impingement $d\gamma$, $d\varphi$ are seen as parallel. Hence we conclude that $A(\gamma) = \sigma M(\gamma)$, and thus that the binocular metric must have the form

$$ds^2 = M^2(\gamma)\,(\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi\, d\theta^2) \qquad (4.66)$$

which reduces our problem to the problem of finding only one function $M(\gamma)$, the size factor of a line element with no disparity $d\gamma$.

The constant σ is a constant depending upon the individual observer; it determines the sensitivity of depth perception through disparity $d\gamma$ as compared with size perception through lateral and vertical angles $d\varphi$ and $d\theta$. We know that the angular threshold of disparity which gives the sensation of depth is considerably smaller than the angular threshold for recognition of size difference. This means that σ must be expected to be considerably greater than 1.

The apparent distance D of two points P_0 and P_1 on the x-axis (or any other line to the origin) is given by the integral

$$D = \sigma \int_{\gamma_0}^{\gamma} M(\gamma)d\gamma \qquad (4.67)$$

which establishes the relation of apparent distance to apparent size in binocular vision.

We mention finally a simple demonstration which shows that line elements attached to points of the x-axis with the same angular coordinates $d\gamma$, $d\varphi$ are seen as parallel. We construct a simple chessboard pattern of the type shown in Fig. 38 and place it normal to the horizontal plane so that an angle $\omega \neq 90^\circ$ is included with the x-axis. By crossing our eyes at a point P_0 of the x axis so that the squares of the pattern are fused in pairs (1, 2). (2, 3), (3, 4), etc., a smaller pattern is seen at some other position (not necessarily P_0). However, the resulting pattern appears to be parallel to the original one. The angular impingement $d\gamma$, $d\varphi$ obviously is unchanged by this observation, the angles γ of convergence, however, are different.

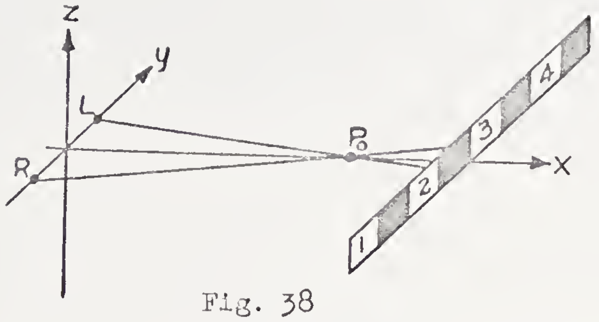


Fig. 38

4.7. The result (4.66) leads us to an interesting demonstration. We construct on a Vieth-Muller circle a number of small vertical rods. Since, for these rods, $d\gamma = d\varphi = 0$, we have by (4.66) for their apparent size the expression

$$ds = M(\gamma) \cos \varphi \, d\theta \tag{4.71}$$

Since $M(\gamma)$ is constant on the Vieth-Muller circle, it follows from (4.71) that the apparent size of the rods is proportional to the expression $\cos \varphi \, d\theta$. If now the physical height, h, of the rods is chosen such that $\cos \varphi \, d\theta$ is a constant, then the rods will have the same apparent height and thus appear to form a circular fence of equal height with the observer as center.

Since the physical height of a rod $d\theta$ at the latitude φ on the Vieth-Muller circle is approximately given by the relation

$$h = 2R \cos^2\varphi \, d\theta \tag{4.72}$$

we conclude that our rods must decrease according to the law

$$h = h_0 \cos \varphi \tag{4.73}$$

if they are to have the same apparent height. The constant h_0 is the height of a rod on the x-axis.

4.8. The remaining problem is to determine the size factor $M(\gamma)$ of the quadratic differential (4.66). We shall do this in the following sections by evaluating certain observable facts. In this section, however, we shall first investigate the simplest hypothesis about $M(\gamma)$, namely $M(\gamma) = \text{const.}$ This assumption would mean that the angular impingements $d\gamma$, $d\varphi$, $d\theta$ and their retinal images are the only significant clues for visual sensations; the absolute value of γ, i.e., the convergence of the eyes, is without significance. Without loss of generality, we may assume $M = 1$ and thus have

$$ds^2 = \sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi \, d\theta^2 \tag{4.81}$$

We notice that, in every plane of elevation, $\theta = \text{const.}$, a Euclidean differential

$$ds^2 = \sigma^2 d\gamma^2 + d\theta^2 \tag{4.82}$$

is obtained so that the plane geometries in these planes must be Euclidean. Yet these plane Euclidean geometries do not supplement each other to a Euclidean geometry of the three-dimensional space. In fact, one can show without difficulty that the quadratic differential (4.81) is non-Euclidean.

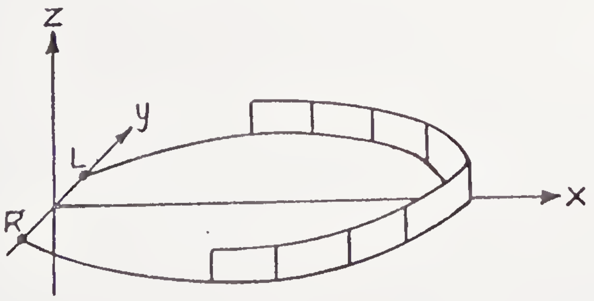


Fig. 39

We notice readily that the differential (4.81) preserves its mathematical form if submitted to the transformation

$$\gamma' = \gamma + \tau$$

$$\varphi' = \varphi \tag{4.83}$$

$$\theta' = \theta + \lambda$$

depending on the arbitrary parameters τ and λ. These transformations form a subgroup of the general iseikonic transformations

$$\gamma' = \gamma + \tau$$
$$\varphi' = \varphi + \sigma^* \qquad (4.84)$$
$$\theta' = \theta + \lambda$$

which we have discussed in § 2.5, and we conclude that the transformations of this subgroup (4.83) represent rigid movements in the non-Euclidean geometry based upon the differential (4.81). However, we see also that the complete group (4.84) of iseikonic transformations does not give the rigid movements of our geometry. Indeed the differential (4.81) changes its form in case $\sigma^* \neq 0$.

It is not difficult to determine the complete group of rigid transformations which belong to the quadratic differential (4.81). It consists of the special iseikonic transformations

$$\gamma' = \gamma + \tau \qquad (4.85)$$

and those transformations of φ, θ into φ', θ' which do not change the quadratic differential $d\varphi^2 + \cos^2\varphi\, d\theta^2$. Since $d\varphi^2 + \cos^2\varphi\, d\theta^2$ is the line element on a sphere of radius one, and since this line element preserves its form by any rotation of the sphere about its origin, we recognize that the desired transformations of φ, θ into φ', θ' must be isomorphic to the three-parameter group of spherical rotations. We thus obtain a four-parameter group of rigid transformations for our geometry (4.81). Complete movability of objects requires, however, the existence of a six-parameter group of rigid movements, as we have seen in §3.8. Hence we conclude that the geometry based upon (4.81) does not provide complete movability. This already seems to indicate that our above hypothesis $M(\gamma) = 1$, which denies the significance of convergence for space perception, has to be revised: It is contradictory to the conviction which accompanies our visual sensations that objects can be moved to any position without changing their metric characteristics.

We can easily find another indication that the assumption $M(\gamma) = 1$ cannot be defended. The geodesic lines in the horizontal plane, i.e., the solutions of the minimum problem

$$\int \sqrt{\sigma^2\, d\gamma^2 + d\varphi^2} = \text{Minimum}$$

are given by the curves

$$\varphi = a\gamma + b \qquad (4.86)$$

where a and b are arbitrary constants.

Since geodesic lines have the property that all their line elements have the same unchanged direction, we conclude that a number of marks arranged on such a curve will give the impression of arrangement on an apparent straight line. Curves of this type are called Horopters; in particular those curves which are symmetrical to the x-axis are known as <u>Frontal plane horopters</u>. From (4.86) it follows readily that these frontal plane horopters are given by the curves γ = const., by the Vieth-Muller circles. But we know already that Vieth-Muller circles appear flatter than they are but not yet plane, namely, as circles with the observer as center. Furthermore, it is well known that the actually observed horopters are curves of the type shown in Fig. 40. We conclude that the horopter phenomenon cannot be explained by the hypothesis $M(\gamma) = 1$.

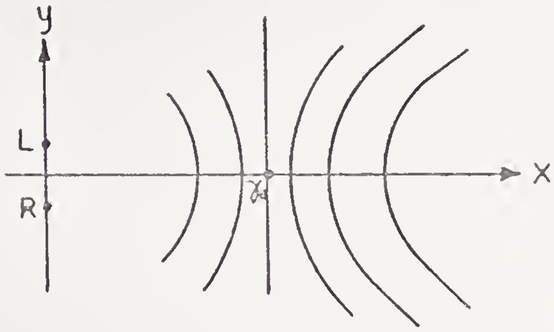


Fig. 40

4.9. We base our next hypothesis about $M(\gamma)$ upon the psychological conviction that the visual shape and size of objects can be repeated in any position and orientation. This conviction makes us consider shape and form of an object as qualities which are independent of localization. We also may say that we are convinced that any object can be moved as a visually rigid body to any desired position and orientation and the result of this movement is an object metrically congruent to the original object. We must, however, not assume that such a movement of an object is necessarily the same as a physical movement in a Euclidean space. In fact, we know the opposite is true: apparent shape and size of an object do not remain unchanged, when it is moved physically. Consequently we prefer to formulate the conviction about the complete movability of visual objects as follows: <u>To any configuration of objects there exists a six-parameter set of configurations which, in all their metrical characteristics, are visually congruent to the original configuration.</u> This psychological fact, that metrical form and localization of objects are considered as independent, would be hard to understand if the geometry of our visual sensations did not itself provide what we have called above complete movability of objects. This, however, is possible only if the quadratic differential (4.66) represents a geometry of constant curvature. It follows from §3.79 that we should expect the factor $M(\gamma)$ to be one of the three functions

$$M(\gamma) = \frac{1}{\cosh \sigma (\gamma + \mu)}$$

$$M(\gamma) = 2e^{-\sigma(\gamma+\mu)} \qquad (4.91)$$

$$M(\gamma) = \frac{1}{\sinh \sigma(\gamma + \mu)}$$

In the first case the geometry is elliptic, in the second, Euclidean, in the last, hyperbolic. In addition to σ another individual constant, μ, enters these expressions. It determines the limit which the size factor $M(\gamma)$ approaches if $\gamma \to 0$, i.e., if the object is physically moved far away. It also is closely related to the apparent distance at which we place objects of "infinite" physical distance.

We have yet no means of deciding which one of the three functions (4.91) has to be chosen for our problem. We shall, however, in the following, accumulate evidence obtained by theoretical considerations and by experimental results that $M(\gamma)$ is given by the last function

$$M(\gamma) = \frac{1}{\sinh \sigma (\gamma + \mu)} \qquad (4.92)$$

so that the metric of our space sensations is hyperbolic, namely, given by the quadratic differential

$$ds^2 = \frac{1}{\sinh^2 \sigma(\gamma + \mu)} (\sigma^2 \cdot d\gamma^2 + d\varphi^2 + \cos^2\varphi \, d\theta^2) \tag{4.93}$$

We may take as a first indication the fact that the hyperbolic metric is best suited for approximate coincidence of physical and apparent size in a sufficiently large interval.

Let us assume that an observer, by systematic training or experience, is in the position to change his constants σ and μ. We shall demonstrate that the observer can improve his size judgment by changing the constant μ from positive values gradually to zero if the metric is hyperbolic but not if it is elliptic or Euclidean.

It is clear that the limiting value $M(0)$ for objects at infinity is the greater in all three cases the smaller μ is. Its greatest value is assumed for $\mu = 0$, and obviously this represents the best approximation of apparent size to physical size. Let us therefore consider this best case in the three metrics (4.91). We have

$$M(\gamma) = \frac{1}{\cosh \sigma \gamma}$$

$$M(\gamma) = \frac{2}{e^{\sigma\gamma}} \tag{4.94}$$

$$M(\gamma) = \frac{1}{\sinh \sigma\gamma}$$

The limits at $\gamma = 0$ are $M(0) = 1$ in case of the elliptic, $M(0) = 2$ in case of the Euclidean, but $M(0) = \infty$ in case of the hyperbolic geometry. Only the hyperbolic metric thus may approach the physical situation that the size of objects subtending equal angles increases beyond all bounds.

To illustrate this, let us consider the apparent size ds of an object of constant physical size dh if moved in the physical space. We have $d\theta = dh \tan\gamma/2$ and hence, assuming $\varphi = 0$, $d\varphi = d\gamma = 0$:

$$ds = dh \frac{\tan \gamma/2}{\cosh \sigma \gamma}$$

$$ds = 2dh \frac{\tan \gamma/2}{e^{\sigma \gamma}} \tag{4.95}$$

$$ds = dh \frac{\tan \gamma/2}{\sinh \sigma \gamma}$$

We conclude from (4.95) that the ratio ds/dh of apparent to physical size converges to zero if the object is moved towards infinity and if the metric is elliptic or Euclidean. In the hyperbolic case, however, it reaches asymptotically a constant value $ds/dh = \frac{1}{2\sigma}$, so that ds/dh remains constant in a great interval. This situation is shown in Fig. 41 where the three curves ds/dh as function of $x = \cot\gamma/2$ are drawn. In case $\mu \neq 0$ all three functions approach zero, but only the hyperbolic metric allows by variation of μ to retard this approach effectively and thus to improve subjective size estimation in regard to agreement with objective physical size.

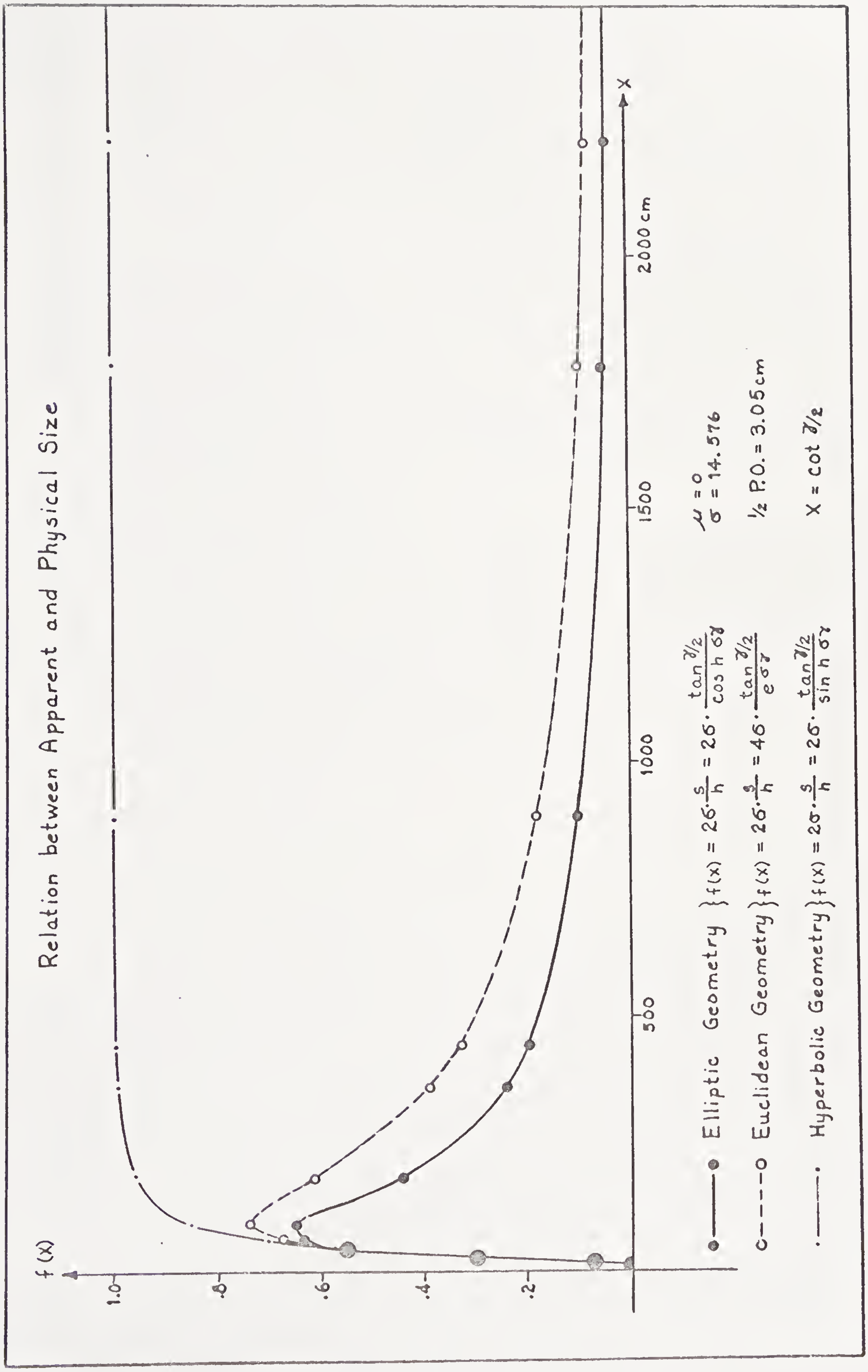

Relation between Apparent and Physical Size
f(x)
1.0
.8
.6
.4
.2
500
1000
1500
2000 cm
X
Elliptic Geometry } f(x) = 2σ · s/h = 2σ · tan γ/2 / cos h σγ
Euclidean Geometry } f(x) = 2σ · s/h = 4σ · tan γ/2 / e^σγ
Hyperbolic Geometry } f(x) = 2σ · s/h = 2σ · tan γ/2 / sin h σγ
μ = 0
σ = 14.576
½ P.O. = 3.05 cm
X = cot γ/2

Fig. 41

Section 5

DERIVATION OF THE HYPERBOLIC METRIC OF VISUAL SENSATIONS

5.1. We shall discuss in this section facts of observation which support the hypothesis that the psychometric of visual sensations is the hyperbolic metric of constant curvature. Our problem is to determine the factor $M(\gamma)$ in the quadratic differential

$$ds^2 = M^2(\gamma)(\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi d\theta^2) \tag{5.11}$$

Without loss of generality we may confine ourselves to observation of objects distributed in the horizontal plane $\theta = 0$. The apparent geometry in this plane is represented by the differential

$$ds^2 = M^2(\gamma)(\sigma^2 d\gamma^2 + d\varphi^2) \tag{5.12}$$

It determines the apparent size ds of a horizontal line element given by the bipolar differentials $d\gamma$, $d\varphi$ and observed at a point (γ, φ) of the horizontal plane.

5.2. Observations on Vieth-Muller circles. Our first demonstration is based upon certain observations referring to properties of Vieth-Muller circles. We construct two horizontal bundles of straight lines from the two eye points $y = \pm 1$ of the y-axis. The angles $\Delta\alpha$ and $\Delta\beta$ between neighboring lines shall be constant, i.e., $\Delta\alpha = \Delta\beta =$ constant. (Fig. 42) For the actual demonstration it is advisable to use two bundles of illuminated threads of different color (for example red and green) stretched out in a dark room. However, the effect described in what is to follow can already be observed by simply drawing two bundles of red and green lines on a plane sheet of paper. Let us assume in the following that the red bundle has its vertex at the left eye L and the green bundle at the right eye R. We have indicated in Fig. 42 the green bundle by solid lines and the red bundle by dotted lines. The lines of the two bundles intersect each other in pairs on a set of Vieth-Muller circles which determine on the x-axis points P_0, P_1, P_2, P_3, P_4, etc., accumulating near the origin.

We now bring our eyes into a position exactly above the points R and L, and observe the bundles from this position by converging at the points of the Vieth-Muller circle through P_0. Instead of two bundles we see one bundle of fused lines which intersect the Vieth-Muller circle through P_0 at regular distances. However, these lines do not lie in the horizontal plane; they are space curves intersecting the horizontal plane on points of the Vieth-Muller circle. This effect can be made very striking by observing in a dark room the illuminated threads with a green filter in front of the right eye and with a red filter in front of the left eye. This, of course, has the result of weakening the stimulus of the red lines on the right eye and the stimulus of the green lines on the left eye. If, on the other hand, we observe the colored lines against a white background--as is the case when the lines are drawn on a white sheet of paper--a red filter has to be used in front of the right eye and a green filter in front of the left eye.

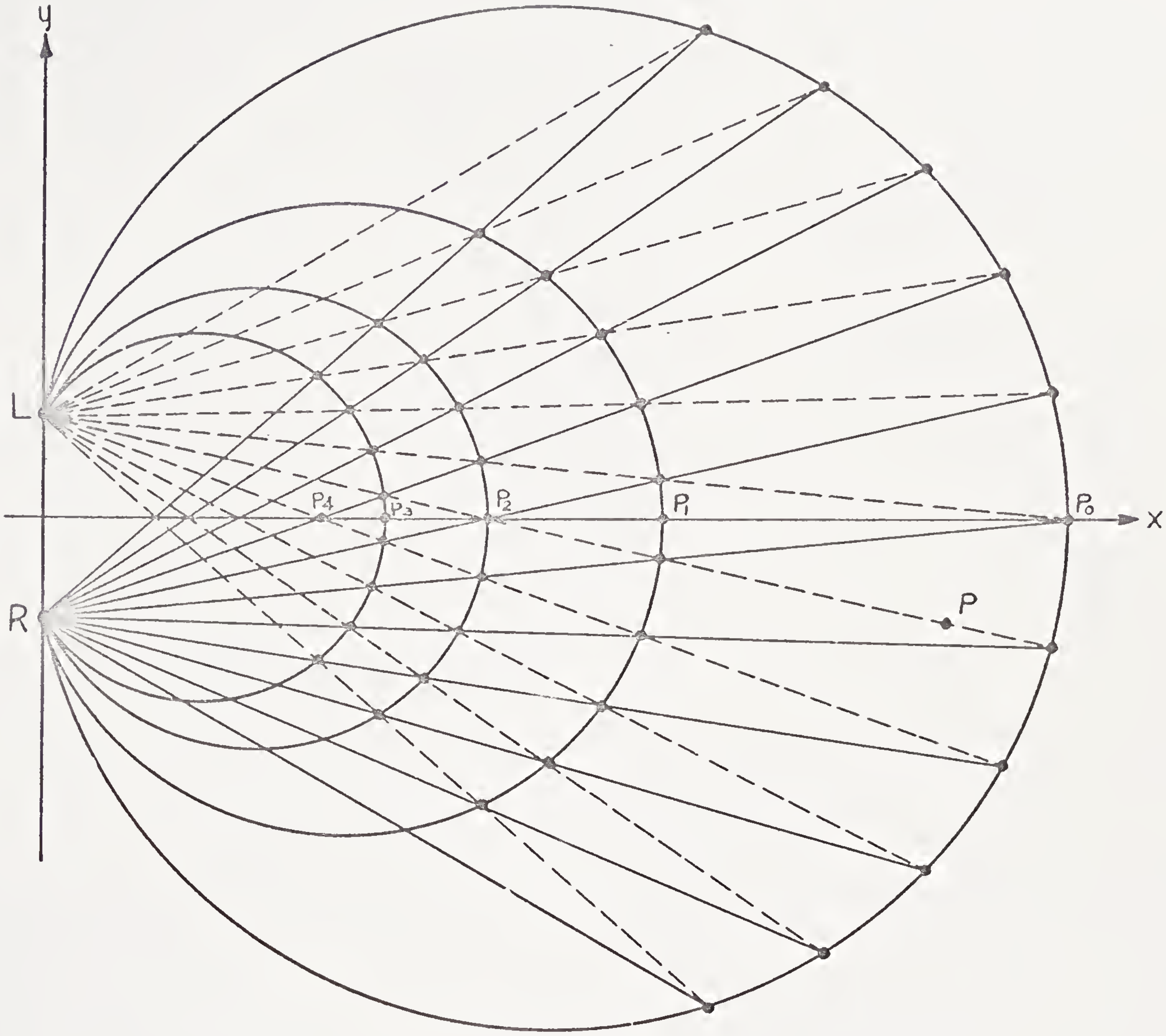


Fig. 42

The appearance of the fused lines is that of straight lines arranged on a circle around the observer. In fact, we know that the Vieth-Muller circle itself appears to the observer as a circle around the point O of the horizontal plane vertically below the apparent center C of observation. We also observe that these straight lines are perpendicular to the lines of view connecting the "egocenter" C with the points of the Vieth-Muller circle. Now, the lines of view form a cone with C as vertex and the Vieth-Muller circle as base; consequently our fused lines are normal to this cone. The surface formed by them must also be a cone with the Vieth-Muller circle as base and a vertex O at some position above the origin. This result is illustrated in Fig. 43. We have to interpret this figure as a Euclidean map of an apparent surface formed in a visual sensation. It represents the <u>interpretation</u> which we give to our sensation, namely, that of a circular cone with its vertex on the Z-axis of a Euclidean, X, Y, Z space. We denote the Cartesian coordinates in this space intentionally X, Y, Z in order to indicate that it is by no means identical with the physical space.

Let us now converge at the points of the Vieth-Muller circle through P_1. We observe the same phenomenon with the difference that the fused lines intersect

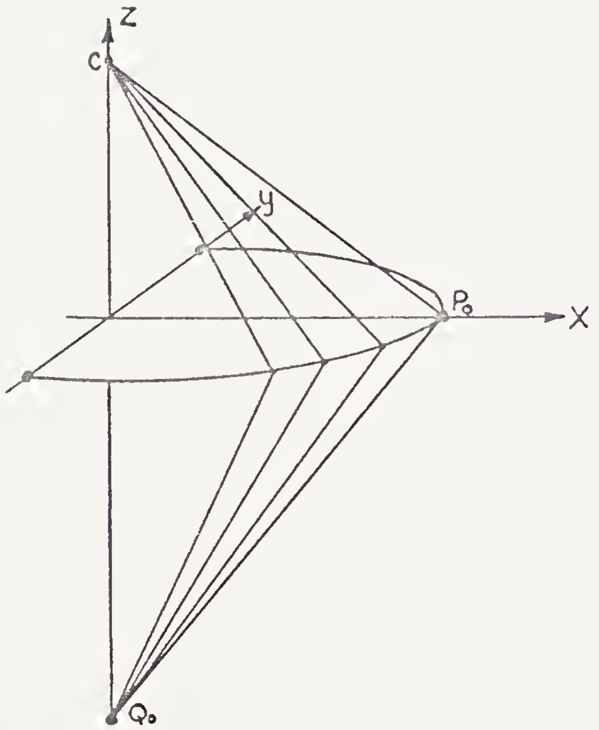


Fig. 43

the horizontal plane at the Vieth-Muller circle through P_1. Again they form a surface interpretable as a Euclidean circular cone with its vertex Q_1 on the Z-axis. This cone is normal to the cone through the observation center C and the Vieth-Muller circle. However, the apparent radius of the base circle has decreased with the effect that the fused lines appear to have a somewhat smaller distance from each other. By changing the convergence of the eyes in succession to the points of the Vieth-Muller circles through P_0, P_1, P_2 . . . . we observe a sequence of circular cones with decreasing apparent radii R_0, R_1, R_2 . . . . and distances of the fused lines decreasing in the same ratio. We point out that this sequence of apparent radii is not proportional to the sequence of physical radii of the Vieth-Muller circles.

The significance of a cone artificially created in the above manner is that it provides the observer with a background towards which he may project the horizontal plane. It establishes the possibility of observing objects in the horizontal plane without changing the convergence of the eyes simply by taking care that the background does not change its position. With a little practice it is possible to scrutinize even objects which are far away from the Vieth-Muller circle of convergence without varying the position of the background, i.e., without changing the convergence of the eyes. The impression associated with this manner of observation is that of a uniocular projection of the horizontal plane towards the conical background, the point C being the center of projection. We shall refer in the following to this peculiar projection as Cyclopean projection. Binocular vision has a twofold office in Cyclopean projection: To create a conical background for uniocular projection, and to determine on the background a linear scale of size which is proportional to the apparent radius R of the base circle of the cone.

In general, any actual point of the horizontal plane is seen as double when observed by Cyclopean projection, i.e., we see two images on the background. Only the points of the basic Vieth-Muller circle itself are seen as single. There exist, however, certain extended curves which do not appear doubled. The Vieth-Muller circles themselves are such curves, since their two projected images coincide on the background. Yet every individual point of these curves has two different points of the image curve as projection. The Vieth-Muller circles appear on the conical background as a system of circles of latitude, which--at least in the neighborhood of the base circle--are nearly equidistant.

In addition to the Vieth-Muller circles there are two other sets of curves which appear single: The two bundles of red and green lines in Fig. 42. Though each line of these bundles has two separated images, we fuse one of them with a generating line of the background and thus see only the other image as an object of the horizontal plane projected upon the background. It becomes a curve on the background which intersects the fused lines at equal apparent angles. The result of viewing the total system of curves drawn in Fig. 42 in Cyclopean projection--namely, the Vieth-Muller and the two bundles through the eyes--is shown in Fig. 44.

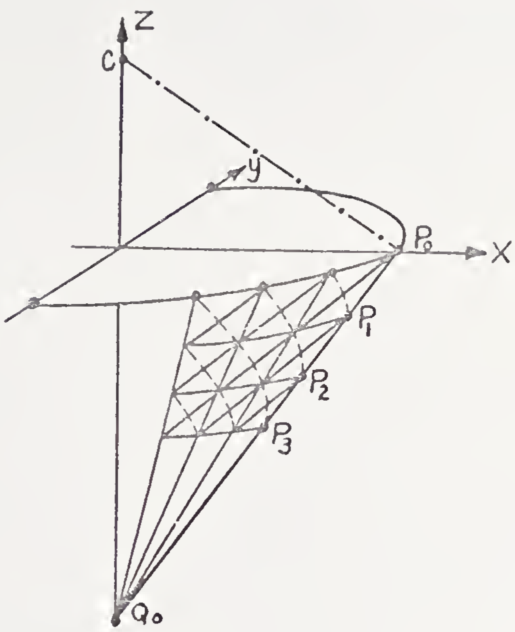


Fig. 44

We remark that the oblique red and green lines in Fig. 44 are the result of a uniocular projection. The green lines are obtained by projecting the green lines of Fig. 42 with the left eye and similarly the red lines by projecting the red lines of Fig. 42 with the right eye. The impression of the observer, however, is in both cases that of uniocular projection from the same apparent projection center C.

In general, it is possible to interpret the above curves also as curves of the horizontal X,Y-plane and not as curves on the cone. However, this alternative disappears more and more when other clues are eliminated in the observation, i.e., when the experiment is carried out in a dark room. We also notice that the compulsion to see the curves on the background becomes greater and greater the lower our eyes are placed with respect to the horizontal plane. Moreover, we base our judgment of size of subdivisions of these lines more and more on their apparent size upon the background, the greater this compulsion becomes to localize our curves on the cone.

The above observations indicate that we may consider the act of assigning size to the horizontal line elements by binocular observation in the horizontal plane as a limiting case of Cyclopean observations from above the plane.

This intimate relationship of binocular vision and Cyclopean projection is more readily understood by the following consideration. Consider the two points P_o and P in Fig. 42. The point P is chosen intentionally on one of the lines through the left eye. By Cyclopean projection we obtain, on the background through P_o, one image of P_o but two images Q and Q' of P (Fig. 45). The image Q' is produced by the left eye; it lies on one of the fused lines of the cone and thus helps to establish the background. Let us now assume that the right eye dominates in localizing objects relative to the observer. Then the image Q will be interpreted as the apparent projection of the point P onto the background, and thus the line element P_oQ on the cone as the projection of the line element P_oP. Since this projection is uniocular, we determine the size of P_oP by the size of its projection P_oQ on the background (Emmert's law).

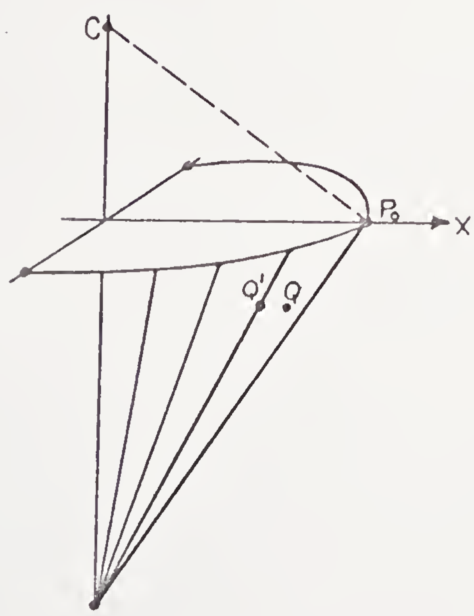


Fig. 45

The above analysis of the mechanism of assigning size to horizontal line elements by binocular vision can be applied to any such line element. Let us consider two arbitrary neighboring points P_o and P of the physical x,y-plane. We connect the two points with the left eye and the point P_o with the right eye. Next we construct the Vieth-Muller circle through P_o and draw a line from its intersection point A with LP towards the right eye (Fig. 46). This configuration is viewed from above with the effect that the point P is seen as

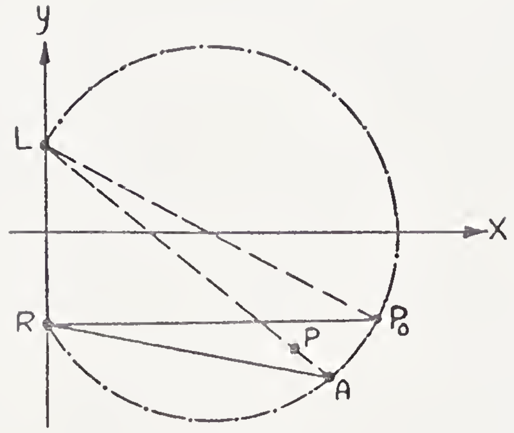


Fig. 46

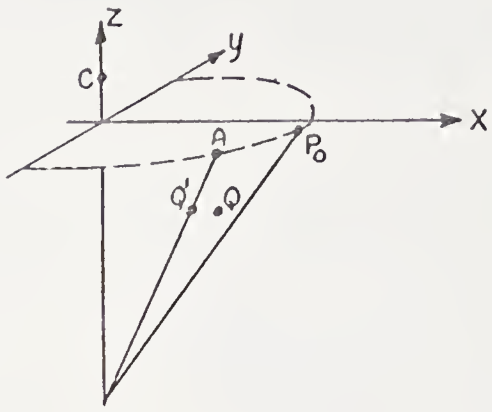


Fig. 47

a point Q on a Cyclopean background. This background is indicated by two fused lines which intersect the horizontal plane at the points P_0 and A of the Vieth-Muller circle. Indeed, as above, the image Q' of P produced by the left eye lies on the fused line through A and thus supports the establishment of the background. The image Q from the dominating right eye determines the Cyclopean projection of P onto the background, and the distance of P_0 from P is estimated by the distance of Q from P_0. A similar construction can obviously be made if the left eye should be the dominating eye and determines the localization of objects relative to the observer. By gradually lowering our point of observation, the result of Cyclopean size estimation becomes gradually the binocular size ds of P_0P based upon the associated differentials $d\gamma$ and $d\varphi$.

We shall now demonstrate that the above interpretation of binocular vision as uniocular projection against a variable conical background induces in the horizontal plane a hyperbolic metric. First we introduce, on the cone, certain angular coordinates ω and ψ. The declination ω is the angle which the line through C and a point Q on the cone includes with the horizontal plane. The azimuth ψ is the angle which a plane through the Z-axis and the point Q includes with the X,Z-plane (Fig. 48). Since ω and ψ are actually spherical coordinates determining the direction of the projection lines through C, we can also characterize the points P of the X,Y-plane by these coordinates. The angle ω determines the apparent declination and ψ the apparent azimuth of a point P of this plane if observed from C.

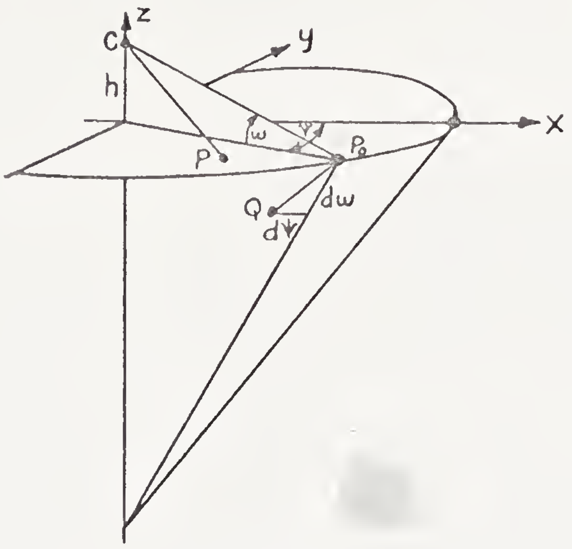


Fig. 48

It is easy to express the distance ds of two neithboring points P_0 and Q on the cone by the above coordinates. We assume that P_0 lies on the base circle of the cone. We find the formula

$$ds^2 = h^2 \left(\frac{d\omega^2}{\sin^2\omega} + \cot^2\omega \, d\psi^2\right) \qquad (5.21)$$

where h is the apparent height of C above the X,Y-plane. Since the distance of P_0Q measures also the apparent distance of the points P_0P in the horizontal plane, we find that the metric introduced in this plane by Cyclopean projection has the form

$$ds^2 = \frac{d\omega^2}{\sin^2\omega} + \cot^2\omega \, d\psi^2 \qquad (5.22)$$

The constant factor h^2 in (5.21) has been dropped, since it is insignificant for metric relations. This means, in other words, that the Cyclopean metric has, independent of the position of the center C of observation, the same mathematical form (5.22) if expressed in the coordinates ω, ψ of the plane.

We transform the differential (5.22) by introducing the apparent distance D:

$$D = \int \frac{d\omega}{\sin \omega} = \log \tan \frac{\omega}{2} \tag{5.23}$$

The result is

$$ds^2 = dD^2 + \sinh^2 D d\psi^2 \tag{5.24}$$

and thus the line element of the hyperbolic geometry. Since the binocular metric of the horizontal plane can be considered as a limit of metrics of Cyclopean projections with decreasing projection heights, and since these Cyclopean metrics all have the same form (5.24), it follows that <u>the binocular metric itself must have the form of the hyperbolic geometry</u>.

It remains to express the differential (5.24) by the bipolar coordinates γ, φ of the x,y-plane. The apparent azimuth ψ, obviously, is identical with the bipolar latitude φ, so that we may write

$$ds^2 = dD^2 + \sinh^2 D d\varphi^2 \tag{5.25}$$

We know, on the other hand, the general form of ds^2 in terms of γ and φ, namely,

$$ds^2 = M^2(\gamma)(\sigma^2 d\gamma^2 + d\varphi^2) \tag{5.12}$$

By identifying the two differentials (5.12) and (5.25) we obtain the relations

$$\begin{aligned} M(\gamma) &= \sinh D \\ \frac{dD}{d\gamma} &= -\sigma \sinh D \end{aligned} \tag{5.26}$$

The last equation yields

$$\tanh \tfrac{1}{2}D = e^{-\sigma(\gamma+\mu)} \tag{5.27}$$

where μ is a certain constant. Since

$$M(\gamma) = \sinh D = \frac{2 \tanh \frac{1}{2} D}{1 - \tanh^2 \frac{1}{2} D}$$

we find by (5.27) that

$$M(\gamma) = \frac{1}{\sinh \sigma(\gamma+\mu)} \tag{5.28}$$

and hence the differential

$$ds^2 = \frac{1}{\sinh^2 \sigma(\gamma+\mu)} (\sigma^2 d\gamma^2 + d\varphi^2) \tag{5.29}$$

This implies that the three-dimensional differential (5.11) must have the form

$$ds^2 = \frac{1}{\sinh^2\sigma(\gamma+\mu)} (\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi d\theta^2) \tag{5.291}$$

and this means that binocular vision establishes a hyperbolic manifold of sensations.

5.3. The interpretation of the Cyclopean background as a cone is not essential for the result of the preceding section. Only a small part around the base circle of the background is used for Cyclopean projection in any instance. Thus any surface which coincides with this part of the cone would lead to the same result, i.e., to the same metric differential

$$ds^2 = \frac{d\omega^2}{\sin^2\omega} + \cot^2\omega\, d\psi^2 \tag{5.31}$$

expressed in apparent declination ω and apparent azimuth ψ. For example, a sphere around the center of observation, C, which intersects the X,Y-plane along the base circle serves our purpose equally well. The fused lines would be interpreted as meridians on this sphere and the Vieth-Muller circles as circles of latitude. This is readily understood by the fact that in a uniocular view such a system of curves allows a multitude of interpretations, for example, that of a sphere or a cone. As we have seen, Cyclopean projection is interpreted by the observer as such a uniocular view.

In order to transform the quadratic differential (5.31) into the final form (5.29), we have made use of a former result, namely, that in the bipolar system γ, φ, it must have the general form (5.12). Independent of this result we can derive the desired transformation as follows. We remark as before that the apparent azimuth is identical with the bipolar latitude, i.e., we have $\psi = \varphi$. We also know by observation that the Vieth-Muller circles γ = const. are imaged as circles of latitude ω = const. upon the background. Hence it follows that ω must be a function of γ alone: $\omega = \omega(\gamma)$. The problem is to determine this relation of the apparent declination ω to the bipolar parallax γ.

We consider for this purpose the two bundles of red and green lines in Fig. 42. Since α = const. and β = const. in these bundles, we conclude that their equations in the γ, φ-plane are

$$\begin{aligned} \gamma - 2\varphi &= \text{const.} \\ \gamma + 2\varphi &= \text{const.} \end{aligned} \tag{5.32}$$

In differential form we may write

$$\frac{d\gamma}{d\varphi} = \pm 2 \tag{5.33}$$

In the γ, φ-plane these curves thus are represented by two systems of equidistant parallel lines (Fig. 49). If the two red and green bundles of Fig. 42 are observed by Cyclopean projection from above the plane, then they appear, as we have seen, as curves on the background which intersect the meridians φ = const. at constant apparent angles. By interpreting the background as a sphere around C,

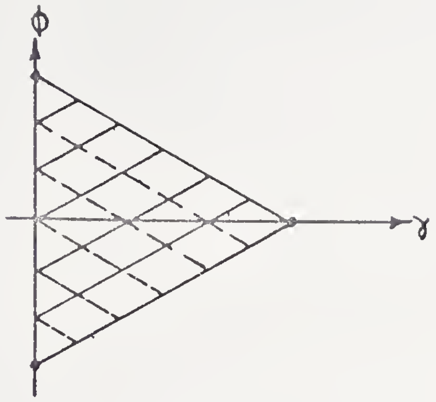


Fig. 49

the apparent angles are equal to the corresponding angles on the sphere itself and our curves may be interpreted as loxodromes on the sphere. These are the curves on which a ship travels without changing its course, i.e., its angle with the meridians. They are characterized by the condition

$$\frac{d\omega}{\cos\omega\, d\varphi} = \text{const.} \tag{5.34}$$

which states nothing but that the tangent of the angle with the meridians is a constant along the curve. If the background is interpreted as a cone, the apparent angle with the meridians is not equal to actual angles on the cone but still its tangent is given by the expression (5.34)

We may formulate our observation as follows: A physical curve on which $\frac{d\gamma}{d\varphi} = \pm 2$ is seen in Cyclopean projection as a curve on which $\frac{d\omega}{\cos\omega\, d\varphi} = \text{const.}$ As ω is a function of γ alone, it follows from

$$\frac{d\omega}{\cos\omega\, d\varphi} = \frac{\omega'(\gamma)}{\cos\omega(\gamma)}\,\frac{d\gamma}{d\varphi} \tag{5.35}$$

that on any particular straight line $\frac{d\gamma}{d\varphi} = 2$ of Fig. 49, the expression $\frac{\omega'(\gamma)}{\cos\omega(\gamma)}$ remains constant. Since this expression is a function of γ alone, it follows that it must be independent of γ, i.e., we have

$$\frac{\omega'(\gamma)}{\cos\omega(\gamma)} = \sigma \tag{5.36}$$

where σ is a constant.

By integration it follows that

$$\log\tan\left(\frac{\pi}{4} - \frac{\omega}{2}\right) = -\sigma(\gamma+\mu) \tag{5.37}$$

where μ is another constant. We may also write

$$\tan\left(\frac{\pi}{4} - \frac{\omega}{2}\right) = e^{-(\gamma+\mu)}$$

and hence

$$\tan\left(\frac{\pi}{2} - \omega\right) = \cot\omega = \frac{1}{\sinh\sigma(\gamma+\mu)} \tag{5.38}$$

By (5.31) and (5.38) it follows that the Cyclopean line element (5.31) expressed in γ and φ must have the form

$$ds^2 = \frac{1}{\sinh^2\sigma(\gamma+\mu)}\,(\sigma^2\, d\gamma^2 + d\varphi^2) \tag{5.39}$$

in agreement with our former result.

5.4. The hyperbolic metric of visual sensations and the relativistic metric of time-space manifolds. In the preceding discussion we have tacitly assumed that the physical space is a Euclidean manifold. We are certainly justified in making this assumption for the domain in which binocular vision is of practical importance to us. Still, our conclusion that visual sensations form a non-Euclidean manifold seems to imply a result which is unsatisfactory: Visual and physical perception of our external environment are contradictory to each other. From a principal point of view it is therefore significant that it is possible to remove this contradiction. We shall show in this section that, in fact, the two perceptions are not contradictory: Physical observation leads to a hyperbolic metric if discussed from a relativistic point of view. On the other hand, this result will give us new evidence for the hyperbolic metric of binocular vision.

We confine ourselves to observations of the horizontal plane. Let us assume that a uniocular observer is placed at the height h above the horizontal plane. Objects in this plane are revealed to him by light signals emitted from the point P of the plane. The light signals which reach C at the time $t = 0$ give to the observer the clues for his visual sensation and are combined to an apparent distribution of objects in this plane at one particular time moment of his individual time scale.

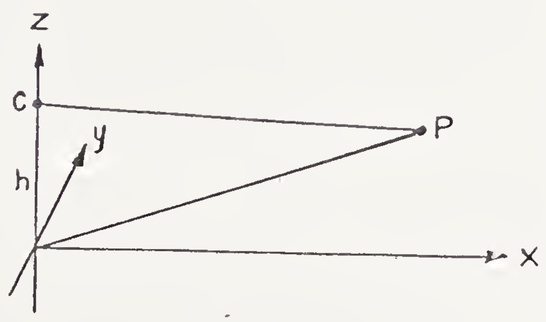


Fig. 50

It is clear that in our assumed x, y, z, coordinate system this combination does not refer to simultaneous events. The light signal which reaches the observer C at the time $t = 0$ was emitted from P at the time

$$t = -\frac{1}{c}\sqrt{x^2 + y^2 + h^2} \tag{5.41}$$

where c is the velocity of light.

Consequently: The physical space-time manifold which is knitted together by the observer at $t = 0$ to one sensation of space is not the manifold $z = 0$, $t = 0$ in our physical coordinate system. However, it is the two-dimensional manifold.

$$\begin{aligned} c^2t^2 - x^2 - y^2 &= h^2 \\ z &= 0 \end{aligned} \tag{5.42}$$

of the physical x, y, z, t space.

The metric relations of the four-dimensional space-time world are now determined by the general line element

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2 \tag{5.43}$$

according to the theory of relativity. This is a "pseudo-Euclidean" line element; it has, like Euclidean line elements, the form of an algebraic sum of the squares of the differentials. It would be truly Euclidean if $-c^2dt^2$ could be replaced by $+c^2dt^2$.

The general line element (5.43) now determines the metric of any manifold of less dimensions which is embodied into the four-dimensional space-time world. It does this in a similar way as the Euclidean line element of the three-dimensional space

$$ds^2 = dx^2 + dy^2 + dz^2$$

in the case of surfaces or curves embodied in the Euclidean space.

Which metric is now introduced by (5.43) in the two-dimensional manifold (5.42)? We first write this manifold in the following parametric form

$$\begin{aligned} ct &= -h \cosh D \\ x &= h \sinh D \cos \varphi \\ y &= h \sinh D \sin \varphi \\ z &= 0 \end{aligned} \qquad (5.44)$$

where D and φ are the variable parameters.

We have

$$\begin{aligned} cdt &= -h \sinh D\, dD \\ dx &= h(\cosh D \cos \varphi\, dD - \sinh D \sin \varphi\, d\varphi) \\ dy &= h(\cosh D \sin \varphi\, dD + \sinh D \cos \varphi\, d\varphi) \\ dz &= 0 \end{aligned} \qquad (5.45)$$

and hence

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2 = h^2(dD^2 + \sinh^2 D d\varphi^2) \qquad (5.46)$$

i.e., the metric of the hyperbolic geometry.

We thus cannot be surprised that the geometry in which our visual sensations form themselves is the non-Euclidean hyperbolic geometry. Indeed, our result means that already the physical events which are the direct basis for our instantaneous sensation are forming in the physical world a manifold with a hyperbolic metric.

5.5. The preceding theoretical considerations give us strong evidence that the psychometric of our spatial sensations is determined by the hyberbolic differential

$$ds^2 = \frac{1}{\sinh^2\sigma(\gamma+\mu)} (\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi d\theta^2)$$

The evidence is certainly strong enough to justify deriving the mathematical implications of the above metric for binocular vision. If these consequences then represent observable and measurable phenomena we are given the means of obtaining additional evidence for the support of our theory.

We shall apply our theory in the following to three significant phenomena of binocular vision; (1) The problem of the Frontal Plane Horopter where curves which appear as straight are physically curved. (2) Hillebrand's Alley Problem where walls which are physically not parallel are seen as a parallel alley. (3) The problem of the distorted room which appears to be rectangular.

We shall see that all these phenomena are a direct consequence of our theory and that they provide us with the possibility of testing the theory by the results of measurements. We shall also find, by comparing the results of measurements with the implications of the elliptic and Euclidean metrics that only the hyperbolic geometry can explain certain facts of observation.

Section 6

GEODESIC LINES: THE HOROPTER PROBLEM

6.1. With the aid of the quadratic differential

$$ds^2 = M^2(\gamma)(\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2\varphi d\theta^2) \qquad (6.11)$$

we are in a position to assign a linear size to the impinging characteristics $d\gamma$, $d\varphi$, $d\theta$ of a small section of a line. If we converge at a point P_0 of the physical space, a rectilinear Cartesian coordinate system in the neighborhood of P_0 is established in which $\sigma M d\gamma$, $M d\varphi$, $M\cos\varphi d\theta$ are the coordinates of a neighboring point P_1. To this neighborhood of P_0 we then apply a Euclidean metric which leads to the expression (6.11). By moving our eyes to another point P a similar Euclidean system is formed for the neighborhood of this point, but different yardsticks are used for the evaluation of the physical quantities $d\gamma$, $d\varphi$, $d\theta$. Observation with eye movements thus requires connecting the results of these different measurements to a unity, i.e., to a sensation of a finite section of space.

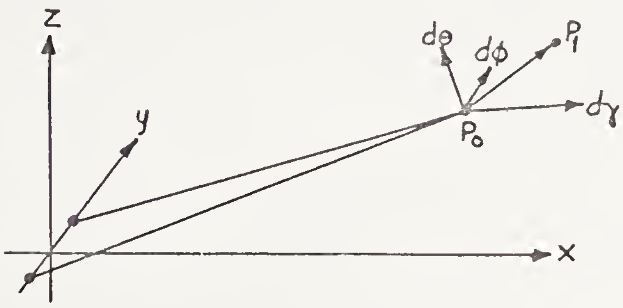


Fig. 51

Mathematically this leads to the problem of integration of the above differential (6.11). Since mathematical integration always gives a set of solutions, i.e., contains parameters which may be assigned arbitrarily, we recognize the possibility of different interpretations of the same impinging characteristics. The former experience of the observer, his present purposes and other psychological factors will influence him in the choice of such arbitrary elements in integrating the primitive sensations (6.11) to a unity. It thus can be seen that the hypothesis of immutable elements in our sensations is not contradictory to an unlimited variety of results in integrating a manifold of immutable primitive sensations to a total unity.

There are, however, certain integration processes in which arbitrary interpretation has no place. The curves which determine the shortest connection between two points or the curves which are the result of attaching line elements to each other without change of apparent direction are uniquely determined by the differential (6.11). These geodesic lines of the metric (6.11), i.e., the apparent straight lines in our sensations, are invariant elements and do not allow other interpretations. For the absolute localization in the apparent space we have still a free choice. However, the impression that they are straight, i.e., that they are geodesic lines, is enforced by the metric differential (6.11) itself and thus beyond our control.

Curves which are apparently straight are of great interest in visual science. Helmholtz noticed the fact that vertical threads arranged by an observer in fixed head position to form an apparent frontal parallel plane are not actually arranged on a physical plane. The shape of the physical surface which appears to be plane varies with the distance from the observer. The shape of their cross section with the horizontal plane is schematically shown in Fig. 40. We shall apply our theory to this problem, and by identifying the horopter curves with geodesic lines of our metric (6.11), we shall find that the geodesic lines of the horizontal plane which are symmetrical to the x-axis form a set of curves exactly of the observed type.

Another classical problem is also closely related to the geodesic lines of our metric: Hillebrand's Alley Problem. If we determine on the frontal plane horopters points which have the same geodesic distance from the x-axis, two curves are obtained (Fig. 52) which have the same apparent (geodesic) distance from each other. They thus should give the impression of an alley formed by equidistant walls. Again we shall find that the curves which follow from our theory are of the type observed by Hillebrand, who first made the experiment.

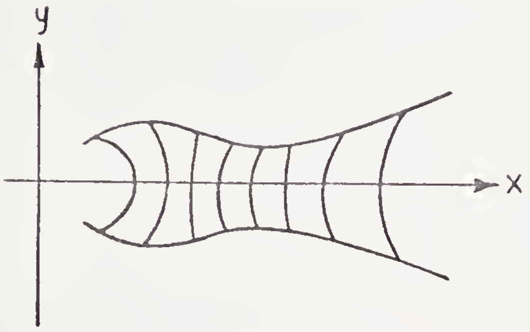


Fig. 52

Instead of constructing two equidistant curves, we may also ask for curves which are straight and parallel to each other and symmetric to the x-axis. We shall see that these two sets of curves are only identical if our metric (6.11) is Euclidean. They are different curves in the elliptic and hyperbolic geometry. Our mathematical result will be that in the hyperbolic geometry the curves of equal distance lie outside the curves of parallel direction, but that the situation is reversed in the elliptic geometry. Blumenfeld, who repeated Hillebrand's experiments with greater precision, found that the two instructions lead the observer to different curves, with the curves of equal distance being outside the curves of parallel direction. This experimental result we then have to consider as additional strong evidence that the metric of binocular space sensations is given by the hyperbolic differential (6.11) with

$$M(\gamma) = \frac{1}{\sinh \sigma(\gamma + \mu)} \tag{6.12}$$

6.2. We determine the geodesic lines of the metric (6.11) as solutions of the problem of variation

$$M(\gamma) \sqrt{\sigma^2 \, d\gamma^2 + d\varphi^2 + \cos^2\varphi \, d\theta^2} = \text{Minimum} \tag{6.21}$$

We may solve this problem simultaneously for all the three geometries by writing $M(\gamma)$ in the form

$$M(\gamma) = \frac{2}{e^{\sigma(\gamma+\mu)} + \varepsilon e^{-\sigma(\gamma+\mu)}} \tag{6.22}$$

where $\varepsilon = -1, 0, +1$ for the hyperbolic, Euclidean, and elliptic geometry respectively.

For the discussion of the solutions it will be advantageous to use, instead of γ, φ, θ other variables

$$\xi = \rho \cos\varphi \cos\theta$$
$$\eta = \rho \sin\varphi \qquad (6.23)$$
$$\zeta = \rho \cos\varphi \sin\theta$$

where

$$\rho = \sqrt{\xi^2 + \eta^2 + \zeta^2} = e^{-\sigma(\gamma+\mu)} \qquad (6.24)$$

The problem (6.21), in these variables, assumes the spherically symmetrical form

$$2\int \frac{\sqrt{d\xi^2 + d\eta^2 + d\zeta^2}}{1 + \varepsilon\rho^2} = \text{Minimum} \qquad (6.25)$$

related to Riemann's normal form

$$ds^2 = \frac{4}{(1 + \varepsilon\rho^2)^2} \, (d\xi^2 + d\eta^2 + d\zeta^2)$$

of the metric of manifolds of constant curvature.

We see from (6.25) that, in the ξ, η, ζ space, our problem is formally identical to finding the light rays in a medium of index of refraction

$$n = \frac{1}{1 + \varepsilon\rho^2} \qquad (6.26)$$

It is a medium in which the optical substances are arranged in concentric spherical layers around the origin.

In case $\varepsilon = 0$ we have a homogeneous medium $n = 1$, and the light rays are the straight lines of the ξ, η, ζ space.

In case $\varepsilon = 1$ (elliptic geometry) we obtain

$$n = \frac{1}{1 + \rho^2} \qquad (6.27)$$

a medium well known as "Maxwell's Fisheye."

Finally the hyperbolic geometry, $\varepsilon = -1$, may be represented optically by a medium

$$n = \frac{1}{1 - \rho^2} \qquad (6.28)$$

in which the index of refraction increases from 1 to ∞ if ρ varies from 0 to 1. (Poincaré's model of hyperbolic geometry.)

The interpretation of the three geometries as an optical medium of spherical symmetry makes it quite obvious that, in the ξ, η, ζ space, the geodesic lines (light rays) are plane curves. Indeed, a light ray in such a medium will remain in the plane which is determined by the origin and any one of the line elements of the ray. Thus if we have found all the geodesic lines in the ξ, η-plane, we can find all other geodesic lines simply by rotating this plane around the origin into any other position. In other words, we may consider our problem as solved if we know the geodesic lines in the ξ,η-plane.

We remark, however, that the actual horopters in the physical x, y, z space are in general not plane. We obtain these horopters by expressing the equation of the geodesic lines in the ξ, η, ζ space first by the angular coordinates γ, φ, θ and then with the aid of the relations (2.23) or (2.31) by the physical coordinates x, y, z. The resulting curves are plane only if, in the ξ, η, ζ space, the geodesic line lies in a plane through the η-axis; indeed, in this case, they lie in the same plane of elevation θ = const. However, if a geodesic line is sought which connects two points in different planes of elevation, then the resulting curve is plane only in exceptional cases (for example, if the two points lie in the median plane $y = 0$).

We shall confine ourselves in the following to the plane geodesic lines, in particular to those which lie in the horizontal plane; the frontal plane horopter and the alley curves are curves of this type.

6.3. Before solving the problem of variation (6.21) or (6.25), let us study the transformation of the horizontal x, y-plane into the ξ, η-plane which is expressed by (6.23). As in §2, we do this by constructing the domain in the ξ, η-plane into which the half-plane $x \geqq 0$ of the horizontal plane is transformed.

Since γ can vary from π to zero, it follows that $\rho = \sqrt{\xi^2 + \eta^2}$ may vary from

$$\rho_0 = e^{-\sigma(\pi+\mu)}$$

to

$$\rho_1 = e^{-\sigma\mu} \tag{6.31}$$

Consequently the half-plane $x > 0$ must be imaged into some domain inside the circular ring enclosed by two circles of radius ρ_0 and ρ_1. The latter circle represents the infinity of the horizontal plane ($\gamma = 0$). Both ρ_0 and ρ_1 are smaller than one.

We have seen that the two eyes, in the γ, φ-plane, are given by the straight lines

$$\begin{aligned} 2\varphi + \gamma &= \pi \\ -2\varphi + \gamma &= \pi \end{aligned} \tag{6.32}$$

By $\rho = e^{-\sigma(\gamma+\mu)}$ it follows that, in the ξ, η-plane, these eyes are represented by the <u>logarithmic spirals</u>

$$\begin{aligned} \rho &= \rho_0 e^{2\sigma\varphi} \\ \rho &= \rho_0 e^{-2\sigma\varphi} \end{aligned} \tag{6.33}$$

This leads to the boundary coordination illustrated in Fig. 53. The half-plane $x > 0$ is imaged into a sickle-shaped figure which lies inside the unit circle of the ξ, η-plane.

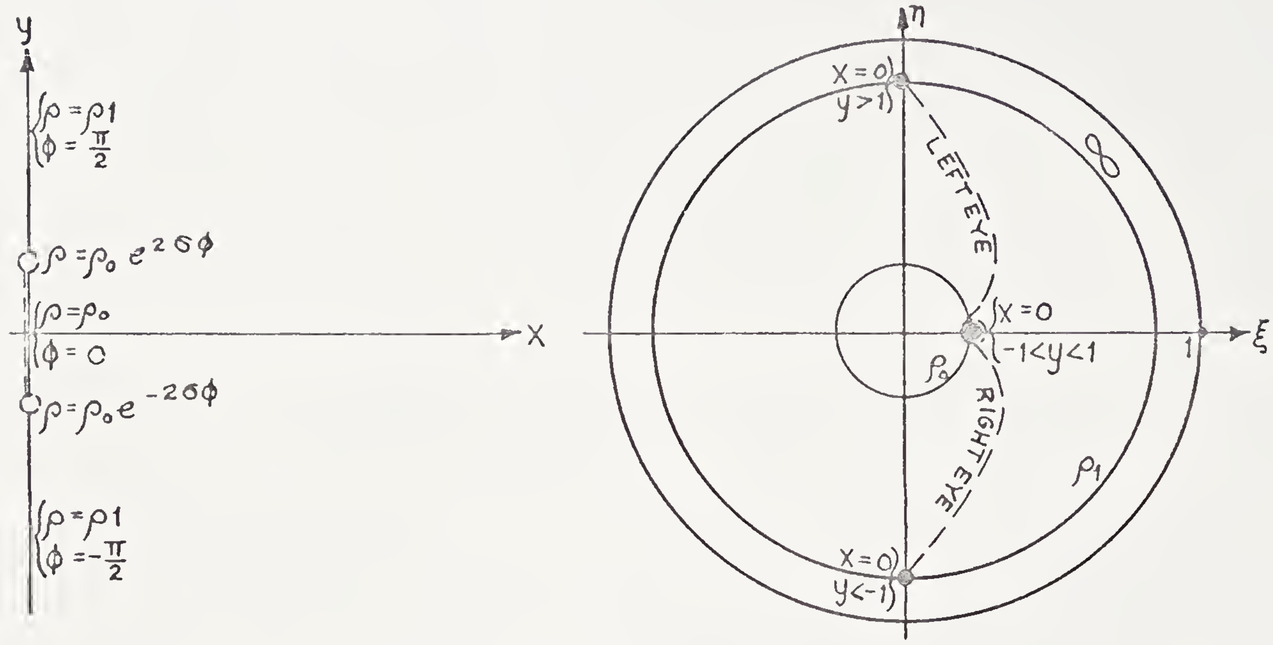


Fig. 53

6.4. The geodesic lines of the horizontal plane (Horopters). The geodesic lines of the horizontal plane are the solutions of the problem of variation

$$\int M(\gamma)\ \sqrt{\sigma^2\ d\gamma^2 + d\varphi^2} = \text{Minimum} \tag{6.41}$$

or, in terms of the coordinates ξ, η:

$$\int \frac{\sqrt{d\xi^2 + d\eta^2}}{1 + \varepsilon\rho^2} = \text{Minimum} \tag{6.411}$$

It is mathematically a little simpler to treat the problem in the first form (6.41). For this purpose we introduce the variable

$$\tau = \sigma(\gamma + \mu) \tag{6.42}$$

and consider φ as function of τ. Then it follows

$$\int M(\tau)\ \sqrt{1 + \varphi'^2}\ d\tau = \text{Minimum} \tag{6.43}$$

where

$$M(\tau) = \frac{1}{e^{\tau} + \varepsilon e^{-\tau}}$$

The solutions of the problem, i.e., the geodesic lines, must satisfy Euler's differential equation which in our case has the simple form

$$\frac{d}{d\tau}\left(\frac{M\varphi'}{\sqrt{1 + \varphi'^2}}\right) = 0$$

It follows

$$\frac{M\varphi'}{\sqrt{1 + \varphi'^2}} = \text{const.} = C$$

or

$$\varphi' = C\ \frac{\frac{1}{2}(e^{\tau} + \varepsilon e^{-\tau})}{\sqrt{1 - \frac{1}{4}C^2(e^{\tau} + \varepsilon e^{-\tau})^2}} \tag{6.44}$$

With the aid of the relations

$$\tfrac{1}{2}(e^{\tau} + \varepsilon e^{-\tau})' = \tfrac{1}{2}(e^{\tau} - \varepsilon e^{-\tau})$$

$$\tfrac{1}{2}(e^{\tau} - \varepsilon e^{-\tau})' = \tfrac{1}{2}(e^{\tau} + \varepsilon e^{-\tau}) \qquad (6.45)$$

$$\tfrac{1}{4}(e^{\tau} + \varepsilon e^{-\tau})^2 - \tfrac{1}{4}(e^{\tau} - \varepsilon e^{-\tau})^2 = \varepsilon$$

we can write (6.44) in the form

$$\varphi' = \frac{\tfrac{1}{2}(e^{\tau} - \varepsilon e^{-\tau})'}{\sqrt{k^2 - \tfrac{1}{4}(e^{\tau} - \varepsilon e^{-\tau})^2}} \qquad (6.46)$$

where

$$\frac{1}{k} = -\frac{C}{\sqrt{1 - \varepsilon C^2}}$$

is again a constant. The condition (6.46) can be integrated directly and leads to the equation

$$\varphi = \varphi_0 + \text{arc cos}\,\frac{1}{2k}(e^{\tau} - \varepsilon e^{-\tau})$$

or

$$2\,k\cos(\varphi - \varphi_0) = e^{\tau} - \varepsilon e^{-\tau} \qquad (6.47)$$

in which k and φ_0 are arbitrary constants of integration. We thus obtain, reintroducing γ by $\tau = \sigma(\gamma + \mu)$ the two-parameter set of geodesic lines

$$2\,k\cos(\varphi - \varphi_0) = e^{\sigma(\gamma+\mu)} - \varepsilon e^{-\sigma(\gamma+\mu)} \qquad (6.48)$$

depending upon the parameters k and φ_0.

By now giving to ε the values -1, 0, +1, we have the result that the horizontal geodesic lines of the three geometries of constant curvature are given by the following curves:

$$\begin{aligned} &\text{Hyperbolic geometry:} && \cosh\sigma(\gamma+\mu) = k\cos(\varphi - \varphi_0) \\ &\text{Euclidean geometry:} && e^{\sigma(\gamma+\mu)} = 2k\cos(\varphi - \varphi_0) \\ &\text{Elliptic geometry:} && \sinh\sigma(\gamma+\mu) = k\cos(\varphi - \varphi_0) \end{aligned} \qquad (6.49)$$

Since $k = \infty$ is a permissible constant, we conclude that the hyperbolae φ = const. are geodesic lines in all three geometries.

6.5. We are especially interested in those geodesic lines which are symmetrical to the x-axis, since these curves represent the <u>frontal plane horopters</u> of binocular vision. Since symmetry to the x-axis means that the two values $\pm\varphi$ must belong to any value of γ, it follows that the parameter φ_0 in (6.49) must have the value $\varphi_0 = 0$.

The frontal plane horopters thus are given by the curves

$$\begin{aligned} &\text{Hyperbolic geometry:} && \cosh\sigma(\gamma+\mu) = k\cos\varphi \\ &\text{Euclidean geometry:} && e^{\sigma(\gamma+\mu)} = 2\,k\cos\varphi \\ &\text{Elliptic geometry:} && \sinh\sigma(\gamma+\mu) = k\cos\varphi \end{aligned} \qquad (6.51)$$

The constant k characterizes the point x_0 where the horopter intersects the x-axis. Indeed, for $\varphi = 0$ we have in the three cases

$$\begin{aligned} k &= \cosh \sigma(\gamma_0 + \mu) \\ k &= \tfrac{1}{2}e^{\sigma(\gamma_0 + \mu)} \\ k &= \sinh \sigma(\gamma_0 + \mu) \end{aligned} \tag{6.52}$$

where γ_0 then determines the point x_0 by

$$x_0 = \cot \tfrac{1}{2}\gamma_0 \tag{6.53}$$

With the aid of (6.52) we may write the frontal plane horopters in the form

$$\begin{aligned} &\text{Hyperbolic geometry:} && \frac{\cosh \sigma(\gamma + \mu)}{\cosh \sigma(\gamma_0 + \mu)} = \cos \varphi \\ &\text{Euclidean geometry:} && e^{\sigma(\gamma - \gamma_0)} = \cos \varphi \\ &\text{Elliptic geometry:} && \frac{\sinh\sigma(\gamma + \mu)}{\sinh\sigma(\gamma_0 + \mu)} = \cos \varphi \end{aligned} \tag{6.54}$$

Since the right side cannot be larger than one, and since on the left side we have functions of γ which are monotically increasing, we conclude that on all of these curves

$$\gamma \leq \gamma_0 \tag{6.55}$$

In other words: <u>The frontal plane horopters lie outside the Vieth-Muller circle through the median point</u> x_0 <u>of the horopters</u> (Fig. 54).

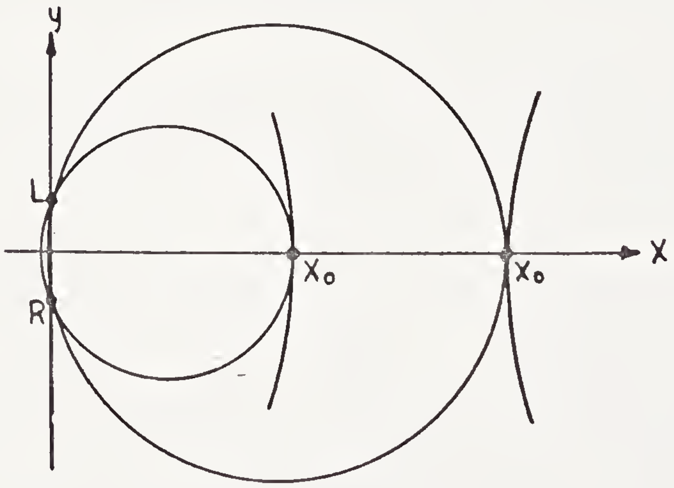


Fig. 54

6.6. The geodesic lines of the horizontal plane assume a remarkably simple form if the ξ, η map of our geometries is chosen. We know that these geodesic lines then are given by the light rays in a medium of index of refraction

$$n = \frac{1}{1 + \varepsilon\rho^2} \tag{6.61}$$

We obtain these "light rays" by introducing $\rho = e^{-\sigma(\gamma+\mu)}$ in the general equation (6.48). This gives

$$2k\ \rho \cos(\varphi - \varphi_o) = 1 - \varepsilon\rho^2 \tag{6.62}$$

or

$$\xi(\xi^2 + \eta^2) + 2(\xi.\xi_o + \eta.\eta_o) = 1 \tag{6.63}$$

where

$$\xi_o = k \cos \varphi_o,\ \eta_o = k \sin \varphi_o.$$

Obviously these curves are circles in case of the hyperbolic or elliptic geometries ($\varepsilon = \pm 1$), and straight lines in the Euclidean case ($\varepsilon = 0$).

Without loss of generality we may assume $\eta_o = 0$ and thus discuss only the frontal plane horopters symmetrical to the ξ-axis. Indeed, all other geodesic lines may be obtained from these special geodesic lines simply by rotation around the origin $\xi = \eta = 0$.

In the hyperbolic case, $\varepsilon = -1$, we write the resulting equation in the form

$$(\xi - \xi_o)^2 + \eta^2 = \xi_o^2 - 1 \tag{6.64}$$

and recognize a circle around a point $\xi_o > 1$ of the ξ-axis which intersects the unit circle at right angles (Fig. 55). The geodesic lines of the hyberbolic geometry thus are resresented in the ξ, η map of this geometry by the circles which intersect the unit circle at right angles.

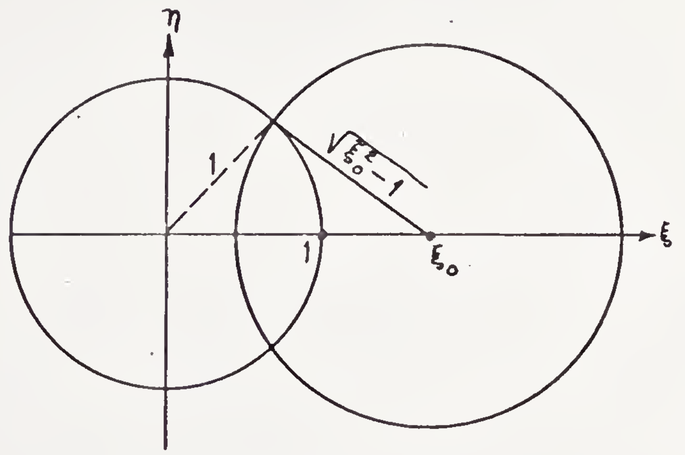


Fig. 55

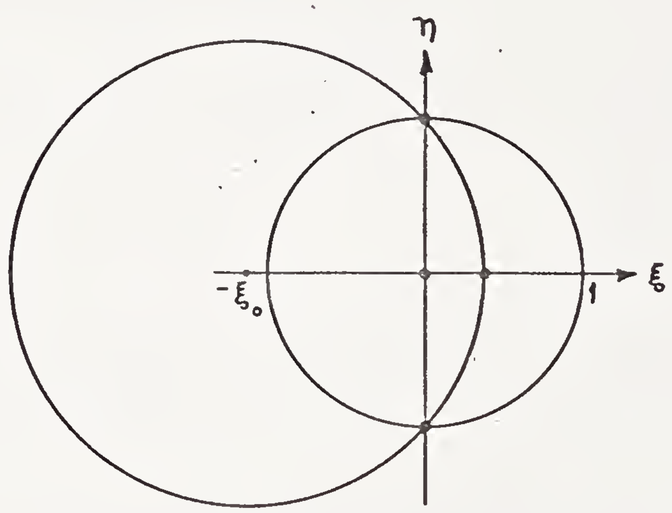


Fig. 56

In the elliptic case we have

$$(\xi + \xi_o)^2 + \eta^2 = 1 + \xi_o^2 \tag{6.65}$$

and recognize a circle around the point $-\xi_o$ which goes through the points $\eta = \pm 1$ of the unit circle (Fig. 56). The geodesic lines of the elliptic geometry thus are given by the circles which intersect the unit circle at two points at opposite ends of a diameter.

The geodesic lines of the Euclidean geometry finally are simply the straight lines of the ξ, η-plane (Fig. 57).

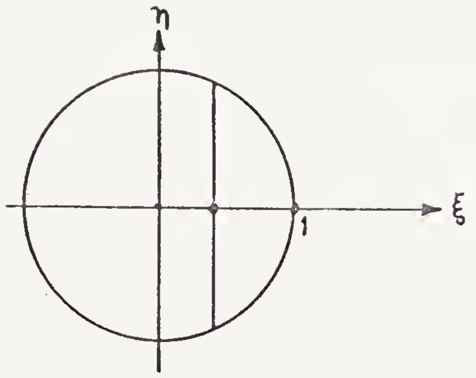


Fig. 57

6.7. The above results can be used in a simple way to determine the general form of the horopter curves in the physical x,y-plane. We have seen (Fig. 53) that the half-plane $x \geq 0$ is represented in the ξ, η-plane by a moon-shaped domain. If we draw in this domain the horopter curves of the hyperbolic geometry, for example, we observe immediately that these curves are divided into two groups (Fig. 58). The curves of the first group go from the median point, ξ_m, directly towards the boundary circle $\rho = \rho_1$ (infinity of the x,y-plane) without leaving the domain. The curves of the second group, however, cross first the red lines (eyes), leave the domain, cross the same

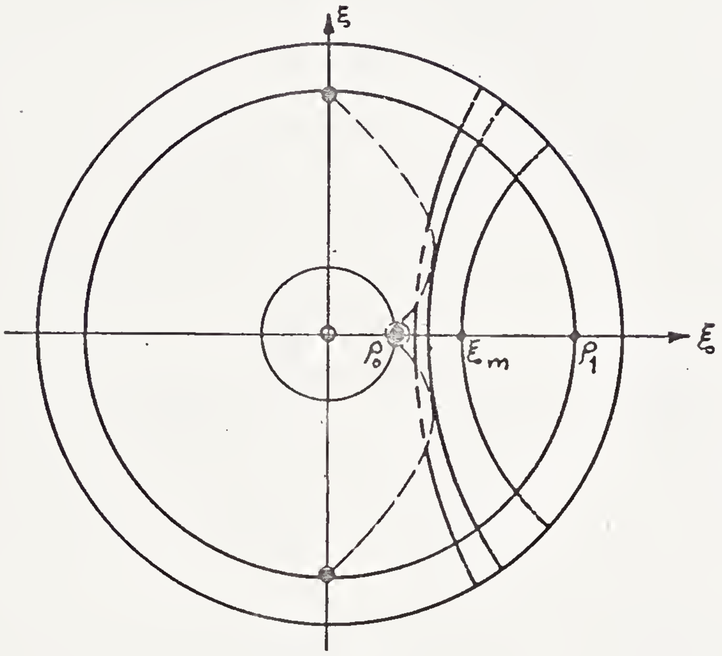


Fig. 58

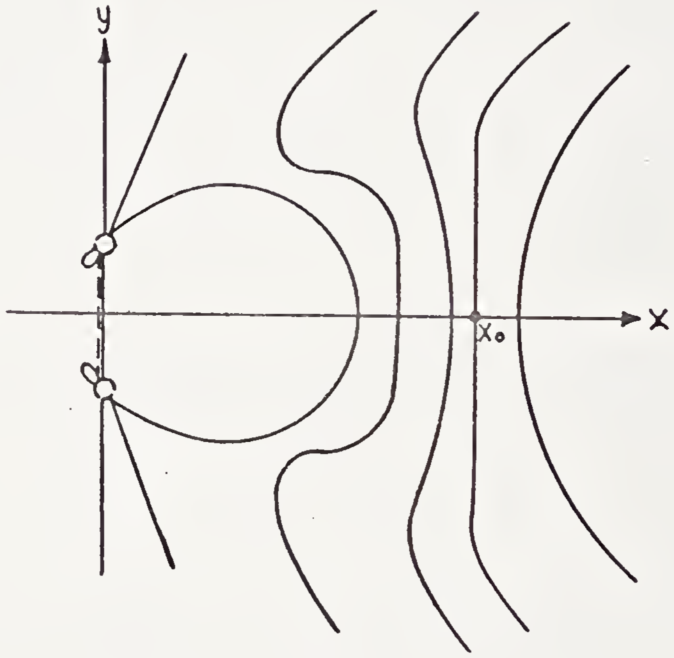


Fig. 59

red lines again, and then reach $\rho = \rho_1$ in the original domain. These two groups are separated by one horopter which just touches the red lines.

This means, in the x,y-plane, that the horopter curves of the first group go from a point x_m of the x-axis without leaving the half-plane $x \geq 0$, towards infinity which is reached asymptotically at a certain angle

$$\varphi_\infty < \pi/2.$$

The curves of the second group go from their median point x_m to the eyes, cross into the left half-plane $x < 0$, loop back through the eyes into the right half-plane, and approach infinity asymptotically. This general behavior implies that, for great values of x_m, the horopters must be convex towards the left and for small values of x_m concave to the left. At a certain median point $x_m = x_o$, the horopter will be nearly flat. This is illustrated in Fig. 59, where a schematic drawing of these curves is given. For a numerically correct picture, refer to Fig. 71.

The geodesic lines of our hyperbolic geometry thus have, in the x,y-plane, a general form which agrees perfectly with the horopter curves determined by experiment in the neighborhood of the x-axis. However, we cannot consider this as evidence for the hyperbolic geometry, since the two other geometries furnish horopter curves with the same phenomenon: changing, in the neighborhood of a point x_0 of the x-axis, from concave to convex curves. This can be seen by replacing in Fig. 58 the geodesic lines of the hyperbolic geometry by the proper geodesic lines of the Euclidean and elliptic geometry (Fig. 60).

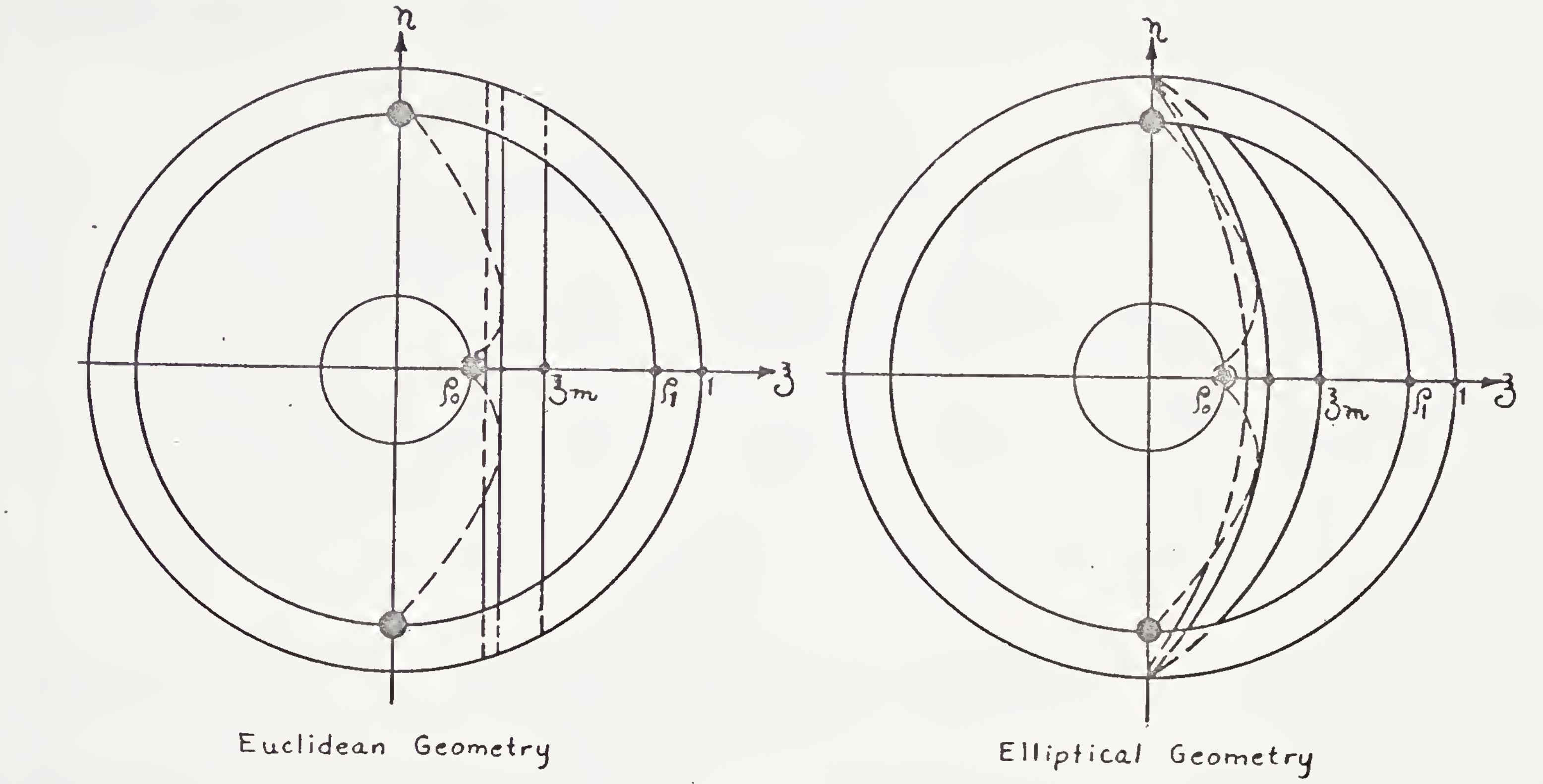


Fig. 60

6.8. The phenomenon that frontal plane horopters are concave in the neighborhood of the observer but convex at greater distances can obviously be used for a numerical test of the theory. We determine for this purpose in this section the curvature of the horopter curves on the x-axis. We carry out the calculation only for the hyperbolic metric.

The frontal plane horopters of the hyperbolic metric are given by the first equation (6.54):

$$\frac{\cosh \sigma(\gamma+\mu)}{\cosh \sigma(\gamma_0+\mu)} = \cos \varphi \tag{6.81}$$

In the neighborhood of $\varphi = 0$, we have

$$\cosh \sigma(\gamma+\mu) = \cosh \sigma(\gamma_0+\mu) + \sigma \sinh \sigma(\gamma_0+\mu)\,\Delta\gamma + \ldots..$$

and

$$\cos\varphi = 1 - \tfrac{1}{2}\varphi^2 + \ldots..$$

Hence we may replace (6.81) by

$$\Delta\gamma = -\frac{1}{2\sigma}\coth \sigma(\gamma_0+\mu)\varphi^2 \tag{6.82}$$

if we are interested only in the shape which the curves have in the immediate neighborhood of the x-axis.

We next have to translate the relation (6.82) into Cartesian x,y coordinates. From the transformation formulae

$$x = \frac{\cos 2\varphi + \cos \gamma}{\sin \gamma} \qquad y = \frac{\sin 2\varphi}{\sin \gamma} \tag{6.83}$$

we conclude for small values of φ that

$$x = \frac{1 - 2\varphi^2 + \cos \gamma_o - \sin \gamma_o \Delta\gamma}{\sin \gamma_o + \cos \gamma_o \Delta\gamma} + \ldots\ldots$$

$$= \frac{1 + \cos \gamma_o}{\sin \gamma_o} \left\{ 1 - \frac{2\varphi^2}{1 + \cos \gamma_o} - \frac{\Delta\gamma}{\sin \gamma_o} + \ldots\ldots \right\} \tag{6.84}$$

The expression on the right side contains all terms of second order in φ since $\Delta\gamma$ is by (6.82) of second order in φ so that $(\Delta\gamma)^2$, etc., may be omitted for our purpose.

By subtracting $x_o = \frac{1 + \cos \gamma_o}{\sin \gamma_o}$ from (6.84), we find

$$\Delta x = - \frac{1 + \cos \gamma_o}{\sin \gamma_o} \left[\frac{2\varphi^2}{1 + \cos \gamma_o} + \frac{\Delta\gamma}{\sin \gamma_o} \right] \tag{6.85}$$

Now, if using (6.81),

$$\Delta\gamma = C\varphi^2 \tag{6.86}$$

we obtain

$$\Delta x = - \frac{1 + \cos \gamma_o}{\sin^2 \gamma_o} (2 \tan \tfrac{1}{2}\gamma_o + C)\varphi^2 \tag{6.87}$$

The second equation (6.83) gives to a sufficient approximation

$$\varphi = \tfrac{1}{2} \sin \gamma_o \, y$$

and thus, by introducing this in (6.87), we have

$$\Delta x = -\tfrac{1}{2} \cos^2 \frac{\gamma_o}{2} (2 \tan \tfrac{1}{2}\gamma_o + C) y^2 \tag{6.88}$$

This is a parabola which has the same curvature K as the original curve (6.86).

We thus may formulate the theorem: <u>If a curve which is symmetrical to the x-axis has, for small values of φ, the development</u>

$$\Delta\gamma = C\varphi^2 + \ldots\ldots$$

<u>then the curvature of this curve on the x-axis is given by the expression</u>

$$K = -\cos^2 \tfrac{1}{2} \gamma_o (2 \tan \tfrac{1}{2}\gamma_o + C) \tag{6.89}$$

With the aid of this result we obtain the curvature of the horopter curves (6.81) by introducing from (6.82) the value

$$C = -\frac{1}{2\sigma}\coth\sigma(\gamma_{\dot{o}} + \mu)$$

It follows

$$K = +\cos^2\tfrac{1}{2}\gamma_o\left\{-2\tan\tfrac{1}{2}\gamma_o + \frac{1}{2\sigma}\coth\sigma(\gamma_o + \mu)\right\} \qquad (6.891)$$

in which γ_o characterizes the point $x_o = \cot\frac{1}{2}\gamma_o$ where the horopter intersects the x-axis.

The sign of K is given by the bracket in (6.891); this bracket approaches $-\infty$ if $\gamma_o \to \pi$, i.e., if $x_o \to 0$. It approaches asymptotically the positive value $\frac{1}{2\sigma}\coth\sigma\mu$ if $x_o \to \infty$, i.e., $\gamma_o \to 0$. Consequently it must be zero at some point x_o in between. For the computation of this special value x_o, we have the equation $K = 0$, i.e., by (6.891):

$$\tan\frac{\gamma_o}{2}\tanh\sigma(\gamma_o + \mu) = \frac{1}{4\sigma} \qquad (6.892)$$

Since $x_o = \cot\frac{1}{2}\gamma_o$, we may write this equation as follows

$$x_o = 4\sigma\tanh\sigma(\gamma_o + \mu) \qquad (6.893)$$

and conclude immediately that

$$x_o < 4 \qquad (6.894)$$

We can use the equation (6.892) and, more generally, the observable curvature relation (6.891) in order to determine the individual constants σ and μ of our hyperbolic metric.

In the next section we shall find another method based on the alley experiments. For one of Blumenfeld's observers (Lo) the constants σ and μ have been calculated on the basis of Blumenfeld's data:

$$\sigma = 14.58$$
$$\mu = 0.0809 \qquad (6.895)$$

We use these values for the calculation of γ_o from the transcendental equation (6.892).

The result is

$$\gamma_o = 2.^\circ 36$$

and hence

$$x_o = \cot\gamma_o/2 = 48.7$$

The interpupillary distance of Blumenfeld's subject is 30.5 mm. Hence we find that for this observer a frontal plane horopter at x_o = 148.2 cm. should have the curvature $K = 0$, i.e., coincide practically with a curve which is physically straight.

6.9. It is easy to carry out similar calculation for the Euclidean and elliptic geometry. From (6.54) it follows

$$\Delta\gamma = -\frac{1}{2\sigma}\varphi^2$$
$$\Delta\gamma = -\frac{1}{2\sigma}\tanh\sigma(\gamma_0+\mu)\varphi^2 \qquad (6.91)$$

for the Euclidean and elliptic horopters respectively. We introduce the constants C which follow from (6.91) in our general theorem (6.89) and obtain formulae for the curvature K of the horopters:

Euclidean Geometry: $K = \cos^2\tfrac{1}{2}\gamma_0\left(\frac{1}{2\sigma} - 2\tan\tfrac{1}{2}\gamma_0\right)$

Elliptic Geometry: $K = \cos^2\tfrac{1}{2}\gamma_0\left(\frac{\tanh\sigma(\gamma_0+\mu)}{2\sigma} - \tan\tfrac{1}{2}\gamma_0\right)$ (6.92)

The curvature K is zero for a value γ_0 which satisfies the equations

$$\tan\tfrac{1}{2}\gamma_0 = \frac{1}{4\sigma}$$
$$\tan\tfrac{1}{2}\gamma_0\coth\sigma(\gamma_0+\mu) = \frac{1}{4\sigma} \qquad (6.93)$$

respectively. By introducing $x_0 = \cot\tfrac{1}{2}\gamma_0$, we obtain

Euclidean Geometry: $x_0 = 4\sigma$

Elliptic Geometry: $x_0 = 4\sigma\coth\sigma(\gamma_0+\mu)$ (6.94)

The difference between the three geometries thus reflects upon the relation of the constant σ to the position x_0 where the frontal plane horopter is physically plane. Indeed, we have the inequalities

$x_0 < 4\sigma$ (Hyperbolic geometry)

$x_0 = 4\sigma$ (Euclidean geometry) (6.95)

$x_0 > 4\sigma$ (Elliptic geometry)

Section 7

THE ALLEY PROBLEM

For the application of our theory to the alley problem, we distinguish with Blumenfeld*) two types of alleys: alleys which are formed by walls of equal apparent distance, and alleys formed by apparently parallel walls. The curves where the walls intersect the horizontal plane are, for brevity, called Distance Curves and Parallel Curves.

Our first problem is to interpret these curves mathematically, i.e., to find mathematical conditions which describe adequately the psychological impressions of equality of distance and of parallelism. This problem is easily solved in the case of the distance curves, but we shall find that the concept of parallelism involves certain difficulties. The reason for this difficulty can be seen in the fact that parallelism of lines has an immediate and intuitive meaning only in the Euclidean geometry.

7.1. Distance Curves. The basis for the construction of aidistance alley are the frontal plane horopters, i.e., the geodesic lines symmetrical to the x-axis. Consider, for example, the geodesic through P_0 and determine a point Q_0 on this line which has a given geodesic distance S from P_0. Similarly, on another geodesic through P_1 we can find a point Q_1 which has the same geodesic distance from P_1. By doing this for the whole set of geodesic lines, a curve is obtained by the points Q which has a constant geodestic distance S from the x-axis. If we determine a similar curve symmetrically located on the other side of the x-axis, an alley is found in which the two walls have the constant geodesic distance 2S from each other. We expect that these alleys are the mathematical expression for the observed Distance Alleys.

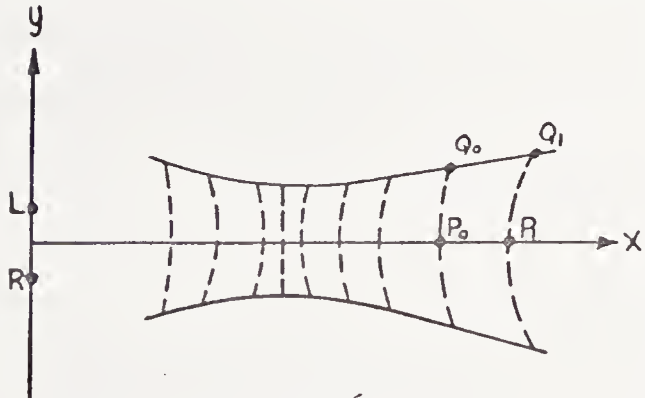


Fig. 61

We determine the equations of the distance curves simultaneously for all three geometries in question. On account of (6.48) we have for the frontal plane horopters the equation

$$2\ k \cos\varphi = e^{\sigma(\gamma+\mu)} - \varepsilon\, e^{-\sigma(\gamma+\mu)} \tag{7.11}$$

where $\varepsilon = -1,\ 0,\ +1$ for the hyperbolic, Euclidean, and elliptic geometry respectively. By introducing temporarily the variable

$$\tau = \sigma(\gamma + \mu)$$

we may write

$$2\ k \cos\varphi = e^{-\tau} - \varepsilon e^{-\tau} \tag{7.12}$$

*Blumenfeld, Walter: Untersuchungen über die scheinbare Grösse im Sehraume. Z.f. Psychol. u. Physiol. d. Sinnesorg., 65: 241-416, 1913.

The geodesic distance S of a point Q on such a curve from the median point P ($\varphi = 0$) is given by the integral

$$S = 2\int_0^{\varphi} \frac{\sqrt{d\tau^2 + d\varphi^2}}{e^{\tau} + \varepsilon e^{-\tau}}\, d\varphi \tag{7.13}$$

We consider τ as function of φ (given by 7.12), and thus have to evaluate the integral

$$S = 2\int_0^{\varphi} \frac{\sqrt{1 + \tau'^2}}{e^{\tau} + \varepsilon e^{-\tau}}\, d\varphi \tag{7.131}$$

Now, from (7.12) it follows that

$$-2\, k \sin\varphi = (e^{\tau} + \varepsilon e^{-\tau})\, \tau'$$

and hence

$$1 + \tau'^2 = \frac{k^2 \sin^2\varphi + \frac{1}{4}(e^{\tau} + \varepsilon e^{-\tau})^2}{\frac{1}{4}(e^{\tau} + \varepsilon e^{-\tau})^2} \tag{7.14}$$

With the aid of the identity

$$\tfrac{1}{4}(e^{\tau} + e^{-\tau})^2 = \tfrac{1}{4}(e^{\tau} - \varepsilon e^{-\tau})^2 + \varepsilon \tag{7.141}$$

and by taking (7.12) into account, we may write (7.14) as follows

$$1 + \tau'^2 \; \frac{k^2 + \varepsilon}{\frac{1}{4}(e^{\tau} + \varepsilon e^{-\tau})^2} \tag{7.15}$$

so that our integral (7.131) becomes

$$S = \sqrt{k^2 + \varepsilon}\int_0^{\varphi} \frac{d\varphi}{\frac{1}{4}(e^{\tau} + \varepsilon e^{-\tau})^2}\,. \tag{7.16}$$

From (7.141) it follows that

$$\tfrac{1}{4}(e^{\tau} + \varepsilon e^{-\tau})^2 = k^2 \cos^2\varphi + \varepsilon$$

and hence

$$\begin{aligned} S &= \sqrt{k^2 + \varepsilon}\int_0^{\varphi} \frac{d\varphi}{k^2 \cos^2\varphi + \varepsilon} \\ &= \frac{1}{\sqrt{\varepsilon}} \arctan\left(\frac{\sqrt{\varepsilon}}{\sqrt{k^2 + \varepsilon}} \tan\varphi\right) \end{aligned} \tag{7.161}$$

We write this last result in the form

$$\frac{1}{\sqrt{\varepsilon}} \tan\sqrt{\varepsilon}\, S = \frac{1}{\sqrt{k^2 + \varepsilon}} \tan\varphi \tag{7.17}$$

It determines the geodesic distance S of a point of latitude φ which lies on the horopter determined by the constant k.

Finally we eliminate the constant k with the aid of (7.11). We first write (7.17) a little differently by using the identity

$$\sin\sqrt{\varepsilon}\ S = \frac{\tan\sqrt{\varepsilon}\ S}{1+\tan^2\sqrt{\varepsilon}\ S}$$

It follows

$$\frac{1}{\sqrt{\varepsilon}}\sin\sqrt{\varepsilon}\ S = \frac{\sin\varphi}{\sqrt{\varepsilon + k^2\cos^2\varphi}}$$

and hence by (7.11) and (7.141):

$$\frac{1}{\sqrt{\varepsilon}}\sin\sqrt{\varepsilon}\ S = \frac{2\sin\varphi}{e^{\tau}+\varepsilon e^{-\tau}} \qquad (7.18)$$

or by reintroducing γ by $\tau = \sigma(\gamma+\mu)$:

$$\frac{1}{\sqrt{\varepsilon}}\sin\sqrt{\varepsilon}\ S = \frac{2\sin\varphi}{e^{\sigma(\gamma+\mu)}+\varepsilon e^{-\sigma(\gamma+\mu)}} \qquad (7.181)$$

The equation (7.181) determines the geodesic distance of any point of the horizontal plane from the x-axis in terms of its bipolar coordinates γ, φ. Finally, we assign to ε its values -1, 0, + 1 and have the result that the geodesic distance of a point Q from the x-axis is given by the functions:

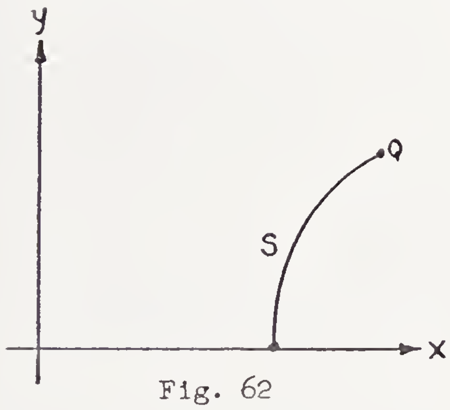


Fig. 62

$$\text{Hyperbolic Geometry: } \sinh S = \frac{\sin\varphi}{\sinh\sigma(\gamma+\mu)}$$

$$\text{Euclidean Geometry: } S = 2\sin\varphi\, e^{-\sigma(\gamma+\mu)} \qquad (7.19)$$

$$\text{Elliptic Geometry: } \sin S = \frac{\sin\varphi}{\cosh\sigma(\gamma+\mu)}$$

The distance curves are characterized by the condition that the geodesic distance from the x-axis is constant. Thus these curves are given by the equations:

$$\text{Hyperbolic Geometry: } \sinh\sigma(\gamma+\mu) = \sin\varphi$$

$$\text{Euclidean Geometry: } e^{\sigma(\gamma+\mu)} = 2\,C\sin\varphi \qquad (7.191)$$

$$\text{Elliptic Geometry: } \cosh\sigma(\dot{\gamma}+\mu) = C\sin\varphi$$

where C is a constant.

7.2. For the discussion of the distance curves (7.191), we introduce again the ξ, η map of the three geometries obtained by the transformation

$$\xi = e^{-\sigma(\gamma+\mu)}\cos\varphi$$

$$\eta = e^{-\sigma(\gamma+\mu)}\sin\varphi \qquad (7.21)$$

$$\rho = \sqrt{\xi^2+\eta^2} = e^{-\sigma(\gamma+\mu)}$$

We write the three distance curves (7.191) in the unified form

$$e^{\sigma(\gamma+\mu)} + \varepsilon e^{-\sigma(\gamma+\mu)} = 2\ C \sin\varphi \tag{7.22}$$

and find, in the ξ, η-plane:

$$1 + \varepsilon\rho^2 = 2\ C\ \rho \sin\varphi$$

or

$$\varepsilon(\xi^2 + \eta^2) - 2\ C\ \eta + 1 = 0 \tag{7.23}$$

For $\varepsilon = -1$ (Hyperbolic geometry) this is the equation of circles

$$\xi^2 + (\eta + C)^2 = 1 + C^2 \tag{7.231}$$

which all go through the points $\eta = 0$, $\xi = \pm 1$ of the unit circle and thus have their center $-C$ on the η-axis.

In case $\varepsilon = 0$ (Euclidean geometry), we find the straight lines

$$\eta = \frac{1}{2C} \tag{7.232}$$

parallel to the ξ-axis.

In case of $\varepsilon = +1$ (Elliptic geometry) we obtain

$$\xi^2 + (\eta - C)^2 = C^2 - 1 \tag{7.233}$$

i.e., circles around points C of the η-axis which intersect the unit circle at right angles.

The three types of curves are shown in Fig. 63. We conclude immediately that the actual alleys in the x,y-plane must go to the eyes and approach infinity asymptotically with a certain direction

$$\varphi_\infty < \pi/2$$

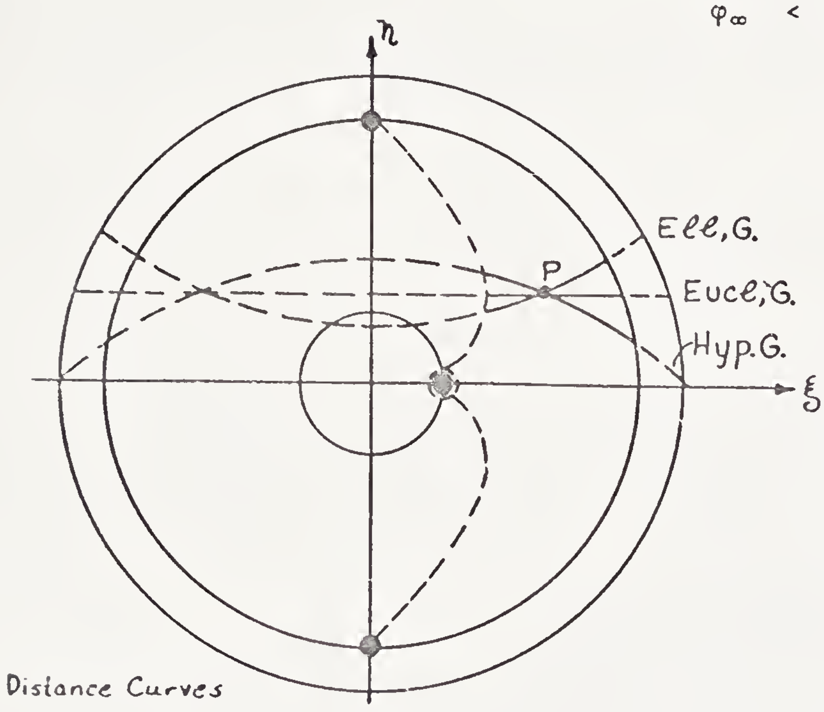


Fig. 63

It follows that these curves must have the general form shown in Fig. 64, which is in agreement with experimental results.

7.3. Parallel Curves. The mathematical formulation of the psychological impression that an alley has parallel walls is not immediately as clear as in case of distance alleys. In fact, we can easily give different mathematical concepts of parallelism and defend them with equally good arguments.

The statement that two lines elements not attached to the same point are parallel has no absolute meaning in non-Euclidean geometries.

It is true that Levi-Civita's concept of parallel transfer allows us to move a given line element $d\ell_0$ along a given curve C to another position P_1 "parallel with itself," i.e., without apparent change of direction. However, the result

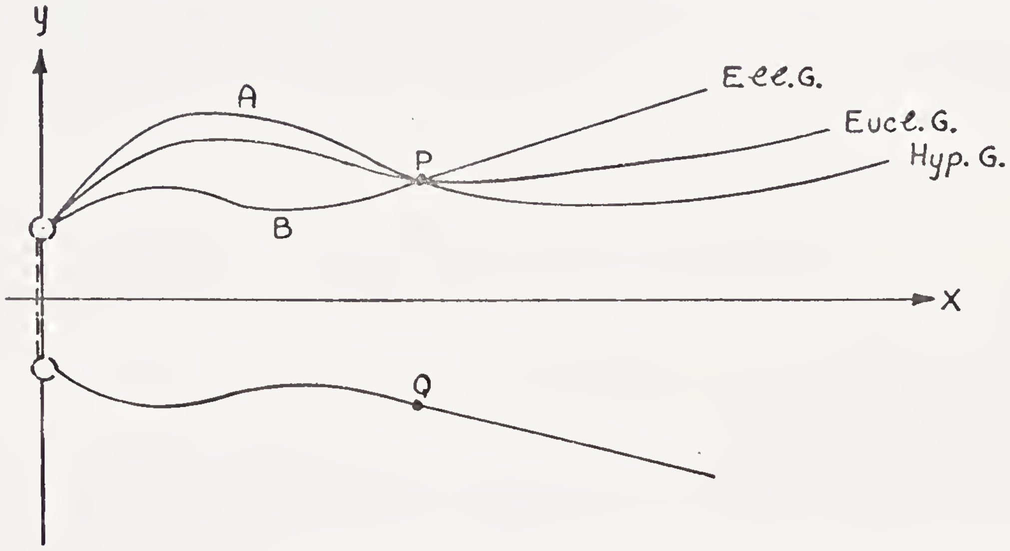


Fig. 64

of this transfer depends on the chosen curve C. A different result will be obtained if another curve is chosen on which the line element is transferred. The Euclidean geometry is the only one in which parallel transfer is independent of the path C so that a statement regarding the parellelism of two line elements not attached to the same point has an absolute meaning only in this special geometry.

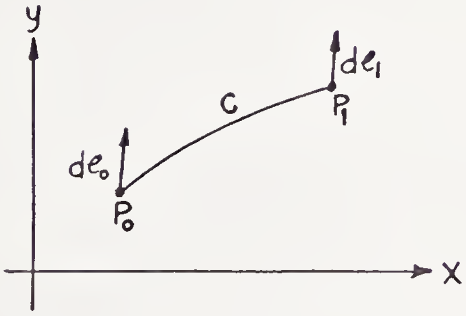


Fig. 65

From this consideration it follows that only in the Euclidean geometry will the instruction to construct alleys with parallel walls without paying attention to the distance of the walls lead the observer to a uniquely determined reaction. If, however, as we have good reason to assume, the metric of binocular vision is non-Euclidean, a uniquely determined reaction cannot be expected, and thus a more specified instruction must be given. We shall discuss in the following two different types of "Parallel Alleys" which may be considered as two different interpretations of Blumenfeld's experimental parallel alleys. We are forced to this ambiguity since the experiments have apparently been carried out without such specified instructions. We shall consider the second of the following interpretations (7.5) as the one which represents Blumenfeld's experiments with greater probability.

7.4 Parallel Curves (1st type). We assume that P_0 and Q_0 are two fixed points symmetrical to the x-axis. Let $\varphi = \varphi_0$ be the "hyperbola of sight" through P_0 so that the straight line OP_0 includes an angle with the x-axis practically equal to φ_0. The observer is asked to move a point P_1 in such a position that the line P_0P_1 includes with OP_0 the same apparent angle φ_0 as OP_0 with the x-axis (Fig. 66). The process is continued with P_1, Q_1 as fixed, and two other points P_2, Q_2 moved into position. As a result two alleys are obtained with walls which are parallel in the sense that two individual opposite wall sections seem

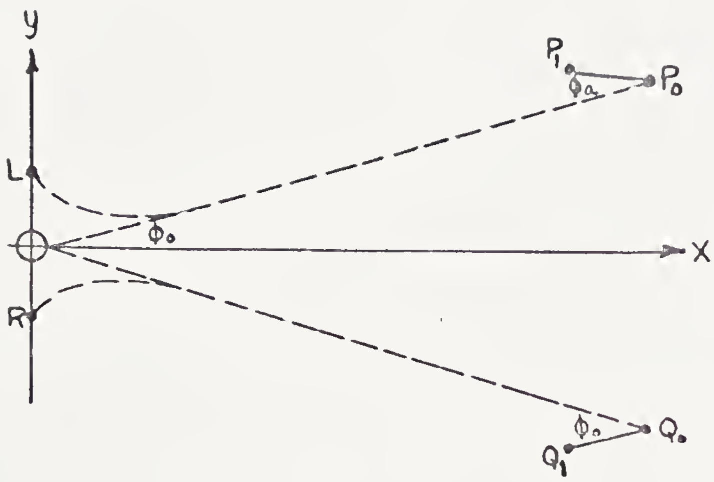


Fig. 66

to be parallel to the median plane. Note that nothing is required with regard to the impression which the curves give in their total extension.

It is easy to find the equations for these curves. Let ($d\gamma$, $d\varphi$) be the line element of the curve at P_0 and ($\delta\gamma$, $\delta\varphi = 0$) the line element of the curve $\varphi = \varphi_0$ at P_0. The two line elements include an apparent angle ω with each other. This angle is determine by the quadratic differential

$$ds^2 = M^2(\gamma)(\sigma^2 d\gamma^2 + d\varphi^2) \qquad (7.41)$$

namely, according to (3.24) given by the formula

$$\cos\omega = \frac{M^2(\gamma)(\sigma^2 d\gamma\delta\gamma + d\varphi\delta\varphi)}{\sqrt{M^2(\gamma)(\sigma^2 d\gamma^2 + d\varphi^2)}\ \sqrt{M^2(\gamma)(\sigma^2 \delta\gamma^2 + \delta\varphi^2)}} \qquad (7.42)$$

We notice immediately that $M(\gamma)$ cancels in this formula so that the apparent angle ω is independent of the choice of M. It is the same not only for the three geometries of constant curvature, but also for any other non-Euclidean geometry which has a metric of the type (7.41).

Since $\delta\varphi = 0$, we find

$$\cos\omega = \frac{\sigma d\gamma}{\sqrt{\sigma^2 d\gamma^2 + d\varphi^2}} \qquad (7.43)$$

as expression for the apparent angle of the curve with the lines of sight φ = const. If we require that $\omega = \varphi_0$, or in general $\omega = \varphi$, we obtain from (7.43) the cpndition

$$\cos\varphi = \frac{\sigma d\gamma}{\sqrt{\sigma^2 d\gamma^2 + d\varphi^2}} \qquad (7.44)$$

or

$$\frac{d\varphi}{d\gamma} = \sigma\tan\varphi \qquad (7.45)$$

The solution of this differential equation is

$$e^{\sigma\gamma} = C\sin\varphi \qquad (7.46)$$

where C is a constant of integration.

We point out again that these "Parallel Curves" are entirely independent of the function $M(\gamma)$. We also notice that they are identical with the distance curves (7.191) of the Euclidean geometry.

In the Euclidean geometry Distance Alleys and Parallel Alleys are identical.

We shall find a similar result in the next section where a different mathematical formulation of parallel alleys is discussed.

By introducing the variables ξ, η by (7.21) in the equation (7.46) we obtain, of course,

$$\eta = \text{const.}$$

so that the straight lines of the ξ, η-plane parallel to the ξ-axis may be interpreted as representing parallel alleys in all three geometries.

This interpretation of the straight lines η = const. in the ξ, η-plane leads us to an important conclusion. Let us assume that a distance alley and a parallel alley of the above type have been constructed, both alleys starting from a fixed pair of end-points P and Q (Fig. 64). The parallel alley, imaged into the ξ, η-plane (Fig. 63), is given by the straight line through P parallel to the ξ-axis, and is identical with the distance curve through P only if the metric is Euclidean. The very fact that distance and parallel alleys are found to differ points out that the visual metric cannot be Euclidean.

Moreover, we see in Fig. 63 that--going from P to the eyes--the distance curve lies above the parallel curve if the geometry is hyperbolic and below this curve if the geometry is elliptic.

This means in the x,y-plane that the distance curve lies outside the parallel curve if the geometry is hyperbolic and inside if the geometry is elliptic.

The alley experiments thus give us the possibility of deciding by experiment which metric is characteristic for our spatial sensations. Blumenfeld's data show clearly that the distance curves lie outside the parallel curves. Thus if we are allowed to identify his parallel alleys with the above-discussed alleys, we must conclude that the psychometric of space sensations is hyperbolic.

In the next section we shall consider another type of parallel alleys which, if identified with Blumenfeld's alleys, leads to the same conclusion.

7.5. Parallel Curves (2nd type). Neither the distance curves nor the parallel curves of 7.4 are geodesic lines if the metric is hyperbolic or elliptic. They are geodesic lines, however, in the Euclidean geometry.

Thus these curves will not appear straight if we observe them by paying attention to their total depth extension. In view of this fact, we now change the instruction for the observer: By paying special attention to the walls as a whole he is to arrange the points so that they appear on two straight lines parallel to the x-axis. Consequently the resulting curves must be given mathematically by geodesics, i.e., they must be two curves of the set (7.43).

$$\varepsilon(\xi^2 + \eta^2) + 2(\xi\xi_0 + \eta\eta_0) = 1 \qquad (7.51)$$

in the ξ, η map of our geometries.

We also know that the second curve must be the mirror image of the first curve with regard to the ξ-axis.

Obviously there are infinitely many pairs of geodesics of this type; in fact, we may choose the first curve arbitrarily from (7.51) and then pair it with its mirror image

$$\varepsilon(\xi^2 + \eta^2) + 2(\xi\xi_0 - \eta\eta_0) = 1 \qquad (7.52)$$

to an alley with walls erected on two straight lines.

In order to pick out among the geodesic lines pairs which give the appearance of being parallel to the median plane, we use a sort of principle of correspondence. It is clear that in case of the Euclidean geometry ($\varepsilon = 0$) the desired pairs are given by $\xi_0 = 0$ in (7.51) so that

$$\eta = \pm \frac{1}{2\eta_0} \tag{7.53}$$

where η_0 is an arbitrary constant.

We now uphold the same principle in case of the other geometries: Two alley curves appear to be parallel to the median plane if they are found from the general set (7.51) by the same principle ($\xi_0 = 0$) which determines parallel alleys in case of the Euclidean geometry.

This consideration leads us to the result that the special geodesic lines

$$\varepsilon(\xi^2 + \eta^2) + 2\,\eta\,\eta_0 = 1 \tag{7.54}$$

in which η_0 is an arbitrary constant will give the impression of an apparently straight line "parallel" to the median plane. Obviously these curves are circles in the ξ, η-plane symmetrical to the η-axis if $\varepsilon \neq 0$.

In case of the Hyperbolic Geometry ($\varepsilon = -1$), we get the circles

$$\xi^2 + (\eta - \eta_0)^2 = \eta_0^2 - 1 \tag{7.55}$$

normal to the unit circle and centered around a point η_0 of the η-axis.

In case of the Elliptic Geometry we have the circles

$$\xi^2 + (\eta + \eta_0)^2 = \eta_0^2 + 1 \tag{7.56}$$

through the points $\xi = \pm 1$ of the ξ-axis and thus also centered around a point $-\eta_0$ of the η-axis. For the graphical illustration of these curves we can use Fig. 63 and Fig. 64, but we have to interchange the legends "Hyperbolic Geometry" and "Elliptic Geometry."

By introducing, instead of ξ, η the variables γ, φ, we obtain from (7.55) and (7.56) the following result:

Parallel alleys according to the above instructions are given by the curves

$$\begin{aligned} &\text{Hyperbolic Geometry: } \cosh \sigma(\gamma + \mu) = C \sin \varphi \\ &\text{Euclidean Geometry: } e^{\sigma(\gamma + \mu)} = 2\,C \sin \varphi \\ &\text{Elliptic Geometry: } \sinh \sigma(\gamma + \mu) = C \sin \varphi \end{aligned} \tag{7.57}$$

where C is an arbitrary constant.

For the interpretation of Blumenfeld's parallel alleys we shall consider the mathematical definition of parallel curves given in this section to be the one which has the greatest probability. We shall justify this point of view in the next section.

We may collect the results of 7.3 and 7.5 with regard to Distance-and Parallel Alleys in the following table.

Hyperbolic Geometry:

Distance Curves A: $\sinh\sigma(\gamma + \mu) = C \sin\varphi$

Parallel Curves B: $\cosh\sigma(\gamma + \mu) = C \sin\varphi$

Euclidean Geometry:

Distance Curves

Parallel Curves $e^{\sigma(\gamma + \mu)} = 2\,C \sin\varphi$

Elliptic Geometry:

Distance Curves B: $\cosh\sigma(\gamma + \mu) = C \sin\varphi$

Parallel Curves A: $\sinh\sigma(\gamma + \mu) = C \sin\varphi$

By disregarding the Euclidean case, we thus recognize that only two different types of curves are involved, namely,

$$\text{Type A: } \sinh\sigma(\gamma + \mu) = C \sin\varphi \qquad (7.58)$$

$$\text{Type B: } \cosh\sigma(\gamma + \mu) = C \sin\varphi$$

Curves of the type A represent the Distance Curves of the Hyperbolic Geometry and the Parallel Curves of the Elliptic Geometry. Curves of the type B represent the Parallel Curves of the Hyperbolic Geometry and the Distance Curves of the Elliptic Geometry.

In the ξ, η-plane these curves A and B are given by the circles shown in Fig. 67. Starting from a point P we see that the A-curve lies always above the B-curve in the region between P and the eyes. This means in the x,y-plane that the A-curves lie always outside the B-curves and that both curves approach one of the eyes (Fig. 64).

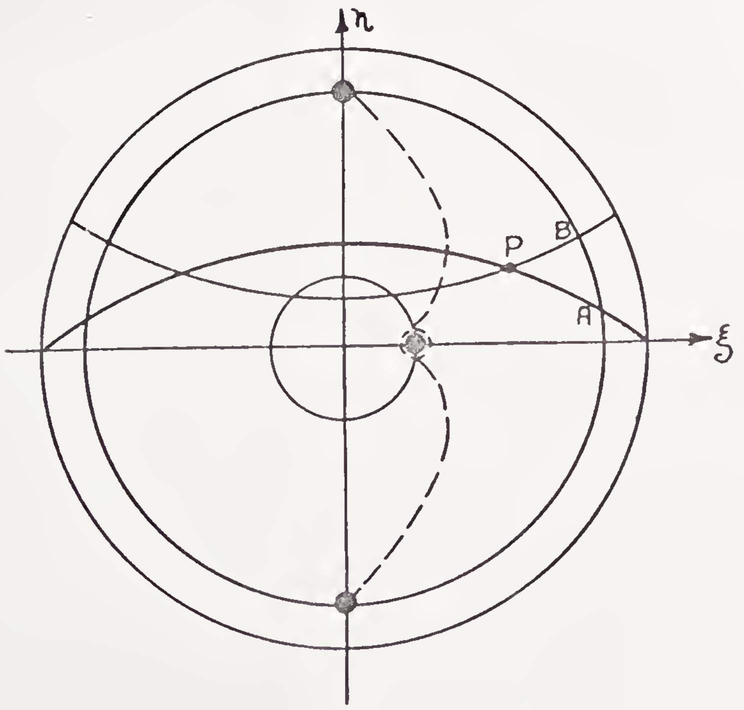


Fig. 67

This means for our alley problem: Consider a Distance Alley and a Parallel Alley both starting at the same fixed point P (and its mirror image Q).

If the geometry is hyperbolic, then the Distance Alleys lie outside the Parallel Alleys.

If the geometry is elliptic, then the Parallel Alleys lie outside the Distance Alleys.

If the geometry is Euclidean, then Parallel Alleys and Distance Alleys are identical.

Blumenfeld's data demonstrate that Parallel Alleys and Distance Alleys are different, and that the Distance Alleys lie outside the Parallel Alleys. Hence we are forced again to the conclusion that the proper geometry for our spatial sensations is the Hyperbolic Geometry.

7.6. It is not difficult to support the principle of correspondence used in the previous section by arguments of a more principal nature. We may consider the often used ξ, η maps of the apparent horizontal plane as a plane map of the actual sensation of objects in the physical horizontal plane. It is of course not a perfect map since we must require from a perfect map that it must give immediately the correct linear and angular relations of the sensational manifold. We know that such a map is impossible. However, the ξ, η map illustrates directly certain features of our sensations, for example, the fact that the Vieth-Muller circles γ = const. are seen as circles around the observer. Indeed, these circles are represented by the circles

$$\sqrt{\xi^2 + \eta^2} = \rho = e^{-\sigma(\gamma + \mu)} = \text{const.} \tag{7.61}$$

in the ξ, η-plane. Furthermore, the "hyperbolae of sight" φ = const. are given by the radial line η/ξ = const. in the ξ, η-plane, and thus the ξ, η map describes directly the apparent significance of these hyperbolae.

Even more important is the fact that the ξ, η map is <u>conformal</u>, i.e., that the Euclidean angles on the map are equal in size to the non-Euclidean apparent angles given to us by the metric of our space perception. Indeed, consider two line elements $(d\xi, d\eta)$ and $(\delta\xi, \delta\eta)$ attached to a point P. The Euclidean angle on the ξ, η map is given by the formula

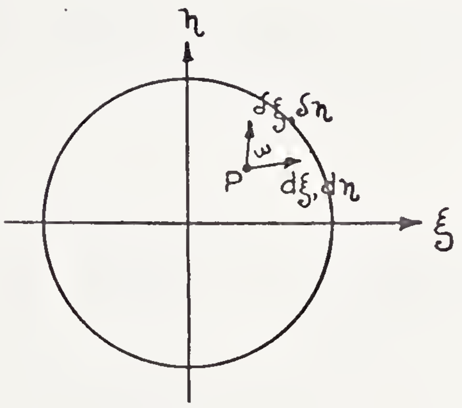


Fig. 68

$$\cos\omega = \frac{d\xi\,\delta\xi + d\eta\,\delta\eta}{\sqrt{d\xi^2 + d\eta^2}\,\sqrt{\delta\xi^2 + \delta\eta^2}} \tag{7.62}$$

The apparent angle ω^* of the two corresponding line elements (dx, dy) $(\delta x, \delta y)$ in the physical plane, however, is determined by the metric differential

$$ds^2 = n^2(\xi, \eta)\,(d\xi^2 + d\eta^2) \tag{7.63}$$

where

$$n = \frac{2}{1 + \varepsilon\rho^2}$$

as we have seen in 6.2.

By applying the general formula (3.24) for the angle ω^*, we get

$$\cos\omega^* = \frac{n^2(d\xi\,\delta\xi + d\eta\,\delta\eta}{\sqrt{n^2(d\xi^2 + d\eta^2)}\,\sqrt{n^2(\delta\xi^2 + \delta\eta^2)}} = \frac{d\xi\,\delta\xi + d\eta\,\delta\eta}{\sqrt{d\xi^2 + d\eta^2}\,\sqrt{\delta\xi^2 + \delta\eta^2}} \tag{7.64}$$

and we see immediately that

$$\cos\omega^* = \cos\omega, \text{ i.e., } \omega^* = \omega$$

This proves our statement, that the ξ, η map is conform, i.e., that the apparent, non-Euclidean angles are directly given by the Euclidean angles of the map.

The apparent place of the observer is the center $\xi = \eta = 0$ from which the lines of sight φ = const. seem to emerge and around which the Vieth-Muller circles γ = const. are apparently located. He observes the horizontal half-plane, i.e., the often illustrated sickle-shaped part of the ξ, η map, through the receiving set represented by the logarithmic spirals

$$\rho = \rho_0\, e^{\pm 2\sigma\varphi} \tag{7.65}$$ 2

The ξ-axis of the map represents the median line of the observer and, since the map is conform, the η-axis must represent a direction apparently normal to the median line. This direction, of course, no longer lies in the field of view, but, without doubt, the observer is conscious of it. An alley with straight walls thus will be interpreted as parallel to the median line if the imagined extensions of the walls approach the apparent Y-axis of the observer at right angles. In the conformal ξ, η map such alleys thus must be given by geodesic lines which are normal to the η-axis. By (7.51) it follows that such geodesic lines must have the form (7.54), namely

$$\varepsilon(\xi^2 + \eta^2) + 2\eta\,\eta_0 = 1 \qquad (7.66)$$

which leads us back to the parallel curves of the preceding section.

7.7. The imperfection of the ξ, η map has to be seen in the fact that it gives directly only the correct apparent angles. We know that an isometric Euclidean map which gives directly the size (and thus also angles) is impossible. However, we can easily construct another significant map which, at least, represents the quality of "straightness" directly. This means that the Euclidean straight lines of the map represent the geodesic lines of the non-Euclidean manifolds in question. This so-called projective map (Klein's Model) is obtained by submitting the ξ, η map to a "radial distortion."

$$R = \frac{2\rho}{1 - \varepsilon\rho^2} \qquad (7.71)$$

In case $\varepsilon = 0$ this new map is a similarity transformation of the ξ, η map.

In case $\varepsilon = -1$ (Hyperbolic geometry) the unit circle $\rho < 1$ is transformed into the unit circle $R \leqq 1$.

However, for $\varepsilon = 1$ (Elliptic geometry), the unit circle $\rho < 1$ is expanded over the whole plane.

By introducing

$$X = R\cos\varphi \qquad Y = R\sin\varphi \qquad (7.72)$$

we may express the transformation (7.71) as a transformation of the physical x,y-plane into the X,Y-plane, namely, in terms of bipolar coordinates γ, φ:

$$X = \frac{2\cos\varphi}{e^{\sigma(\gamma+\mu)} - e^{-\sigma(\gamma+\mu)}} \qquad (7.73)$$

$$Y = \frac{2\sin\varphi}{e^{\sigma(\gamma+\mu)} - e^{-\sigma(\gamma+\mu)}} \qquad (7.73)$$

which takes the place of the transformation (7.21).

It can be seen that the geodesic lines in all three geometries are given by the straight lines of the X, Y map. Indeed, in ξ, η coordinates, we have found in (6.62) the equation for geodesic lines

$$\frac{1 - \varepsilon\rho^2}{\rho} = 2k\cos(\varphi - \varphi_0) \qquad (7.74)$$

Hence, it follows with the aid of (7.71) that their equation in the X,Y-plane is

$$R \cos (\varphi - \varphi_0) = \text{const.}$$

or

$$X \cos \varphi_0 + Y \sin \varphi_0 = \text{const.} \qquad (7.75)$$

which is the equation of a straight line.

The metric quadratic differential in these new coordinates X, Y or R, φ has now the form

$$ds^2 = \frac{1}{4} \frac{dX^2 + dY^2 + \varepsilon (YdX - XdY)^2}{[1 + \varepsilon (X^2 + Y^2)]^2} = \frac{1}{4}\left(\frac{dR^2}{(1 + \varepsilon R^2)^2} + \frac{R^2}{1 + \varepsilon R^2} d\varphi^2\right) \qquad (7.76)$$

as one easily verifies.

If we consider the X,Y map as another attempt to interpret our sensations in a Euclidean realization, we may say that it represents truly the quality of apparent straightness. However, it is not in the least conform and thus does not allow us to determine apparent angles from the angles of the map.

Except for the angles φ of the lines of sight at the origin! Indeed, the hyperbolae of sight φ = const. are still given by the straight lines through the origin in the X,Y-plane, and their angles with each other are equal to the apparent, non-Euclidean angles. Also the property of Vieth-Muller circles to appear as circles around the observer is truly represented on our map. Thus if any visual curve includes an apparent right angle with a radial line, then its image on the X,Y map will do the same. As before, the "egocenter" of the observer has to be identified with the point X = Y = 0 on the map. He observes a section of the X,Y-plane through a "receiving set" of a similar form as in the case of the ξ, η map. The physical half-plane $x \geq 0$ lies between two circles of radius R_0 and R_1 given by

$$R_0 = \frac{2}{e^{\sigma(\pi + \mu)} - \varepsilon e^{-\sigma(\pi + \mu)}}$$

and

$$R_1 = \frac{2}{e^{\sigma\mu} - \varepsilon e^{-\sigma\mu}}$$

the latter corresponding to the infinity of the physical space. The two eyes are represented by the two curves of spiral type

$$R = \frac{2\rho_0 e^{\pm 2\sigma\varphi}}{1 - \varepsilon\rho_0^2 e^{\pm 4\sigma\varphi}} \qquad (7.77)$$

as shown in Fig. 69.

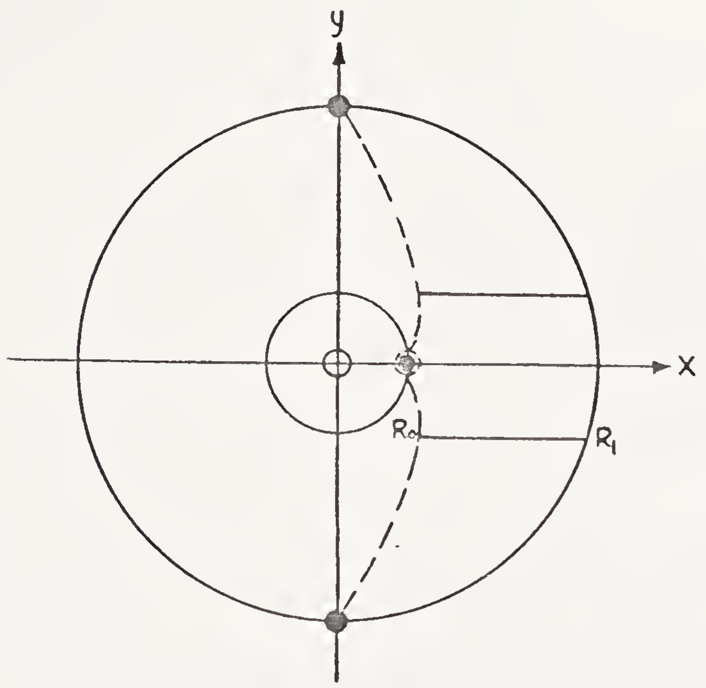


Fig. 69

The X-axis determines the median line of the observer and the Y-axis, since the map is conform at the point 0, a direction in the consciousness of the observer normal to the median line.

Alleys with straight walls are given by a pair of straight lines in the X,Y-plane symmetrically located to the X-axis. Among these pairs we have the special ones $Y = \pm C$ which are, on the X,Y map, parallel to the X-axis and thus seem to approach the Y-axis at right angles. We suspect with great probability that it is this special property which is interpreted by the observer as "parallel to the median plane." (Fig. 69).

The curves in the physical x,y-plane which correspond to the parallel alleys $Y = \pm C$ of the X,Y map can readily be found by (7.73), namely

$$e^{\sigma(\gamma + \mu)} - \varepsilon e^{-\sigma(\gamma + \mu)} = \pm \frac{2}{C} \sin \varphi \qquad (7.78)$$

These curves, however, are identical with the parallel curves of § 7.5.

7.8. We conclude our discussion of the Alley Problem by outlining a simple method of calculating the constants σ and μ from alley observations. We determine for this purpose first the asymptotes of the curves A and B given by equation (7.58). For small values of γ we may write

$$\begin{aligned} A&: \sinh \sigma\mu + \sigma \cosh \sigma\mu\gamma = C \sin \varphi_o + C \cos \varphi_o \Delta\varphi \\ B&: \cosh \sigma\mu + \sigma \sinh \sigma\mu\gamma = C \sin \varphi_o + C \cos \varphi_o \Delta\varphi \end{aligned} \qquad (7.81)$$

Hence it follows that, in this region of small γ, the curves follow the approximate equations

$$\begin{aligned} A&: \Delta\varphi = \frac{\sigma}{t} \tan \varphi_o \gamma \\ B&: \Delta\varphi = \sigma t \tan \varphi_o \gamma \end{aligned} \qquad (7.82)$$

where t is the constant

$$t = \tanh \sigma\mu \qquad (7.83)$$

and φ_o the angle of the asymptote with the x-axis.

In order to translate the conditions (7.82) into x,y coordinates, we develop the equation

$$\frac{y}{x} = \frac{\sin 2\varphi}{\cos 2\varphi + \cos \gamma} \qquad (7.84)$$

with respect to γ, considering only terms of order zero and one.

We find readily

$$\frac{y}{x} = \tan \varphi_o + \frac{\Delta\varphi}{\cos^2 \varphi_o} \qquad (7.85)$$

This gives by (7.82) for curves of the type A:

$$\frac{y}{x} = \tan \varphi_o + \frac{\sigma \tan \varphi_o}{t \cos^2 \varphi_o} \gamma \qquad (7.86)$$

From the equation

$$x = \frac{\cos 2\varphi + \cos \gamma}{\sin \gamma}$$

we find with sufficient approximation that

$$x = \frac{1 + \cos 2\ \varphi_0}{Y} = \frac{2 \cos^2\ \varphi_0}{Y}$$

and hence by introducing $\gamma = \frac{2 \cos^2\ \varphi_0}{x}$ in (7.86) that the equations of the asymptotes of the curves (7.58) are

$$\begin{aligned} A&: y = x \tan \varphi_0 + \frac{2\sigma}{t} \tan\ \varphi_0 \\ B&: y = x \tan \varphi_0 + 2\sigma t \tan \varphi_0 \end{aligned} \qquad (7.87)$$

The latter is found from the first one by replacing t by 1/t.

Let us now assume that a distance and a parallel alley through a point P have been set up and that the distance alley lies outside the parallel alley, so that the metric is hyperbolic. Let us assume that the distance curve has the asymptote (found by graphing the observation data)

$$y = ax + b \qquad (7.88)$$

and that the parallel curve has the asymptote

$$y = a'x + b' \qquad (7.881)$$

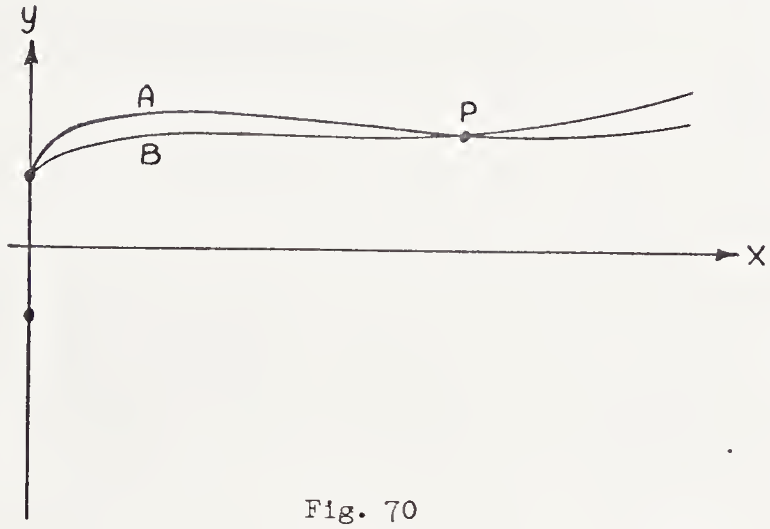


Fig. 70

Then it follows from (7.87) that

$$\begin{aligned} \frac{2\sigma}{t} &= \frac{b}{a} \\ 2\sigma t &= \frac{b'}{a'} \end{aligned} \qquad (7.89)$$

By multiplication

$$4\sigma^2 = \frac{b}{a} \quad \frac{b'}{a'} \qquad (7.891)$$

By division

$$t^2 = \tanh^2 \sigma\mu = \frac{b'}{a'} \div \frac{b}{a}$$

i.e., two simple equations for the individual constants σ and μ of the observer.

An evaluation of Blumenfeld's experiments on this basis has given the following values of σ and μ for one of his subjects (L_0)

$$\begin{aligned} \sigma &= 14.58 \\ t &= 0.827 \\ \mu &= 0.0809 \end{aligned} \qquad (7.892)$$

Since the experiments are not quite sufficient for a trustworthy determination of the above constants, the above values should be considered only as giving the order of magnitude we have to expect.

The horopter and alley curves illustrated in Fig. 71 are based on the above constants; nothing more is intended here than to give a general impression of the shape of our theoretical curves.

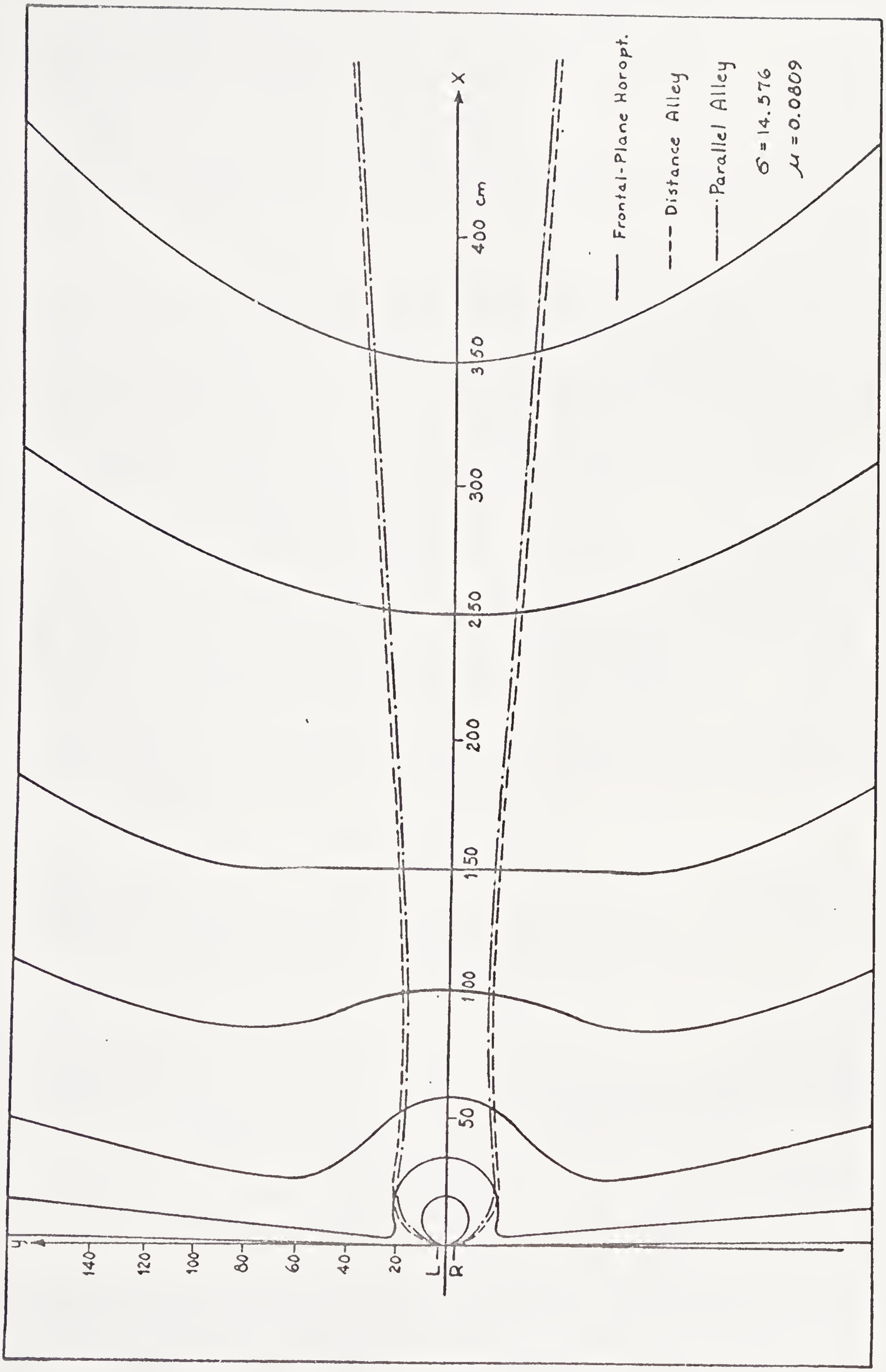

Frontal-Plane Horopt.
Distance Alley
Parallel Alley
σ = 14.576
μ = 0.0809
x
400 cm
350
300
250
200
150
100
50
y
140
120
100
80
60
40
20
L
R

Fig. 71

Section 8

RIGID TRANSFORMATIONS OF THE HYPERBOLIC SPACE

8.1. We shall derive in this final section the rigid transformations of the hyperbolic space. We have considered the existence of ∞^6 transformations of this type as the mathematical expression of the conviction associated with visual sensations that objects have size and form independent of their localization. An object, like a solid body, can be moved in space to any desired position without changing its geometrical form. This possibility of a solid geometry is well known to us in the realm of the Euclidean geometry.

It is, however, not restricted to this geometry; both the elliptic and hyperbolic metric are distinguished by the fact that just as many transformations exist to move a given object as a solid body to any desired position.

In the following, we shall confine ourselves to the hyperbolic geometry, since we may consider it now as the proper geometry for binocular visual sensations. We shall recognize the existence of ∞^6 hyperbolic "movements"; it is not difficult to show that these ∞^6 movements are the only ones which exist, but, for the proof, we refer the reader to the literature on this subject. Detailed formulae will be given for a special one-parameter group of hyperbolic movements which are of importance for the construction of distorted rooms.

8.2. Hyperbolic rotations. The simplest description of the hyperbolic movements can be given in the ξ, η, ζ space, where the hyperbolic metric has the form

$$ds^2 = \frac{d\xi^2 + d\eta^2 + d\zeta^2}{[1 - (\xi^2 + \eta^2 + \zeta^2)]^2} \qquad (8.21)$$

Our problem is to find point transformations

$$\xi' = f(\xi, \eta, \zeta)$$
$$\eta' = g(\xi, \eta, \zeta) \qquad (8.22)$$
$$\zeta' = h(\xi, \eta, \zeta)$$

such that the metric differential (8.21) preserves its form. This means the validity of the identity

$$\frac{d\xi'^2 + d\eta'^2 + d\zeta'^2}{[1 - (\xi'^2 + \eta'^2 + \zeta'^2)]^2} \equiv \frac{d\xi^2 + d\eta^2 + d\zeta^2}{[1 - (\xi^2 + \eta^2 + \zeta^2)]^2} \qquad (8.23)$$

It is immediately clear that the ∞^3 rotations of the ξ, η, ζ space around the origin are rigid movements of the hyperbolic geometry. Indeed, these rotations are given by the linear transformations

$$\xi' = A_{11}\xi + A_{12}\eta + A_{13}\zeta$$
$$\eta' = A_{21}\xi + A_{22}\eta + A_{23}\zeta \qquad (8.24)$$
$$\zeta' = A_{31}\xi + A_{32}\eta + A_{33}\zeta$$

where the matrix (A_{ik}) is orthogonal, i.e., satisfies the six orthogonality condi-conditions

$$\sum_{\alpha=1}^{3} A_{i\alpha}A_{\alpha k} = \begin{cases} 0, & i \neq k \\ 1, & i = k \end{cases} \tag{8.25}$$

Six conditions between nine coefficients allow a three-parameter degree of freedom which expresses the fact that ∞^3 different rotations are possible

A direct consequence of the conditions (8.25) is

$$\begin{aligned} \xi'^2 + \eta'^2 + \zeta'^2 &\equiv \xi^2 + \eta^2 + \zeta^2 \\ d\xi'^2 + d\eta'^2 + d\zeta'^2 &\equiv d\xi^2 + d\eta^2 + d\zeta^2 \end{aligned} \tag{8.26}$$

and hence we recognize by (8.23) that these ∞^3 rotations are rigid transformations of the metric (8.21).

We also notice that the interior of the unit sphere $\xi^2 + \eta^2 + \zeta^2 = 1$ is transformed into the interior of the unit sphere $\xi'^2 + \eta'^2 + \zeta'^2 = 1$. In particular, the points on the sphere $\xi^2 + \eta^2 + \zeta^2 = 1$ remain on this sphere.

The rotations (8.24) of the ξ, η, ζ space around the origin determine a group of transformations in the physical x, y, z space which is not quite as obvious. We obtain this physical group by using the relations of the coordinates ξ, η, ζ to the physical bipolar coordinates γ, φ, θ.

$$\begin{aligned} \xi &= e^{-\sigma(\gamma+\mu)} \cos\varphi \cos\theta; & \xi' &= e^{-\sigma(\gamma'+\mu)} \cos\varphi' \cos\theta' \\ \eta &= e^{-\sigma(\gamma+\mu)} \sin\varphi; & \eta' &= e^{-\sigma(\gamma'+\mu)} \sin\varphi' \\ \zeta &= e^{-\sigma(\gamma+\mu)} \cos\varphi \sin\theta; & \zeta' &= e^{-\sigma(\gamma'+\mu)} \cos\varphi' \sin\theta' \end{aligned} \tag{8.27}$$

The transformation (8.24) then assumes the form

$$\begin{aligned} \gamma' &= \gamma \\ \varphi' &= F(\varphi, \theta) \\ \theta' &= G(\varphi, \theta) \end{aligned} \tag{8.28}$$

The first relation states that the points on a Vieth-Muller Torus γ = const. remain on this torus but are shifted into other positions.

Already these simple rotations (in the ξ, η, ζ space) will transform a rectangular room into a distorted room, and we expect that the distorted room will be interpreted as rectangular if suitable clues are provided.

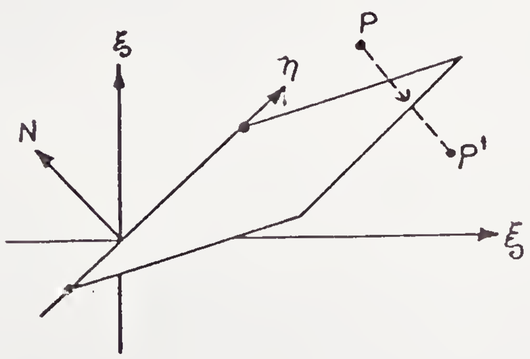


Fig. 72

8.3. Hyperbolic reflections. Another group of simple rigid transformations can be found directly. Let us consider a plane $a\xi + b\eta + c\zeta = 0$ through the origin of the ξ, η, ζ space. Let a, b, c be the direction cosines of its normal. By "reflecting" a point P on this plane we obtain an associated point P' on the other side of the plane. A configuration of points P_i thus is replaced by its "mirror image" P'_i. In order to express these point transformations analytically we use the vector notation

$$H = (\xi, \eta, \zeta)$$

$$H' = (\xi, \eta', \zeta') \tag{8.31}$$

$$N = (a, b, c)$$

and can express our transformation by the vector relation

$$H' = H - 2(H.N)\, N \tag{8.32}$$

By reintroducing coordinates, we find

$$\xi' = (1 - 2a^2)\xi - 2ab\eta - 2ac\zeta$$

$$\eta' = -2ab\xi + (1 - 2b^2)\eta - 2bc\zeta \tag{8.33}$$

$$\zeta' = -2ac\xi - 2bc\eta + (1 - 2c^2)\zeta$$

as the general transformation which represents the reflection on a plane normal to the unit vector $N = (a, b, c)$.

It is quite obvious that these transformations are rigid, i.e., that the identity (8.23) is satisfied. We can, of course, verify this fact formally with the aid of the relations (8.33).

The associated transformations of the physical x, y, z space can be found by replacing the variables ξ, η, ζ and ξ', η', ζ' in (8.33) by the bipolar coordinates γ, φ, θ and γ', φ', θ' with the aid of (8.27). The resulting "hyperbolic" reflections of the physical x, y, z space are only in exceptional cases actual physical reflections, namely, only if the "mirror" is a plane of elevation or the median plane. In all other cases, "hyperbolic" reflection will introduce distortion of the physical shape.

The mirror transformations (8.33) are not hyperbolic movements in the proper sense. This is because a right-hand coördinate system will be transformed into a left-hand system. Two mirror transformations in succession on two different mirrors, however, determine a true hyperbolic movement, namely, a rotation of the type discussed in 8.2.

We note again that the interior of the unit sphere is transformed into this interior and that the points on the unit sphere remain on this sphere.

8.4. Inversions. Hyperbolic translatory shifts. The only true hyperbolic movements we have found thus far are the ∞^3 rotations of 8.2. We know that ther must exist ∞^6 rigid movements, and thus we have to find the remaining ∞^3 transformations which, in the Euclidean geometry, are known as translatory shifts.

We remark first that the ξ, η, ζ map of the hyperbolic geometry is conformal (refer to 7.6), i.e., the Euclidean angles on the map are equal to the non-Euclidean hyperbolic angles. Since rigid transformations do not change the non-Euclidean angles, we conclude that the associated transformations (8.22) in the Euclidean ξ, η, ζ map must be conformal transformations of the ξ, η, ζ space.

This result allows us to narrow down the range considerably. In fact, according to Dupin's Theorem:

Rotations, Reflections and Inversions on spheres are the only conformal transformations of the three-dimensional space.

Of course, any composition of these three different types is again a conformal transformation.

We have considered the rotations and reflections already in the preceding sections. We have found that those particular rotations and reflections which transform the interior of the unit sphere $\xi^2 + \eta^2 + \zeta^2 = 1$ into the interior of the unit sphere $\xi'^2 + \eta'^2 + \zeta'^2 = 1$ are rigid transformations of the hyperbolic geometry. We also found that points on the unit sphere remain on the unit sphere. Indeed, the rotations around the origin and the reflection on planes through the origin are the only transformations of this type which have the above two characteristic properties.

This consideration leads us to the suspicion that the particular inversions which transform the interior of the unit sphere into itself will be again rigid transformations of the hyperbolic geometry. In general, an inversion is obtained as follows. We consider a sphere of radius D around a point C with coordinates (a, b, c). To any point P a point P' is associated (its conjugate point with respect to the sphere) by the geometrical construction illustrated in Fig. 73. Obviously P' lies on the line CP, and one verifies easily the relation

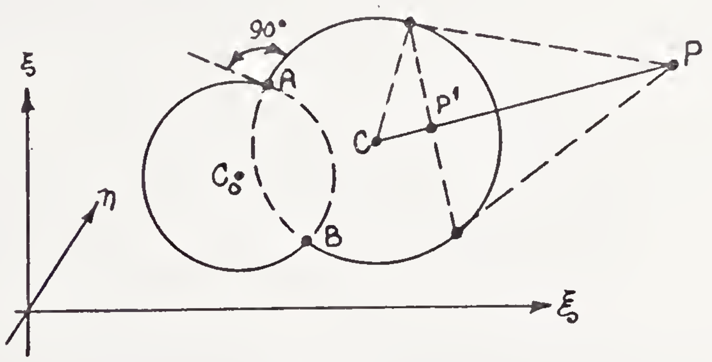


Fig. 73

$$(CP)\,(CP') = D^2 \tag{8.41}$$

This leads to the analytical relation between coordinated points

$$\begin{aligned}
\xi' - a &= D^2 \frac{\xi - a}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2} \\
\eta' - b &= D^2 \frac{\eta - b}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2} \\
\zeta' - c &= D^2 \frac{\zeta - c}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2}
\end{aligned} \tag{8.42}$$

We observe that the interior of the sphere around C is transformed into the exterior by these transformations and vice versa. We mention as a characteristic property of these inversions that any sphere in the ξ, η, ζ space has a sphere in the ξ', η', ζ' space as an image.

We next determine the particular inversions among the ∞^4 transformations (8.41) which transform the interior of the unit sphere into itself. We consider for this purpose in Fig. 73 a sphere around C_0 which intersects the sphere around C at right angles along a circle through AB. The points on this circle obviously remain fixed by the transformation. Furthermore, since the transformation is conformal, the image of the sphere around C_0 must be again a sphere which intersects the basic sphere around C in the same circle and also at right angles. Since

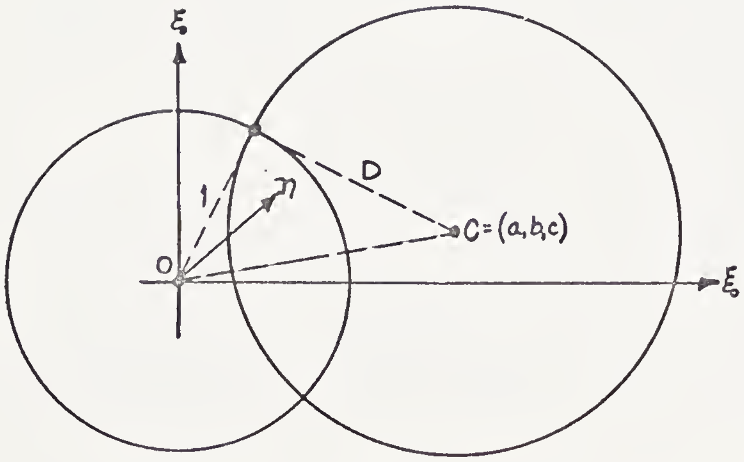


Fig. 74

there exists only one sphere with these properties, we conclude that the image is identical with the original. We thus may formulate the theorem:

Any sphere which intersects the basic sphere of the inversion (8.42) at right angles is transformed into itself.

Now, if we wish that the unit sphere be transformed into itself, we must choose the inversion sphere such that it intersects the unit sphere at right angles. This means that its radius D and the coordinates (a, b, c) of its center are related by the equation (Fig. 74)

$$D^2 = a^2 + b^2 + c^2 - 1 \tag{8.43}$$

We thus obtain the ∞^3 particular inversions

$$\begin{aligned}
\xi^* - a &= (a^2 + b^2 + c^2 - 1)\frac{\xi - a}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2}\\
\eta^* - b &= (a^2 + b^2 + c^2 - 1)\frac{\eta - b}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2}\\
\zeta^* - c &= (a^2 + b^2 + c^2 - 1)\frac{\zeta - c}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2}
\end{aligned} \tag{8.44}$$

which transform the interior of the unit sphere into itself.

These transformations (8.44) are indeed rigid transformations of the hyperbolic geometry. One verifies by elementary calculation that

$$1 - \xi^{*2} - \eta^{*2} - \zeta^{*2} = \frac{a^2 + b^2 + c^2 - 1}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2}(1 - \xi^2 - \eta^2 - \zeta^2)$$

and (8.45)

$$d\xi^{*2} + d\eta^{*2} + d\zeta^{*2} = \frac{(a^2 + b^2 + c^2 - 1)^2}{[(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2]^2}(d\xi^2 + d\eta^2 + d\zeta^2)$$

It follows that

$$\frac{d\xi^{*2} + d\eta^{*2} + d\zeta^{*2}}{(1 - \xi^{*2} - \eta^{*2} - \zeta^{*2})^2} \equiv \frac{d\xi^2 + d\eta^2 + d\zeta^2}{(1 - \xi^2 - \eta^2 - \zeta^2)^2} \tag{8.46}$$

and thus the rigid character of (8.44).

Our transformations (8.44) are, however, not hyperbolic movements. If, for example, C lies on the ξ-axis, then we readily see that the directions of the original ξ, η, ζ-axis are transformed into directions which form a left-hand coordinate system. We can remove this defect simply by an additional reflection.

We choose as a mirror plane--quite arbitrarily--the η, ξ-plane. A point ξ^*, η^*, ζ^* is transformed, by reflection on this plane, into

$$\xi' = -\xi^* \qquad \eta' = \eta^* \qquad \zeta' = \zeta^* \tag{8.47}$$

By submitting the transformed points ξ^*, η^*, ζ^* to this additional reflection, we obtain the following three-parameter set of true hyperbolic movements

$$\begin{aligned}
\xi' + a &= -(a^2 + b^2 + c^2 - 1) \frac{\xi - a}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2} \\
\eta' + b &= (a^2 + b^2 + c^2 - 1) \frac{\eta - b}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2} \\
\zeta' - c &= (a^2 + b^2 + c^2 - 1) \frac{\zeta - c}{(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2}
\end{aligned} \tag{8.48}$$

depending upon the three parameters a, b, c. It supplements the three-parameter group of rotations 8.2 to the complete six-parameter group of rigid movements of the hyperbolic geometry.

8.5. With regard to the practical problem of the distorted room, we are interested especially in those transformations which leave the position of the horizontal plane and of the median plane unchanged. Rotations of the type 8.2 thus are excluded. For similar reasons it follows that $b = c = 0$ in (8.44) and (8.48). Hence we find the one-parameter group of transformations determined by the inversion

$$\begin{aligned}
\xi^* &= a + (a^2 - 1) \frac{\xi - a}{(\xi - a)^2 + \eta^2 + \zeta^2} \\
\eta^* &= (a^2 - 1) \frac{\eta}{(\xi - a)^2 + \eta^2 + \zeta^2} \\
\zeta^* &= (a^2 - 1) \frac{\zeta}{(\xi - a)^2 + \eta^2 + \zeta^2}
\end{aligned} \tag{8.51}$$

and the reflection on the η, ζ-plane:

$$\xi' = -\xi^* \qquad \eta' = \eta^* \qquad \zeta' = \zeta^* \tag{8.52}$$

The center of the inversion sphere lies at the point a of the ξ-axis; it intersects the unit sphere at right angles and thus has the radius

$$D = \sqrt{a^2 - 1}$$

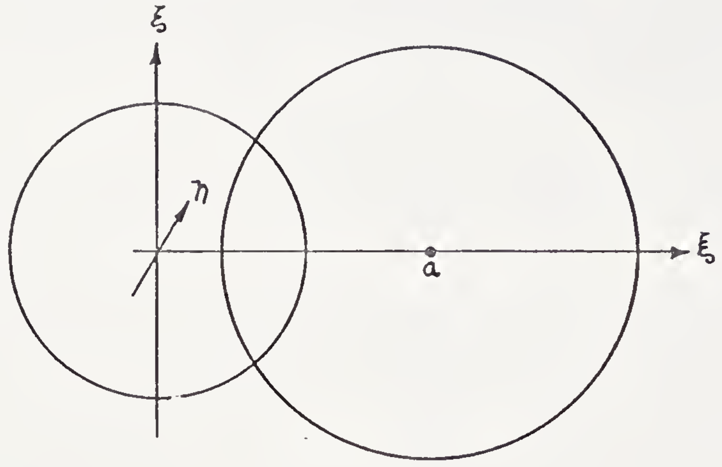


Fig. 75

By combining the equations (8.51) and (8.52), we obtain the transformation

$$\begin{aligned} \xi' &= -a - (a^2 - 1) \frac{\xi - a}{(\xi - a)^2 + \eta^2 + \zeta^2} \\ \eta &= (a^2 - 1) \frac{\eta}{(\xi - a)^2 + \eta^2 + \zeta^2} \\ \zeta' &= (a^2 - 1) \frac{\zeta}{(\xi - a)^2 + \eta^2 + \zeta^2} \end{aligned} \tag{8.53}$$

which is the basis for the construction of distorted rooms in the realm of hyperbolic congruence.

8.6. The simplicity of the equations (8.53) in the ξ, η, ζ space is, unfortunately, not preserved in the physical x, y, z space. The equations between bipolar coordinates γ, φ, θ and γ', φ', θ' are already complicated enough. We obtain these equations from (8.53) with the aid of the relations (8.27). Introduction of Cartesian coordinates x, y, z and x', y', z' then complicates the formulae even more.

In view of this, a numerical procedure for the design of a distorted room is indicated which is based upon the transformation (8.53) in the ξ, η, ζ space.

For this purpose we establish first direct relations between the physical coordinates x, y, z and the coordinates ξ, η, ζ.

The variables ξ, η, ζ and the bipolar coordinates γ, φ, θ are connected by the formulae

$$\begin{aligned} \xi &= e^{-\sigma(\gamma+\mu)} \cos\varphi \cos\theta \\ \eta &= e^{-\sigma(\gamma+\mu)} \sin\varphi \\ \zeta &= e^{-\sigma(\gamma+\mu)} \cos\varphi \sin\theta \end{aligned} \tag{8.61}$$

For the relation of Cartesian coordinates x, y, z and bipolar coordinates γ, φ, θ we may, in our present problem, safely use the simplified formulae (2.31 ...):

$$\gamma = \frac{2\sqrt{x^2 + z^2}}{x^2 + y^2 + z^2}$$

$$\tan\varphi = \frac{y}{\sqrt{x^2 + z^2}} \qquad (8.62)$$

$$\tan\varphi = \frac{z}{x}$$

We notice immediately that

$$\tan\varphi = \frac{y}{\sqrt{x^2 + z^2}} = \frac{\eta}{\sqrt{\xi^2 + \zeta^2}}$$

$$\tan\theta = \frac{z}{x} = \frac{\zeta}{\xi}$$

from which it follows that

$$\frac{y}{x} = \frac{\eta}{\xi}$$

$$\frac{z}{x} = \frac{\zeta}{\xi} \qquad (8.63)$$

We write these last equations in the form

$$\xi = \lambda x$$

$$\eta = \lambda y \qquad (8.64)$$

$$\zeta = \lambda z$$

where λ is a certain function of x, y, z or of ξ, η, ζ.

In order to determine λ we use the relation

$$\sqrt{\xi^2 + \eta^2 + \zeta^2} = e^{-\sigma\mu} e^{-\frac{2\sigma\sqrt{x^2 + z^2}}{x^2 + y^2 + z^2}} \qquad (8.65)$$

which follows from (8.61) and (8.62)

We first introduce the expressions (8.64) on the left side of (8.65). It follows

$$\lambda = \frac{e^{-\sigma\mu} e^{-\frac{2\sigma\sqrt{x^2 + z^2}}{x^2 + y^2 + z^2}}}{\sqrt{x^2 + y^2 + z^2}} \qquad (8.66)$$

Next we replace x, y, z on the right side of (8.65) by $\frac{1}{\lambda}\xi$; $\frac{1}{\lambda}\eta$; $\frac{1}{\lambda}\zeta$ and obtain

$$\sqrt{\xi^2 + \eta^2 + \zeta^2} = e^{-\sigma\mu} e^{-2\sigma\lambda \frac{\sqrt{\xi^2 + \zeta^2}}{\xi^2 + \eta^2 + \zeta^2}}$$

This leads to the expression of λ in ξ, η, ζ:

$$\lambda = -\frac{1}{2\sigma}\frac{(\xi^2+\eta^2+\zeta^2)(\sigma\mu + \log\sqrt{\xi^2+\eta^2+\zeta^2})}{\sqrt{\xi^2+\zeta^2}} \tag{8.67}$$

With these expressions for λ we get from (8.64) the desired formulae:

$$\xi = e^{-\sigma\mu} x \frac{e^{-\frac{2\sigma\sqrt{x^2+z^2}}{x^2+y^2+z^2}}}{\sqrt{x^2+y^2+z^2}}$$

$$\eta = e^{-\sigma\mu} y \frac{e^{-\frac{2\sigma\sqrt{x^2+z^2}}{x^2+y^2+z^2}}}{\sqrt{x^2+y^2+z^2}} \tag{8.68}$$

$$\zeta = e^{-\sigma\mu} z \frac{e^{-\frac{2\sigma\sqrt{x^2+z^2}}{x^2+y^2+z^2}}}{\sqrt{x^2+y^2+z^2}}$$

and the reversed formulae

$$x = -2\sigma\xi \frac{\sqrt{\xi^2+\zeta^2}}{(\xi^2+\eta^2+\zeta^2)(\sigma\mu + \log\sqrt{\xi^2+\eta^2+\zeta^2})}$$

$$y = -2\sigma\eta \frac{\sqrt{\xi^2+\zeta^2}}{(\xi^2+\eta^2+\zeta^2)(\sigma\mu + \log\sqrt{\xi^2+\eta^2+\zeta^2})} \tag{8.69}$$

$$z = -2\sigma\zeta \frac{\sqrt{\xi^2+\zeta^2}}{(\xi^2+\eta^2+\zeta^2)(\sigma\mu + \log\sqrt{\xi^2+\eta^2+\zeta^2})}$$

With the aid of the formulae (8.68) and (8.69) we can now design the distorted room as follows. We choose on the rectangular walls of the original room a suitable network of points (x_1, y_1, z_1). We determine the associated points (ξ_1, η_1, ζ_1) in the ξ, η, ζ space with the aid of the formulae (8.68). We next apply the transformation formulae (8.53), and obtain the transformed points $(\xi_1', \eta_1', \zeta_1')$. The parameter, a, thereby is determined by the position P_1 into which the frontal point P_0 is to be moved. Finally by (8.69) the transformed points (x_1', y_1', z_1') in the physical space are determined.

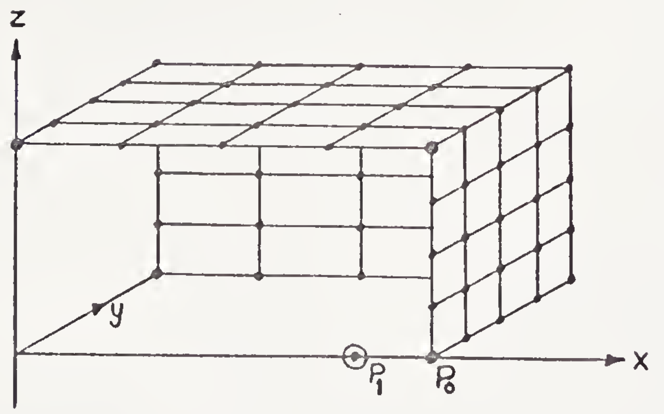


Fig. 76

This distorted network of points (x_1', y_1', z_1') then is the frame of the mechanical construction. Additional points can be found by interpolation.

The above outline shows that the design of a distorted room which is hyperbolically congruent to a rectangular room requires a considerable amount of numerical computations. We notice also that it can only be carried out if the constants σ and μ have been determined before. Theoretically a particular distorted room will be seen as a rectangular room only by a person who has the constants σ and μ used for the construction of the room. Whether or not individual differences are significant can, of course, be found only by the actual experiment.

8.7. Topological discussion of distorted rooms congruent to a rectangular room. The general form of distorted rooms congruent to a given rectangular room can be derived by a method similar to that in § 2.6. We determine for this purpose the curves in the ξ, η-plane which correspond to the rectangular cross section of the given room with the x,y-plane (Fig. 13). The front wall $x = x_0$ is represented by a curve symmetrical to the ξ-axis which arrives at infinity, i.e., on the circle $\rho = \rho_1$ with the azimuth $\varphi = \pm \pi/2$. The side wall $y = y_0$ can be described as a curve connecting a point $y_0 > 1$ of the y-axis with a point at infinity for which $\varphi = 0$. Thus, its image in the ξ, η-plane is a curve from the point $(0, \rho_1)$ of the η-axis to the point $(\rho_1, 0)$ of the ξ-axis. This leads to Fig. 77 where the shaded region indicates the interior of the room. We mention that, in an exact drawing, the corner angles on the map give directly the apparent non-Euclidean angles of the visual sensation related to the given rectangular room. This follows from the fact that the ξ, η map is a conformal image of the visual sensations.

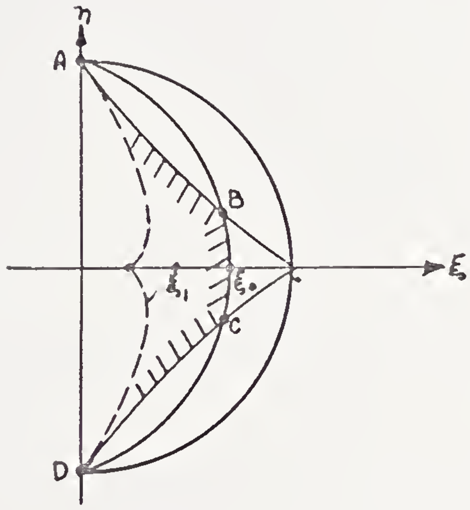


Fig. 77

We wish to construct a congruent distorted room such that the front-wall intersects the x-axis at a given point x_1 different from the original frontal point x_0. The corresponding points ξ_0 and ξ_1 in the ξ, η-plane can be found by the relations

$$\xi_0 = e^{-\sigma(\gamma_0+\mu)} \qquad x_0 = \cot \gamma_0/2$$
$$\xi_1 = e^{-\sigma(\gamma_1+\mu)} \qquad x_1 = \cot \gamma_1/2 \qquad (8.71)$$

which show that $\xi_1 > \xi_0$ if $x_1 > x_0$ and $\xi_1 < \xi_0$ if $x_1 < x_0$.

We next determine the parameter a in (8.53) by the condition that the frontal point ξ_0 be transformed into the point ξ_1. This means, on account of (8.53) that these points are related by the equation

$$\xi_1 = -a - \frac{a^2 - 1}{\xi_0 - a} \qquad (8.72)$$

from which it follows that

$$a = \frac{1 - \xi_0 \cdot \xi_1}{\xi_0 - \xi_1} \qquad (8.73)$$

Since both ξ_0 and ξ_1 are smaller than one, we conclude that $a > 0$ if $\xi_1 < \xi_0$, i.e., $x_1 < x_0$, and $a < 0$ if $\xi_1 > \xi_0$, i.e., if $x_1 > x_0$. The center, a, of the inversion sphere thus lies on the positive ξ-axis if the frontal point x_1 of the distorted room is closer to the observer than the point x_0. It lies on the negative ξ-axis if the frontal point x_1 is farther away than x_0.

The geometric interpretation of a as the abscissa of the center of a sphere which intersects the unit sphere at right angles implies that always $|a| > 1$. We prove that analytically by (8.73) as follows. Since $|\xi_0| < 1$ and $|\xi_1| < 1$, we conclude that always $(1 - \xi_0)(1 + \xi_1) > 0$ and also that $(1 + \xi_0)(1 - \xi_1) > 0$. We write these inequalities in the form

$$1 - \xi_0 \cdot \xi_1 > \xi_0 - \xi_1$$

$$1 - \xi_0 \cdot \xi_1 > \xi_1 - \xi_0$$

which means that

$$1 - \xi_0 \cdot \xi_1 > |\xi_0 \quad \xi_1|$$

and hence

$$|a| = \frac{1 - \xi_0 \xi_1}{|\xi_0 - \xi_1|} > 1$$

With the aid of this result we can formulate our information about the parameter, a, more precisely: In case of an interior distorted room $x_1 < x_0$, we have $a > 1$; in case of an exterior room $x_1 > x_0$, we have $a < -1$.

Let us now investigate in which direction the points of the curve ABCD in Fig. 77 move if the frontal point ξ_0 is transformed into ξ_1. For this purpose we write the first equation (8.53) in the form

$$\xi - \xi' = \frac{a(1 + \eta^2 - \xi^2) - (1 - \xi^2 - \eta^2)}{(\xi - a)^2 + \eta^2} \tag{8.74}$$

remembering that $\xi = 0$ in the horizontal plane. Since the denominator $(\xi - a)^2 + \eta^2$ is always positive, it follows that the sign of $\xi - \xi'$ is determined by the sign of the numerator

$$N = a(1 + \eta^2 - \xi^2) - \xi(1 - \xi^2 - \eta^2) \tag{8.75}$$

The second term is never positive on the boundary curve of Fig. 77. The sign of the first term is given by the sign of the parameter aa since the bracket $1 + \eta^2 - \xi^2$ cannot be negative.

Consider now the case $\xi_1 > \xi_0$ of an exterior room. We know that $a < -1$ and hence that $N < 0$, i.e., $\xi - \xi' < 0$. This means that every point of the curve ABCD moves with ξ_0 towards the right.

Consider next the case $\xi_1 < \xi_0$ of an interior room. Since $a > 1$, in this case, and $1 + \eta^2 - \xi^2 \geq 1 - \eta^2 - \xi^2$, we conclude by (8.75) that

$$N > (a - \xi)(1 - \xi^2 - \eta^2)$$

However, the quantity $a - \xi$ is certainly positive, since $a > 1 > \xi$. Consequently $N > 0$ and thus $\xi - \xi' > 0$. Thus every point of the curve ABCD moves with ξ_0 to the left.

We summarize the above results: If the curve ABCD is submitted to a rigid transformation of the special type (8.53), then all the curve points move simultaneously either to the right or to the left, depending on the direction in which the frontal point ξ_0 is moved.

The transformed curves thus must be localized in the manner indicated in Fig. 78 and Fig. 79.

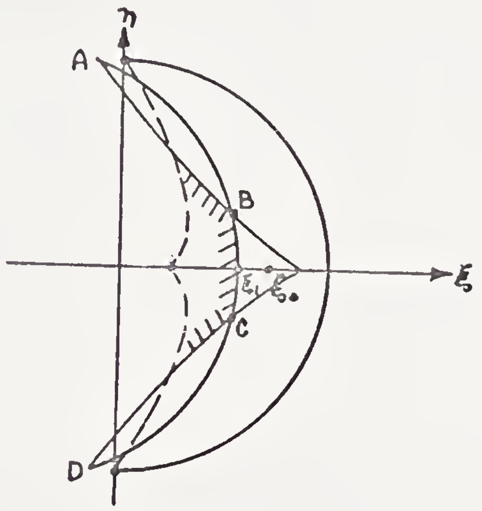


Fig. 78

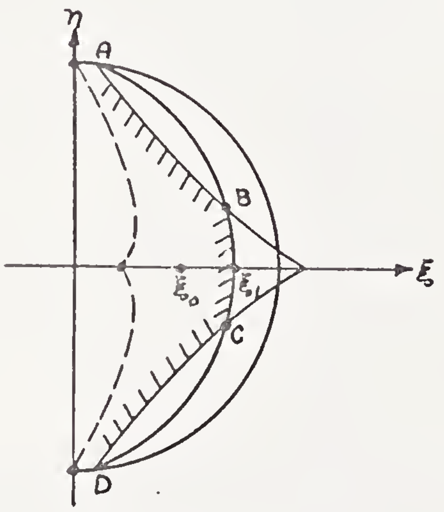


Fig. 79

Fig. 78 represents the ξ, η image of an interior distorted room. It allows us to determine the general form of such a room in the physical x,y-plane. The front wall is a curve which approaches the two eyes if extended beyond the corner points B and C. The side walls go from the corner points directly to the eyes. If extended beyond B and C they meet each other at a finite point of the x-axis (Fig. 80).

In a similar manner we find the general form of an exterior distorted room by Fig. 79. The front wall goes from the point x_1 asymptotically towards infinity. The side walls reach the infinity of the plane at two different asymptotic angles, and thus are curves which, like hyperbolae, have two asymptotes (Fig. 81).

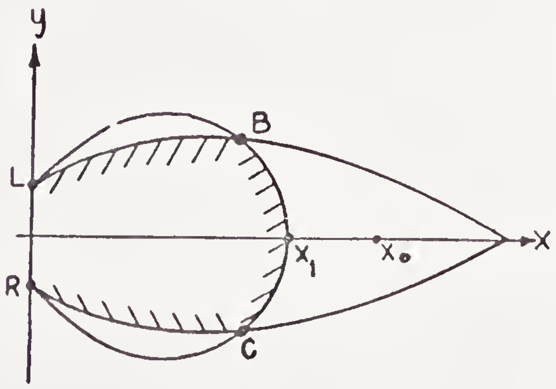


Fig. 80

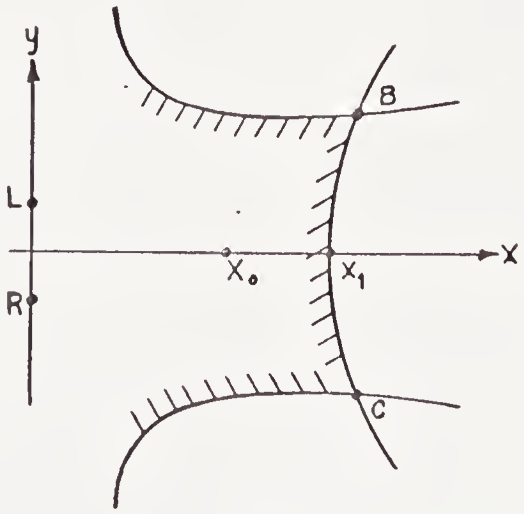


Fig. 81

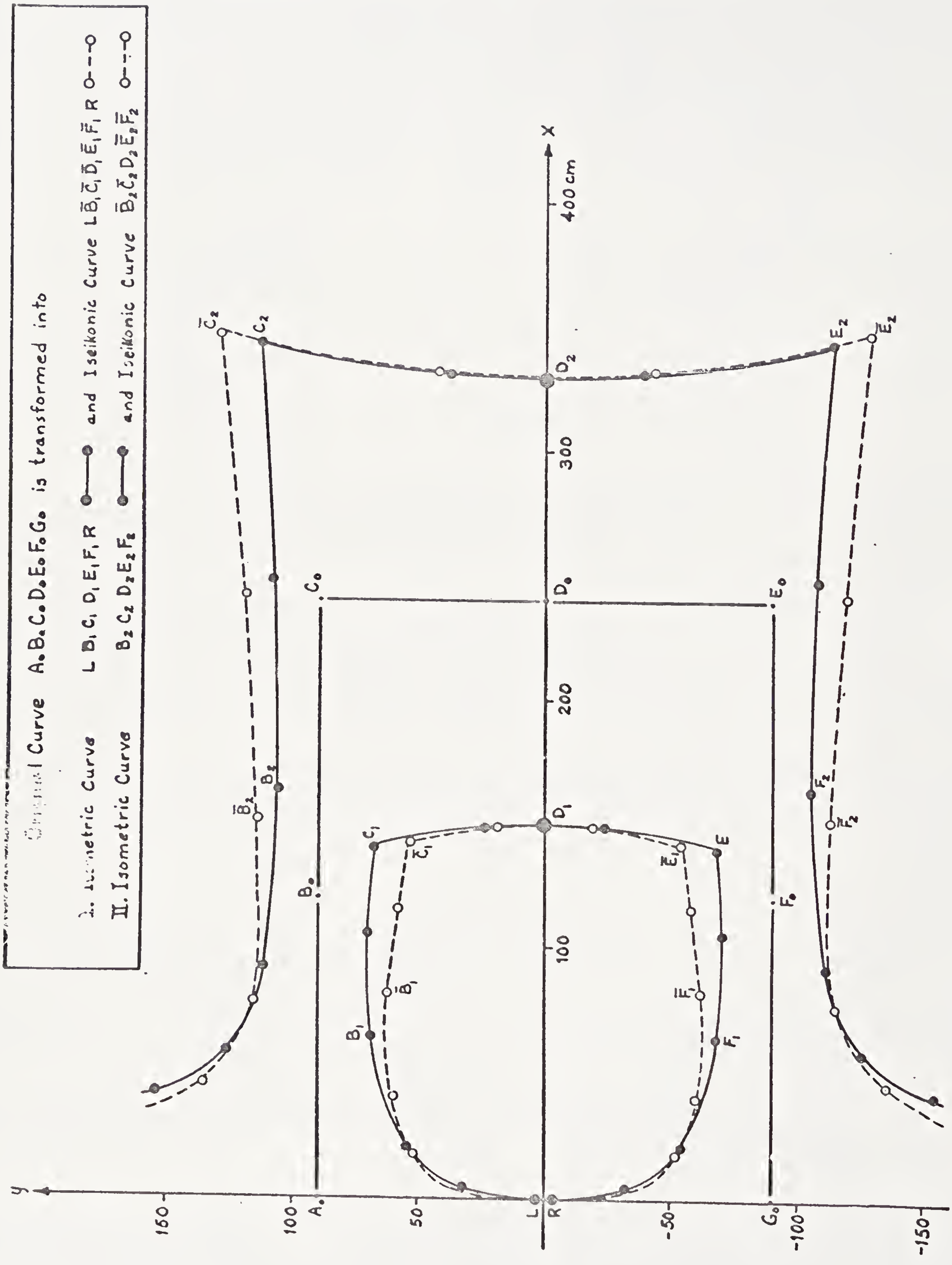

Curve A0 B0 C0 D0 E0 F0 G0 is transformed into
and Iseikonic Curve
II. Isometric Curve B2 C2 D2 E2 F2 and Iseikonic Curve
400 cm
300
200
100
150
100
50
-50
-100
-150
x
y

Fig. 82

8.8. We see by the above discussion that congruent rooms determined by the rigid hyperbolic movement (8.53) have the same general form as the equivalent rooms discussed in § 2.6. This is a rather surprising result in view of the fact that two entirely different mathematical principles of construction have been employed. It explains, on the other hand, why actual rooms, constructed on the basis of iseikonic equivalence, approach the impression of rectangularity already to a considerable degree.

It is not difficult to explain this interesting result. We have seen in § 4.8 that the iseikonic transformation

$$\begin{aligned} \gamma' &= \gamma + \tau \\ \varphi' &= \varphi \\ \theta' &= \theta \end{aligned} \qquad (8.81)$$

is a rigid transformation of the metric

$$ds^2 = M_0^2\,(\sigma^2\, d\gamma^2 + d\varphi^2 + \cos^2\varphi\, d\theta^2) \qquad (8.82)$$

where M_0 is a constant. This transformation is the basis of the construction of the above equivalent rooms. The transformation (8.53), on the other hand, is a rigid transformation of the differential

$$ds^2 = M^2(\gamma)(\sigma^2\, d\gamma^2 + d\varphi^2 + \cos^2\varphi\, d\theta^2) \qquad (8.83)$$

where $M(\gamma) = \dfrac{1}{\sinh \sigma(\gamma + \mu)}$. In a domain where $M(\gamma)$ varies only slowly we thus expect approximately the same result from both transformations. Such a domain, however, is a region of small γ-values, i.e., a region sufficiently far away from the observer. The limits of this region depend on the size of the constant μ. Observers with a small μ-value will notice the differences between both types of rooms sooner than observers with large μ-values.

In general, we must expect the differences to be greater in the neighborhood of the eyes, since $M(\gamma)$ varies rapidly in this region. Thus, differences between equivalent and congruent rooms will be more noticeable the smaller the original rectangular room is assumed to be or the nearer to the eyes the new frontal point x_1 is chosen.

The close numerical approximation of congruent rooms by equivalent rooms is illustrated in Fig. 82, which shows an actual computed example.* The dimensions of the original room are 1.8 x 2.4 m. The drawing shows the two different types of rooms by two pairs of horizontal cross sections belonging to the interior frontal point x_1 = 1.5 m. and to the exterior frontal point x_1 = 3.3 m., respectively.

*This example is taken from a detailed mathematical investigation of the problem of distorted rooms by Dr. A. Stein, which is to to be published in the near future.

CONCLUSION

The applications discussed in the last three chapters give ample evidence about the fruitfulness of the idea of considering the visual space as a Riemannian manifold. They also show the primary importance of the problem of finding the mathematical form of the metric of the visual space. We have demonstrated how this problem can be solved by an analysis of suitably chosen observations. Our conclusion has been that the geometry of the visual space is the hyperbolic geometry of Lobachevski, represented by the quadratic differential

$$ds^2 = \frac{1}{\sinh^2 \sigma(\gamma+\mu)} (\sigma^2 \, d\gamma^2 + d\varphi^2 + \cos^2\varphi \, d\theta^2)$$

This formula establishes, on the other hand, a relation between the visual and physical space. It expresses the general hypothesis that, for an individual observer, the apparent size, ds, of a line element is uniquely determined by its physical coordinates, i.e., by the bipolar differentials $d\gamma$, $d\varphi$, $d\theta$ and by the parallax γ of the base point of the line element. According to the above formula, judgment of visual size is not only based upon the angular differentials $d\gamma$, $d\varphi$, $d\theta$. It also depends upon the localization, γ, of the line element. Our scale of size seems to contract with increasing values of γ, namely, proportional to the expression $1/\sinh \sigma(\gamma + \mu)$. Such a well-defined contraction can probably be understood only if there is a physiological basis for it, either in the dioptric system of the eyes, on the retina, in the transmission to the brain, or in the cortex itself. However, we do not attempt to maky any hypothesis about this question.

The above mathematical form of the visual metric can, of course, not be considered as absolutely insured. Though our discussion has given a considerable amount of evidence for the hyperbolic geometry, careful quantitative tests have yet to be made. Any prediction derived mathematically from the metric differential gives us such a quantitative test. Our above applications represent only a small part of the implications of our theory. The laws of pictural reproduction of visual sensations, i.e., of the aesthetic impression of an environment, can be derived without principal difficulty. This would give us a theory of visual perspective quite different from the perspective based on projection from one center. Furthermore, it should be possible to find new directives for the design of binocular instruments such as binocular microscopes or range finders. The final test of our theory then rests upon the consistency of such applications.

It is clear that any specific application requires the knowledge of the constants σ and μ in our metric. These constants are probably quite different for different observers and even for the same observer at different periods of his life. It is well possible that the values of these constants are related to the individual differences of persons with regard to certain technical abilities.

For the problem of measuring these constants it is of interest that the above metric can be expressed in a more general form. We write

$$\sinh \sigma(\gamma + \mu) = \tfrac{1}{2} e^{\sigma\mu} \, (e^{\sigma\gamma} + K \, e^{-\sigma\gamma})$$

where $K = -e^{-2\sigma\mu}$. By introducing this expression and suppressing the constant factor $e^{-2\sigma\mu}$, we obtain

$$ds^2 = \left(\frac{2}{e^{\sigma\gamma} + K e^{-\sigma\gamma}}\right)^2 (\sigma^2 d\gamma^2 + d\varphi^2 + \cos^2 \varphi \, d\theta^2)$$

depending on the individual constants σ and K, instead of on σ and μ. If we introduce new coordinates ξ, η, ζ by the transformation

$$\xi = 2e^{-\sigma\gamma} \cos \varphi \cos \theta$$

$$\eta = 2e^{-\sigma\gamma} \sin \varphi$$

$$\zeta = 2e^{-\sigma\gamma} \cos \varphi \sin \theta$$

we obtain Riemann's normal form (3.72)

$$ds^2 = \frac{d\xi^2 + d\eta^2 + d\zeta^2}{[1 + \frac{1}{4} K (\xi^2 + \eta^2 + \zeta^2)]^2}$$

This shows that $K = -e^{-2\sigma\mu}$ is the <u>constant Riemannian curvature of the visual space</u>.

In the case of the hyperbolic geometry we have of course $K < 0$. However, since σ and K are constants to be determined experimentally, we may abandon the hypothesis $K < 0$ and leave the answer to the experiment. In other words, we need not exclude that K may have values $K \geqq 0$. Since in this case the geometry is Euclidean or elliptic, we thus consider it possible in principle to find observers with a Euclidean or elliptic visual space. We still assume that the visual space is a space of constant curvature. But whether hyperbolic, Euclidean or elliptic is decided by the experimental result of K being negative, zero, or positive.

From this point of view, we recognize that a survey of a great number of persons with respect to the constants σ and K must be in itself an interesting statistical investigation. It will give us information about the question whether or not there exist significant personal differences in the relation of physical and visual perception. For personal K-values near -1 sensed size and physical size approach each other so that visual and physical space are practically identical in a large domain. For negative K-values near zero or even more for positive K-values, this approximation must be poor: Visual and physical space are far apart.

By the evaluation of the alley experiments we find for Blumenfeld's observer Lo the curvature $K = -0.095$. However, this result is only a rough estimate which we have used for illustration. We are certainly not justified in basing any practical application of our theory on it. For such applications it is necessary to obtain, as a first step, a wide range of careful experimental measurements of the constants σ and K.