DOCUMENT RESUME

ED 143 546 SE 023 030

AUTHOR Kutuzov, B. V.
TITLE Studies in Mathematics, Volume IV. Geometry.
INSTITUTION Stanford Univ., Calif. School Mathematics Study Group.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE 60
NOTE 582p.; For related documents, see SE 023 028-041; Contains occasional light and broken type

EDRS PRICE MF-$1.16 HC-$31.47 Plus Postage.
DESCRIPTORS *Calculus; College Mathematics; *Geometry; *Instructional Materials; Mathematics; Reference Materials; *Secondary School Mathematics; *Textbooks
IDENTIFIERS *School Mathematics Study Group

ABSTRACT

This book is a translation of a Russian text. The translation is exact, and the language used by the author has not been brought up to date. The volume is probably most useful as a source of supplementary materials for high school mathematics. It is also useful for teachers to broaden their mathematical background. Chapters included in the text are: (1) Geometric Figures as Point Sets; (2) Geometric Constructions; (3) The Transformation of Figures; (4) Parallel Translations; (5) Rotation; (6) Symmetry; (7) Similarity; (8) Inversion; (9) The General Problem of Measuring Lengths, Areas, and Volumes; (10) Euclid's "Elements"; (11) The Geometry of Lobachevskii; (12) The Axiomatic Structure of Geometry; and (13) The Idea of an Interpretation of a Geometric System. A selected bibliography is included. (RH)

SCHOOL MATHEMATICS STUDY GROUP

STUDIES IN MATHEMATICS

VOLUME IV

GEOMETRY

by B. V. KUTUZOV

Translated by Louis I. Gordon and Edward S. Shater

This is a complete translation of a Russian text done by the CHICAGO SURVEY OF RECENT EAST EUROPEAN MATHEMATICAL LITERATURE. This book is being made available because the Panel on Teacher Training Materials of the School Mathematics Study Group feels that it will be of great value to teachers in this country readying themselves to teach an improved geometry course. In making this exact translation available, there is, of course, no endorsement of the historical or other extraneous statements made.

Prepared by the Chicago Survey of Recent East European Mathematical Literature

Reproduced and Distributed by the SCHOOL MATHEMATICS STUDY GROUP

PRINTED IN THE U.S.A.

FOREWORD

The series "Studies in Mathematics" is devoted to material useful to high school mathematics teachers and directly related to high school mathematics courses. In general, the volumes in this series contain about the amount of material that can be covered in a summer or in-service course. Earlier volumes in this series were devoted to set theory, geometry and algebra.

This volume is a slight departure from the established pattern in two ways. First, it contains more material than can be covered in a short course and some of this material is somewhat more sophisticated than that in the earlier volumes. Hence, this particular volume will probably be used chiefly as a source of supplementary material.

Second, this is a translation of a Russian text. The SMSG Panel on Teacher Training Materials has examined a number of Russian texts for teachers and will examine more. So far, only a few of these, of which this is one, have seemed to the Panel to be directly useful to teachers in this country. We hope to be able to make available in this series additional translations which do seem to be useful.

The translation of this text was carried out under the auspices of the Survey of Recent East European Mathematical Literature, located at the University of Chicago. We are very pleased that the Survey has been willing to make this translation available through this series.

One warning to the reader--the translation is exact, and the language used by the author has not been brought up to date. In particular, where we would speak of the "product", the author speaks of the "sum" of transformations.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

PART TWO - ELEMENTARY GEOMETRICAL TRANSFORMATIONS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

FOREWORD TO THE FIRST EDITION

The present text is based entirely on secondary school course content and is intended to serve the professional needs of future teachers, in particular the broadening of their mathematical culture and the cultivation of the required skills.

We begin with the study of the fundamental set-theoretic concepts, the application of the general concept of the function (according to Lobachevskii) in geometry, and the concept of a group of transformations; the rudiments of the general theory of groups are presented. It is not possible today to teach geometry successfully without mastering the elements of the geometry of Lobachevskii.

The text contains a considerable number of suggested exercises. Independent work by the students is indispensable, and in each chapter topics for such work are given and reference is made to the appropriate literature, which, notwithstanding the novelty of the program, facilitate rapid orientation for both teacher and students.

Especial importance is attached to the exercises connected with the structure of secondary school textbooks. The most progressively organized school course in geometry was the creation of the outstanding Russian savant and pedagogue H. A. Glagolev. The present work is intended to aid in putting into effect the ideas on which his textbook is founded. The tendency to put off the beginning of a systematic course in geometry until the upper classes of the middle school, and in the sixth and seventh classes to substitute for it some sort of amorphous potpourri of logic and "intuitive deduction" must be regarded as altogether indefensible. Within the framework of the existing school program in geometry the teacher has inexhaustible opportunities to inculcate important ideas of contemporary science - set-theoretic concepts, the general concept of the function, geometrical transformations, the concept of group, the axiomatic method, and so on. From our point of view such a change in the school program as, for example, the introduction of differential and integral calculus without a sufficient foundation, appears to be at present outdated.

The present course together with secondary school texts will almost completely provide for the study of all parts of the program. Lack of space has forced me to omit two chapters devoted to surveying and the application of trigonometry to geometry. This gap can be readily made good by the availability of other existing textbooks. The material in small type* is not to be considered optional.

Moscow, 1st August 1950.

B. Kutuzov

* Material which appeared in the original in small type is here indicated by the use of wider margins.

FOREWORD TO THE SECOND EDITION

In the second edition only slight additions have been introduced, which take into account suggestions from the reviews of the first edition. The ideas of projective geometry would require a precise and detailed construction of projective space, and are not expounded in this book.

Moscow, 25 October 1954.

B. Kutuzov

PART ONE

INTRODUCTION

Chapter I

GEOMETRIC FIGURES AS POINT SETS

In the beginning of Chapter I, the fundamental concepts of axiom and theorem are briefly discussed. These questions will be taken up in greater detail in part IV of this book which is devoted to the foundations of geometry and the geometry of Lobachevskii.

In the main, Chapter I deals with the notion of a geometric figure as a point set and the application in geometry of basic set-theory notions (the operations of intersection and union of figures, and the identification of elements of a figure).

1. BASIC CONCEPTS AND DEFINITIONS IN GEOMETRY

There cannot be two opinions concerning the origin of mathematical concepts.

"The concepts of number and figure are derived, in fact, from the real world ... The concept of figure, as well as that of number, is derived exclusively from the external world and did not at all originate in the mind as a result of pure thought. Before men could arrive at the concept of figure there had to exist material objects possessing various forms which were compared one with another. Pure mathematics has as its subject matter spatial forms and quantitative relations in the objective world, i.e. an entirely real content. The fact that this content appears in extremely abstract form hardly conceals its origin in the external world. In order to study these forms and relations in their pure aspect, it is necessary to divorce them completely from their context, removing this context as something irrelevant to the matter at hand. In this manner are obtained points... lines..."(1)

In geometry, as in every mathematical science, we meet with concepts which are defined in terms of other concepts. These, in turn, are frequently defined with the aid of still other concepts and so on. Let us consider an example. A circle is defined as the locus of points lying in a plane such that the segments joining them with a point O in the same plane are equal to a given segment R.

Here the concept "circle" is defined with the aid of a series of other concepts: "point," "plane," "lie on," "locus of points," "segment," "joining," "equal." In order to define a circumference, these latter concepts have to be regarded as already known. If, however, these concepts also are unknown, we must either define them in terms of still other concepts or else leave them without definition, limiting ourselves to explanations or simple illustrations only.

Since a "chain of definitions" can not be continued without end, we must, clearly, <u>begin</u> with concepts which remain undefined.

In geometry, we choose certain concepts as <u>basic</u>, trying at the same time to keep their number as small as possible. Basic concepts remain undefined. They may be chosen in a variety of ways.

---

(1) F. Engels Anti-Deuring
K. Marx and F. Engels, Sochinenia, Vol. 14, page 39.
Marx-Engels Institute, 1939.

Most frequently the choice falls on such undefined concepts as "point," "line," "plane," "belong to" or "lie on," "lie between," "be equal."

All the remaining concepts met with in geometry should be defined with precision in terms of these basic concepts.

For instance, the concept of segment is defined as the set of "points" "lying between" two given points A and B. The points A and B are called the "end points" of the segment AB.

Expressions such as "to define correctly" or "to possess a certain property" belong to the domain of logic. In the study of geometry, the fundamentals of logic as well as certain basic concepts of arithmetic are regarded as known.

2. AXIOMS AND THEOREMS

Just as a "chain of definitions" in geometry cannot be continued endlessly, so must also a "chain of proofs" be finite. Ordinarily, any proposition in geometry, for instance the Pythagorean theorem, is proved on the basis of certain preceding propositions; the latter are proved with the aid of others proved still earlier, and so on. This series of references to "previously proven" propositions cannot be continued endlessly.

Geometry as a mathematical discipline begins with propositions which are accepted without proof. These propositions are called geometric axioms. The basic axioms of geometry are derived from experience, from practice.

Examples of axioms: "On a straight line, there are at least two points. In a plane, there exist at least three points which do not lie on one straight line."

"If two planes α and β have a common point A, then they have at least one other common point B."

Euclid's Axiom. "Let a be an arbitrary line, and A a point not on a; then in the plane determined by the line a and the point A there does not exist more than one line passing through the point A and not intersecting the line a."

Let us note that in the last axiom - the axiom of parallels - a statement is made concerning the existence of "not more than one line passing through the point A and not intersecting the line a" (fig. 1). The axiom does not state whether even one such line exists. If there existed no lines at all not intersecting the line a, the axiom as stated would still be valid.

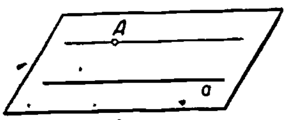


Fig. 1

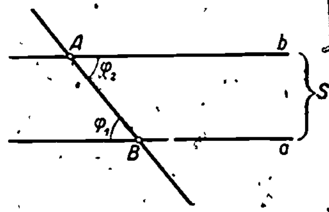


Fig. 2

That there exists, in the plane α determined by the line a and the point A at least one line passing through the point A and not intersecting the line a, is something which may be proved. This proposition is a theorem. A simple demonstration of it is known:

Let B be an arbitrary point on line a (fig. 2). Let us join the point B to the point A by means of the line AB, and let us construct the line b so that the alternate interior angles ϕ_1 and ϕ_2 are equal. Then line b does not intersect line a. In fact, if we assumed that lines a and b intersected in the point S we should find, that in the triangle SAB the exterior angle ϕ_1 was equal to the non-adjacent interior angle ϕ_2. Consequently, the point of intersection S of the indicated lines a and b cannot exist.

Combining the assertions of Euclid's axiom and the theorem just proved, we can state that there exists one and only one line b which, under the conditions considered above, does not intersect the line a.

The lines a and b, which are co-planar and non-intersecting are called parallels.

Every proposition in geometry which is not in the list of axioms is a theorem and needs to be proven. The simplicity or "self evidence" of the proposition is here immaterial.

Many propositions possessing a high degree of "self evidence" are nevertheless not in the list of axioms and, consequently, should be proven as theorems. On the other hand, there are propositions which are not at all "self evident" but, being in the list of axioms, are acceptable without proof.

In a textbook, it is possible to show only individual examples of proofs carried out in full logical detail and ultimately based only upon the axioms. An exhaustive axiomatic exposition of geometry could only be taken up in bulky scientific treatises.

A full list of axioms, composed only at the end of the XIXth and the beginning of the XXth centuries, will be given in part IV of our text. In the following exposition, we shall consider the contents of the mathematical course in the secondary schools as already known to the reader.

3. THE CONCEPT OF A GEOMETRIC FIGURE AS A POINT SET

Any set of points is called a figure. The notion of set is a primary concept which may only be explained but is not defined. In place of the word "set" other, equivalent, words are used: "aggregate," "class," "collection," "system," etc.

Sets consisting of points are known as point sets or figures. The points belonging to a set are the elements of this set.

One speaks of the set of points of a segment, the set of points of a line, the set of points of a sphere and so on.

One may also speak of the set of all lines in space, the set of all spheres, the set of all planes tangent to a given sphere and so on. In these latter cases the elements of the sets are themselves sets.

A finite set is often given by simply pointing out its elements, for instance, the set of spokes in the wheels of a given bicycle.

An infinite set is given by indicating a property possessed by its elements. After indicating this characteristic property of the elements of a set, it is possible to decide with regard to any object, whether or not it belongs to the given set.

Let us consider some examples of figures.

A point, a line, a plane as well as the whole of space are figures. A point is a set consisting of one element; a line, a plane and the whole of space are infinite sets of points.

When we speak of a segment we should distinguish three figures. We shall call a set consisting of two points A and B a zero-dimensional segment. The set consisting of the points A and B and all the points lying between A and B will be called a one-dimensional closed segment or simply a segment.

Here we should understand the word "closed" to mean that the ends of the segment, i.e. the points A and B, belong to the segment AB.

The set consisting only of the points lying between the points A and B, will be called a one-dimensional open segment or interval.

Here we should understand the word open to mean that the end points of the interval, i.e. the points A and B, do not belong to the interval.

An interval is a segment with the end points excluded; a zero-dimensional segment is the set of end points of a segment.

A conventional representation of these figures is given in fig. 3.

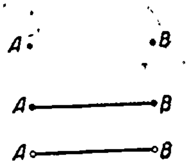


Fig. 3

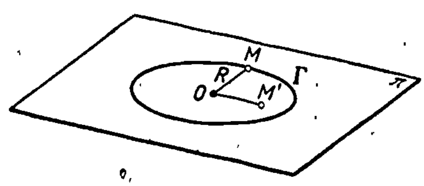


Fig. 4

Let us also distinguish between the following figures.

A circle Γ is the set of points M of a plane π whose distances from a given point O in the plane π (fig. 4) are equal to a given segment R, i.e. such that

$$OM = R = \text{const.}$$

A closed disk, or simply a disk, is a set consisting of the points M' of the plane π lying inside the circle Γ, together with the points of the circle Γ itself (fig. 4), i.e. such that

$$OM' \leqslant R.$$

The circle Γ is called the boundary of the disk. The circle is a curve.

An open disk is a set consisting of the points M'' interior to the circle Γ, i.e. such that

$$OM'' < R.$$

An open disk is a disk without its boundary.

A sphere is the set of all points which are at a given distance R from a given point O, i.e. such that

$$OM = R = \text{const}.$$

A sphere is a surface.

A solid sphere or globe is the set of all points M' at a distance not greater than R from a given point O (fig. 5), i.e. such that

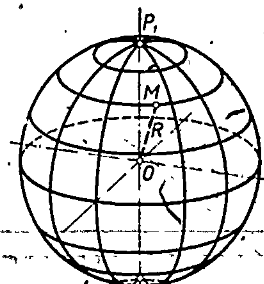
$$OM' \leqslant R.$$

A globe is a solid. Its boundary is a sphere.

An open globe is a globe without its boundary, i.e. the set of all points M'' such that

$$OM'' < R.$$

Fig. 5



The set of all points M''' exterior to a given solid sphere ($OM''' > R$) is also a figure.

The equator, parallels and meridians on a sphere are figures. The two poles P_1 and P_2 of a sphere form a zero-dimensional diameter of the sphere.

Let us consider now the notion of a triangle. We shall distinguish the following figures.

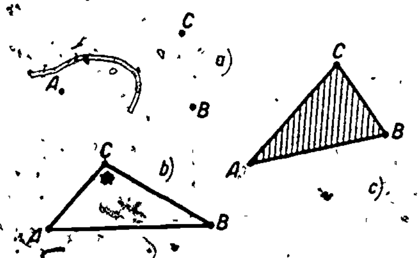


Fig. 6

A zero-dimensional triangle ABC is a set consisting of three non-collinear points A, B, and C (fig. 6a).

A one-dimensional triangle ABC is the set of points belonging to the segments AB, BC, and CA where the end points A, B, and C are non-collinear (fig. 6b). Such a triangle is frequently called a linear triangle.

A two-dimensional triangle ABC consists of all points of the one-dimensional triangle ABC together with the points of the plane ABC interior to this linear triangle (fig. 6c). The one-dimensional triangle ABC is the boundary of the two-dimensional triangle ABC. A linear triangle has no boundary. A two-dimensional triangle with the boundary excluded is called an open two-dimensional triangle.

The necessity of distinguishing these three figures ABC appears not only in geometry. In mechanics, for instance, when we speak of finding the center of gravity of a triangle ABC, the problem has no meaning unless we make clear which of the three forms of the triangle we have in mind.

Let us find the center of gravity of each of the triangles ABC.

Let equal masses m be concentrated at each of the points A, B, and C of a zero-dimensional triangle (fig. 7). The center of gravity D of the two masses concentrated at the points A and B is at the midpoint of the segment AB. Let us concentrate both of these masses at the point D. This leaves the center of gravity of the zero-dimensional triangle unchanged. Dividing then the median CD by means of the point E in inverse ratio to the masses, we find

$$\frac{CE}{ED} = \frac{2m}{m} = 2.$$

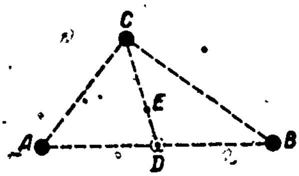


Fig. 7

Thus E, the point of intersection of the medians, is the center of gravity of the zero-dimensional triangle.

Let us consider now a two-dimensional triangle. Here it is necessary to find the center of gravity of a homogeneous triangular plate ABC. It is impossible to find with mathematical exactness the center of gravity of such a plate without the aid of the theory of limits. We shall limit ourselves to a heuristic discussion.

Let us divide the plate ABC into very thin strips parallel to the base AB. The center of gravity of each strip PQ (fig. 8) will be at the center E of that strip. If we were to concentrate the mass of each strip PQ at its center of gravity E, we should obtain a non-uniform distribution of mass along the median CD. Consequently, the center of gravity of the triangular plate lies on the median CD. This is true for any median, and it follows that the desired center of gravity of a two-dimensional triangle is at the point of intersection of the medians. The result is the same as for a zero-dimensional triangle.

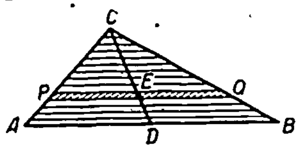


Fig. 8

A different result, however, is obtained for a one-dimensional triangle.

Let us consider a homogeneous "wire" triangle ABC (fig. 9). The center of gravity of each side of the triangle is at its midpoint. The masses of the sides are proportional to the lengths of the sides. Let γ be the linear density of the one-dimensional triangle, i.e. the mass of a unit of length. Let us concentrate the mass of each side at its midpoint. We obtain a zero-dimensional triangle $A_1 B_1 C_1$ with masses γa, γb, γc at the vertices. The center of gravity, C_2, of the two masses γa and γb divides the segment $A_1 B_1$ in a ratio inversely proportional to the masses, i.e.

$$\frac{A_1 C_2}{C_2 B_1} = \frac{\gamma b}{\gamma a} = \frac{b}{a} = \frac{\frac{1}{2} b}{\frac{1}{2} a} = \frac{A_1 C_1}{C_1 B_1}.$$

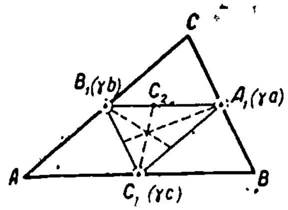


Fig. 9

Thus the point C_2 divides the side of $A_1 B_1$ of the triangle $A_1 B_1 C_1$ into parts proportional to the adjoining sides of this triangle. This implies that $C_1 C_2$ is the bisector of the angle C_1 of the triangle $A_1 B_1 C_1$. Concentrating the masses γa and γb at their center of gravity C_2, we reduce the problem to finding the center of gravity of two masses $\gamma(a + b)$ and γc concentrated at the ends of the segment $C_2 C_1$. The desired center of gravity, consequently, lies on the angular bisector $C_1 C_2$ of the triangle $A_1 B_1 C_1$. But there is no reason for singling out $C_1 C_2$ from the other bisectors. It follows that the center of gravity of a one-dimensional triangle is at the point of intersection of the angular bisectors of the triangle whose vertices are at the mid-points of the sides of the given triangle (fig. 9).

Only in the case of an equilateral triangle do the centers of gravity of all three figures coincide.

In some elementary geometry textbooks the word triangle is used to denote a one-dimensional triangle. The intersection of the medians is, at the same time, called the center of gravity of the triangle, which is, of course, incorrect.

A convex figure is a figure which together with every pair of points A and B belonging to it, contains all points of the segment AB.

All regular polyhedrons, considered as solids, together with their faces, edges, and vertices are convex figures. There are five regular polyhedrons.

Refining our definition as we did in the case of the triangle, we have to distinguish, for instance, the figures: a zero-dimensional tetrahedron - the set of vertices A, B, C and D of a regular tetrahedron; a one-dimensional (or linear) tetrahedron - the set of vertices and edges of a regular tetrahedron; a two-dimensional tetrahedron - the set of points belonging to the faces, edges and vertices; and, finally, a three-dimensional tetrahedron - the set of interior and boundary points of a regular tetrahedron (fig. 10).

Such a differentiation can also be carried out for the cube (fig. 11), octahedron (fig. 12), dodecahedron (fig. 13) and icosahedron (fig. 14).

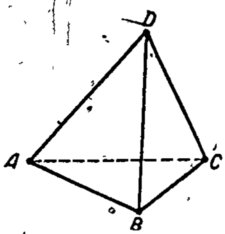


Fig. 10

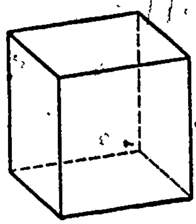


Fig. 11

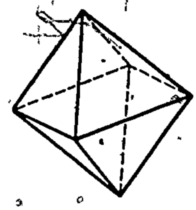


Fig. 12

If we take any two points A and B belonging to one of the above-mentioned three-dimensional figures, then the entire segment AB with all its points will also belong to the figure.

Two-dimensional, one-dimensional and zero-dimensional regular polyhedrons are not convex figures.

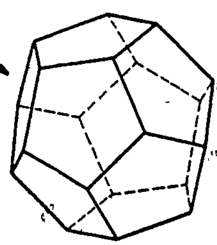


Fig. 13

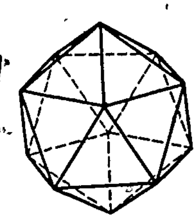


Fig. 14

Each separate face, separate edge and separate vertex of a regular polyhedron is a convex figure.

A disk and a solid sphere are convex figures. However, their boundaries, the circle and the sphere, are not convex.

A point, a segment, a line, a plane, the whole of space are convex figures.

Considering for instance a plane, if we join any two points of the plane by a segment, we obtain a segment belonging entirely to the plane, which proves that the plane is convex.

We pause to consider the assertion: "a point M is a convex figure." Let us follow closely the definition of convexity. Take any two points A and B of the figure; they coincide with the point M. The segment AB consists of the coincident end points A and B. There are no inner points on this segment. Thus all the points of the segment AB belong to our figure - the point M. It follows that a point is a convex figure.

Out of formal logical considerations we are led to introduce the so-called "empty set." This is a set which does not contain any elements. The empty set of points is also considered a figure. As an example of an empty set we may take the set of inner points of a segment whose end points coincide. We shall call such a segment a null segment.

It is necessary to differentiate between the notion of "null" segment and that of "zero-dimensional" segment.

It frequently turns out that there do not exist any elements with some given property. The set of such elements is empty.

One may usually picture an empty set as an empty bag containing no objects. We must also distinguish between a set containing a single element and the element itself. It is useful to have before one's eyes a "visual model" of such a set: a bag containing one object.

We may say that "a set is a multitude conceived as one" (36). This "multitude," however, may consist of one object or may even contain no objects at all.

Frequently statements containing the notion of empty set are formulated by the use of "double negation." For instance, we shall prove that "a point M is a convex set" as follows: Let us take any two points A and B belonging to the given figure. There are no points of the segment AB which do not belong to the given figure M. This is what was required to show. This manner of argument using double negation should be mastered.

If a figure has no boundary then the figure should be considered as closed, since we may say that the set of points forming the boundary is empty, and consequently there are no boundary points which do not belong to the figure. It is indeed in this sense that the circle and the sphere are closed figures.

The empty point set should be considered as a convex set. In fact, if the ends of any segment belong to such a figure, then there are no points on the segment which do not belong to the given figure. Speaking plainly, there is neither such a segment nor any points belonging to it. The need to introduce the empty set is dictated by the same considerations as the introduction of zero in arithmetic. (36).

In concluding this section, let us point out a large and important class of figures - figures of revolution. Let there be given a line ℓ and a figure ϕ (fig. 15). The point M belongs to the figure of revolution Ω if we can find in the set ϕ a point

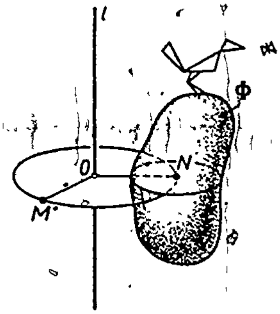


Fig. 15

N so that the foot, $\mathcal{O}$, of the perpendicular from M to ℓ coincides with that of the perpendicular from N to ℓ, and the distances OM and ON of the points M and N from the line ℓ are equal:

$$OM = ON$$

In other words, every point N of the set ϕ "describes" a circle with its center O on the "axis of revolution" ℓ and radius ON. The points of all such circles form the figure of revolution. The figure ϕ, generating the figure of revolution Ω, is called the revolving figure.(1)

Let us consider some examples of figures of revolution. If we rotate a <u>line m</u> parallel to the axis of rotation ℓ then we obtain the <u>surface</u> of a circular cylinder (fig. 16). Only the part of this surface described by the segment PQ of the line m is shown in the drawing.

If we revolve the entire <u>strip</u> between the parallel ℓ and m including the points on the lines ℓ and m (fig. 16), then we obtain a right circular <u>solid</u> cylinder.

(1) In figure 15, the "solid," the "bird," and the "butterfly" constitute one figure ϕ.

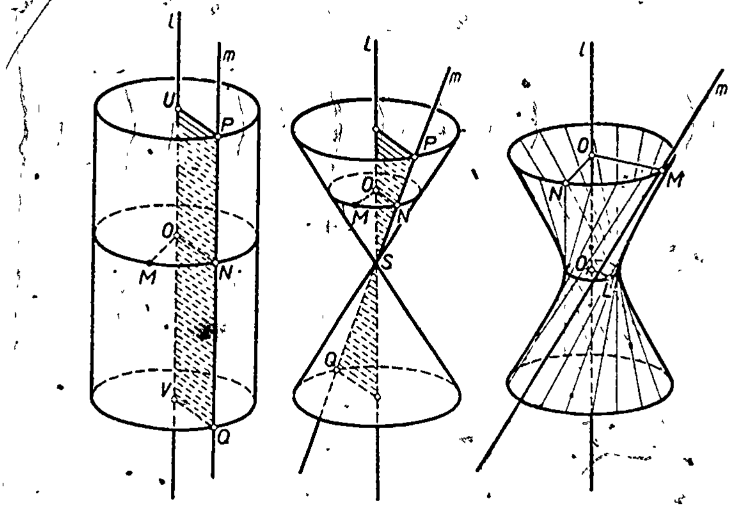


Fig. 16 Fig. 17 Fig. 18

If we revolve a line m not perpendicular to the axis of rotation ℓ and intersecting ℓ in a point S (fig. 17), we obtain a conical surface of revolution. Only the part of this surface described by the segment PQ of the line m is shown in the drawing.

If we revolve the part of the plane between the lines ℓ and m, i.e. the angle mSL and its vertical angle, we obtain a solid circular cone.

Let ℓ and m be non-perpendicular skew lines. Revolving m around ℓ we obtain a surface called a hyperboloid of revolution of one sheet (fig. 18).

The set of points belonging to the disks described by the segments OM constitute a solid of revolution, whose boundary is the hyperboloid of one sheet. Only part of the surface is shown in the drawing. There are also represented the successive positions of the revolving line m. The circle described by the endpoint L of the segment O_1 L, where O_1 L is the shortest distance between the lines ℓ and m, is called the minimal circle (throat circle) of the hyperboloid of one sheet.

If the line m forms with ℓ an angle approximating a right angle, the parts of the surface on either side of the minimal circle will be close to the plane of this circle. If the line m is perpendicular to ℓ, the figure of revolution is the plane with the exclusion of an open disk (fig. 19). As the angle between ℓ and m diminishes, the surface approaches more and more the form of a cylinder. As the radius O_1L of the minimal circle decreases, the hyperboloid of one sheet approaches the form of a cone.

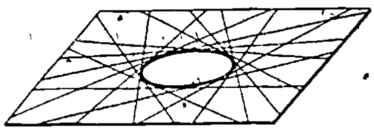


Fig. 19

So far we have used a straight line as the generator of figures of revolution. Let us now revolve a circle about an axis.

We shall limit ourselves to the case in which the revolving circle lies in a plane passing through the axis of revolution ℓ, and itself does not intersect the axis ℓ. Such a figure of revolution is called a torus (fig. 20). If we revolve a disk, we shall obtain a solid whose boundary is a torus.

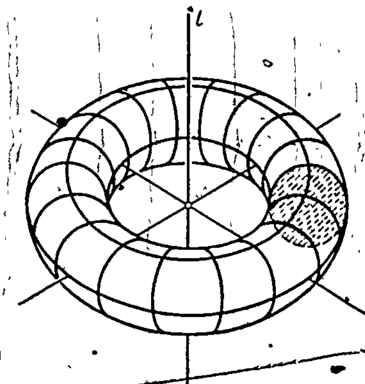


Fig. 20

All of space is also a figure of revolution, generated, say, by itself.

If we revolve a plane which is parallel to the axis of revolution ℓ, the figure of revolution will be "space with a channel," i.e. all of space with the exclusion of an open cylinder whose axis is ℓ and whose radius is equal to the distance from ℓ to the rotating plane. Fig. 19 represents the cross section of such a figure by a plane perpendicular to the axis ℓ.

4. THE INTERSECTION OF FIGURES. EXAMPLES. CONIC SECTIONS.

By the intersection of a given collection of figures we mean the set of all points common to all the given figures. Thus, the intersection of figures is also a figure.

Let us consider some examples. Let there be given a plane π and a triangle ABC in it. Let us take a point P on the side AB and a point Q on AC. Let S denote the line PQ (fig. 21). If the given figures are the two-dimensional triangle and the line S, then the intersection of these figures is the segment PQ.

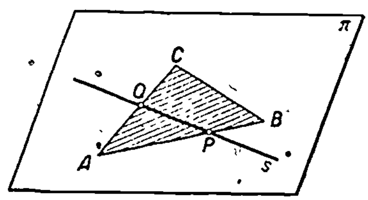


Fig. 21

If we take as our first figure the one-dimensional triangle ABC, then we obtain as the intersection a figure consisting of the two points P and Q - the zero-dimensional segment PQ.

The intersection of the line S (fig. 21) with the plane π is the line S itself.

The intersection of a triangle ABC of any dimension with the plane of this triangle is the triangle ABC itself with the same dimension.

The intersection of two different planes is either a line or the empty set. The latter occurs when the planes are parallel.

The intersection of a sphere with a line is either a pair of points P and Q, one point L, or the empty set (fig. 22).

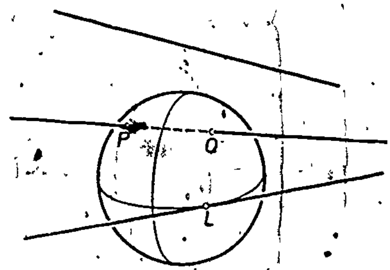


Fig. 22

If instead of a sphere we consider a solid sphere, then the intersections with the same lines will be the segment PQ, the point L and the empty set.

The intersection of the disk PQS (fig. 23) lying in the plane π with the lines s_1, s_2 and s_3 will be respectively the point S, the segment PQ and the point L.

The intersection of a sphere by a plane is, as we already know, either a circle Γ, a point P or the empty set (fig. 24). The intersection of a solid sphere with a plane is either a disk, a point or the empty set (fig. 24).

The intersection of the surface of a circular cylinder with a plane parallel to a generator of the cylinder is either the empty set, a line, or two lines (fig. 25); if the intersecting plane is perpendicular to the generator, then the intersection is a circle (fig. 26).

The case in which the intersecting plane is not perpendicular to the generators of the cylinder is of special interest. The intersection obtained is <u>an ellipse</u> (fig. 27).

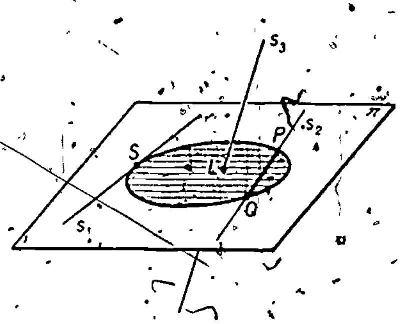


Fig. 23

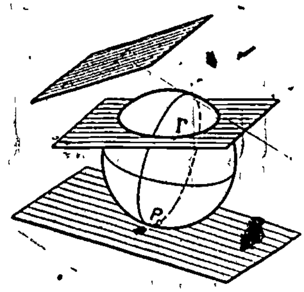


Fig. 24

Let us establish a characteristic property of the ellipse.

Let us put into the cylinder, on opposite sides of the plane π, two spheres, with radii equal to the radius of the cylinder, so that they touch the plane π at the points F_1 and F_2. Figure 28 shows a cross section by a plane passing through the axis of the cylinder and perpendicular to the plane π. The spheres are tangent to the cylinder in circles passing through the points A and B (fig. 27).

Let us take an arbitrary point M of the ellipse. Let us pass through it a segment AB of the generator of the cylinder. In addition, let us join the point M with the points F_1 and F_2. The lines MA and MF_1 are tangent to the sphere (1), and consequently the segments MA and MF_1 are equal.

$$MF_1 = MA$$

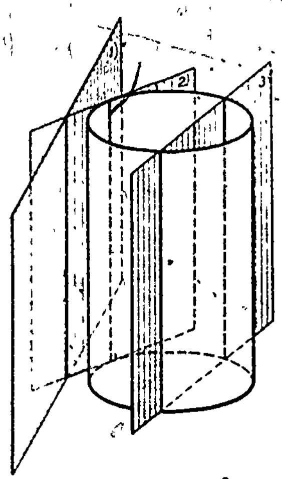


Fig. 25

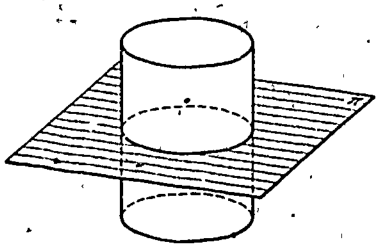


Fig. 26

Similarly the segments MB and MF_2 of the tangents to the sphere (2) are equal.

$$MF_2 = MB$$

From these equalities we conclude that

$$MF_1 + MF_2 = MA + MB = AB = \text{constant}$$

An ellipse[(1)] is the set of points M in the plane π, the sum of whose distances from two given points F_1 and F_2 in the same plane is a constant.

The points F_1 and F_2 are called the foci of the ellipse.

Usually, this property is taken as the definition of an ellipse.

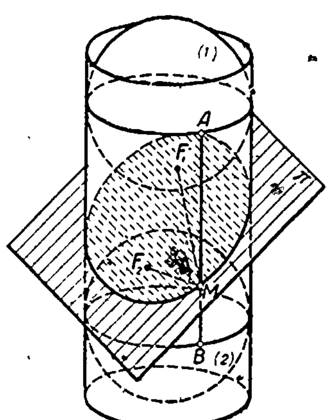


Fig. 27

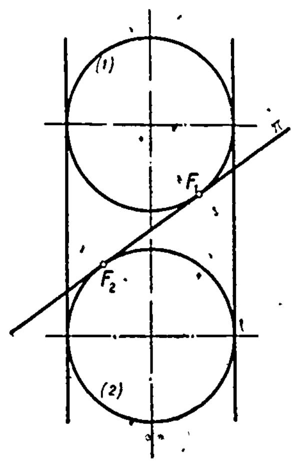


Fig. 28

If we tilt the plane π so that it approaches a position perpendicular to the generator of the cylinder, the points F_1 and F_2 at which the spheres (1) and (2) are tangent to the plane π will tend to coincide, and when they do so the ellipse will become a circle (fig. 26).

The definition of the ellipse can be applied without change to the circle, provided we postulate that $F_1 F_2$ is a null segment.

(1) Footnote: In the interest of simplicity the proof that all points equidistant from F_1 and F_2 are on the ellipse is omitted. --Translators

The simplest surface of revolution after the cylinder is the circular cone. Let us examine the intersection of the cone with a plane not passing through its vertex S.

If the plane π intersects all the generators of the cone, the intersection is an ellipse (fig. 29). The proof is entirely a repetition of the case of the cylinder considered above. The notation in figs. 29 and 30 is the same as in figs. 27 and 28.

If the cutting plane intersects both nappes of the cone and does not pass through S, i.e. if the plane π is parallel to two generators of the cone, the intersection is a hyperbola. That the indicated plane is parallel to two generators of the cone can easily be seen by passing a plane through the vertex S of the cone parallel to the plane π. A part of the plane π in fig. 31 is shaded.

As in the case of the ellipse, we shall find a characteristic property of the hyperbola.

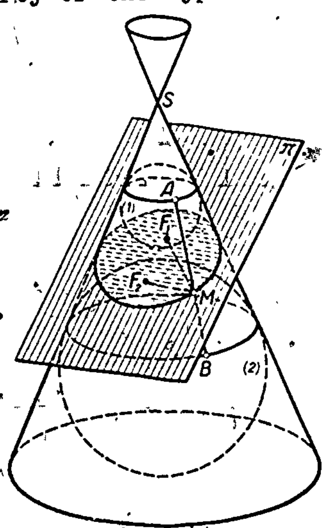


Fig. 29

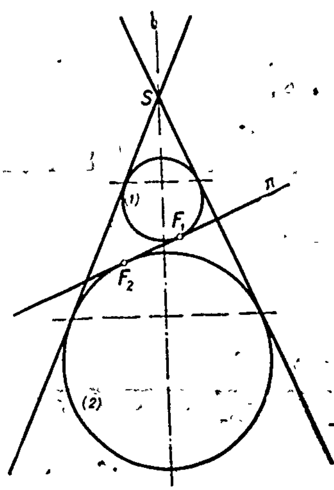


Fig. 30

Let us place in each nappe of the cone one of the spheres(1) and (2) so that the spheres are tangent to plane π at the points F_1 and F_2 respectively, and to the surface of the cone in the circles passing through points A and B (in fig. 31). It is easy to construct great circles of these spheres in a plane perpendicular to the plane π and passing through the vertex S of the cone (fig. 32).

Taking an arbitrary point M of the hyperbola (fig. 31), we find:[1]

$$MF_1 = MA \quad \text{and} \quad MB = MF_2$$

since the tangents MA and MF_1, say, drawn from the point M to the sphere (1) are equal.

From the equalities just obtained we conclude:

$$MF_1 - MF_2 = MA - MB = AB$$

But the length of AB does not depend on the choice of the point M on the hyperbola, and consequently

$$|MF_1 - MF_2| = \text{constant}$$

<u>A hyperbola is the set of points of the plane π, the difference of whose distances from two given points in the same plane is constant in absolute value.</u>

We say "absolute value" because if the point M is taken on the second branch of the hyperbola the difference obtained will be negative but will have the same absolute value.

The points F_1 and F_2 are called the <u>foci</u> of the hyperbola.

---

(1) Points A and B are the intersections of the generator passing through M with the circles of tangency of the spheres (1) and (2) to the cone. --Translators

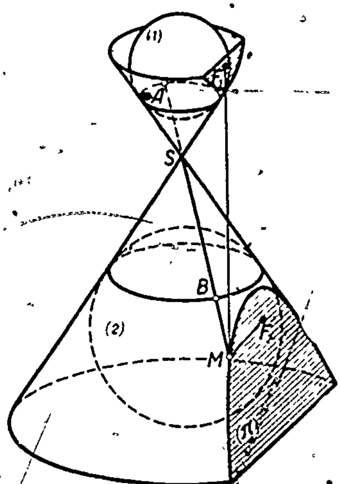


Fig. 31

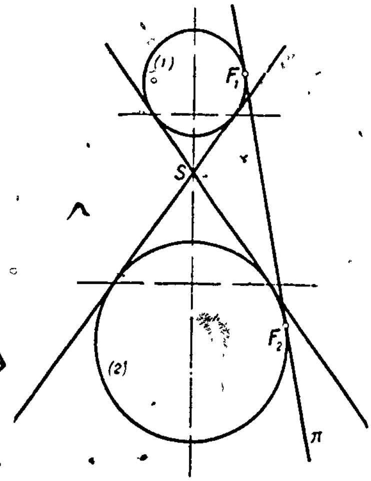


Fig. 32

There still remains to be considered the intersection of a cone with a plane parallel to <u>one generator</u>. This plane π intersects only one nappe of the cone (fig. 33) in a curve which is known as <u>a parabola</u>.

In establishing a characteristic property of the parabola we shall not be able to follow the method used for the ellipse and hyperbola; here the second sphere used in the previous constructions does not exist.

Let us pass a plane π parallel to only one generator SQ of the cone (figs. 33 and 34). Let us construct a sphere (1) so that it is tangent ~~to the~~ cone in a circle AP and to the plane π at the point F. Let $a \equiv CD$ be the line of intersection of the plane π with the plane α of the circle AP.

We shall show that an arbitrary point M of the parabola is equidistant from point F and line CD, i.e. that

$$MF = MD$$

For proof, we pass through the point M a generator MS of the cone and a plane MQ parallel to the plane α.

(Note revision to fig. 34. The point B is in general not on the axis of the cone. --Translators)

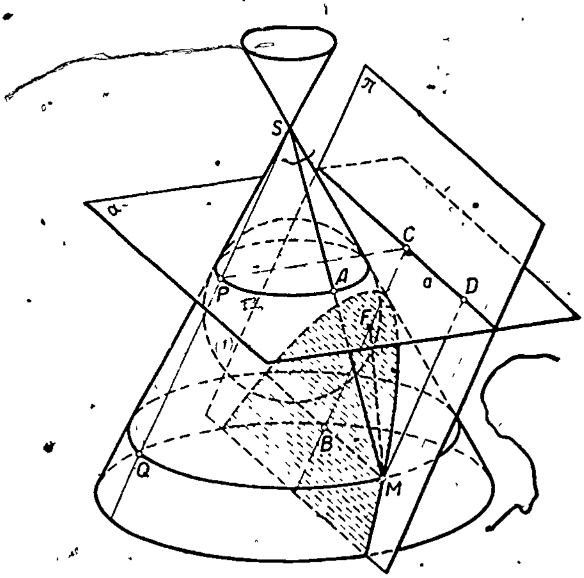


Fig. 33

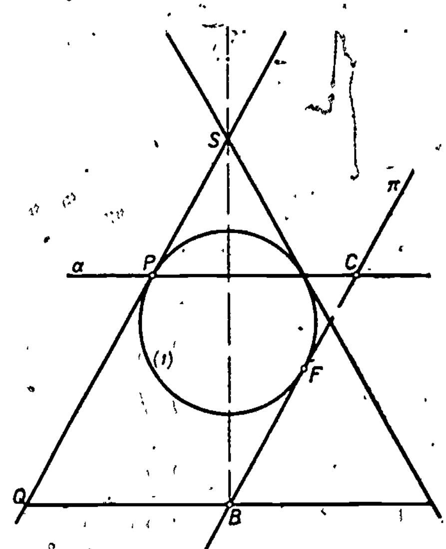


Fig. 34

Let A be the point of intersection of this generator MS with the circle AP.

Since MF and MA are tangents to the sphere (1) drawn from the same point M, we have

$$MF = MA$$

But MA = QP, both being generators of the frustrum of the cone whose bases are the planes α and MQ. Furthermore, QP = BC, since they are parallel segments (1) between parallel planes (figs. 33 and 34). For the same reason, BC = MD. Placing all these equalities side by side, we get

$$MF = MA = QP = BC = MD$$

or

$$MF = MD$$

A parabola is the set of points M of a plane π equidistant from a given point F and a given line a in the same plane π, where the point F is not on the line a.

The point F is called the focus of the parabola; the line a is the directrix.

The equality MF = MD may be written in the form of a ratio

$$\frac{MF}{MD} = 1$$

This property of the parabola could be established, in almost the same form, for the ellipse and the hyperbola.

(1) Although not stated explicitly, it is evident from figs. 33 and 34 that BC is defined as that line parallel to PQ which passes through F. Similarly, MD is defined as the line through M perpendicular to CD.

To make the proof complete one would have to show that BC is perpendicular to CD. This is left as an exercise for the reader. --Translators

The ellipse and the hyperbola also have directrices; each curve has two of them.

These directrices are the intersections of the planes of the circles of tangency of the spheres (1) and (2) with the plane π (figs. 27, 28, 29, 30, 31, 32).

Since the ellipse, hyperbola and parabola appear as sections of a cone by a plane, they are called conic sections.

A common characteristic property of these curves, i.e. a property which could be taken as their definition, is the following:

A set of points M of a plane π is called a conic section if the ratio of the distance MF of each point M from a given point F to the distance MD from a given line L is a constant, different from zero.

$$\frac{MF}{MD} = e = \text{constant}$$

It is understood that the focus F and directrix L are in the plane π and F does not lie on L.

When $e < 1$, we obtain an ellipse; when $e = 1$, a parabola; $e > 1$ yields a hyperbola.

It is not very difficult, and will serve as a useful exercise to prove these facts, using the same method as in the proofs above.

The conic sections play an exceptionally important role in science. As we have seen, familiarity with them requires only a knowledge of elementary geometry. Further on, the properties of conic sections just established will find important use.

In concluding this section we shall prove a theorem concerning convex figures:

The intersection of convex figures is a convex figure.

We shall carry through the proof for the case of only two figures. The proof may be carried over without change to the case of any collection of figures.

Let there be given two convex figures ϕ_1 and ϕ_2 (fig. 35). Their intersection, shaded in the diagram, is the set of points belonging to both figures. Let A and B be two arbitrary points in the intersection. Since points A and B belong to figure ϕ_1, the entire segment AB belongs to ϕ_1. Since the points A and B belong to figure ϕ_2, the entire segment AB belongs to figure ϕ_2.

But, belonging to both figures, the segment AB also belongs to their intersection. Consequently, the intersection is a convex figure, q.e.d.

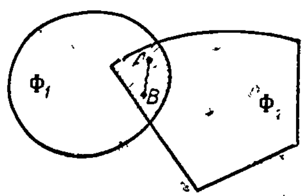


Fig. 35

If the convex figures have only one point in common, then, as we already know, their intersection --- a point --- is a convex figure.

We see that the proof given is completely general.

Example: Let there be given an infinite set of concentric solid spheres (a solid sphere is a convex figure) of decreasing radii:

$$R_1 = 1,\ R_2 = \frac{1}{2},\ R_3 = 2\frac{1}{2},\ \ldots,\ R_n = \frac{1}{2^{n-1}},\ \ldots$$

The intersection of all these solid spheres is a point --- their common center; and a point is a convex figure.

5. THE UNION OF FIGURES

The set of all points belonging to at least one of a given collection of figures is called the union (or sum) of the given figures.

The union of figures is a figure. Let us consider examples of unions of figures.

The union of two points is a zero-dimensional segment.

The union of an interval AB and its boundary, consisting of the zero-dimensional segment AB, is a closed segment, or simply a segment.

The union of a segment and its ends is again the same segment.

This example shows that we may take the union of figures which have a non-empty common part, which, of course, belongs to the union.

The union of an open disk and its boundary (circle) is a disk.

The union of an open solid sphere and its boundary (sphere) is a solid sphere.

The union of a solid sphere and its boundary (sphere) is again the same solid sphere.

The union of a one-dimensional triangle ABC with the two-dimensional triangle ABC is the same two-dimensional triangle ABC.

The union of the parallels on a sphere yields the sphere. We could also obtain the sphere from its meridians (fig. 5).

Each line tangent to a circle is a set of points.

The union of all such sets (lines, tangent to the given circle) is the plane in which the given circle lies, with an open disk excluded (fig. 19).

The union of all the generating lines of a cylinder is the cylinder. Analogously, the union of all the lines m in fig. 18 is a hyperboloid of one sheet.

The union of the circles in fig. 20 gives us a torus.

The latter cases are examples of the union of an infinite set of figures.

The union of two concentric circles Γ_1 and Γ_2 (fig. 36) is a figure. This union must be considered as <u>one</u> figure! One can give a characteristic property for it: This union is the set of points in the plane of the given circles which are at a given distance from a circle Γ in this plane (fig. 36).

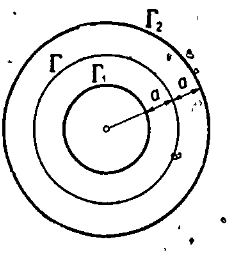


Fig. 36

In solving construction problems, for example by the method of loci, we frequently encounter such figures formed of a union of two figures. For instance, the set of points in a plane at which a segment lying in this plane subtends the same angle α is a figure consisting of the two circular arcs (1) and (2) (fig. 37) <u>with their end points A and B excluded</u>.

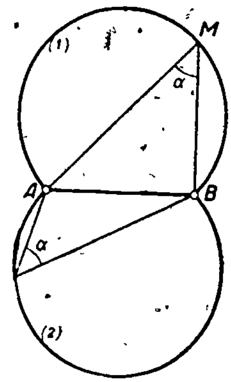


Fig. 37

The set of all points at which the given segment AB subtends the given angle α is a figure of revolution generated by revolving the pair of open arcs just considered about the axis AB.

This surface of revolution is analogous in construction to a torus; here the circle revolved intersects the axis of revolution. Moreover, in this case the surface generated by the second arc AB of the same circle must be excluded.

6. PASTING TOGETHER, OR IDENTIFICATION

We shall now acquaint ourselves with yet another mathematical operation, namely the operation of pasting together or identification.

We shall begin with some examples.

Let there be given two equal two-dimensional right triangles (fig. 38). It is possible to obtain from these right triangles a rectangle ABCD by pasting together the sides AB as indicated in the drawing..

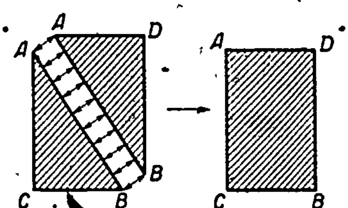


Fig. 38

Mathematically this means that we consider the segments AB not as two but as one. In pasting together any two points (for instance the points A and the points B as well as the corresponding interior points of the segments) we consider them as one. When such an identification (of the points) of the sides AB is carried out, the union of the two triangles gives us the rectangle ACBD.

Let there be given four equilateral and for simplicity two-dimensional congruent triangles (fig. 39). Carrying out the operation of union we already think of all the four triangles as one figure. Observe that all points of each pair of segments whose end points are correspondingly lettered in the diagram are pairwise pasted together or identified. As a result of this taking of the union and pasting together we obtain a two-dimensional tetrahedron.

Note that the three points D are identified, i.e. are considered as one.

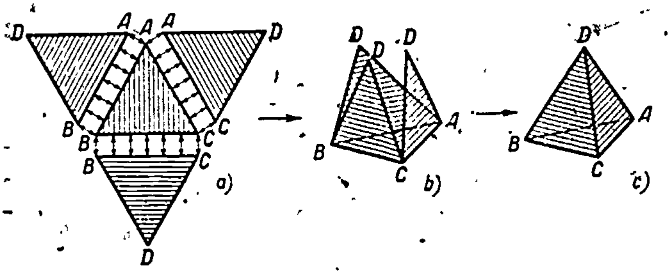


Fig. 39

In fig. 39a the shaded triangles are thought of as united into one figure already in the separated positions in which they have been drawn, while we merely indicate with double arrows the sides AB, BC and CA which are going to be pasted together (fig. 39b). We deal in the same way with the sides AD, BD, and CD (fig. 39c).

With the aid of cut-out patterns of regular polyhedrons (figs. 40, 41, 42, 43, 44) it is easy to trace how the identification of the edges of the polygons is carried out to obtain the corresponding polyhedrons.

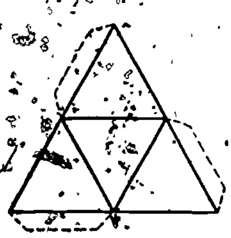


Fig. 40

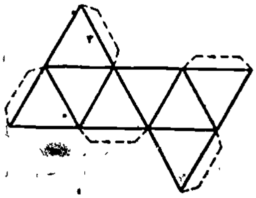


Fig. 41

It is useful actually to paste together all regular polyhedrons from heavy paper, constructing the patterns accurately to scale.

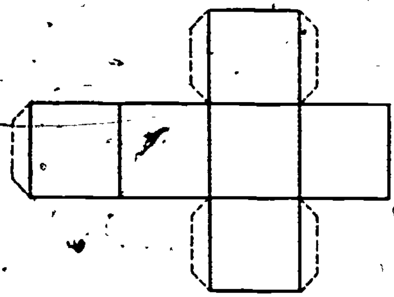


Fig. 42

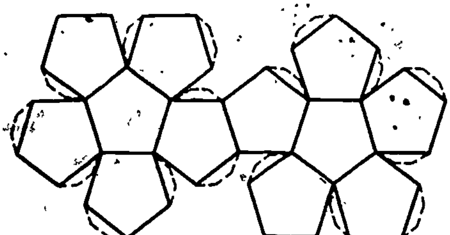


Fig. 43

These patterns are, in a sense, maps of the polyhedral surfaces. If we disregard the disposition of the polyhedron in space and investigate only the structure of the polyhedral surfaces themselves, then the patterns just mentioned, together with the given identification of points, give us all that is needed for the complete investigation of the surface of the polyhedron.

Identification of points of figures, or pasting together, is met with in geographical maps of the world. On the map (fig. 45) the equivalent points of the two representations of the meridian 20^{o}W are to be identified, and the same is to be done for the representations of the meridian 160^{o}E. Those points which represent the same points on the earth's surface are considered equivalent.

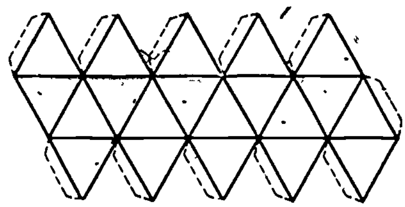


Fig. 44

Each pair of equivalent points is considered as one. The two "hemispheres" are united into one figure, with the use of appropriate pasting together.

The sheets of large geographic maps also have entire regions near the edges which it is necessary to identify in the sense indicated, i.e. equivalent points which one may encounter on different sheets have to be regarded as one.

In common usage we say that a map (figure) represents the surface of the earth. On it, we are able to trace a 'round-the-world journey passing from one hemisphere (or sheet) to another through equivalent points.

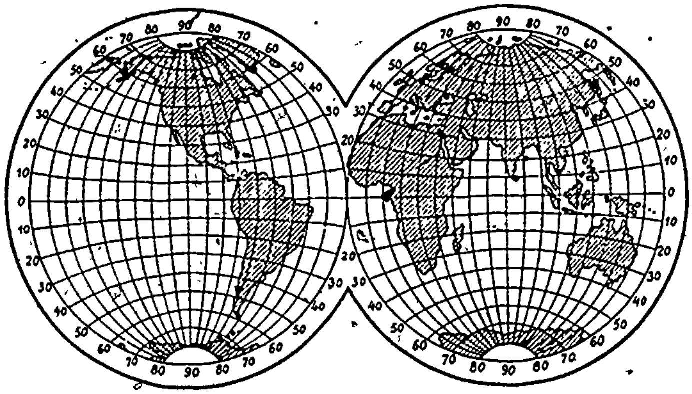


Fig. 45

With the aid of such a method we may draw and study any curve on the surface of a polyhedron, using only the pattern or map of the polyhedron. However, when measuring on a map, say, the length of a curve, one must know the <u>scale</u> of the map at each point and in each direction issuing from that point (see 14).

As an exercise, let the reader number all the vertices of the polyhedrons on the maps, assigning the same number to equivalent vertices. It would be of interest, starting with any point on the surface of the polyhedron, to trace a path around it which passes through each of the faces and to draw this path on its map.

In the same way in which we construct a representation of a sphere out of two disks by taking their union and pasting together (fig. 45), we can form a cylindrical surface by pasting together the edges of the closed strip between two parallels, considering a pair of points lying one on each of these parallels as equivalent if they lie on one of the common perpendiculars (fig. 46).

If we wish to study, on the surface of a circular cylinder, a curve which cuts all the straight-line generators at the same angle α, it is easy to do so on such a strip between two parallels, which "depicts" the cylinder.

If $\alpha = \frac{\pi}{2}$ this will turn out to be a circle, for instance AA, MM, BB etc. These circles are represented in the strip by the lines AA, MM, BB perpendicular to the parallel, identified edges of the strip (fig. 46).

If α is an acute or obtuse angle we get a <u>helix</u> on the circular cylinder, which as is well known has important practical applications.

On the "map," all the linear generators are represented by lines parallel to the two mutually equivalent lines and, clearly, the representation of the helix on the map will consist of the linear segments AB, BC, CD, DE etc., parallel to each other and cutting any linear generator in the angle α (fig. 47).

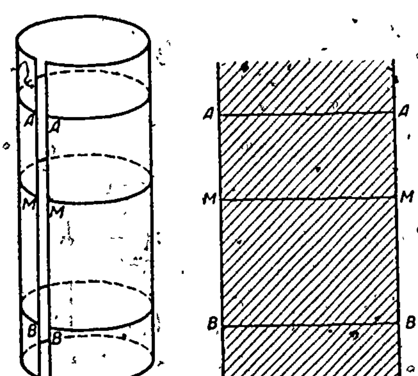


Fig. 46

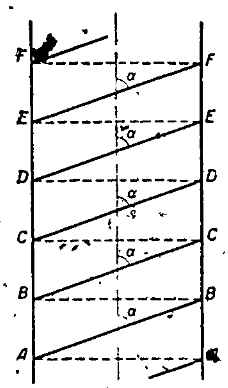


Fig. 47

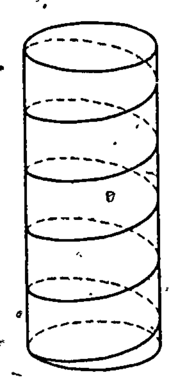


Fig. 48

Pasting together the edges of this strip in the manner of fig. 46, we easily produce the helical curve on the cylinder (fig. 48).[1].

An exact definition of the operation of identification or pasting together is as follows: Having established by means of any rule whatsoever which points in a figure are considered equivalent, we then regard all equivalent points as one point.

It is not necessary to carry out the actual pasting together, although indeed it may be possible to do it with great gain in graphic clearness; but it is then sometimes necessary to "bend" or "stretch" the figure as if it were made of rubber.

If the equivalent points to be considered as one form an infinite set, we cannot physically carry out the pasting together.

In concluding this section let us consider two cases each involving a rectangular figure with identified points.

In the first rectangle (fig. 49) we consider as equivalent any pair of points on opposite sides, lying on a line perpendicular to these sides. Equivalent points are denoted by the same letter. The equivalent points of this figure are identified, or are pasted together (mentally, of course, taking into consideration only the exact mathematical meaning of this operation), i.e. they are considered as one.

In this figure there is represented a closed continuous path originating at the point A, passing through the points E and D (without a break) and finally back to the starting point A.

Figure 49 is, by itself, geometrically fully defined. However, we may regard it also as the map of a torus.

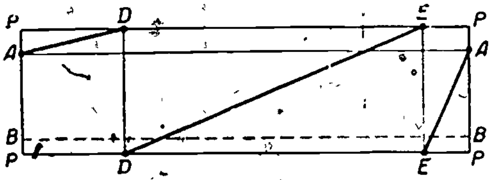


Fig. 49

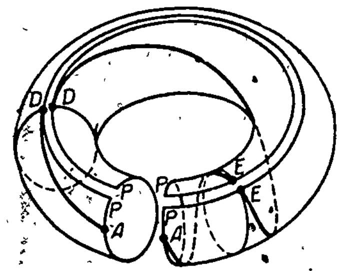


Fig. 50

In figure 50 the torus has been prepared for actual pasting together. It was necessary to "bend" and to "stretch" the antecedent rectangle to some extent. The dashed line BB is not shown.

In the second rectangle (fig. 51) we shall regard as equivalent those points of one pair of opposite sides lying symmetrically with respect to the center of the rectangle; they are denoted by the same letter. Equivalent points are to be identified. This defines the figure exactly.

Let us examine certain remarkable properties of this figure.

We observe that the figure has only one closed boundary. That is, starting from the boundary point P in the direction of the first arrow, we arrive at the point A and then, continuing the unbroken path along the boundary as indicated by the second arrow, we arrive at the point B. The third arrow indicates the return along the boundary to the point of origin P.

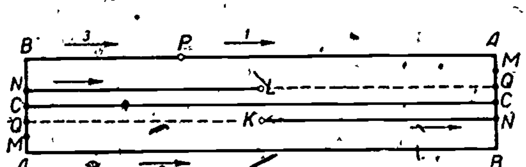


Fig. 51

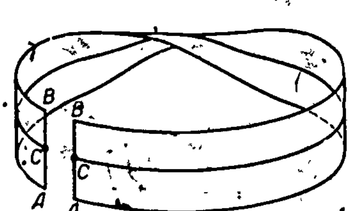


Fig. 52

Let us imagine further that the closed line CC represents a canal. Finding ourselves on one bank of the canal at the point K, let us walk along the bank following the line KN; continuing our unbroken path along the bank in the line NL, we shall arrive at the point L on the "other" bank of the canal without crossing the canal.

In the case of a closed canal on the surface of the earth, such a phenomenon would be impossible.

In our figure, as we have clearly seen, the canal CC has no "other" bank.

We can "imbed in space" the figure under consideration, carrying out an actual pasting together. Namely, let us rotate the right hand segment AB (fig. 51) of our figure through 180° and then paste it to the left hand segment, as shown in figure 52.

The figure thus pasted together -- one of the most remarkable of surfaces -- is called a Moebius strip.

It is highly instructive to verify the properties of this figure on a model pasted together from heavy paper in the manner indicated above.

The Moebius strip is a one-sided surface, a fact of which we can easily convince ourselves by painting a model of this surface with a brush. The Moebius strip cannot be a part of the boundary of a solid.

If we were to "stand up" at a point C on the Moebius strip and walk around the surface along the curve CC, upon returning to the point C we should find ourselves "upside down."

It is useful as an exercise to "cut" the Moebius strip along the line CC, along the line KNLQK etc., studying these "cuts" at first on the map (fig. 51) and then verifying the results on a paper model. It is interesting to see the results of a second cut following the first one, and so on.

The exact meaning of the notion of a cut of a surface along a curve is that each point M of the indicated curve is regarded as two or more points.

The operation of cutting a figure is the inverse, in a sense, of the operation of pasting together.

For instance, one can cut a circle as follows: One point M of the circle is to be considered as two points M' and M'' (fig. 53); one can cut a sphere along a parallel, that is, count every point of the parallel as two (fig. 54).

A solid sphere can be cut by removing an equatorial disk and then adding a disk to each open hemisphere thus obtained, thereby closing the hemispheres.

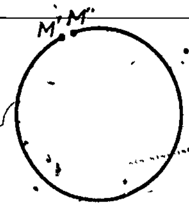


Fig. 53

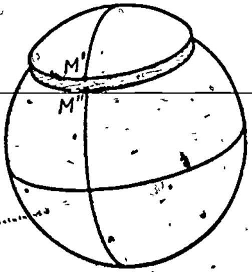


Fig. 54

It is suggested as an exercise that the reader study the closed cuts represented in diagrams 55 and 56. It is impossible to obtain the figure in diagram 56 by performing an actual pasting together; this figure cannot be imbedded in space without self-intersections.

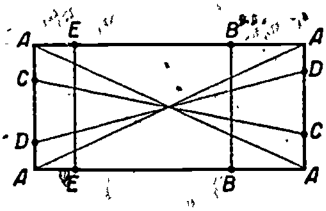


Fig. 55

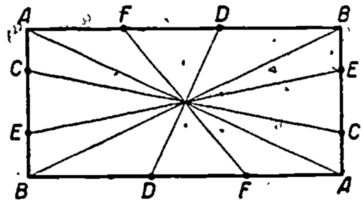


Fig. 56

A simple figure is obtained from a rectangle by regarding as equivalent all points of the four sides, which are thus considered as one point.

A no less simple figure may be obtained by regarding all the points on a pair of opposite sides of a rectangle as one point.

By the methods indicated one could, for instance, paste together opposite faces of a solid cube and study in the resulting figure various types of cuts and curves. [6] [7] [16].

Basic notations. In order to indicate that a point A belongs to a figure ϕ, we use the notation $A \in \phi$, in words: A belongs to the figure ϕ, or, the point A is a point of the figure ϕ.

Sometimes we write $\phi \ni A$ and we say: the figure ϕ contains the point A.

Both symbols express the same fact, that the element A belongs to the set ϕ.

Usually, the elements x or any set M are denoted by small letters while the set itself is denoted by a capital. The notation $x \in M$ signifies that x is an element of the set M. The set M may be of any kind whatsoever and need not be a figure.

It frequently becomes necessary to regard one figure as a part of a second.

Thus the boundary of a two-dimensional triangle (a one-dimensional triangle) is a part of the triangle. The set of points of a conic section is a part of the set of points of the cone in which the conic section lies (figs. 29, 30, 31). The sphere is a part of the solid sphere which it bounds.

When every point of the figure ϕ is at the same time a point of the figure ψ, the figure ϕ is called a part of the figure ψ.

We also say that figure ϕ is included in figure ψ.

Figure ϕ may coincide completely with figure ψ, and in this case also figure ϕ is considered a part of figure ψ, since in fact every point of figure ϕ is at the same time also a point of figure ψ.

In defining the notion of a part of a figure, we could have used another, equivalent mode of expression: a figure ϕ is a part of figure ψ if there are no points in figure ϕ which do not belong to figure ψ.

In this connection it becomes clear that the empty set must be regarded as a part of every figure. This is so because there are no points in the empty figure which do not belong to any given figure, since the empty figure has no points in the first place.

The empty set of points and the figure ϕ itself are called improper parts of the figure ϕ. All other parts of the figure are called proper.

When figure A is a part of figure B we write $A \subseteq B$ and we say: figure A is a part of figure B, or, figure A is contained in figure B.

If the figure A is only a proper part of figure B, we write

$$A \subset B$$

The symbols $\subseteq$, $\subset$ are called symbols of inclusion of one figure in another, or of one set in another.

In general, if two sets M and N of any sort are given and if every element of the set M belongs to the set N, we write $M \subseteq N$, or, if M is not empty and does not coincide with N, $M \subset N$. If $M \subseteq N$ then M is a <u>subset</u> of N.

The <u>intersection</u> of given figures consists of the set of points belonging to all the given figures. The intersection of figures A and B is denoted by the sign of intersection $\cap$ and we write $A \cap B$. The symbol $\cap$ is a sign of taking the common part of the figures.

If for instance A is the set of points belonging to the cone and B the set of points of the plane π (fig. 33), then $A \cap B$ denotes the set of points of the conic section.

The intersection of sets is frequently called their <u>product</u>.

The intersection of figures was discussed in detail in 4.

The <u>union</u> or <u>sum</u> of given figures is the figure consisting of all points M belonging to at least one of the given figures. The union of figures A and B is denoted by the sign $\cup$ and we write

$$A \cup B.$$

For instance, if A denotes an open sphere and B its bounding sphere, then $A \cup B$ denotes the solid sphere.

The union of sets is frequently called the <u>sum</u> of the sets.

The union of sets was discussed in detail in 5.

50.

Let us now consider the operation of subtraction of figures. By the difference of the sets A and B we mean the set of those elements of set A which do not belong to set B. The difference of the sets A and B is denoted by A - B.

For example, if A is a disk and B its circumference, then A - B is an open disk; if A is a sphere and B a one-dimensional closed diameter PQ then A - B is a sphere "perforated" at two diametrically opposite points P and Q. B does not necessarily have to be entirely contained in A. For instance, $\phi_1 - \phi_2$ in figure 35 is the figure ϕ_1 with the shaded part excluded.

Chapter II

GEOMETRICAL CONSTRUCTIONS

In Chapter II we shall study the most important loci, the method of loci in the solution of problems, and the theory of geometrical constructions using the straight-edge and compass, as well as other instruments. The mathematical basis of such constructions is examined; examples are given of problems which can not be solved solely with the aid of some particular combination of instruments; and proof of the impossibility of certain constructions is adduced. Algebraic methods of solving construction problems are examined.

7. LOCI

A locus of points is the same thing as a set of points, and since every point set is called a figure, the concept of a locus of points and the concept of a figure coincide.

We may speak, for instance, of the locus of points whose distance from a given point is equal to the length of a given segment R (which would be a sphere); or of the locus of points whose distance from a given point is not less than R (the whole of space minus an open solid sphere).

The locus of points equidistant from all points on a given circle is the straight line passing through the center of the given circle and perpendicular to its plane.

The locus of points equidistant from all points on a given sphere is the center of this sphere.

The locus of points equidistant from all points of a given plane is the empty set.

The locus of points in space at which a given segment AB subtends a right angle is a sphere having AB as a diameter with points A and B excluded.

In order to specify any set, it is necessary to indicate a characteristic property of its elements. Every object possessing this property is an element of the set in question; every object not possessing the specified property is not an element of the given set.

In order to specify a locus of points, it is necessary to specify a property which its points must possess. Those and only those points possessing the indicated property will then belong to this locus of points.

Elementary geometry usually deals with loci consisting of circles, straight lines, spheres, planes, etc., and also of parts of these figures and combinations of them.

In analytical geometry, a locus of points is defined as the set of all points whose coordinates satisfy one or more equations or inequalities; the fulfillment of this requirement constitutes the characteristic property which unites the points into a set.

It is immaterial whether this set of points resembles curves and surfaces familiar to us, or whether it is some other sort of collection of points.

The indication of a property which associates points into a set is all that is required to set up a locus.

Let us consider some examples of finding loci in a plane.

1. To find the locus of points, the ratio $\frac{a}{b}$ of whose distances from two given points A and B is a constant not equal to unity.

If the ratio $\frac{a}{b}$ were set equal to 1, the locus would be the perpendicular bisector of the segment AB.

With A and B as centers we describe two circles whose radii are either equal to or proportional to a and b. Let M be one of the points of intersection of these circles (fig. 57). Drawing the bisectors MC and MD of the supplementary angles AMB and BML, we observe that their points of intersection C and D with the line AB belong to the required locus, since

$$\frac{AC}{CB} = \frac{AM}{MB} = \frac{a}{b}, \text{ and } \frac{AD}{DB} = \frac{AM}{MB} = \frac{a}{b}.$$

Angle CMD is a right angle and consequently M lies on the circle having CD as a diameter. Thus, any point M of the locus lies on this circle.

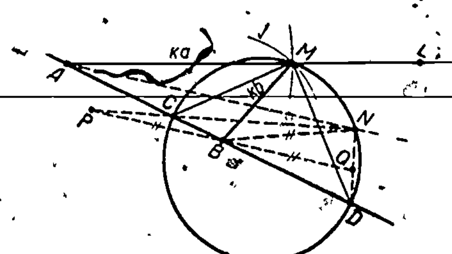


Fig. 57.

We shall now show that every point N of this circle possesses the property $\frac{AN}{NB} = \frac{a}{b}$, i.e., that any point that lies on this circle satisfies the condition of the locus. We draw through B a line parallel to AN, and mark the points P and Q of its intersection with the lines NC and ND.

From the similarity of triangles ANC and BCP we have:

$$\frac{AN}{BP} = \frac{AC}{CB} \quad (= \frac{a}{b}),$$

and from the similarity of triangles AND and BQD:

$$\frac{AN}{BQ} = \frac{AD}{DB} \quad (= \frac{a}{b}) .$$

From these proportionalities it follows that

$$BP = BQ .$$

But triangle PNQ is a right triangle with right angle at N and consequently B is the midpoint of its hypotenuse PQ. Hence:

$$BP = BQ = BN ,$$

which, in turn, means that triangle PBN is isosceles.

Accordingly, $\angle BPN = \angle BNP$; furthermore, $\angle BPN = \angle PNA$, being alternate interior angles formed by the parallel lines AN and PB and the transversal PN.

From the equality of angles BNP and PNA it follows that NP is the bisector of angle ANB, i. e.,

$$\frac{AN}{NB} = \frac{AC}{CB} = \frac{a}{b} .$$

This circle (fig. 57) is known as the circle of Apollonius.

2. Given the secants drawn through a given point S to a given circle, to find the locus of the midpoints of the chords cut off by the circle on the secants.

Let A and B be the points of intersection of a secant with the circle (fig. 58), and M the midpoint of the chord AB. Joining the point M with the center O of the given circle, we find that $\angle OMS$ is a right angle. Consequently, every point M of the desired locus lies on the circle having OS as a diameter.

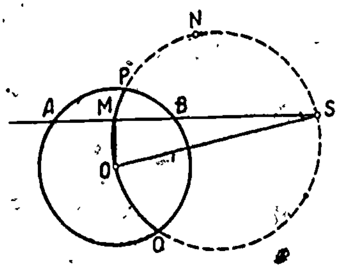


Fig. 58.

A point N on this circle can be proved to belong to the desired locus when and only when the line SN intersects the given circle.

If the point S lies outside the given circle, intersections will not occur on all lines of the pencil whose center is S.

All points N of the circle with diameter OS lying <u>within</u> or <u>on</u> the given circle will belong to the desired locus, which in the case where the point S is exterior to the given circle will be the <u>arc</u> POQ of the constructed circle.

This example emphasizes that after proving that any point of the desired locus belongs to a certain curve it is necessary to make clear whether or not every point of this curve belongs to the desired locus, and if not, to define exactly which points of the curve do belong to the locus.

3. <u>The locus of points M at which a given segment AB subtends a right angle is the circle having a diameter AB, with the exclusion of the points A and B.</u>

We made use of this locus in dealing with loci 1 and 2 above. In this connection it is necessary to examine in each case whether or not the end points of the diameter belong to the desired locus.

4. The locus of points at which a fixed segment AB subtends a given angle consists of two circular arcs with the exclusion of points A and B (fig. 37).

The question as to what angle is subtended by the segment AB at one of its ends, for example A, is meaningless.

Let us examine a problem frequently encountered:

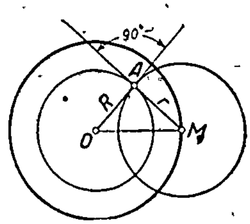


Fig. 59.

5. To find the locus of the centers of the circles having a given radius r, which orthogonally intersect a given circle.

Let O be the center of the given circle and R its radius (fig. 59). At an arbitrary point A we draw a tangent to the given circle and lay off upon it the segment AM = r. The point M is the center of a circle of radius r which orthogonally intersects the given circle.

But

$$OM = \sqrt{R^2 + r^2} = \text{const.},$$

and consequently the desired locus is a circle with center at O and radius $\sqrt{R^2 + r^2}$ ...

6. To find the locus of points at which two given circles subtend the same angle.

Let A and B be the centers of the given circles, and R and r their radii (fig. 60).

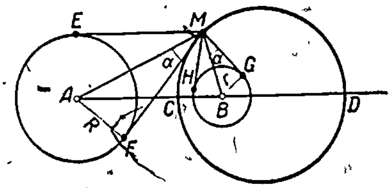


Fig. 60.

If M is a point of the desired locus, then it is required that $\angle$ EMF = $\angle$ GMH, where ME and MF are the tangents from M to the first given circle, and MG and MH the tangents to the second. The right triangles are similar since each has the acute angle α which is half of the angle subtended by either circle. It follows that

$$\frac{AM}{MB} = \frac{AF}{BG} = \frac{R}{r} = \text{const.}$$

Thus, all points M of the desired locus lie on a circle of Apollonius if $R \neq r$, and on a straight line if $R = r$.

Let M be a point of the circle thus found. From M it is possible to draw tangents to both of the given circles if M is exterior to both of them. Drawing tangents MF and MG, we find by virtue of the properties of a circle of Apollonius that in the right triangles MAF and MGB:

$$\frac{AM}{MB} = \frac{R}{r}.$$

If in two right triangles the hypotenuses are proportional to a pair of corresponding legs, the triangles are similar.

Consequently, $\angle AMF = \angle BMG \ (= \alpha)$,

q.e.d.

The desired locus consists of those points of a circle having CD as a diameter, where C is the interior and D the exterior center of similarity of the given circles, which are exterior to the given circles.

If the given circles intersect, the locus passes through the points of intersection. If one of the given circles lies within the other, the desired locus is the empty set. If the given circles are tangent internally, the locus consists of one point. If the given circles coincide, the desired locus is the entire plane with the exclusion of the open disc whose boundary is the given circle.

Remark. Every locus of points is characterized by some definite property of its elements. But any such figure, for example a circle, is itself the possessor of numerous other properties which may in their turn be characteristic, i.e., may completely determine the figure. We shall illustrate this statement with examples. It is possible to define a circle either as the locus of points (in a plane) at a given distance from a given point, or as the locus of points at which a given segment subtends a right angle (in which case the ends of this segment must be excluded from the circle), or as the locus of points the ratio of whose distances from two given points is a constant ($\neq 1$); and so on.

The more characteristic properties of a figure one knows, the greater is the possibility of recognizing the figure as one encounters it in various problems. For example, in determining the locus of points the ratio of whose distances from two given points is a constant ($\frac{AM}{MB} = \frac{a}{b} \neq 1$), we made use of that property of the points of a circle whereby a diameter subtends a right angle at any of them. In determining the locus of points at which two given circles subtend equal angles, we made use of the property of the circle of Apollonius. Let us examine some loci involving a straight line.

7. <u>The locus of points equidistant from two given lines.</u>

If the given lines intersect, this locus will consist of a pair of lines perpendicular to each other and constituting the bisectors of the angles formed by the given lines (fig. 61).

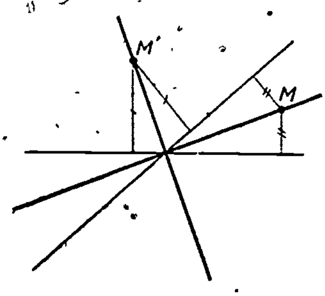


Fig. 61.

When this well-known locus is discussed in secondary school courses, the sides of an angle are usually given instead of two entire lines, a procedure which is inadvisable.

If the given lines are parallel, the locus will be a line parallel to the given lines (fig. 62).

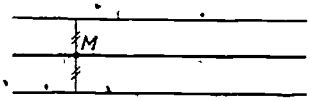


Fig. 62.

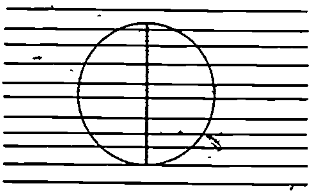


Fig. 63.

8. The locus of the midpoints of the parallel chords cut off by a given circle on a pencil of parallels is the diameter of the circle perpendicular to the chords (fig. 63).

9. To find the locus of points the difference of the squares of whose distances from two given points is a constant.

Let A and B be the given points and M a point of the desired locus (fig. 64).

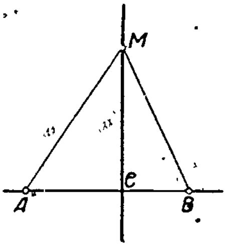


Fig. 64.

Dropping a perpendicular MC from M to the line AB, we have:

$$AM^2 = MC^2 + AC^2 \text{ , and } BM^2 = MC^2 + BC^2$$

Subtracting one equation from the other, we obtain:

$$AM^2 - BM^2 = AC^2 - BC^2$$

which is equal to the given constant. The point M of the desired locus lies on the perpendicular to the line AB, which passes through the completely determined point C of that line. It is easy to prove the converse, that every point on this perpendicular CM is a point of the desired locus.

Thus, the desired locus is a determinate line, perpendicular to the line AB joining the given points A and B.

10. <u>To find the locus of points the lengths of the tangents from which to two given circles with distinct centers are equal.</u>

Let A and B be the centers of the given circles, R and r their radii, and M a point of the desired locus (fig. 65). The tangents from M to the circles are equal:

$$MT_1 = MT_2 .$$

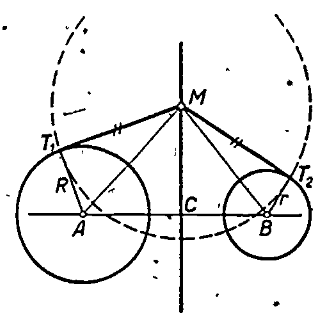


Fig. 65.

We join M with the centers A and B and draw the radii $AT_1 = R$ and $AT_2 = r$. In the right triangles MT_1A and MT_2B we have:

$$AM^2 = MT_1^2 + R^2 \text{ and } BM^2 = MT_2^2 + r^2 ,$$

whence

$$AM^2 - BM^2 = R^2 - r^2 = \text{const.}$$

But this means that point M belongs to the locus examined under number 9. above, that is, it lies on the perpendicular MC to the line AB and passes through the point C lying on the segment AB and determined by the equation

$$AC^2 - BC^2 = R^2 - r^2$$

We shall now make clear which points of the line MC belong to the desired locus.

Let M be a point of this perpendicular MC exterior to the given circles. From this point tangents MT_1 and MT_2 can be drawn. We join M and T_1 with A, and M and T_2 with B. In the right triangles AT_1M and MT_2B we have:

$$MT_1^2 = AM^2 - R^2 \quad \text{and} \quad MT_2^2 = BM^2 - r^2 ,$$

whence

$$MT_1^2 - MT_2^2 = (AM^2 - BM^2) - (R^2 - r^2) = 0 ,$$

which means that $MT_1 = MT_2$, that is, every point M of the previously determined line MC which is exterior to the given circles belongs to the desired locus.

The entire line CM is called the radical axis of the two circles.

The locus of points the tangents from which to two given circles with distinct centers are equal consists of those points of the radical axis of these circles which are exterior to the circles.

If the circles intersect, the radical axis passes through the points of their intersection, and the desired locus consists of two rays of this line (fig. 66). If the given circles are tangent, the locus coincides with the radical axis (fig. 67, a and b). For two concentric circles the locus in question is the empty set.

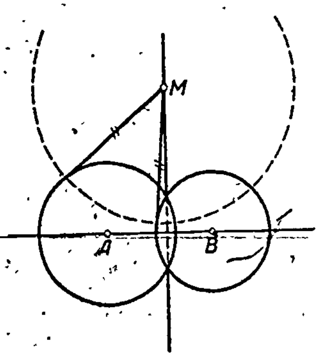


Fig. 66.

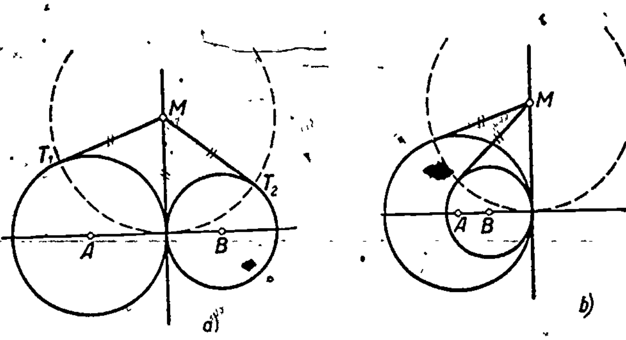


Fig. 67.

11. The locus of the centers of the circles intersecting two given circles orthogonally consists of those points of the radical axis which are exterior to the given circles. (figs. 65, 66, 67).

We shall turn now to more complex loci.

12. The locus of points the sum of whose distances from two given points is a constant is an ellipse:

$$MF_1 + MF_2 = \text{const.}$$

(figs. 27 and 29).

13. The locus of points the absolute magnitude of the difference between whose distances from two given points is a constant is an hyperbola:

$$|MF_1 - MF_2| = \text{const.}$$

(fig. 31).

14. The locus of points equidistant from a given point and a given line is a parabola:

$$MF = MD$$

(fig. 33).

We are familiar also with the following general definition of conic sections:

15. The geometric locus of points the ratio of whose distances from a given point and a given line is a constant is a conic section:

$$\frac{MF}{MD} = e = \text{const.}$$

If $e < 1$ we have an ellipse; if $e = 1$, a parabola; and if $e > 1$, an hyperbola.

In conclusion let us consider briefly some loci which have played important roles in the history of the development of mathematics.

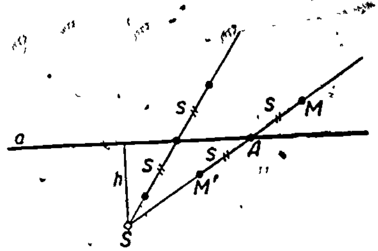


Fig. 68.

16. Let a be a given line and S the center of a given pencil of rays (fig. 68). The locus of points at equal distances s from the line a measured along the rays of the pencil with center at S is a curve known as the conchoid of Nikomedes.

The point S is called the pole of the conchoid, the line a its base and the constant segment s = AM the interval of the conchoid.

It is easy to construct as many points of the conchoid as desired. If h be the distance of the pole S from the base a, three forms of the conchoid can be distinguished according to which of the three relationships

$$s < h \;,\quad s = h \;,\quad s > h$$

holds good (figs. 69, 70, 71).

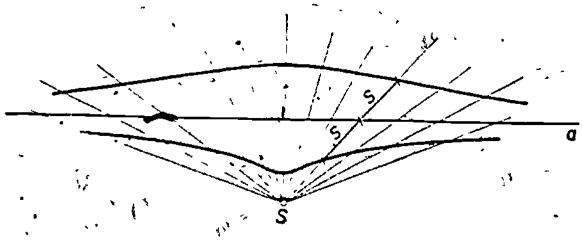


Fig. 69.

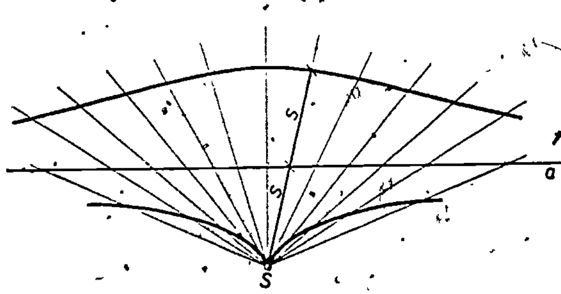


Fig. 70.

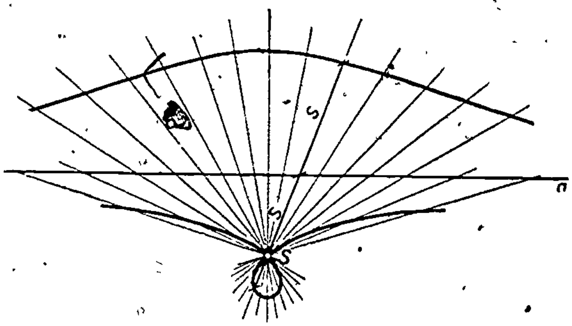


Fig. 71.

With the aid of instruments for drawing conchoids it is possible to construct the roots of cubic equations, and consequently to trisect any angle and to duplicate the cube.

It was for precisely these latter purposes that this curve was devised by Nikomedes (about 150 B.C.E.)

17. Let there be given a circle, the tangent to it at one end of a diameter, and the pencil of rays having its center at the other end of this diameter (fig. 72).

Laying off from the center S of the pencil along a ray SQ the segment SM equal to the segment PQ, where P is the point of intersection of the ray with the circle and Q its intersection with the tangent through T, i.e. SM = PQ, we determine the locus of points M.

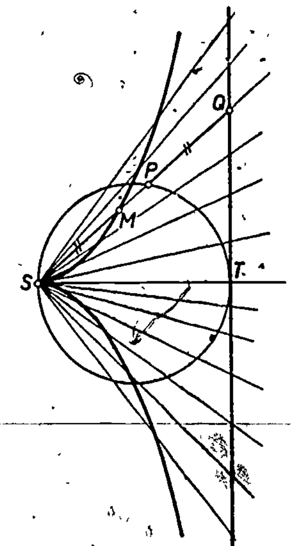


Fig. 12.

The curve so obtained is known as the scissoid of Diocles. With the aid of this curve Diocles (second century B.C.) trisected any arbitrary angle and duplicated the cube.

8. THE METHOD OF LOCI IN THE SOLUTION OF CONSTRUCTION PROBLEMS

The method of loci in the solution of problems in geometric construction consists, as is well-known, of the following.

If the problem is reduced to the determination of one or more points in a plane which are to satisfy several given conditions, we discard one of the conditions and obtain a locus. Discarding another of the conditions we obtain a second locus, and so on. At the intersection of these loci lies the required point or aggregate of points.

Sometimes one of the loci is given directly in the statement of the problem itself. The method of loci also facilitates the analysis of the problem.(1)

We now turn to the consideration of some examples.

Problem. To find on a given line a a point equidistant from two given lines b and c. Lines a, b, and c are assumed distinct.

The given line a is the first locus of the given point. If the condition that the desired point belongs to the given line a is suspended, the point could be anywhere on the locus of points equidistant from the two given lines b and c. The required point will actually lie at the intersection of these loci. Two cases are possible:

1. Lines b and c intersect. The second locus is a pair of bisectors, and the required point is found at the points of inter-

(1) Footnote: i.e. the investigation of the conditions under which a solution to the problem exists and the number of such solutions. --Translators

section of these bisectors with the given line a .

In this case the problem has, in general, two solutions. However, if the line a is parallel to one of the bisectors there is only one solution. If the line a coincides with one of the bisectors there is an infinite set of solutions. If line a passes through the intersection of the bisectors the problem has one solution.

2. The lines b and c are parallel. The second locus is the line g , parallel to lines b and c.

The problem has one solution if line a intersects the equidistant parallel g ; no solution if a and g are parallel; an infinite set of solutions if g coincides with a .

Problem. To construct a circle having a given radius a and tangent to two given circles.

Let O_1 and O_2 be the centers of the given circles and R_1 and R_2 their radii. It is sufficient to find the center of the required circle. Suspending the condition of the tangency of the required circle to the second of the given circles, we obtain the first locus, consisting of the pair of concentric circles with center at O_1 and radii $R_1 + a$ and $R_1 - a$. Suspending the condition of tangency to the first circle, we obtain another locus -- the pair of concentric circles with center at O_2 and radii $R_2 + a$ and $R_2 - a$.

The required centers will be found at the intersection of these loci. Not more than eight solutions are possible. Drawing the figure and carrying out the analysis is suggested as an exercise.

Sometimes the method of loci is not applied in so pure a form.

<u>Problem.</u> <u>To construct a triangle given the altitude AH = h, the median AM = m and the bisector AD = d, all drawn from the same vertex.</u>

The triangles ADH and AMH can be constructed immediately (fig. 73). Let us imagine that a circle has been circumscribed about the required triangle ABC. The bisector AD must pass through the midpoint E of the arc of this circle subtended by the chord BC.

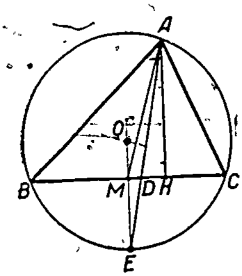


Fig. 73.

The points M, E and the center O of the circumscribed circle lie on the perpendicular to the side BC at its midpoint M.

This perpendicular to MH at M can be constructed. Its intersection E with the bisector AD can likewise be constructed. The center O of the circumscribed circle is equidistant from the points A and E and therefore can be found as the intersection of the locus of points equidistant from A and E with the already constructed line EM. Thus, the position of the center O is determined. From O with radius OA we construct the circumscribed circle. Its intersection with the line MH yields the vertices B and C of the required triangle. For the construction to be possible it is necessary that $m > d$, since the angular bisector always lies within the angle formed by the median and the altitude.

We shall now solve an auxiliary problem.

Problem. To construct the radical axis of two circles.

If the two circles intersect, the line joining their points of intersection will also be the radical axis. If the circles are tangent, the radical axis is the perpendicular at the point of tangency to the line joining their centers.

There remains to be considered the case of non-intersecting circles. Let O_1 and O_2 be the centers of the two circles (fig. 74). We draw a circle of arbitrary radius and arbitrary

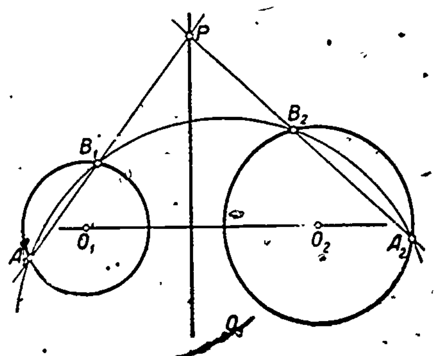


Fig. 74.

center so that it intersects both of the given circles: the one in the points A_1, B_1 and the other in the points A_2, B_2. The point P of intersection of the lines A_1B_1 and A_2B_2 belongs to the required radical axis. For proof of this we observe that the tangents drawn from P to the circles (O_1) and (O_3) are equal, since P is an exterior point of their radical axis. The tangents from P to circles (O_2) and (O_3) are likewise equal. It follows that the tangents from P to circles (O_1) and (O_2) are equal, q.e.d.

Letting fall a perpendicular from P to the line of centers O_1O_2 or else constructing analogously a second point P', we find the required radical axis.

Problem. To construct a circle orthogonal to three given circles.

Let O_1, O_2 and O_3 be the centers of the given circles. Suspending the condition of the orthogonal intersection of the required circle with circle (O_3), we find the locus of the centers of the circles orthogonally intersecting two of the given circles, (O_1) and (O_2), namely, the set of exterior points of their radical axis. Suspending the condition of the orthogonal intersection of the required circle with circle (O_2), we find the locus of the centers of the circles orthogonally intersecting circles (O_1) and (O_2).

The center of the required circle lies at the intersection of the loci thus found. Its radius is equal to the common length of the tangents from this center to the given circles. The center thus found is called the radical center of the three circles.

The construction and analysis of the problem are suggested as an exercise.

The excellent collection of geometrical construction problems by I. I. Alexandrov [3] is recommended for additional practice.

Methodological pointers especially intended for teachers but useful also to students will be found in the brochure by Prof. D. I. Perepiolkin [42], in the well-known problem collection of P. S. Modenov [38] and in reference [35].

9. ON PROBLEMS IN GEOMETRIC CONSTRUCTION AND THEIR SOLUTION

The most general definition of the term "problem" is as follows. A problem is the statement of a demand "to find", on the basis of "given" objects other "required" objects bearing specified relationships to each other and to the given things [2].

Thus in every problem there is some class of things given, and a class of things to be found (both classes consisting of points, lines, segments, triangles, circles and so on). Besides this, there must be specified the rules, procedures, conditions to be fulfilled in order that the things sought for shall be considered to have been found. Without this the problem becomes indefinite and the expression "to find" has no meaning. Whatever operations of construction were carried out, it would never be possible to say that the thing had been found, that is, transferred from the class of things sought to the class of things given.

Without precisely established and formulated conditions whose fulfillment is necessary before an object can be considered as transferred from the class of things sought to the class of known, given things, it is not possible to speak even of a problem, much less of its solution.

Let it be required, for example, to inscribe in a circle a regular seventeen-sided polygon.

Here the given thing is understood to be the circle. It is commonly considered that the center and the radius of this circle are also given without explicit mention, but this is not entirely correct.

If neither the center nor the radius of the given circle were given, it would be necessary to find them - supposing, of course, that they were necessary for the solution of the problem. Center and radius would then be things sought. Thus it is evident that precise indication of the things given is absolutely indispensable. Otherwise, instead of one problem we risk having a different one.

Further, as the object sought we understand a regular seventeen-sided polygon, that is, a polygon whose sides (and angles) bear the relationship of equality to each other. The vertices of this required polygon must bear to the given circle the relationship of being points belonging to it. The problem reduces to the finding of two determinate points - vertices on the circle, or to the finding, of a determinate segment.

Those conditions under which the polygon sought will be considered to have been found play a fundamental role. It is possible, for example, to consider the polygon as found if its construction is carried out with a pair of compasses only without the use of a straight-edge. What is meant by "carried out with a pair of compasses only" must be precisely elucidated. Does it mean that the required side of the polygon will be considered to have been constructed if its ends are found as the points of intersection of circles upon whose radius no limitation of any sort is imposed, or is this radius given? In the latter case we should have a construction with one setting of the compasses.

The solution of a problem will be different as the specification of the means to be used is changed. It can even happen that with the means specified the problem can not be solved at all. It is also possible to consider the polygon as found if it is constructed with the aid of a straight-edge only, without the use of compasses. The straight-edge, in turn, may be single-edged and without markings, that is, only the drawing of straight lines is admissible. In this case, the polygon will be considered constructed if its vertices are found as points of intersection of the given circle with straight lines. The straight-edge may be double-edged, that is, a pair of parallel lines. Finally, there may be on the straight-edge a scale of length, of which essential use is made in the construction.

In each of these cases different conditions are indicated upon whose fulfillment the polygon will be considered as found or constructed.

If the conditions chosen are indeed fulfilled, but the actual construction demands such a complex fabric of lines that in practice the required polygon will not even close because of inescapable errors in the drawing, then regardless of the theoretical irreproachibility of the construction, the problem must be acknowledged to be insoluble practically.

It is well-known that with a not very large number of empirical trials it is possible actually to find the side of an inscribed seventeen-sided polygon with a high degree of accuracy which is not approached by any theoretically correct construction of the figure.

Furthermore, it is sometimes possible to indicate a regular chain of constructions such that with each successive link in the chain we obtain the required figure with greater and greater accuracy. If it is agreed to consider a problem solved when one or several constructed points of the required figure can be made to fall within any given neighborhoods, no matter how small, of the corresponding theoretically determined points, then here also we have a correctly formulated mathematical problem.

Finally, it is possible, without insisting upon any regular chain of constructions converging toward the required figure, to seek approximate constructions with the aid of instru s chosen in advance. In this case the problem is also proper y stated and can be considered a mathematical problem.

In secondary-school geometry, by virtue of an historically developed tradition, a construction problem is considered solved if the construction is accomplished with the aid of compasses, i.e. by the drawing of circles of any radius and with any centers, and of a one-sided straight-edge, i.e., by the drawing of straight lines. In this connection it is considered possible to draw with complete accuracy a straight line through two points and a circle of any desired radius from a given center.

It must be stressed that any other mathematical instruments can likewise be regarded as ideally accurate, and constructions performed with their use as equally irreproachable.

In such cases the construction problem will remain just as much a proper mathematical problem as if the construction were by compasses and straight-edge. There is only one invariable requirement: the precise specification of the instruments to be used and the rules which must be observed before objects sought are to be considered as found. For example, using the ellipsograph it is considered to be possible to construct any desired ellipse with foci at any points F_1 and F_2 and a given sum of the distances MF_1, and MF_2 from the foci to every point M of the ellipse.

10. CONSTRUCTIONS WITH COMPASSES AND STRAIGHT-EDGE

In plane geometry a construction problem is considered solved if it is reduced to a finite number of the following five basic (simplest) problems:

1. To draw a line or a segment through two given points.

2. To draw a circle of given radius and with a given point as center, or to draw a circular arc given its endpoints and center.

3. To find the point of intersection of two given lines.

4. To find the points of intersection of a given line and a given circle.

5. To find the points of intersection of two circles.

The solution of these simplest basic problems is considered to be known.

Arbitrary elements are often used in constructions.

The possibility of adding arbitrary points to those given or already constructed required special stipulations which are customarily formulated in the following two statements:

Stipulation A. It is possible to construct an arbitrary point in the plane exterior to a given line.

Stipulation B. It is possible to construct on a given line an arbitrary point not coinciding with any of the points already constructed on the line.

Stipulations A and B are essential in the solution of many problems. For example, if it is required to draw a circle intersecting a given line a at right angles having only one given line and not a single given or constructed point we are deprived of the possibility of using problems 1 through 5. Without stipulations A and B we are likewise deprived of the possibility of constructing a line through a given point exterior to a given line and parallel to the latter. Having only one given point and one given line, we could not draw any circle and construct any line (since, for example to draw a circle we must have two points determining its radius).

In proving the correctness of any construction making use of arbitrary elements we must not, of course, base our argument on any special properties of these elements, but must consistently regard the elements as arbitrary.

We might say that a figure has been "constructed" when each part of the figure has been obtained as the end result of a suitable sequence of the five basic constructions and the two stipulations A and B.

The constructions themselves may be carried out on paper, with compasses and straight-edge in hand, or only verbally. Which of these modes of procedure we choose is immaterial. For example, there exists no geometrical instrument for drawing a sphere with given center and radius.

It is necessary to seek out the simplest, most economical and most accurate construction. Complicated constructions turn out to be, as a rule, the least precise.

A different mode of employment of the compasses and straight-edge or the use of other instruments requires a different list of simplest problems.

Problems 1 through 5 and stipulations A and B may be called the conditions corresponding to the free use of compasses and straight-edge.

Since it would be extremely tiresome to carry out for every construction problem its complete reduction to the indicated five simplest problems, this reduction is carried out once and for all for a certain range of moderately complex problems. All other problems are thereafter reduced to these already solved problems. To such problems solved "once and for all" belong, for example, the following:

Problem 1. To draw through a given point A not on a given line a a line parallel to a.

Solution. We select on the line a an arbitrary point B not coinciding with the foot of the perpendicular from A to line a (fig. 75).

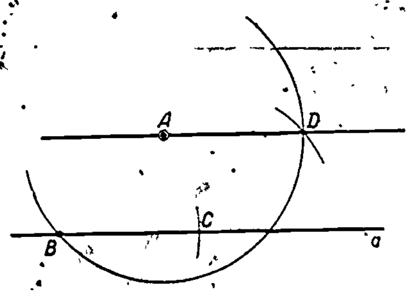


Fig. 75.

The selection of point B is in practice effected by means of an arbitrary setting of the compasses greater than the distance from point A to line a, one point of the compasses being placed at A and the other at B. The choice of a compass setting greater than the distance from point A to line a is not imposed by any theoretical necessity and merely serves the end of greater precision of the drawing.

The construction of a circle with center A and radius AB is permissible as an instance of simplest problem number 2.

From point B with the same radius AB, we describe a circle, using again the same problem 2. In practice, of course, only a small arc intersecting line a is drawn. We mark the point of intersection C of this circle with line a, which we are permitted to do by problem 4. With the same radius CB and center at C, making use of problem 2, we draw a circle and, in accord with problem 5, mark the point of intersection D of this circle with the circle first drawn. We draw a line joining points A and D, on the basis of problem 1. The constructed line AD is the required line.

The proof of the correctness of the construction is given in the secondary school geometry course.

The line AD parallel to the line a has been constructed with compasses and straight-edge "once and for all", and may in future be drawn by means of, say, a triangle and straight-edge, as is done in practice. The important thing for us is that the problem of drawing the parallel can be solved by means of compasses and straight-edge.

Problem II. Given a segment AB on the line a, it is required:

a) to find a second segment which is a given multiple of the first;

b) to divide the given segment AB into any given number of equal parts;

c) to find a second segment bearing a given ratio to the given segment AB.

Solution. a) With radius BA we draw a circle with center at B (problem 2) and from the point C (distinct from A) of the intersection of this circle with line a (problem 4) as center we draw a circle of the same radius (problem 2), marking its intersection D with line a, and so on until the desired multiple has been reached (fig. 76).

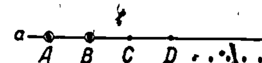


Fig. 76.

In intuitive terms, we "pace with compass-legs along the line". Usually the problem is solved with a pair of dividers "alternating" its legs.

b) We draw through point A an arbitrary line b, which can be done by taking an arbitrary point M not on line AB and joining it with point A (fig. 77).

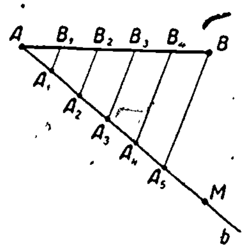


Fig. 77.

In constructions with compasses and straight-edge, such arbitrary elements can in the majority of cases be replaced by constructed ones. In the present problem circles can be drawn from A as center with radius AB and from B as center with radius BA, and one of their points of intersection can be taken as point M.

On line AM we lay off from point A an arbitrary segment a number of times equal to the number of parts into which it is required to divide segment AB, in the manner just described under part a) of the present problem. We join the endpoint A_n of the last segment laid off with point B (problem 1) and draw through the points $A_1, A_2, \ldots, A_n$ lines parallel to A_nB (problem 1).

Points $B_1, B_2, \ldots, B_{n-1}$ of the intersections of these parallels with line AB (problem 4) will be the required points of division of segment AB into n equal parts (in the diagram, n = 5).

The proof of the correctness of the construction is common knowledge; it is independent of the choice of point M.

For the division of a segment into halves or into an even number of parts other constructions, familiar from secondary-school geometry, can be used.

c) Let it be required that the segment CD shall bear to the given segment AB the ratio of m to n, where m and n are positive integers. From the definition of a ratio it follows that $n \cdot CD = m \cdot AB$, whence the construction of segment CD is obvious. We take segment AB m times in accord with part a) of the present problem, and divide the result into n equal parts as indicated in part b).

Problem III. Through a given point A to draw a line perpendicular to a given line a.

Solution 1. Point A lies on line a (fig. 78). On line a we select an arbitrary point B distinct from A (stipulation B), and with radius AB describe a circle with center at A (problem 2), that is, with an arbitrary setting of our compasses we simply "mark off points" on line a. We denote by C that point of intersection of the circle with line a which is distinct from B (problem 4) and with radius BC we describe circles with centers B and C (problem 2). Taking one of the points D of of the intersection of these last circles (problem 5), we join A and D by a line AD (problem 1). The line AD thus constructed is the required perpendicular to line a erected at point A.

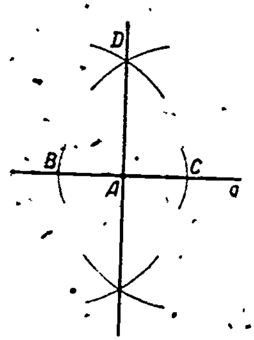


Fig. 78.

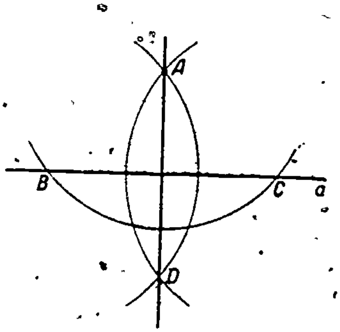


Fig. 79.

The proof of the correctness of the construction is given in secondary school geometry.

2. Point A lies outside line a (fig.79). We select on line a an arbitrary point B, which must be distinct from the foot of the perpendicular from point A to line a (stipulation B). From A as center and with radius AB we describe a circle (problem 2) and denote by C the second point of its intersection with line a (problem 4). From B and C as centers and with the same radius we describe circles (problem 2) and join by a line AD (problem 1) the points A and D of the intersection of these latter circles (problem 5). Line AD will be the required perpendicular from point A to line a.

The proof of the correctness of the construction should be known to the reader.

Problem IV. To draw through a given point A a line which shall form with a given line a an angle equal to a given angle.

Solution 1. Point A lies on line a (fig. 80).

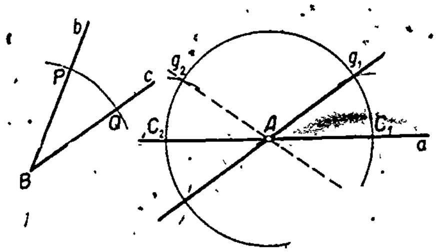


Fig. 80.

Let bBc be the given angle. On a side of this angle we take an arbitrary point P (stipulation B) and with radius BP from B as center describe a circle (problem 2). With the same radius we describe a circle from A as center (problem 2) and denote by

C_1 and C_2 the points of intersection of this circle with line a (problem 4). We then construct circles with centers C_1 and C_2 having equal radii PQ (problem 2) and denote by g_1 and g_2 their points of intersection with the circle whose center is A (problem 5). We draw lines Ag_1 and Ag_2 (problem 1), which are the required lines.

Here we have two solutions. The proof of the correctness of the construction is based, of course, on the fact that central angles subtending equal arcs of equal circles are equal.

2. Point A lies outside line a. We take an arbitrary point A_1 on line a (stipulation B) and construct a line intersecting line a at point A_1 at the given angle, as in solution 1 above. We then draw through point A a line parallel to this line (problem 1). The parallel so constructed will be the required line. Here again we have two solutions. The proof of the correctness of the construction is obvious.

Problem V. a) To bisect a given angle; b) to construct an arbitrary multiple of a given angle.

Solution. a) Let aAb be the given angle (fig. 81). On line a we select an arbitrary point B (stipulation B) and describe a circle with A as center and radius AB (problem 2). In practice this circle of arbitrary radius is drawn at once, without first choosing an arbitrary point B. We denote by C the point of intersection of the circle with line b (problem 4) and with the same radius AB we describe circles from centers B and C (problem 2). We denote by D that point of intersection of these circles which is distinct from A (problem 5) and join points A and D by a line (problem 1). The line AD so constructed is the required bisector of the given angle.

The proof of the correctness of the construction is based upon the equality of triangles ADC and ADB (corresponding sides are equal).

b). Let aAb be the given angle (fig. 82).

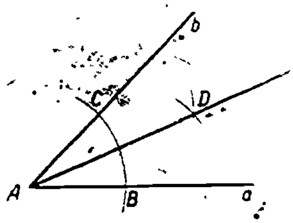


Fig. 81.

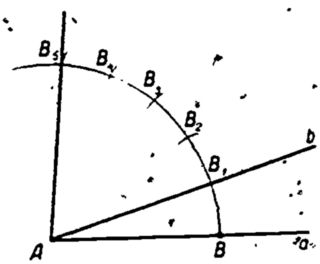


Fig. 82.

Taking an arbitrary point B on ray a (stipulation B), we describe a circle with center A and radius AB (problem 2). We denote by B_1 the point of intersection of the circle with line b (problem 4), and with B_1 as center and radius B_1B we describe a circle (problem 2) and denote by B_2 the intersection of this circle with the one previously drawn (problem 5). With B_2 as center and the same radius B_1B we describe a circle (problem 2) and establish the point B_3 (problem 5). Continuing thus we construct the angle BAB_n which is the required multiple n of angle aAb. In the diagram $n = 5$.

The proof of the correctness of the construction is based, of course, on the equality of arcs subtended by equal chords and on the equality of central angles subtended by equal arcs.

With this we shall conclude the list of problems solved "once and for all" with compasses and straight-edge. The list could, of course, be extended, and in secondary school geometry, for example, under the heading of basic constructions, we have a rather large list of problems.

It should be noted that in solving any particular construction problem with the use of compasses and straight-edge the problems I through V just solved are made use of in the same way as the basic (simplest) problems 1 through 5. For example, it is immaterial whether we draw a perpendicular to a given line by repeating the entire construction given in problem III or by making use of a draftsman's triangle. Such use of the triangle does not constitute the introduction of a new geometrical instrument, but simply replaces a chain of construction steps carried out with compasses and straight-edge.

Remark. Under part a) of problem V the angle was bisected, but not divided into any number n of equal parts like the segment in part b) of problem II.

This brings us to the fact that not all geometric constructions can be executed with compasses and straight-edge, or, more exactly, on the basis of problems 1 through 5 and stipulations A and B set forth above.

Thus, for example, it is impossible with compasses and straight-edge to divide arbitrary angles into three equal parts. This means that the drawing of only lines and circles is not sufficient for obtaining the rays which would divide an arbitrary angle into three equal parts. It must be emphasized that an angle equal to one-third of any given angle exists. Likewise, rays exist which divide any given angle into any given number of equal parts, but for some numbers of parts (e.g., n = 3) it is not possible to bring these rays from the class of things sought into the class of things given by means only of basic problems 1 through 5 and stipulations A and B, that is, "with the aid of compasses and straight-edge". These instruments are "not powerful enough" for such problems.

By means of certain other instruments, that is, with the aid of other basic problems the division of any angle into three equal parts -- the trisection of the angle -- can be carried out without difficulty and with unimpeachable logic. Furthermore, the practical accuracy of such constructions often considerably exceeds the accuracy of constructions with compasses and straight-edge.

It should be noted that it is possible to divide some specific angles into three equal parts by means of compasses and straight-edge. Among such angles, for example, is the right angle. The problem arises of precisely determining the entire set of angles admitting of trisection with compasses and straight-edge, and of proving that the remaining angles (including in particular, the angle 60°) cannot be divided into three equal parts with these instruments, that is, on the basis of these basic problems.

Proofs of impossibility of this kind form some of the most difficult problems of mathematics and for the most part are beyond the scope of elementary methods. At the end of the nineteenth century the question of the solvability of construction problems by means of compasses and straight-edge was investigated completely. The classes of problems solvable and not solvable by these means were precisely specified. The complete solution of the question is taken up in higher algebra and mathematical analysis.

In 12. we shall become acquainted with one elementary proof of impossibility.

11. ON MEANS OF SOLVING CONSTRUCTION PROBLEMS OTHER THAN THE FREE USE OF COMPASSES AND STRAIGHT-EDGE

The free use of compasses and straight-edge is mathematically expressed by means of simplest problems 1 through 5 and the stipulations A and B. This means that a geometrical object may be considered to have been trasferred from the class of things sought to the class of things given only if the procedure involved may be reduced to the above mentioned basic problems and stipulations. These latter may be regarded as the logical or theoretical instruments for solving problems.

To other combinations of geometrical instruments there will correspond other logical means of solution, formulated in the shape of basic problems. We shall consider some examples.

Let us take as the only allowable geometrical instrument a pair of free compasses, that is, compasses by means of which we can draw circles of any radius. This is the case of construction with compasses only. Mathematically this means that only basic (simplest) problems 1, 2 and 5 may be used.

The application of problem 1 has now to be understood only in the sense that a straight line is determined by two points; but we can not, for example, select an arbitrary point on that line. A point on a given line other than the two points which determine it must be obtained as the intersection of two circles.

The exclusive use of problems 1, 2 and 5 does not prohibit the physical drawing of a line in a diagram with a straight-edge; it simply prohibits finding a point as the intersection of two lines or of a line and a circle. If lines are actually drawn rather than simply imagined, problems 1, 2 and 5 do not permit us to consider the points of their intersection in the diagram as transferred from the class of things sought to the class of things given. A point of intersection of drawn lines continues to belong to the class of things sought. Such a point does, however, enter the class of things given if it may be obtained as the point of intersection of constructed or given circles.

For construction with compasses alone, stipulation A is not needed, since there are in this case always two points on a line, which determine it and which belong to the class of things given. Having these two points, we can construct with the aid of the compasses as many points exterior to the line as desired.

The use of compasses only changes the way in which a straight line is given. Previously a line could be given (by drawing it with a straight-edge) without any points upon it being identified. A circle, as before, is considered given if "it and its center are drawn", although it may be that on the circle itself there are no identified points. In other words, the circle together with its center belongs to the class of things given, but the points on the circle do not belong to the class of things given. This is analogous to the situation with regard to the straight line in constructions with the ruler and compass.

We shall now add to problems 1, 2 and 5 the following stipulation.

Stipulation B*. It is possible to construct an arbitrary point on a given circle not coinciding with any previously constructed point on the circle.

There is no need for an analogue to stipulation A since there is always given a point not lying on the constructed circle, namely the center of this circle.

The necessity of introducing stipulation B* will be evident from the following problem.

To inscribe in a given circle (the center of which is also given!) a regular hexagon, using only basic problems 1, 2 and 5.

Since we have here no two identified points, but only one -- the center of the given circle -- we can not set up the radius of the given circle, we "can't open the compasses", and are consequently deprived of the possibility of executing any construction whatever.

Making use, then, of stipulation B*, we select an arbitrary point on the circle, which together with the center determines a radius and will itself be the center of the circle with which the usual construction of an inscribed regular hexagon begins.

As we can not consider in detail all the constructions performed with a single pair of compasses, we shall solve only a few selected problems.

a) Given segment AB. Required: to construct a second segment which will be a given multiple of the first (fig. 83).

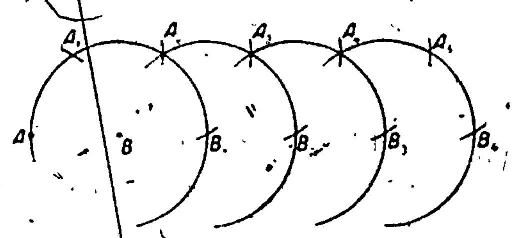


Fig. 83.

Solution. With B as center we describe a circle of radius BA (problem 2). By means of circles with the same radius BA but with centers at A, A_1 and A_2 (problems 2 and 5) we find point B_1. Segment AB_1 (problem 1) is twice the given segment. Continuing, with B_1 as center and the same radius AB we describe a circle (problem 2), and from the point A_2 we again describe a circle of the same radius (problem 2). The circle next described (problem 2) from point A_3 (problem 5) yields point B_2 as an intersection of circles (problem 5). The segment AB_2 (problem 1) is equal to three times segment AB. Continuing in the same manner, we construct a segment which is the required multiple of the given segment.

The correctness of the construction is readily apparent. Furthermore, it is evident from this construction that when the instruments allowed are other than the compasses and straight-edge different procedures and devices are required for the solution of problems.

b) To divide a given segment AB into any given number of equal parts (fig. 84).

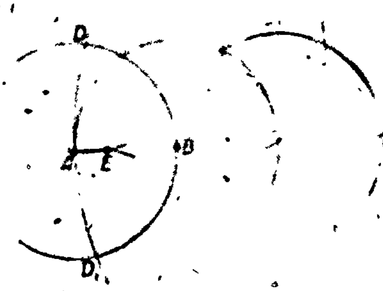


Fig. 84.

Solution. We multiply segment AB a number of times equal to the number of parts into which we require to divide it (problem a) above). We obtain the segment $AC = n \cdot AB$ (in the diagram $n = 3$). From C as center with radius CA we describe a circle (problem 2) and determine its points of intersection D and D_1 (problem 5) with the circle having center at A and radius AB. This latter circle was constructed in connection with finding the segment AC. From centers D and D_1 we describe circles of radius $DA = D_1A$ (problem 2) and find their point of intersection E (problem 5). Segment AE is then the required part of segment AB.

Proof. By virtue of the symmetry of the construction, point E is collinear with points A, B and C. The isosceles triangles DAC and AED are similar, since base angle A is common to both. From the similarity of these triangles it follows that

$$\frac{AE}{AD} = \frac{AD}{AC},$$

but $AD = AB$ and consequently

$$\frac{AE}{AB} = \frac{AB}{AC} = \frac{1}{n},$$

whence $AE = \frac{1}{n} AB$.

In particular, we may take $n = 2$, i.e., we can bisect a segment using only a single pair of compasses. The set of operations with compasses alone constitutes only a part of the set of operations with compasses and straight-edge; nonetheless the following theorem holds good: Every problem solvable with the use of compasses and straight-edge can be solved using only a single pair of compasses.

We shall not stop to prove this (see Chapter VIII).

At the end of the eighteenth century the mathematician Masqueroni dealt systematically with constructions using only a single pair of compasses. It was he who established the theorem just cited [13] [2] [3].

In his constructions Masqueroni had, for example, to solve the following problems.

1) To find the point of intersection of a given line and a given circle.

2) To find the point of intersection of two given lines.

These had to be found by drawing circles only.

The straight-edge alone is a geometrical instrument "less powerful" than the compasses. Many problems solvable with compasses and straight-edge or -- what is the same thing -- with a pair of compasses alone, are unsolvable with the aid of a straight-edge only, that is, using only basic problems 1 and 3 and stipulations A and B. In the next section we shall meet one of these constructions.

Without lessening our ability to solve all construction problems which can be solved by the free use of compasses and straight-edge, we can limit the freedom of use of the compasses by giving them a completely determined setting, that is, limiting ourselves to the drawing of circles of a single fixed radius. More than this, the use of the compasses can be limited to the drawing of a single circle together with its center, then laying aside the compasses performing the remaining construction with the straight-edge [53].

Clearly, these constructions are characterized by basic problems 1 and 3, stipulations A, B and B* together, and the following:

4'. <u>It is considered possible to construct one circle together with its center.</u>

5'. <u>It is considered possible to construct the points of intersection of given or constructed lines with the circle mentioned in stipulation 4'.</u>

But if this single circle were to be given <u>without its center</u>, such a circle together with the straight edge would not be adequate to enable the solution of all problems which can be solved with the free use of compasses and straight-edge. This assertion will be proved in the next section.

We shall turn now to constructions executed with the double-edged ruler, that is, with the aid of two parallel lines separated by a fixed distance a. Such constructions are based on basic problems 1, 2 and 3 and stipulations A and B. In place of problems 4 and 5 we introduce the following:

4". It is considered possible to construct two lines parallel to a given or constructed line and separated from it by the distance a.

This permits us to place one edge of the ruler upon a line and use the other edge to draw another line.

5". It is possible to construct two parallel lines separated by the distance a and passing respectively through two given points, the distance between which is not less than a.

This permits us to place the straight-edge so that one of its edges passes through one given point and the other edge through the other given point.

With respect to basic problem 2 it is necessary to make the same observation as was made regarding lines when we considered constructions using only a single pair of compasses. It is permitted to draw any circles, but it is not permitted to use their points for constructions unless these points are also obtained as intersections of given or constructed lines.

Before adducing examples of constructions using the double-edged ruler, we shall prove the following lemma on the trapezoid.

The line joining the point of intersection of the diagonals of a trapezoid with the point of intersection of its non-parallel sides bisects the base of the trapezoid (fig. 85).

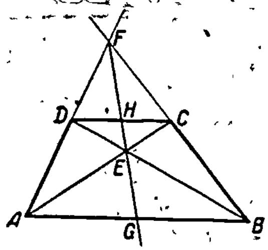


Fig. 85.

Proof. Let ABCD be the trapezoid, E the point of intersection of the diagonals AC and BD, and F the point of intersection of its non-parallel sides AD and BC. We shall denote by G and H the points of intersection of the line EF with the parallel sides AB and DC of the trapezoid. From the similarity of triangles AGF and DHF we have:

$$\frac{AG}{GF} = \frac{DH}{HF},$$

and from the similarity of triangles BGF and CHF we have:

$$\frac{GF}{GB} = \frac{HF}{HC}.$$

Multiplying the corresponding sides of these equalities, we obtain:

$$\frac{AG}{GB} = \frac{DH}{HC}.$$

In exactly the same way it follows from the similarity of triangles AGE and CHE and of triangles BGE and DHE that

$$\frac{AG}{GB} = \frac{HC}{DH},$$

or, after multiplying together the corresponding sides of the last two equations,

$$\left(\frac{AG}{GB}\right)^2 = 1.$$

It follows that AG = GB and, consequently, DH = HC.

Problem. To bisect a given segment AB (fig. 85).

Solution. We draw a line parallel to line AB and separated from it by the distance a (problem 4"). We select two arbitrary points D and C on the constructed line (stipulation B) and construct the trapezoid ABCD, drawing the lines AD and BC (problem 1). Constructing the diagonals of the trapezium AC and BD (problem 1), we find point E of their intersection (problem 3); denoting by F the intersection of the non-parallel sides (problem 3), we draw a line joining points E and F. (problem 1). The point of intersection G of lines EF and AB (problem 3) is the required midpoint of segment AB.

The proof follows directly from the lemma on the trapezoid.

Problem. Through an arbitrary point A exterior to line a to draw a line parallel to line a.

We have solved this problem using compasses and straightedge. We shall now give the solution using the double-edged ruler.

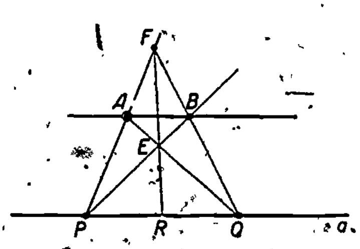


Fig. 86.

Solution. On line a (fig. 86) we select two arbitrary points P and Q (stipulation B) and construct the midpoint R of segment PQ (by the preceding problem).

Selecting an arbitrary point F on line AP (problem 1; stipulation B) outside the segment PA, we join point Q with points A and F by the lines QA and QF (problem 1). Point E, the intersection of lines QA and RF (problems 1 and 3) is joined to P (problem 1) and the intersection of PE with QF is denoted by B (problem 3). The line AB (problem 1) is the required parallel.

The proof, using the lemma on the trapezoid is left to the student.

Problem. To draw through point A of line b a line perpendicular to line b (fig. 87).

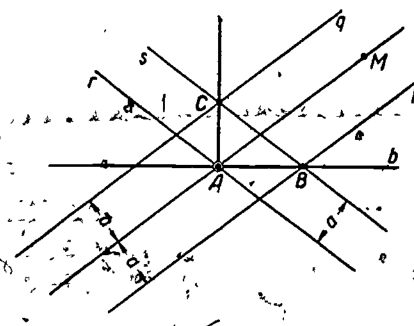


Fig. 87.

Solution. Selecting an arbitrary point M in the plane of the drawing (stipulation A), we draw to line AM (problem 1). We draw lines p and q one on each side of line AM, parallel to AM and separated from it by the distance a (problem 4"). Let B be the point of intersection of lines p and b (problem 3) (the construction being so performed that AB > a). We next draw a pair of parallels r and s separated by the distance a and such that line r passes through A and line s through B (problem 5"). Joining point C, the intersection of lines q and s (problem 3), with point A by line AC (problem 1) we obtain the required perpendicular AC to line b. The proof results from the consideration of the rhombuses obtained during the construction and the properties of their diagonals.

Being able to draw parallels, it is easy to drop a perpendicular from a given point to a given line using only the double-edged ruler. It would be useful actually to execute a construction of this kind.

Problem. a) To bisect a given angle;

b) to find an arbitrary multiple of a given angle.

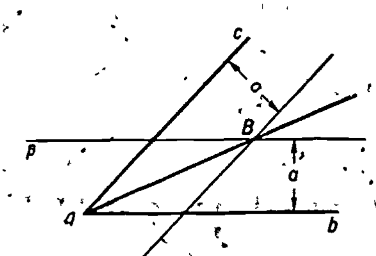


Fig. 88.

Solution. a) Let bAc be the given angle (fig. 88). We draw lines p and q respectively parallel to sides b and c of the angle and separated by the distance a from these sides (problem 4"). We then draw the diagonal AB of the rhombus thus formed (problems 3 and 1). The proof of the correctness of the construction follows from the properties of the diagonal of a rhombus.

b) It is sufficient to be able to double the given angle. Let bAc be the given angle (fig. 89). We draw a parallel p to the line b separated from it by the distance a (problem 4") and denote by C the point of intersection of p and c (problem 3). We then construct a pair of parallels r and s at distance a from each other and passing respectively through points A and C (problem 5"). The angle rAb is twice the angle bAc. The proof follows from the properties of the rhombus.

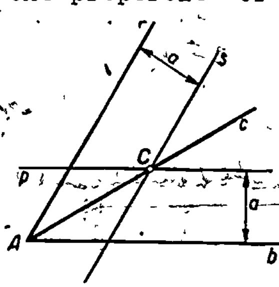


Fig. 89.

The ability to draw parallels also permits the solution of the following

Problem. To draw through a given point a line forming with a given line an angle equal to a given angle.

The carrying out of this construction is suggested as an exercise.

As a final example of the employment of the double-edged ruler, we present the solution of a problem which is highly typical of constructions performed by means of instruments with which it is impossible actually to draw a circle, that is, to make use of the points of a circle.

Problem. To construct the points of intersection of a given line b with a circle of which the center C and radius CM are given (fig. 90).

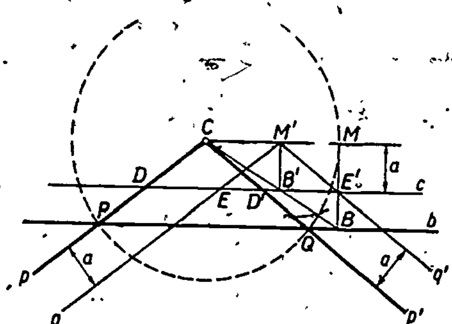


Fig. 90.

Solution. Let us suppose that the radius CM is parallel to the line b. If it is not, it can readily be reproduced in this position by means of the foregoing constructions. (This should be done as an exercise.)

We draw line c parallel to line CM and separated from it by distance a (problem 4"). Selecting an arbitrary point B on the given line b (stipulation B), we draw the line CB (problem 1) and mark the point B' of its intersection with line c (problem 3). We join B to M (problem 1) and draw through B' the line B'M' parallel to BM (a previously solved problem). Through points C and M' we draw a pair of parallel lines p and q, at distance a from each other, which, in general, can be done in two ways (problem 5").

Points P and Q of the intersection of lines p and p' with line b (problem 3) are the required points of intersection of the given line b with the given circle.

Proof. By inspection of the rhombuses CM'ED and CM'E'D' (fig. 90) we find:

$$CD = CD' = CM' .$$

By the theorem on the segments cut off by parallel lines on the sides of an angle, we obtain:

$$\frac{CP}{CD} = \frac{CQ}{CD'} = \frac{CB}{CB'} = \frac{CM}{CM'}$$

From this by virtue of the previous equation we find:

$$CP = CQ = CM ,$$

that is, points P and Q belong to the given circle of radius CM, q.e.d.

On the basis of the foregoing we may consider the points of intersection of a given or constructed line with a given or constructed circle as already constructed.

We performed this operation earlier with the aid of the free use of compasses and straight-edge, namely on the basis of basic problem 4. We now see that the same thing can be done with the double-edged straight-edge. With the same means of construction it is possible to find the points of intersection of two given circles. (Find this construction as an exercise.) Now this means that a point is considered to be constructed if it is a point common to two given or constructed circles. In other words, using the double-edged ruler we can also solve basic problem 5. Thus, basic problems 1 through 5 are solved by constructions using the double-edged ruler.

We conclude from this: all constructions possible with compasses and straight-edge may also be carried out with a double-edged ruler only.

In order to acquire skill in these practically very important constructions, the student should find the center of a circle which has been drawn but whose center is not given in advance, using only the double-edged ruler and without using the points of the drawn circle. It is also suggested as an exercise that the student elucidate the question: is it possible, conversely, to solve with the free use of compasses and straight-edge every problem which can be solved with the double-edged ruler.

We may deal analogously with constructions done with the aid of a right angle, formulating basic problems appropriate to this instrument. This yields results analogous to the foregoing.

Every construction possible with compasses and straight-edge can be executed with a movable right angle alone.

Two movable right angles are more powerful than the compasses and straight-edge. With two right angles it is possible to solve not only the problems solved with compasses and straight-edge, but also problems which the latter instruments cannot solve [2], [51].

It is strongly recommended that, having acquired skill in solving the basic problems with various instrumentations, such as the double-edged ruler, the right angle and so on, the student solve with each of these instrumentations all the construction problems appearing in the basic secondary school geometry textbook.

The examples of geometrical instrumentation considered in this section enable us to draw certain general conclusions.

The choice of geometrical instrumentation can be made in various ways.

The mathematical equivalent of an instrumentation for geometrical construction is a system of basic (simplest) problems. Each geometrical instrumentation has its own system of basic problems; these having been discovered and the list of them established, all constructions may be performed merely in the mind's eye.

Whatever the instrument (or system of instruments) chosen, constructions performed with it will be mathematically rigorous providing only that the appropriate set of basic problems has been established and that all constructions are reduced with logical correctness to these basic problems.

It must once more be remarked that the concept of a problem includes the conditions which must be fulfilled in order that an object may be considered as transferred from the class of things sought to the class of things given. Consequently, if we change these conditions, we pass to a <u>different</u> problem.

12. EXAMPLES OF PROBLEMS NOT SOLVABLE WITH COMPASSES AND STRAIGHT-EDGE

Let us first consider the concept of the unsolvability of a particular construction problem using a particular set of instruments. As already noted, the proof of impossibility of the solution of a problem by a particular means is by no means easy.

The whole complex of questions relating to geometric constructions was finally solved only in the second half of the nineteenth century and only with the help of far from elementary parts of higher algebra and mathematical analysis.

All the more instructive, then, is the study of some elementary proof of impossibility presenting itself in a striking example. Such an example, unquestionably, is the following theorem.

Theorem. It is impossible to find the center of a given circle using only a single-edged ruler.

It is understood, of course, that the center of the circle is not given.

Proof. Before setting forth the proof itself, let us note those properties of central projection, which we shall need in our proof.[1)]

Under central projection (fig. 125) straight lines in plane π are projected into straight lines in plane π'. There is, however, in plane π, a special line a which has no projection in plane π'. On plane π, in turn, there is a line c' which is not the

1) The concept of central projection and an elementary proof of these properties will be introduced in the next chapter.

projection of any line of plane π. The lines a and c' are parallel to the line of intersection p of planes π and π'.

On every line in either plane which is not parallel to line p there is likewise a special point; these points are the intersections of the lines in question with the special lines, that is, with line a in plane π or line c' in plane π'.

Furthermore -- and this is essential for our problem -- it turns out that the center of projection S and the plane π' can be so chosen that a given circle a (fig. 124) of plane π can be projected from S onto a circle a' in plane π'. Moreover, the center S and plane π' can be chosen in an infinity of ways, so that under this projection the center K of circle a is not projected into the center of circle a'. This is the basic proposition which we need for the proof of the theorem.

The proof is by contradiction.

Let us suppose that the center K of the given circle a has been found with the aid of the single-edged ruler.

This is to say that merely by drawing a finite number of single lines through certain points constructed, in their turn, as the intersections of some constructed lines with each other or with the given circle a, we have obtained, finally, two lines p and q such that their intersection K determines the center of circle a.

The construction has thus involved a finite number of points which have served us for drawing lines, and a finite number of lines, the intersections of which have given us points.

We select a center of projection S and a plane of projection π' such that the given circle a will be projected into a circle a' of plane π' and such that none of the lines which we have constructed will be special lines nor will any of the constructed points be special.

Since we have at our disposal an infinite number of choices of the center of projection S and the plane π', while the constructed lines and points are finite in number, this selection of S and π', which we need, can always be accomplished. We can intuitively think of such a central projection as one in which the special line in plane π is located sufficiently far from our circle and constructed points.

Projecting our circle together with the constructed network of lines and points from center S onto plane π', we obtain on plane π' the circle a' and an analogous network of lines and points. We shall now trace step-by-step the construction which gave us lines p and q and their intersection K, the required center of the given circle a.

Each point and each line which we constructed has its image in plane π'. The entire construction on plane π is copied on plane π'.

Finally, there were constructed the last two lines p and q in plane π and, consequently, their images p' and q' in plane π'. The logical basis and all the steps of the constructions in plane π are identically repeated in plane π'. Lines p and q intersect at the center K of circle a; consequently, their images p' and q' intersect in the center K' (the image of K) of circle a'. But this is not true, since we know that K' is not the center of circle a'.

The assumption that there is a construction using a single straight-edge which will give the center of circle a has led to a contradiction. Consequently, such a construction does not exist.

Remark. Students often confuse the question of the existence of a mathematical object with the question of actually finding it, discovering or constructing it, by a particular means. It must not be forgotten that these are two entirely different questions. The foregoing theorem serves to illustrate this.

As an exercise, the student should prove that it is impossible to draw through a given point A, exterior to a given line a, a line parallel to a, using only a single-edged ruler.

From the proof of impossibility just given there may be drawn in particular another important conclusion: It is not possible to solve by means of one fixed circle without its center and a single-edged ruler all construction problems solvable with compasses and straight-edge. This is proved by the fact, for example, that it is impossible, as we have just seen, to construct the center of a given, fixed circle with the straight-edge alone. But with compasses and straight-edge this center can easily be found. As already pointed out, all problems solvable with compasses and straight-edge can also be solved by means of one fixed circle with its center and a single-edged ruler, i.e., by means of the so-called Steiner constructions [53] [2].

We shall consider briefly some general questions relating to the possibility or impossibility of solving a particular construction problem by means of compasses and straight-edge. Such problems, as geometrically formulated, are "mapped" into algebra, and only there do they receive their proper solution.

The fundamentals of the algebraic method of solution of construction problems are familiar from secondary school geometry [17]. We know how to construct with compasses and straight-edge, from given segments a, b, c, ..., certain required segments such as

$$a + b;\ a - b\ (a > b);\ \frac{ab}{c};\ \sqrt{ab};\ \sqrt{a^2 + b^2};\ \sqrt{a^2 - b^2},\ (a > b)\ . \qquad (*)$$

Also familiar are various ways of constructing the roots of quadratic equations with the use of compasses and straight-edge; the coefficients of the quadratic equation in such cases are usually given in terms of segments instead of numbers, and the equations are written in the homogeneous form

$$x^2 + px + q_1^2 = 0;\ x^2 + px - q_1^2 = 0\ ,$$

where p and q_1 are the given segments.

Of course the roots of such quadratic equations can also be found by applying the constructions of the above-mentioned expressions (*). It follows already from this that if a construction problem can be reduced to the construction of a segment and if this segment can be given as a root of an equation of degree not greater that two, then such a problem can be solved with compasses and straight-edge. But this is still a quite trivial result; we have here as yet no indication of a class, however narrow, of problems which it would be <u>impossible</u> to solve with compasses and straight-edge.

Before "mapping" the whole question from the "world of figures" into the "world of numbers", let us make the following observation. Every problem in construction can be reduced to this pattern: given some finite number of segments a, b, c, ..., it is required to construct one or more segments x, y, z, ... The given segments are specified as the sides of triangles and polygons, the radii of circles and so on; the required figures also consist of such elements.

We can most easily convince ourselves of the truth of this assertion by straightway mapping the whole plane upon which the construction is carried out into the field of all complex numbers, using the familiar "geometrical representation of complex numbers":

$$w = u + iv .$$

Here we have an example of a mapping -- which is in fact one-to-one -- of one set onto a second. One set consists of points, the other of numbers.

In this way, upon the introduction of a grid of Cartesian coordinates all the given figures can be represented with the aid of numbers. For example a circle is given by its center $z_0 = x_0 + iy_0$ [or simply by the coordinates (x_0, y_0)] and by its radius r; a line is given by two points $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, alternatively expressed as the two pairs of coordinates (x_1, y_1) and (x_2, y_2), and so on.

The numbers $z = x + iy$ are determined, in turn, by segments of length x and y on the coordinate axes.

Thus, actually every construction problem reduces to the construction, starting with given segments, of certain new segments.

Let the segments ~~a, b, c, ...,~~ as well as x, y, z, ... , be laid off from some point O on line a. Then to every point A of line A there will correspond a segment OA, and vice versa. Segments laid off in one direction we shall consider positive; those laid off in the other direction, negative. If one of the segments is taken as unity, we obtain the familiar "real number line".

Let us take first of all a single segment e and consider it as given.

We shall consider e as the unit segment of the numerical axis and ask ourselves: what points of the numerical axis can be constructed with compasses and straight-edge, starting from the unit segment e ?

Since we have agreed that all segments are to be laid off starting from point O, it will be sufficient to determine their endpoints, i.e., to determine certain points. If we succeed in finding all those, and only those, points which it is possible, given segment e, to construct with compasses and straight-edge (fig. 91), then the problem will have been solved for the indicated special case, namely, where there is only one given segment e. Instead of segments we may speak as is customary of numbers.

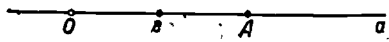


Fig. 91.

Arbitrary elements introduced in the course of constructions can always be considered rational, precisely by virtue of their arbitrary character.

Starting from the unit Oe, we can construct any integer n as the segment

$$n \cdot Oe = Oe + Oe + \dots + Oe .$$

Thus, every element of the ring of integers can be constructed with compasses and straight-edge.

It is easy to construct the number (or segment) $x = \frac{n}{m}$, on the basis of the proportion $\frac{x}{e} = \frac{n}{m}$, where $e = 1$ (the unit segment).

Thus, it is possible to construct every rational number, that is, any element of the field R_o of rational numbers (or segments), with compasses and straight-edge.

The everywhere dense set R_o of rational numbers is, as we know [40], countable, and consequently it does not exhaust the set of all real numbers (or segments).

Starting from any rational number a, we can construct the number $x = \sqrt{a}$, or the segment $x = \sqrt{ae}$, as the mean proportional between segments a and e; that is, $\frac{a}{x} = \frac{x}{e}$. Let us now select some rational number k_o such that $\sqrt{k_o}$ is irrational. By the foregoing, the number $\sqrt{k_o}$ can be constructed.

We can easily, in this instance, construct any number of the form

$$a_o + b_o\sqrt{k_o}, \text{ where } a_o \in R_o \text{ and } b_o \in R_o .$$

Having taken note that k_0 is a <u>fixed</u> rational number, we allow the numbers a_0 and b_0 to run through the entire set R_0 of rational numbers. The sum, difference, product and quotient of such numbers will again be numbers belonging to the same system.

In proof of this we note: if $x_1 = a_0' + b_0'\sqrt{k_0}$ and $x_2 = a_0'' + b_0''\sqrt{k_0}$, then

$$x_1 \pm x_2 = (a_0' \pm a_0'') + (b_0' \pm b_0'')\sqrt{k_0} \;;$$

$$x_1 \cdot x_2 = (a_0' + b_0'\sqrt{k_0})\,(a_0'' + b_0''\sqrt{k_0}) =$$

$$= (a_0'a_0'' + b_0'b_0''k_0) + (a_0'b_0'' + a_0''b_0')\sqrt{k_0} \;;$$

$$\frac{x_1}{x_2} = \frac{a_0' + b_0'\sqrt{k_0}}{a_0'' + b_0''\sqrt{k_0}} = \frac{(a_0' + b_0'\sqrt{k_0})\,(a_0'' - b_0''\sqrt{k_0})}{a_0''^2 - k_0 b''^2} = \alpha_0 + \beta_0\sqrt{k_0} \;;$$

$a_0''^2 - k_0 b_0''^2 \neq 0$, since if the contrary were true $\sqrt{k_0}$ would be rational. It is, of course, assumed that $b_0'' \neq 0$.

The system of numbers under consideration is closed with respect to the rational operations, and therefore constitutes a field R_1. Furthermore,

$$R_0 \subset R_1 \,.$$

The field R_1 is obtained by "joining" the number $\sqrt{k_0}$ to the field R_0. R_1 is an intermediate field. The numbers in the field R_1 can also be constructed with compasses and straight-edge.

Let us next select from field R_1 a number k_1 such that $\sqrt{k_1}$ does not belong to the field R_1, that is, the radical $\sqrt{k_1}$ is irreducible in R_1. Upon "joining" $\sqrt{k_1}$ to field R_1 we obtain a new intermediate field R_2 of numbers of the form

$$a_1 + b_1\sqrt{k_1}\,,$$

where a_1 and b_1 run through the entire set R_1 and k_1 is a fixed element of field R_1.

The proof is seen to be a repetition of the previous computations.

We note that

$$R_1 \subset R_2 ,$$

since for $b_1 = 0$ we have $a_1 + 0 \cdot \sqrt{k_1} \in R_1$. The numbers belonging to field R_2 can also be constructed with compasses and straight-edge.

This process of "extension" of a field by the "joining" of new elements can be indefinitely continued. Each new field will consist of elements which can be constructed with compasses and straight-edge.

If the number of given segments a, b, c, ... is more than 1, we can construct by means of the rational operations the minimal field K_o containing these segments and then by the "joining" of a new number $\sqrt{k_o}$, not belonging to field K_o (whereas $k_o \in K_o$), obtain a new "extension" of the field. We shall now give the pattern of such extensions:

1. To the rational field R we "join" all the given segments a, b, c, ... and thus obtain the field K_o.

2. To the field K_o we "join" the radical $\sqrt{k_o}$ irreducible in K_o, where $k_o \in K_o$, thus obtaining field K_1.

3. To the field K_1 we join the radical $\sqrt{k_1}$ irreducible in K_1, where $k_1 \in K_1$, thus obtaining field K_2.

. . . .

$n + 1$. To field K_{n-1} we "join" the radical $\sqrt{k_{n-1}}$ irreducible in K_{n-1}, where $k_{n-1} \in K_{n-1}$, thus obtaining the field K_n.

Starting from the given segments a, b, c, ..., it will be possible to construct any segment belonging to any such "extended" field.

The question naturally arises: do there exist any segments which can be constructed with compasses and straight-edge starting from the given segments a, b, c, ... but which nevertheless are not elements of any "extension" of the type we have described?

It turns out that every segment which can be constructed with compasses and straight-edge starting from the given segments a, b, c, ... belongs without exception to one of the indicated "extensions".

In order to verify this we shall show that there is no application of the compasses and straight-edge which is capable of taking us outside these "extension" fields. Suppose we are able to construct every number of some field K. Let us examine the application of the straight-edge.

If (x_1, y_1) and (x_2, y_2) are the coordinates of two points and if these coordinates belong to field K, the line drawn through these points will have the equation:

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1},$$

$$(y_2 - y_1)\, x + (x_1 - x_2)\, y + (x_2 y_1 - x_1 y_2) = 0.$$

The coefficients of this equation belong to field K.

If we seek now the point of intersection of two such lines

$$Ax + By + C = 0, \; A'x + B'y + C' = 0 ,$$

where the coefficients of the equations belong to field K, we shall have for the coordinates of this point :

$$x = \frac{C'B - B'C}{AB' - BA'} \text{ and } y = \frac{A'C - AC'}{AB' - A'B'},$$

that is, numbers belonging to the same field K.

The distance between such points, that is, the length of the segment determined by these points, is expressed by the quadratic radical

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

where the expression under the radical sign is a number belonging to field K. The construction of segment d requires the use of the compasses.

Each segment (number) which can be constructed with the straight-edge, starting from the field K, is in one of the previously indicated extensions of field K.

Let us examine the application of the compasses.

If (x_1,y_1) are the coordinates of the center of a circle and (x_2,y_2) the coordinates of one of its points and all these coordinates belong to field K, the equation of the circle will be:

$$(x - x_1)^2 + (y - y_1)^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 ,$$

or

$$x^2 + y^2 - 2x_1x - 2y_1y - x_2^2 + 2x_1x_2 - y_2^2 + 2y_1y_2 = 0 ,$$

that is, an equation with coefficients in the field K.

If we seek the point of intersection of the circles

$$x^2 + y^2 - ax - by + c = 0$$

and

$$x^2 + y^2 - a'x - b'y + c' = 0 ,$$

then subtracting the second equation from the first we obtain the equivalent system

$$x^2 + y^2 - ax - by + c = 0 ,$$

$$(a' - a) x + (b' - b) y + (c - c') = 0 .$$

These are the equations of a circle and a line.

Solving these equations, we obtain a number of the form

$$p + q\sqrt{s} ,$$

where p, q and s belong to the field K, consequently, the coordinates of the points of intersection belong to an extension of field K.

The distance between points having coordinates of this kind, even if they have different values of s, in turn, yields a segment belonging to some "extension" of field K, an assertion which it is not difficult to verify.

We have at the same time also examined the intersection of a circle with a line.

Thus, constructions with compasses and straight-edge cannot yield a segment which does not belong to one of the "extension" fields which can be obtained, proceeding from the originally given segments, in accordance with the procedure specified above.

We shall now prove a theorem concerning the roots of a cubic equation

$$x^3 + ax^2 + bx + c = 0$$

with rational coefficients a, b, c.

Theorem. If a cubic equation with rational coefficients does not have rational roots, then none of its roots can be constructed with compasses and straight-edge starting with the field R of rational numbers.

The proof is by contradiction. Let us assume that the root x of the indicated equation admits of construction with compasses and straight-edge. Then, as has been shown, x belongs to some field R_n, the last in a chain of "extensions" according to the above-mentioned pattern

$$R, R_1, R_2, \ldots, R_{n-1}, R_n .$$

We assume that $x \in R_n$, but $x \notin R_{n-1}$(1) and that none of the roots belongs to R_{n-1}, since if any of the roots belonged to R_{n-1} we would turn our attention to precisely that root and would shorten the chain of extensions. Furthermore, this chain of fields can not consist simply of the one field R since by assumption none of the roots is rational.

Thus, $n \geq 1$ is the least integer such that $x \in R_n$.

We may represent x in the form

$$x = p + q\sqrt{k_{n-1}},$$

where $k_{n-1} \in R_{n-1}$ and p and q likewise belong to field R_{n-1}, but $\sqrt{k_{n-1}}$ does not belong to field R_{n-1}.

We shall now show that if $x = p + q\sqrt{k_{n-1}}$ is a root of the cubic equation under consideration, then the number

$$y = p - q\sqrt{k_{n-1}},$$

belonging to field R_n, is likewise a root of this equation.

(1) Footnote: The symbol $\notin$ denotes "is not an element of".

Since $x \in R_n$, x^3 and x^2 also belong to this field; and consequently the number

$$x^3 + ax^2 + bx + c$$

belongs to this same field R_n. But this means that

$$x^3 + ax^2 + bx + c = r + s\sqrt{k_{n-1}},$$

where r and s belong to field R_{n-1}, while k_{n-1} has the same sense as before.

It is easily calculated that

$$y^3 + ay^2 + by + c = r - s\sqrt{k_{n-1}}.$$

Since x, by assumption, is a root of the cubic equation, we have

$$r + s\sqrt{k_{n-1}} = 0.$$

If s were not equal to zero, it would follow that $\sqrt{k_{n-1}} = -\frac{r}{s}$, and the radical $\sqrt{k_{n-1}}$ would belong to field R_{n-1}; but this is not true, whence $s = 0$, which means that $r = 0$ also.

Thus, if $x = p + q\sqrt{k_{n-1}}$ is a root of the cubic equation, $y = p - q\sqrt{k_{n-1}}$ also is seen to be a root.

These roots are distinct since if they were equal it would follow that

$$x - y = 2q\sqrt{k_{n-1}} = 0 \quad \text{or} \quad q = 0$$

and then x would be equal to $p \in R_{n-1}$, which is not true.

But the property of the roots of cubic equations

$$x_1 + x_2 + x_3 = -a.$$

Thus, the third root

$$x_3 = -a - x - y ,$$

or

$$x_3 = -a - 2p ,$$

that is $x_3 \in R_{n-1}$; but this contradicts the assumption we have made that n is the least number such that a field R_n of our chain contains a root of the cubic equation in question.

The contradiction thus reached proves the theorem.

We are now able to consider some famous problems of antiquity [22], [50].

The trisection of the angle. To divide any arbitrary angle into three equal parts.

We shall show that the problem can not be solved with compasses and straight-edge. We take note beforehand that given an angle it is easy to construct its cosine as a leg of a right triangle whose hypotenuse is equal to the unit length e, and conversely, given the cosine, we can construct the angle.

Let α be the given angle and a its cosine,

$$a = \cos \alpha .$$

It is required to construct $x = \cos \frac{\alpha}{3}$.

According to the familiar formula

$$\cos \alpha = 4 \cos^3 \frac{\alpha}{3} - 3 \cos \frac{\alpha}{3} , \qquad (*)$$

we set up the equation

$$a = 4x^3 - 3x$$

which will be satisfied by the cosine of the angle $\frac{\alpha}{3}$.

The trisection of the angle will be equivalent to constructing a root of the cubic equation (*). For which values of a ($|a| \leq 1$)

this equation will have no rational roots, is a question belonging to algebra. For our purposes it is sufficient to indicate even one example of an angle which it is <u>not</u> possible to divide into three equal parts with compasses and straight-edge. This example will be enough to show that <u>in the general case</u> the trisection of the angle with compasses and straight-edge is impossible.

Let $\alpha = 60^\circ$. Then $\cos \alpha = \frac{1}{2}$, and the cubic equation of the problem will take the form

$$8x^3 - 6x = 1 ,$$

We simplify the equation by the substitution of $z = 2x$:

$$z^3 - 3z = 1. \qquad (**)$$

If it turns out that this equation has no rational roots, then it will not be possible to construct its roots with compasses and straight-edge, and it will therefore be impossible to construct x.

Let us assume that equation (**) has a rational root $z = \frac{p}{q}$, where p and q are integers and prime to each other.

From this assumption it would follow that

$$p^3 - 3pq^2 = q^3 ,$$

that is, that q^3 is divisible by p. But this would mean that p and q have a common factor, providing only that $p \neq \pm 1$. In exactly the same way the number $p^3 = 3pq^2 + q^3$ would be divisible by q, and consequently p and q would have a common factor provided only that $q \neq \pm 1$.

But since p and q are prime to each other they can have no common divisor other than ± 1, and it must be concluded that p and q are equal to ± 1, that is, that $z = \pm 1$.

But $z = 1$ and $z = -1$ are not roots of equation (**). Equation (**) has no rational roots, whence it follows that an angle of 60° can not be divided into three equal parts with compasses and straight-edge, that is, with these instruments it is not possible to construct an angle of 20°.

From this there follows the important proposition: regular nine-sided polygons, eighteen-sided polygons and so on can not be constructed with compasses and straight-edge.

There exist, of course, angles which can be trisected with compasses and straight-edge, for example, the angles of 90°, 120° and so on. These are the angles for which equation (*) has rational roots.

If, in addition to compasses, we use a single-edged ruler upon which are marked two points A and B representing the length of a given segment AB, or a scale of length, then the trisection of any angle is possible with absolute theoretical precision. The rules for the use of such instruments must, of course, be formulated in the shape of appropriate basic problems. One of these, for example, is as follows: a line is considered constructed if it passes through a given or constructed point and if any two of its points A and B separated by the given distance lie on given or constructed lines.

We shall show how to trisect an arbitrary angle with compasses and a straight-edge with the interval AB marked upon it. Let $bOc = \alpha$ be an arbitrary angle (fig. 92). With a radius equal to the distance AB we describe a circle with center at O. We mark the point C of intersection of the circle with side c of the given angle and construct a line passing through point C in such a way that its points A and B lie respectively on the prolongation of side b of the angle and on the constructed circle. Angle BAO is then the required one-third of the given angle bOc.

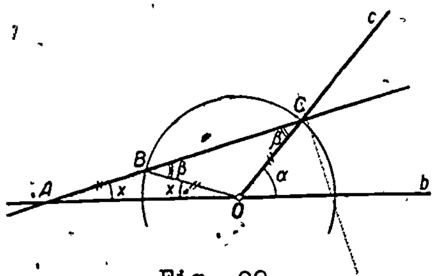


Fig. 92.

Proof. Joining B with O, we have AB = OB = OC = a . From the isosceles triangles and by the theorem on the exterior angle we obtain $\alpha = \beta + x$; $\beta = 2x$; and consequently $x = \frac{\alpha}{3}$, q.e.d.

Archimedes is considered to be the author of this solution. [50]

Let us turn to a second famous problem of antiquity:

Duplication of the cube. Given a cube with edge a. To construct the edge of the cube whose volume will be twice that of the given cube. We shall show that the problem cannot be solved with compasses and straight-edge. The solution reduces to that of the equation

$$x^3 = 2a^3,$$

where x is the edge of the required cube.

Taking the side of the given cube as unity we arrive at the cubic equation

$$x^3 - 2 = 0$$

having rational coefficients. This equation, as can readily be shown, has no rational roots; consequently, it is impossible to construct its roots with compass and straight-edge starting from the segment $a = 1$, q.e.d.

Let us examine a mathematically rigorous solution of the problem with the use of two movable right angles. As a preliminary step we shall prove a lemma on the rectangular trapezoid with perpendicular diagonals (fig. 93): in every rectangular trapezoid with perpendicular diagonals the segments of the diagonals form a geometric progression:

$$\frac{a}{y} = \frac{y}{x} = \frac{x}{b}.$$

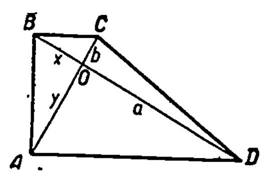


Fig. 93.

In the trapezoid ABCD let the diagonals DB and AC be perpendicular to each other and let the angles A and B be right angles. The remaining designations are shown in fig. 93. From the right triangle ABD with altitude AO we have $\frac{a}{y} = \frac{y}{x}$, and in the same way from the right triangle ABC with altitude x we have $\frac{y}{x} = \frac{x}{b}$, q.e.d.

From the proportions $\frac{a}{y} = \frac{y}{x} = \frac{x}{b}$ we obtain:

$$y^2 = ax \text{ and } x^2 = by, \qquad (*)$$

whence

$$x = \sqrt[3]{ab^2} \text{ and } y = \sqrt[3]{a^2b} .$$

Setting $a = 2$, $b = 1$ we find $x = \sqrt[3]{2}$, that is, a solution of the equation

$$x^3 - 2 = 0 .$$

From the lemma we obtain a method for constructing the segment x with the aid of two right angles.

On two mutually perpendicular lines r and s we lay off from their point of intersection O the segments $OC = b = 1$ and $OD = a = 2$ (fig. 94). We place the movable right angles so that they form a rectangular trapezoid $ABCD$ with lines r and s as its diagonals. Then the segment $OB = x$ is the required segment.

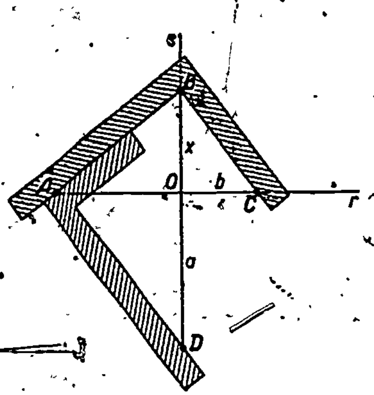


Fig. 94.

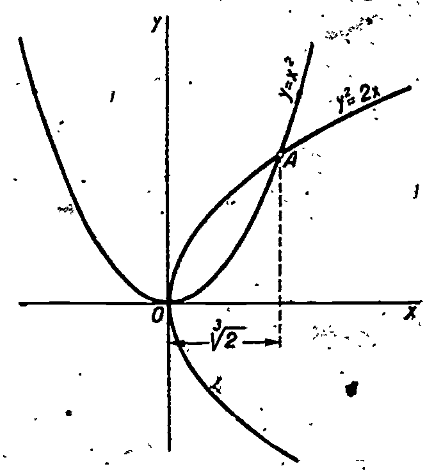


Fig. 95.

The solution of the problem is ascribed to Plato [50].

The equations (*) show also that the required segment occurs as the abscissa of the point of intersection A of two parabolas (fig. 95) having the equations

$$y^2 = 2x \text{ and } x^2 = y .$$

With the aid of two movable right angles it is possible to construct the roots of any equations of the third or fourth degree with rational coefficients [2]. With this we end our study of the problem of duplicating the cube.

The regular seven-sided polygon. To construct a regular seven-sided polygon.

This problem is also not solvable with compasses and straightedge.

We shall prove this. The problem of constructing a regular polygon is, as we know, equivalent to the problem of extracting the corresponding root of unity, or -- what is the same thing -- solving the binomial equation

$$z^n - 1 = 0.$$

In the present case $n = 7$. One root is equal to unity [33]; the remaining roots satisfy the equation

$$z^6 + z^5 + z^4 + z^3 + z^2 + z + 1 = 0 .$$

This is an equation with reciprocal roots. Dividing both sides of the equation by z^3, we obtain:

$$z^3 + \frac{1}{z^3} + z^2 + \frac{1}{z^2} + z + \frac{1}{z} + 1 = 0 .$$

Setting $z + \frac{1}{z} = y$ we reduce the equation to the form

$$y^3 + y^2 - 2y - 1 = 0 . \qquad (\alpha)$$

The seventh root of unity is given by the formula

$$z = \cos \phi + i \sin \phi ,$$

where $\phi = \frac{2k\pi}{7}$ with $k = 0, 1, 2, \ldots, 6$. For us the important value is $k = 1$. We have furthermore

$$\frac{1}{z} = \cos\phi - i \sin\phi$$

and

$$y = z + \frac{1}{z} = 2\cos\phi .$$

Finding the cosine of the required angle has been reduced to the finding of the roots of the cubic equation (α) with rational coefficients.

We shall show that equation (α) does not have rational roots. Let us suppose the contrary, that is, that equation (α) has the rational root $y = \frac{p}{q}$, where p and q are integers ($q \neq 0$) and prime to each other. Setting $y = \frac{p}{q}$ in equation (α) we obtain:

$$p^3 + p^2q - 2pq^3 - q^3 = 0 .$$

From this it follows that p^3 is divisible by q, and q^3 by p. By virtue of the fact that p and q are prime to each other, we find $p = \pm 1$ and $q = \pm 1$, that is, the supposed root y is equal to ± 1. Substitution in equation (α) shows that ± 1 are not among its roots.

Equation (α) has no rational roots; consequently it is impossible to construct a regular seven-sided polygon with compasses and straight-edge, q.e.d.

We have also seen that it is impossible to construct with compasses and straight-edge a regular nine-sided polygon, nor, of course, any of the polygons having 2^n times this number of sides.

Gauss proved the <u>theorem</u> that with compasses and straight-edge it is possible to construct those and only those regular polygons the number of whose sides has the form

$$n = 2^m \cdot p_1 p_2 \ldots p_s ,$$

where the distinct odd prime numbers $p_1, p_2, \ldots, p_s$ enter as factors in the first degree and each of them has the form

$$p_i = 2^{2^\lambda} + 1 .$$

The values $\lambda = 0, 1, 2, 3, 4$ actually yield prime numbers for p_i, namely 3, 5, 17, 257 and 65,537.

But $2^{2^5} + 1$ has the divisor 641 (discovered by Euler) and is, therefore, composite.

The well-known Russian mathematician Pervushin discovered that for $\lambda = 12$ 114,689 is a divisor and for $\lambda = 23$ 167,772,161 is a divisor.

The divisors were found with the aid of extremely subtle theoretical reasoning. In view of the immensity of the numbers

$$2^{2^{12}} + 1 \text{ and } 2^{2^{23}} + 1$$

the mere writing out in full of these numbers demands a colossal amount of time. In the well-known problem of grains on a chessboard we have only to deal with the number $2^{64} - 1$.

For which sufficiently large values of λ additional composite numbers are obtained, and whether or not among the numbers of the form $2^{2^\lambda} + 1$ there exists an infinite set of primes, is not known.

With these remarks we conclude our survey of famous problems of antiquity, to which belongs also the problem of squaring the circle, that is, the problem of constructing with compasses and straight-edge the side x of a square equivalent to a circle of given radius r. The problem reduces to the solution of the equation $x^2 = \pi r^2$, or, if we set $r = 1$, to the construction of the segment

$$x = \sqrt{\pi},$$

which is equivalent to the construction of the number π with compasses and straight-edge. This construction is impossible, but the proof of that fact belongs to mathematical analysis [27], [44].

Chapter III

THE TRANSFORMATION OF FIGURES

In Chapter III we examine the idea of functional dependence in geometry and study the elementary properties of mappings and transformations of figures; in particular, we encounter the concept of transformations of a plane into itself and of space into itself. We consider some special transformations of a plane into a sphere and of a plane into a plane. The properties of these speial transformations find their application in the theory of geometrical constructions, and also in the study of the geometry of Lobachevskii. At the end of the chapter we present the concept of a group of transformations and the general definition of a group.

13. THE MAPPING OF ONE FIGURE INTO ANOTHER

If to each point A of a figure ϕ we assign by means of some rule or law a definite point B in some figure ψ, we then say that the figure ϕ is mapped into the figure ψ, and we write $B = f(A)$.

Point B is called the image of point A under the given mapping. We commonly speak of the value B of the function f corresponding to the value A of the argument, the argument being understood to run through all points of the figure ϕ. This is in accord with the general definition of function given first by Lobachevskii (1834) and subsequently by Dirichlet (1837).

The totality ϕ' of all image points which are assigned by the mapping f to the points of the figure ϕ is called the image of the figure ϕ under the mapping f.

This image of the figure is denoted by $f(\phi)$, that is,

$$\phi' = f(\phi).$$

An important special case of the mapping of a figure ϕ into a figure ψ is that in which <u>every</u> point in figure ψ is the image of at least one point in figure ϕ.

When this is the case we say that figure ϕ is mapped <u>onto</u> figure ψ. A figure ϕ is thus always mapped <u>onto</u> its own image $f(\phi)$.

Let us consider some examples in order to familiarize ourselves with these simple but extremely important concepts.

Let there be given two segments, PQ and KL (fig. 96).

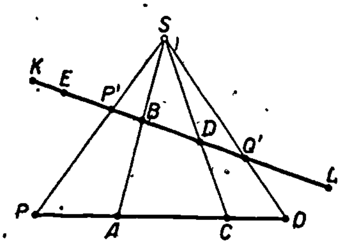


Fig. 96

We select a point S as indicated in the diagram, and assign to each point A of segment PQ that point B of segment KL which lies on the ray SA. In particular, the points P' and Q' will be assigned to the end points P and Q of segment PQ. In this manner the segment PQ is mapped <u>into</u> the segment KL.

Point B is the image of point A; points D, P', Q' are respectively the images of points C, P and Q.

We may express this as: $B = f(A)$; $D = f(C)$; $P' = f(P)$; $Q' = f(Q)$, where f is the symbol of the given mapping.

The image of the entire segment PQ is the segment P'Q', namely: P'Q' = f(PQ).

We emphasize once more that segment PQ is mapped into segment KL; we cannot say in this case that PQ is mapped onto KL since, for example, points E, K and L are not the images of any points whatever of the segment PQ. It is, of course, proper to say that PQ is mapped into P'Q'.

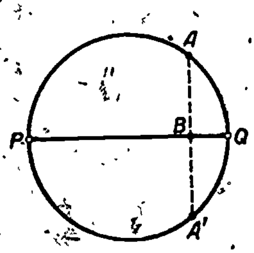


Fig. 97

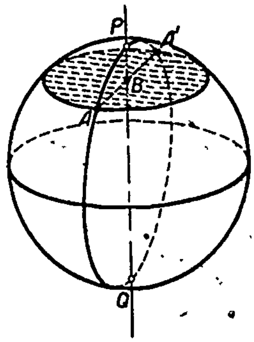


Fig. 98

In this example we have encountered all the concepts which we have introduced concerning the mapping of figures; but it cannot be expected that a single example will bring out all the features of a general concept. If instead of segment KL we should take the point S and assign this point to every point of segment PQ, we should have a case of the mapping of one figure into another.

Since every point S of the second figure (which consists of the point S) is the image of some point of the first figure (segment PQ), in this case segment PQ is mapped not only into the point S but also onto the point S.

The last example brings out the fact that the images of distinct points may coincide; but it must be emphasized that no point A may have under a mapping more than one image B.

Let us map a circle into its diameter, assigning to each point A of the circle the point B at the foot of the perpendicular drawn from A to the diameter (fig. 97). Here the circle is mapped not only into but also onto the diameter. Each interior point B of the diameter is the image of two points: $B = f(A)$ and $B = f(A')$.

If by the same method a sphere is mapped onto its diameter PQ (fig. 98), each interior point of the diameter is the image of all the points of a parallel (considering P and Q as poles).

In exactly the same way the whole of a solid sphere can be mapped onto its diameter, assigning to each point A' of the solid sphere the point B at the foot of the perpendicular A'B upon its diameter PQ; each interior point of the diameter will be the image of an entire disk (fig. 98). Each point on the diameter PQ will be mapped into itself.

If to every point A of a torus is assigned the point B at the foot of the perpendicular AB to the axis of revolution ℓ, the torus will be mapped into the line ℓ; here we cannot say "onto ℓ". The point B is, in general, the image of a pair of circles. The image of the whole torus is a segment of the line ℓ. The torus is mapped onto this segment (and likewise into it).

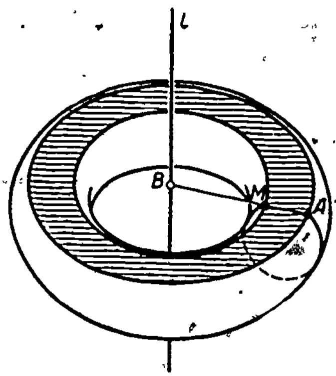


Fig. 99

If the solid bounded by a torus is in the same way mapped into the line ℓ, the point B will, in general, be the image of an entire annulus (fig. 99).

The image of the whole solid is again the same segment of ℓ. The solid in question is mapped <u>onto</u> this segment (and also, of course, <u>into</u> it).

To each point A of the plane π (fig. 100) let us assign that point B of the sphere Σ which lies on the line SA, where S is the north pole of the sphere, and the plane π is tangent to the sphere Σ at its south pole. By this correspondence, called <u>stereographic projection</u>, the plane π is mapped <u>into</u> the sphere Σ.

We cannot say that the plane is mapped <u>onto</u> the sphere, since the point S of the sphere is not the image of any point of the plane. The image of the plane is a sphere "perforated" at the point S. Plane π is mapped <u>onto</u> its image -- the sphere perforated at the point S. And, of course, <u>into</u> its image also.

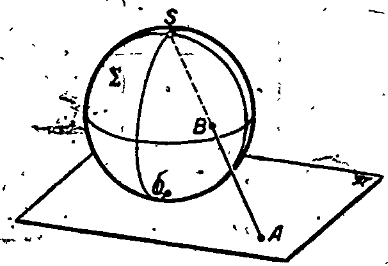


Fig. 100

If to every point M of the parabola (fig. 33) is assigned that point A of the circle which lies on the generator SM of the cone, we obtain a mapping of the parabola into the circle. Here again we cannot say that the parabola is mapped onto the circle. The point P on the circle is not the image of any point on the parabola, since the generator SP is parallel to the plane π . The image of the parabola is the circle with the point P excluded. The parabola is mapped onto as well as into its own image.

Assigning in exactly the same way to each point M of the hyperbola (fig. 31) the corresponding point B′ on the circle of tangency of sphere (2), we map the hyperbola into but not onto the circle.

In this case two points of the circle, lying on the two generators of the cone which are parallel to the plane π , are not images of any point of the hyperbola. The image of the hyperbola is the circle, with the exclusion of the two points on the generators of the cone which are parallel to the plane π . The hyperbola is mapped onto its own image.

If to each point M of the ellipse we assign a point A of the circle as indicated in figures 27 and 29, we obtain a mapping of the ellipse onto, and of course likewise into, the circle. The image of the ellipse will be the circle in question.

We shall use some of these examples to illustrate the following important definition:

Definition: Let there be given a mapping f of the figure ϕ onto the figure ψ, and let B be an arbitrary point of the figure ψ. The set of all those points of the figure ϕ to which under the mapping f there is assigned the given point B is called the inverse image of the point B. This set (or figure) is denoted by $f^{-1}(B)$.

The inverse image of the point B under the mapping of the plane π into the sphere punctured at the point S (fig. 100) is the point A; $f^{-1}(B) = A$.

The inverse image of the interior point B of the diameter of the circle under the mapping of the circle onto its diameter (fig. 97) is the zero-dimensional segment AA'. The inverse image of either end-point of the diameter is that point itself. We may write $f^{-1}(B) = (AA')^{o}$, where the sign ^{o} indicates the dimension of the segment AA'.

Under the mapping of a sphere onto its diameter PQ previously discussed (fig. 98), the inverse image of an interior point B of the diameter PQ is the circumference of the parallel whose plane passes through point B. The inverse images of the poles are the poles themselves.

Under the mapping of a solid sphere onto its diameter (fig. 98), the inverse image of each interior point of the diameter is the disk bounded by the parallel whose plane passes through that point of the diameter.

In the case of the mapping of the torus onto its image (fig. 99) the inverse image of the point B is the pair of circles of the torus, the plane of which passes through the point B.

In the mapping of the solid bounded by a torus into its axis ℓ the inverse image of the point B will be the closed annulus which is shaded in figure 99.

14. THE ONE-TO-ONE MAPPING OF FIGURES

If under a mapping f of the figure ϕ onto the figure ψ the inverse image $f^{-1}(B)$ of each point B in figure ψ consists of only one point A in figure ϕ, then the mapping of figure ϕ onto figure ψ is one-to-one.

We shall consider some examples illustrating this definition.

The mapping of the plane onto the sphere perforated at S is a one-to-one mapping (fig. 100). But the mapping of the plane into the entire sphere is not one-to-one. One of the prerequisites of a one-to-one mapping is that one figure be mapped onto the other.

The mapping of a solid sphere onto its diameter (fig. 98) is not one-to-one because the inverse image of an interior point of the diameter is not a single point of the solid sphere, but consists of an entire disk.

The mappings of an ellipse on a circle in the manner described in connection with figures 27 and 29 are one-to-one mappings.

The previously discussed mappings of the hyperbola onto its image -- the circle with the exclusion of two points (fig. 31) -- and of the parabola onto its image -- the circle with the exclusion of one point (fig. 33) -- are likewise one-to-one mappings. The mappings of the hyperbola and parabola into the full circle are not one-to-one.

Corresponding to the mapping of the ellipse onto the circle (fig. 29), the points interior to the ellipse can be mapped one-to-one onto the points interior to the circle. To each point L interior to the ellipse we can assign that interior point N of the disk which lies on the ray SL. The analogous procedure may be carried out in the cases of the parabola and hyperbola (figs. 33 and 31).

Before going on to other examples, let us note the following general conclusion:

A one-to-one mapping f of figure ϕ onto figure ψ automatically generates a one-to-one mapping f^{-1} of figure ψ onto figure ϕ.

This follows from the fact that the inverse image $f^{-1}(B)$ of each point B of figure ψ is a single point A of figure ϕ, whence it is plain that f^{-1} is a mapping and that it is one-to-one, since $A = f^{-1}(B)$, and the inverse image of the point A under this mapping f^{-1} of figure ψ onto figure ϕ, namely f(A), consists of only a single point.

Under a one-to-one mapping of figure ϕ onto figure ψ each point A of figure ϕ is paired with a definite point $f(A) = B$ of figure ψ; whence it is seen that each point B of the figure ψ is paired with a single and completely determined point A of figure ϕ. This pairwise association is clearly evident in the examples discussed.

The mapping f^{-1} is called the inverse mapping relative to the mapping f. It is evident that the mapping inverse to f^{-1} is the mapping f.

In view of the symmetrical nature of these one-to-one mappings of one figure onto another, we speak, in such a case, of one-to-one correspondence between the two figures.

We shall consider some examples.

Let us perform a one-to-one mapping of the one-dimensional triangle ABC onto the circumscribed circle (fig. 101). We assign to each point M of the triangle the point N of the circle by drawing through M the radius OMN. The inverse mapping of the circle onto the one-dimensional triangle assigns to point N precisely the point M of the one-dimensional triangle. There takes place an association of the points of the two figures into pairs (M,N). The vertices of the triangle correspond to themselves. We have obtained a one-to-one correspondence between the two figures.

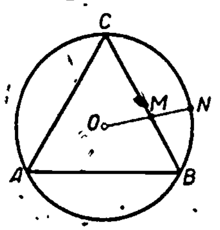


Fig. 101.

If one-to-one correspondence can be established in some manner between two sets, the sets are said to be equivalent or equipotent ([36] and [40]). The greatest possible power of a figure is the continuum; to it corresponds the transfinite number ℵ. The set of points of all of space as well as the set of points of a plane or of a line all have one and the same power ℵ. Every figure is either a finite or countable set of points or it possesses the power of the continuum. We shall not stop to prove this assertion here ([35], [43]). It is for this reason that the most general one-to-one correspondences are of little interest to geometry. Important to geometry is a certain limited class of one-to-one correspondences, which will be discussed in the following section.

Let the diameter PQ of the open semicircle PMQ be parallel to the line a (fig. 102). By means of the ray OA let us map the line a onto the open semicircle; we have $M = f_1(A)$. Further, let us

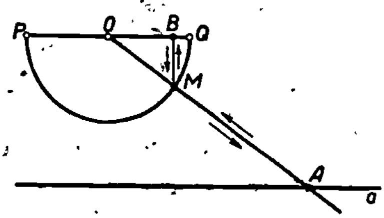


Fig. 102.

map this semicircle onto the interval PQ, drawing from each point M a perpendicular MB to PQ; we have $B = f_2(M)$. By this method we obtain an association of the points of line a and those of the interval PQ into pairs (A,B). In so doing we have established a one-to-one correspondence between the line and the interval.

By an analogous procedure a one-to-one correspondence can be established between a plane and an open hemisphere and then between the plane and a disk minus its boundary (fig. 103).

It is likewise easy to establish a one-to-one correspondence between an open disk and a perforated sphere. For this purpose we map the open equatorial disk onto the plane tangent to the sphere at the south pole (fig. 103) and then, by the method of stereographic projection, we map the plane onto the perforated sphere.

Finally, let us establish a one-to-one correspondence between one segment and another. We shall deliberately do this in a "complicated" and "unnatural" way. The point, of course, is to show that one-to-one mappings are not necessarily simple or natural. If we wish to study only simple and "reasonable" mappings, we must impose some further restrictions beyond merely one-to-oneness.

Let there be given the segment PQ and the interval P'Q', for simplicity equal to PQ, parallel to each other and so positioned that PP'Q'Q is a rectangle (fig. 104).

We take on the segment PQ a sequence of points 1,2,3,4,... such that each successive point shall be one-half as far from a given point O as the preceding one. In exactly the same manner we choose a sequence of points on the interval P'Q'. In the diagram we have chosen for the point O the midpoint of the segment PQ.

We now construct the mapping f of the segment PQ onto the interval P'Q' as follows. Each point A of the segment PQ different from the points of the chosen sequence is mapped into a point A' of the interval P'Q' by dropping a perpendicular AA' upon the interval P'Q'; we have:

$A' = f(A)$, $O' = f(O)$, etc.

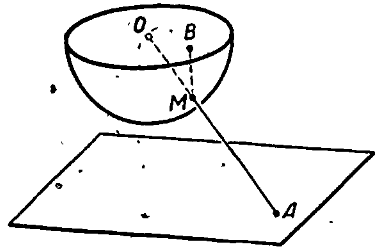


Fig. 103.

The points of the sequence 1,2,3,... of the segment PQ are mapped into the points 3', 4', 5',..., of the other sequence:

$3' = f(1)$, $4' = f(2)$,...., $n' = f(n - 2)$,...

We complete the construction of the mapping by assigning to the points P and Q of the segment PQ the points 2' and 1' of the interval P'Q':

$2' = f(P)$, $1' = f(Q)$.

The construction of the one-to-one mapping of the segment onto the interval is complete. We think the reader will agree that this mapping is neither "simple" nor "natural." It is, however, one-to-one.

Evidently, one-to-oneness is not sufficient to guarantee that a mapping will be a "simple" or "natural" one.

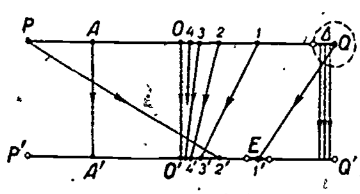


Fig. 104.

In exactly the same manner one may construct a one-to-one correspondence between a disk and an open disk, and between a solid sphere and an open solid sphere. The establishment of these correspondences is suggested as an exercise. It is also useful to establish a one-to-one correspondence between a complete sphere and a plane; between a plane and a plane minus an open or a closed disk; etc.

It is possible to establish a one-to-one correspondence even between figures of different dimensions, for example, between a segment and all of space.

15. CONTINUOUS MAPPINGS OF FIGURES

We introduce the following definitions: By a neighborhood of a point A relative to space, or simply a neighborhood of a point A we mean any open solid sphere with its center at point A. The radius of this sphere is called the radius of the neighborhood. Accordingly, the set of all concentric open solid spheres with center A is the set of all neighborhoods of that point.

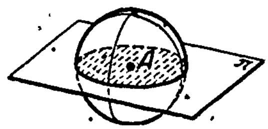


Fig. 105

By a neighborhood of point A of a figure ϕ relative to that figure we mean the intersection of the figure with any open solid sphere having center A. Thus a neighborhood of point A in plane π relative to the plane π will be the intersection of this plane with any open solid sphere having center A, i.e. any open disk of plane π having center A (fig. 105). A neighborhood of point A on the line a relative to this line will be any interval with its midpoint at A (fig. 106).

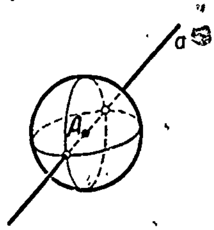


Fig. 106

A neighborhood of point A of a sphere relative to that sphere will be any open spherical disk (fig. 107), i.e. the set of points on the given sphere which lie within any open solid sphere having center A.

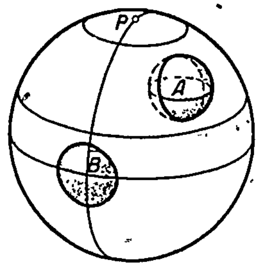


Fig. 107

The sphere perforated at a point diametrically opposite to point A of the said sphere, as well as the complete sphere, are spherical neighborhoods of the point A. The former lies within the open solid sphere with center A and radius equal to the diameter of the given sphere, while the latter lies within a solid sphere with radius greater than the diameter of the given sphere. The diagram also indicates spherical neighborhoods of points B and P.

The neighborhood of a point A of a circle relative to that circle is an arc of the given circle with A as its midpoint. Other neighborhoods, in this case, are constituted by the circle minus a point diametrically opposite A, and by the entire given circle. In fig. 108 dashed lines indicate those open solid spheres, the intersections with which give the respective neighborhoods of point A relative to the given circle.

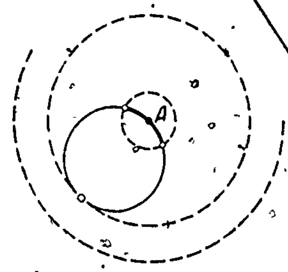


Fig. 108

It is necessary to distinguish, for example, a neighborhood of the point A, an end-point of the segment AB, relative to this segment (fig. 109a), relative to the line AB (fig. 109b), relative to the plane π containing the line (fig. 109c), and relative to all of space (fig. 109d).

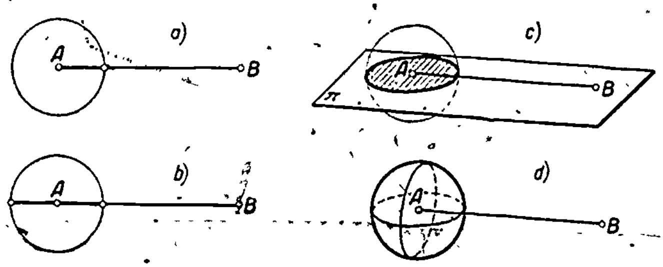


Fig. 109

After having elucidated the concepts <u>neighborhood of a point</u> and <u>relative neighborhood</u>, we proceed to the definition of one of the most important concepts of geometry, that of <u>continuous mapping</u>. Let f be a mapping of figure ϕ into figure ψ, and let A_0 be a point of figure ϕ. Then the mapping

f is called continuous at the point A_0 if for every neighborhood E of the point $f(A_0)$ of figure ψ relative to ψ there is always to be found a neighborhood Δ of the point A_0 relative to ϕ such that the entire neighborhood Δ is mapped into the neighborhood E, that is

$$f(\Delta) \subset E.$$

If the mapping is continuous at all points of figure ϕ, it is called simply a continuous mapping of figure ϕ.

Let us consider some illustrative examples.

Let a plane be mapped onto the line a belonging to it so that each point A of the plane is mapped into the foot B of the perpendicular AB to the line a (fig. 110).

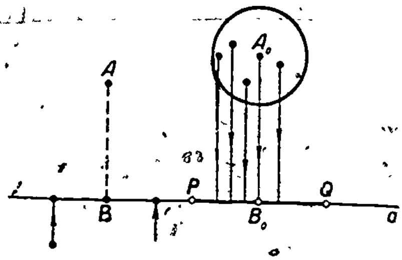


Fig. 110

The points of line a are mapped into themselves. We shall prove that this mapping f of the plane onto the line a is continuous. Let $f(A_0) = B_0$. We select an arbitrary neighborhood of B_0 relative to the line a. This will be some interval PQ. If we select a neighborhood of point A_0 relative to the plane, i.e. an open disk, such that its radius is not greater than the radius of interval PQ, then the whole of this neighborhood is mapped into the interval PQ. Since A can be any point of the plane, the mapping is continuous.

The mappings previously discussed: of the segment PQ into the segment KL (fig. 96), of a circle onto its diameter (fig. 97), of a sphere and likewise of a solid sphere onto its diameter PQ (fig. 98), of a torus and of the solid which it bounds onto its image, a segment of its axis ℓ (fig. 99), -- are all continuous mappings. In exactly the same way the stereographic mapping of a plane into a sphere (fig. 100) is a continuous mapping.

In all these cases, for an arbitrary relative neighborhood E of any point f(A) we can readily find a relative neighborhood $\triangle$ of point A such that

$$f(\triangle) \subset E.$$

It is suggested that the student verify this for himself.

By way of a model, we shall consider the case of the mapping of a one-dimensional triangle ABC onto the circumscribed circle (fig. 101). Let us find for any given neighborhood E a neighborhood $\triangle$ such that $f(\triangle) \subset E$.

Let M be an arbitrary point of the triangle and let N = f(M). We select an arbitrary neighborhood E of the point N relative to the circle (fig. 111). This will be an open arc PQ. Let us join the points P and Q with the center O and mark the points P' and Q' where the segments OP and OQ intersect the triangle.

Any interval of the segment P'Q' with its midpoint at M may be taken as the $\triangle$-neighborhood; then

$$f(\triangle) \subset E.$$

If the neighborhood E of the point N is the circle with the point diametrically opposite N excluded, or if it is the entire circle, then any neighborhood of the point M relative to the triangle which does not contain the point diametrically opposite to M -- in the second case even the entire one-dimensional triangle -- may be taken as a $\triangle$-neighborhood of point M.

If the mapped point be a vertex A and the radius of the E-neighborhood is sufficiently small we may take as the $\triangle$-neighborhood relative to the triangle the intersection of the triangle with the same open solid sphere which generated the E-neighborhood relative to the circle; in the given case (fig., 111) $f(\triangle) \subset E$.

This one-to-one mapping of a triangle onto a circle is also continuous in the inverse direction, i.e., the mapping f^{-1} is also continuous.

We can readily convince ourselves of this since with reference to the mapping f^{-1}, we can find for every given neighborhood of the triangle a neighborhood of the circle which is mapped completely into the given neighborhood of the triangle.

We give the following general definition:

Definition: If the mapping f is one-to-one, continuous and such that the mapping f^{-1} inverse to it is also continuous, then f is called a topological mapping. We might also express this by saying: a mapping is topological if it is one-to-one and bicontinuous.

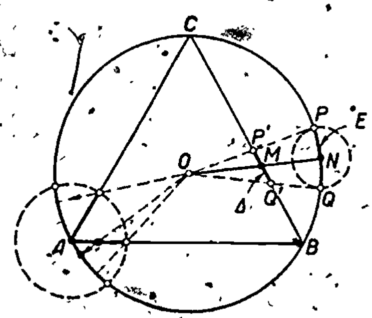


Fig. 111

In the foregoing example we were dealing with a topological mapping of a one-dimensional triangle onto a circle and, inversely, the topological mapping of the circle onto the one-dimensional triangle.

The stereographic mapping of a plane onto a perforated sphere (fig. 100) (and vice-versa) is a topological mapping.

The mapping of a sphere onto its diameter (fig. 98) is not one-to-one, and consequently not topological.

The "unnatural" mapping of the segment PQ onto the interval P'Q' (fig. 104), although one-to-one, is not continuous, even in one direction. This is true because if we take, for example, a sufficiently small neighborhood E of the point 1', then in any neighborhood $\triangle$ of the point $Q = f^{-1}(1')$ we obtain points which are not mapped into the E-neighborhood.

A sphere may be mapped onto a plane one-to-one but not topologically, similarly a segment cannot be mapped onto an interval topologically, i.e. one-to-one and continuously in both directions. This assertion we shall leave without proof. We shall point out also without proof that it is impossible to establish a topological correspondence between figures of different dimensionality, for example between a segment and a cube or between a plane and a circle.

Two figures which can be topologically mapped one upon the other are called topologically equivalent, or homeomorphic.

Thus, it will easily be seen that all circles are homeomorphic with each other and topologically equivalent to any one-dimensional triangle. The torus and the sphere are topologically different, i.e. not homeomorphic, a fact whose proof we omit.

Topological properties are those properties of figures which are preserved, or as we say are invariant, under all topological mappings; in other words, the topological properties of a figure ϕ are those properties which belong not only to figure ϕ but to every topologically equivalent figure ψ. For example, the property of a figure of being a curve, of being a surface, of being a solid, or of being a closed curve -- such as a circle or a one-dimensional triangle -- all these are topological properties.

If a figure ϕ is mapped onto a figure ψ, we also say that the figure ϕ is transformed into the figure ψ. If the figure ϕ is topologically mapped onto the figure ψ, we say that figure ψ is obtained from figure ϕ by a topological transformation.

With the aid of the concepts just studied, it is possible to define rigorously such concepts as an arc of a curve, a closed curve, etc.

A simple arc is a topological image of a segment, a closed simple arc is a topological image of a one-dimensional triangle, etc.

The majority of mappings encountered in geometry are topological mappings.

The specifically topological properties of figures and the properties of topological transformations as such are the subject matter of "the geometry of the continuous" -- topology ([6], [7]).

For elementary geometry we require only the very simple topological concepts which have been set forth above.

16. THE CONCEPT OF THE TRANSFORMATION OF A PLANE INTO ITSELF AND OF SPACE INTO ITSELF

In mapping a figure ϕ onto a figure ψ it is not necessary that figure ψ be distinct from figure ϕ. If figure ψ coincides with figure ϕ we have the mapping of figure ϕ into itself or, in particular, the mapping of figure ϕ onto itself.

In the case of a mapping of a figure ϕ into itself or onto itself we often speak of the transformation of figure ϕ into itself making no distinction between into and onto; the meaning is always clear from the context.

In particular, we speak of the transformation of a plane into itself and of the transformation of space into itself.

The following examples illustrate this general definition.

The mapping of a solid sphere upon its diameter (fig. 98) is at the same time also a mapping of the solid sphere into itself. The diameter PQ will be the complete image of the solid sphere. This mapping of the solid sphere into itself is one of the transformations of the solid sphere into itself.

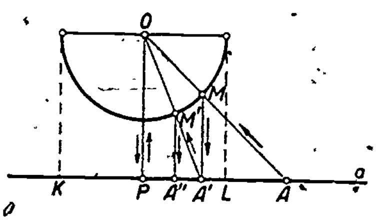


Fig. 112

If we map the line a onto an open semicircle and then by perpendicular projection map this semicircle back into the line a, we obtain the transformation f of line a into itself (fig. 112), under which $f(A) = A'$; $f(A') = A''$,... and $f(P) = P$. The point P is a fixed point of this transformation. The complete image of the line a will be the interval KL. This transformation is topological, as can easily be verified.

Proceeding analogously we can construct a <u>transformation of a plane π into itself</u> (fig. 113). The complete image of the plane will be an open disk of this plane. Such a transformation of a plane into itself is likewise topological.

It is necessary to keep in mind that under the indicated transformation of the plane into itself every figure in plane π is transformed into some figure in that same plane. It is easy to see, for example, that a circle with center at P will be transformed into another circle with the same center. Lines passing through P will be transformed into themselves; the complete image of each such line will be an interval, as is evident from fig. 112. A line not passing through P will be transformed into an arc of some curve. A circle having its center not at P will be transformed into some closed curve; the disk bounded by this circle, into a region of the plane bounded by the image of the circle; etc.

<u>A transformation of a plane into itself induces a transformation of every figure of this plane.</u>

Let us construct a transformation of the plane π into itself. We draw in the plane π the equilateral one-dimensional triangle ABC and its circumscribed circle (fig. 101). Let every point M of the one-dimensional triangle be transformed into a point N of the circle, and let every point N of the circle be taken back into the corresponding point M of the triangle; all other points remain in their places, that is, are mapped into themselves. We then have a transformation of the plane into itself (and in fact onto itself) which is not topological.

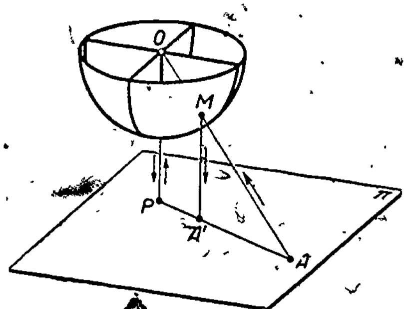


Fig. 113

This transformation of the plane into itself induces a transformation of every figure in the plane π. Thus, triangle ABC is topologically transformed into the circumscribed circle. The circle also is transformed topologically into the triangle. Every figure not containing points of the triangle or circle other than A, B and C remains unchanged and, consequently, its transformation is topological.

Analogously we may construct a transformation of space into itself by circumscribing a sphere about a two-dimensional regular tetrahedron, transforming points of the tetrahedron and sphere lying on the same radius into each other, and leaving all other points in their places.

The topological transformation of the one-dimensional triangle into the circumscribed circle can be extended continuously to the entire plane with the aid of the following procedure. The point O goes into itself, while to every point S distinct from O in the plane π we assign its image S' such that S' lies on the ray OS, and

$$\frac{OS'}{OS} = \frac{ON}{OM}.$$

This determines uniquely the point S' and thus we have constructed a transformation of the plane into itself which is, moreover, topological, as can readily be demonstrated. This mapping of the plane into itself (and in fact onto itself) induces a topological transformation of every figure of this plane.

A ray issuing from point O undergoes a homothetic transformation, but each ray has its own coefficient of similarity $k = \frac{ON}{OM}$. Points on the rays OA, OB and OC remain fixed since for them k = 1. In figure 114 is shown the transformation of the triangle PQR into the figure P'Q'R' and the transformation of a circle.

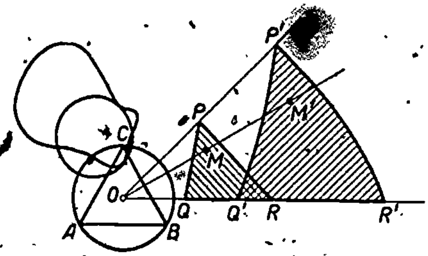


Fig. 114

The triangle ABC is transformed into its circumscribed circle. It is suggested as an exercise that the student draw the figure into which this circle itself is in turn transformed.

An analog of this transformation can be constructed in space with the aid of a regular tetrahedron and its circumscribed sphere.

The transformation of space into itself (and in fact onto itself) thus obtained will be a topological one, as may be readily verified.

Let us recall once more that properties of figures which are preserved under all transformations of a given class are called invariant properties of these figures (relative to that class of transformations). An invariant property of a circle, for example, under topological transformations will be the property of being a closed curve without self-intersection. A break in the topological image is excluded by virtue of the continuity of the transformation, and self-intersection is excluded by virtue of its one-to-one character. If we imagine a circle formed of rubber thread capable of infinite elasticity and flexibility, then by stretching and bending this circle as we please we obtain its topological images.

Such transformations do not preserve, for example, length, convexity, or the property of being a plane figure.

In general, under topological transformations, lengths of lines and angles between lines are not preserved. Nor can we speak of the preservation of area either, since a plane figure may be transformed into a very intricate spatial one, and so on.

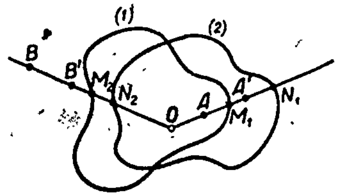


Fig. 115

It is not difficult to find any number of examples of transformations of the plane into itself and of space into itself constructed in the manner just examined. We can take any two curves (1) and (2) (in space, two surfaces) which intersect each ray issuing from a given point O in only one point (fig. 115) and proceed analogously with the foregoing.

That is to say, we transform every point A into a point A' such that A' lies on the ray OA and

$$\frac{OA'}{OA} = \frac{ON_1}{OM_1}$$

where M_1 and N_1 are the points of intersection of the ray with the curves (1) and (2) respectively.

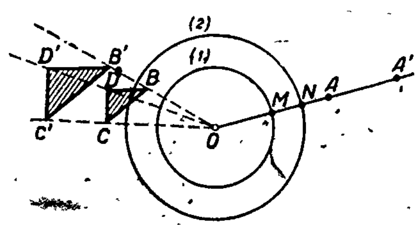


Fig. 116

If the curves (1) and (2) are concentric circles (in space, concentric spheres) the transformation will be the familiar similarity transformation with the center of similarity at point O (fig. 116).

It will be interesting to work out the forms taken by the images of various elementary figures (triangle, square, circle etc.) if a series of simple figures are successively chosen as the curves (1) and (2).

17. STEREOGRAPHIC PROJECTION

Let us consider still another transformation of figures --- central projection. In mapping conic sections into a circle (but in general not onto the circle) (figs. 29, 31, 33), as well as a plane into a sphere (but not onto a sphere)(fig. 100), we were already dealing with central projections, the center being at point S.

Let us investigate the properties of one central projection of plane figures into spherical figures -- the properties of stereographic projection -- which will be indispensable to us later.

Stereographic projection is important not only in mathematics -- where, aside from geometry, we may point out the possibility of the one-to-one mapping of the plane of complex numbers onto a perforated sphere, the utility of which is well known -- but also, for example, in cartography. In cartography we must deal with the mapping of the whole or a part of the sphere into a plane, that is, with the construction of plane maps of the earth's surface regarded as a sphere. The stereographic projection is one of the most important cartographic projections.[1] In our study of stereographic projection (fig. 100) we shall consider that the sphere is perforated at the center of perspective

1) See A. V. Gedymin: Cartography, textbook for teacher training institutes, Uchpedgiz 1946. Examining the instructive diagrams and sketches of this book one may readily get an idea of the distortions of figures under various kinds of transformations of parts of a sphere into a plane [14].

S, and that the point S^* belongs to space but not to the sphere. Every figure lying on the sphere is likewise considered to be perforated at S if it passes through that point.

We observe, in the first place, that any line a of the plane π is transformed by the stereographic projection into a circle passing through the point S (fig. 117). This follows from the fact that all rays SA, projecting points A of line

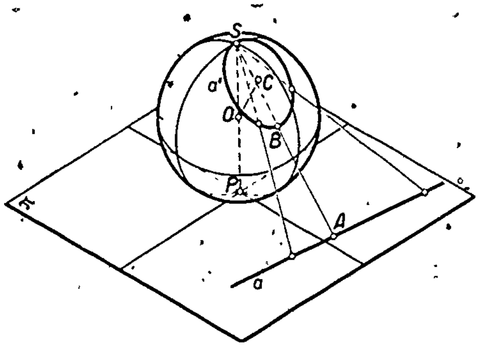


Fig. 117

a into points B of the sphere, lie in one plane which projects the line a and which intersects the sphere in a circle passing through S.

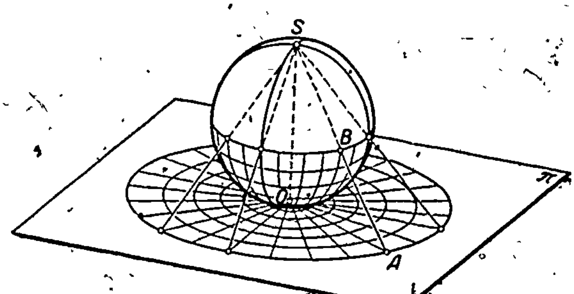


Fig. 118

Thus, any line a of plane π is transformed into circle a' passing through S. Every circle a' belonging to the sphere and passing through S goes over, under the inverse transformation, into a line a of the plane. This is one of the fundamental properties of stereographic projection.

All meridians, in particular, are transformed into straight lines passing through the south pole (fig. 118), and vice versa: every line belonging to the pencil whose center is at the south pole is transformed into a meridian (fig. 118).

By virtue of the one-to-one nature of the transformation, we shall speak sometimes of the transformation of the perforated sphere onto the plane π and sometimes of the inverse transformation of the plane π into the sphere (or onto the perforated sphere). Which of these transformations we have in mind will be evident from the context.

It is easily seen that all the parallels of the sphere, including the equator, are transformed into concentric circles in the plane having their center at the south pole of the sphere (fig. 118). In proof we note that the right circular cone which projects any parallel intersects the plane π in a circle. This is the method used in constructing maps of circumpolar regions; examples of such maps may be found in any geographical atlas.

Let us now inquire what are the images of circles on the sphere other than those we have just considered (figs. 117, 118).

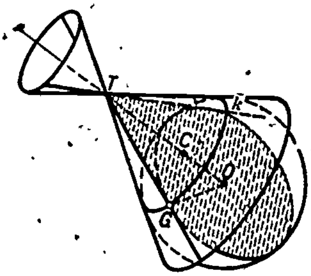


Fig. 119

We find that the image of any circle on the sphere not passing through the center S of the stereographic projection will be a circle in the plane π of that projection.

THIS IS THE FIRST FUNDAMENTAL THEOREM OF STEREOGRAPHIC PROJECTION.

Before proceeding to prove it, let us point out the following facts:

Firstly, in order to find the vertex T of the cone tangent to a sphere in a given circle k of the sphere, we may proceed as follows: We draw from the center O of the sphere (fig. 119) the perpendicular OC to the plane of the given circle, and in any plane passing through this perpendicular (shaded in the diagram) we construct the tangent GT to the great circle of the sphere at the point of its intersection G with the given circle k. The point T of the intersection of this tangent with line OC will also be the desired vertex T of the cone. If circle k is a great circle, the tangent GT will be parallel to OC and the tangent cone degenerates into a cylinder.

Returning to the stereographic projection, we note, secondly, that the vertices T of the cones tangent to the sphere in any circle a' passing through the center S of the projection (fig. 117) lie in a plane α tangent at S to the sphere and consequently parallel to plane π. In figure 120 is shown the section of the sphere by a plane passing through S and containing the perpendicular OC to the plane of the circle a' (see fig. 117).

Thirdly, every line tangent to the sphere is tangent to any circle of the sphere the plane of which contains this tangent.

Proof: Let a be a line tangent to the sphere at M (fig. 121), which is to say, let a be perpendicular to radius OM of the sphere. If through line a we pass a plane intersecting the sphere in the circle k (we suppose here that k is not a great circle, in which case the assertion would be obvious), and if we drop the perpendicular OC to this plane, then OC will also be perpendicular to the line a of this plane. [Here, of course, the two perpendicular lines are skew. --Translators.]

Thus, line a is perpendicular to OM and OC, and hence perpendicular to the plane of these lines, whereby it is also perpendicular to the line CM of this plane. Thus, line a is tangent to circle k.

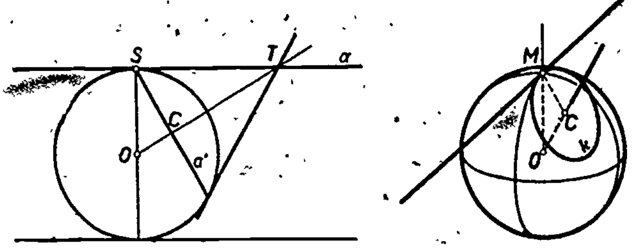


Fig. 121

Keeping these facts in mind, we shall prove one more lemma:

Lemma:

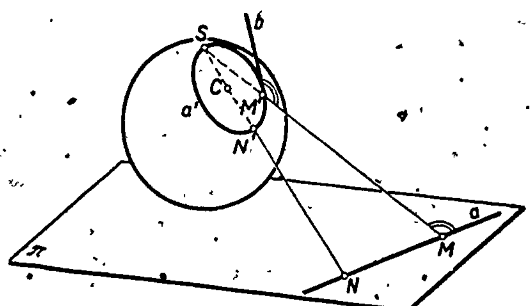


Fig. 122

Let b denote a line tangent to the sphere at point M' (fig. 122). We pass a plane through the line b and the pole S of the stereographic projection. This plane intersects the sphere in the circle a', and intersects the plane π in the image of this circle, the line a. (Fig. 123 shows this plane together with circle a' and lines b and a.) It can easily be verified that line SC (figs. 122, 123), where C is the center of circle a', is perpendicular to line a. We shall prove that lines b and a intersect line SM at the same angle.

Proof: $\angle bM'M = \angle SM'D = \angle SM'C + \frac{\pi}{2}$;

and since $\angle MSN = \angle SM'C$

and $\angle aMM' = \angle MSN + \frac{\pi}{2}$

we get $\angle bM'M = \angle aMM'$,

q.e.d.

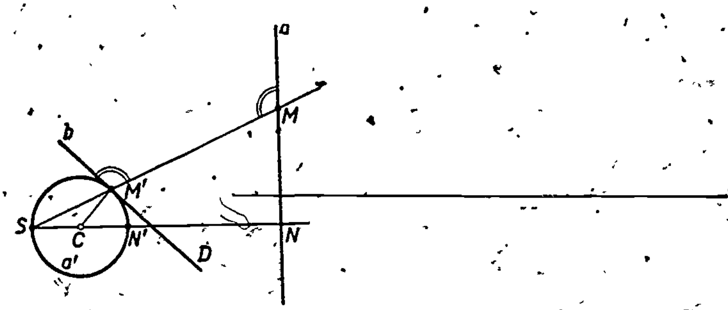


Fig. 123

We now turn to the proof of the first fundamental theorem of stereographic projection.

Let a' be a circle of the sphere not passing through point S and not a parallel of the sphere (fig. 124). Further, let T be the vertex of a cone tangent to the sphere in the circle a', so that if M' is an arbitrary point of circle a', TM' is tangent to the sphere. Drawing the line SM', we denote the image of point M' in the plane π by M. We also denote by K the intersection of the line ST and the plane π. We know that T does not lie in the plane tangent to the sphere at S and parallel to π; consequently ST is not parallel to the plane π.

We shall now show that the image a of circle a' is a circle having center K.

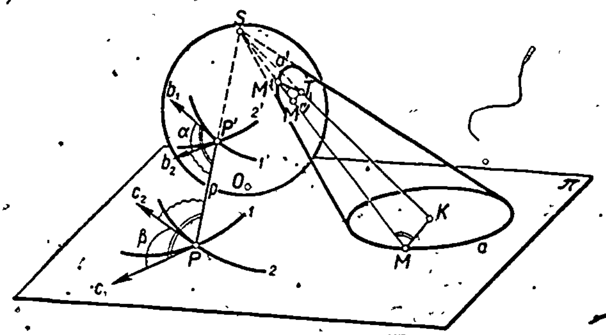


Fig. 124

By the lemma, $\angle TM'M = \angle KMM'$, since TM' is tangent to the sphere and KM is its central projection on plane π.

We draw TM'' parallel to KM. Then either points M'' and M' coincide, or triangle $M'M''T$ has equal angles at the vertices M' and M'', and consequently

$$TM' = TM'';$$

but $\frac{MK}{M''T} = \frac{SK}{ST}$,

or $MK = M''T \frac{SK}{ST} = M'T \frac{SK}{ST}$.

But $M'T$ has the same length for all points M' on the circle a'. Consequently, MK has the same length for all points M on the curve a and the curve a is a circle, q.e.d.

Here again we draw attention to an important fact of which we made use in the preceding chapter. In order to obtain the spherical center of circle a', it is necessary to drop a perpendicular from the vertex T (fig. 124) to the plane of circle a'. The points of intersection of this perpendicular with the sphere will be the two spherical centers of the circle a'. But by virtue of our assumption that the circle a' is not a parallel of the sphere relative to the pole S, ST is not perpendicular to the plane of the circle a', and consequently the center of a' cannot lie on ST. Therefore, the spherical center of a circle a', the plane of which does not pass through the center of the stereographic projection and is not parallel to the plane π of the projection, does not have as its image the center K of the transformed circle a.

Similarly it may be shown that any circle in plane π whose center K is not the point of tangency of the sphere has as its image on the sphere a circle which does not pass through S and whose center does not lie on the projecting ray SK.

WE SHALL NOW PROVE THE SECOND FUNDAMENTAL THEOREM OF STEREOGRAPHIC PROJECTION:

Under stereographic projection the angles between curves on the sphere are equal to the angles between their images.

Let there be given the curves 1' and 2' on the sphere (fig. 124) intersecting at angle α in the point P'. The angle between the curves is the angle between the tangents b_1 and b_2 to these curves, where b_1 and b_2 lie in a plane tangent to the sphere at P'[1]. The images of the curves 1' and 2' are the curves 1 and 2 in the plane π. Curves 1' and 1 lie on the surface of a projecting cone with vertex at S. Curves 2' and 2 similarly lie on a conic surface with vertex at S: SP'P' is a common generator of these conic surfaces.

The plane tangent to the first conic surface along its generator $SP'P \equiv \rho$ contains the tangents b_1 and c_1 to the curves 1' and 1.

By virtue of the lemma (page 171) we have:

$$\angle b_1P'P = \angle c_1PP',$$

and $$\angle b_2P'P = \angle c_2PP'.$$

Thus, in the trihedral angles $P'pb_1b_2$ and Ppc_1c_2 we have a common dihedral angle (between the planes tangent to the cones) and two pairs of corresponding equal face angles. Such trihedral angles are equal or symmetrical, and consequently their third face angles are equal, that is, $\alpha = \beta$. q.e.d.

Stereographic projection is a transformation which preserves angles.

Any transformation under which angles between curves are preserved is called conformal.

Stereographic projection is a conformal projection.

We shall utilize our knowledge of stereographic projection to obtain some transformations of the plane.

If the plane π is stereographically mapped into a sphere, the sphere then rotated about any diameter, not necessarily the one passing through S, and the points of the sphere are projected back onto the plane, then we obtain a transformation of the plane into itself.

Initially, point A of plane π goes over into a point A' of the sphere; when the sphere is rotated, A' goes over to some point A" of the same sphere; and under the inverse projection, A" is transformed into point A_1 of plane π. The result is that point A of the plane π goes over into point A_1 of the same plane.

However, while this is the case generally, it is not so without exception. The perforation S_* in the sphere, for example, takes up after the rotation some new position S', the projection of which does not give an image of a point A in the plane. Furthermore, the point which after the rotation assumes the highest position above the plane π will not have an image after the subsequent projection. Consequently, that point of the plane which after projection and rotation assumes this highest position will have no image in the plane π. The transformation of the plane thus obtained will not be a one-to-one transformation of the plane onto itself. Under each such transformation there will exist on the plane π a point into which no point of the plane is transformed, as well as a point which has no image.

What is important for us is that figures in the plane π and in fact whole regions of the plane, not containing these two exceptional points, are transformed topologically. Under this transformation the family consisting of all lines and circles goes over into itself, that is, it is invariant under the transformation; furthermore, this transformation is conformal.

18. THE CONCEPT OF CENTRAL PROJECTION

A stereographic projection is a special case of central projection of a sphere onto a plane, with a particular arrangement of the center S and the plane of projection π. Sometimes, retaining the center at S, the sphere is centrally projected onto the plane of the equator; and so on. Different kinds of stereographic projections are thus obtained.

Particular importance attaches to the central projection of a plane π onto another plane π. Here again the one-to-one onto property is violated if we consider the planes π and π' in their entirety and if they are not parallel (fig. 125). In the plane π there is a particular line a lying in the plane which is parallel to π' and passes through S. This line has no image. Analogously, there is a line c' on plane π' which is not the image of any line whatever of plane π.

If the planes are cut through along these lines, i.e. if these lines are excluded, the correspondence of points in the figures obtained preserves the one-to-one property. Lines of the plane π are projected into lines of the plane π' and vice versa. On these lines there will be some exceptional points, namely the points of intersection of these lines with the special lines a and c'.

Under central projection of one plane upon another circles go over into conic sections. We are already acquainted with particular cases of such projections.

It is always possible to locate the planes π and π' and select the center of projection S in such a manner that the planes π and π' intersect and a given circle a in plane π will be transformed into a circle a' in plane π'.

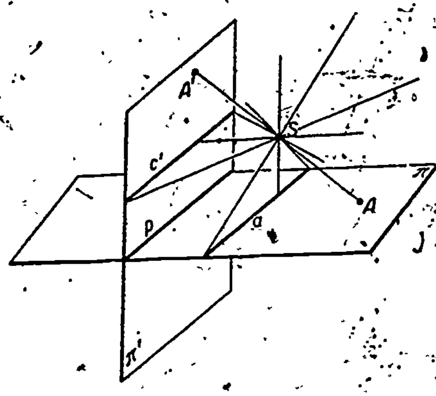


Fig. 125.

How to choose such a central projection is shown by figure 124. The required procedure is to select any sphere tangent to plane π at a point other than the center K of circle a and, taking the stereographic projection a' of the circle a, to consider the plane of circle a' as plane π'.

It is to be particularly noted that, as we already know, the center K of circle a is not projected into the center of circle a'.

We made use of this observation in the preceding chapter.

In general, all other circles of plane π will not be projected into circles by the projection just constructed.

19. THE CONCEPT OF A GROUP OF TRANSFORMATIONS

In geometry an important role is played not only by individual transformations of figures but, preeminently, by sets, classes or aggregates of transformations satisfying prescribed conditions. Central importance attaches to the concept of a group of transformations.

To illuminate this concept let us consider the following example.

Let a be some line in the plane π (fig. 126) and let k be a given positive number. To the point M we assign the point M' such that: firstly, M' lies on the perpendicular MA to line a and on the same side of a as point M; and ~~secondly,~~ $AM' = k \cdot AM$.

The correspondence thus established is a one-to-one transformation ϕ_k of plane π into itself, which is called a uniform stretching of plane π if $k > 1$, or a uniform shrinking if $k < 1$.

Under this transformation the points of line a remain fixed.

In the diagram are shown the transformations of triangle PQR into triangle P'Q'R' and of a circle into an ellipse (with $k = 3$). It can easily be shown that under this transformation a straight line goes over into a straight line and two parallel lines p and q go over into two parallel lines p' and q'.

Taking the given line a and all possible $k > 0$ we have a set of transformations of the plane into itself. Transformation ϕ_1 (with $k = 1$) is the identity transformation; it leaves all points of plane π in their original positions.

Two transformations ϕ_{k_1} and ϕ_{k_2} of the set of transformations under consideration, performed one after the other, will be equivalent to one transformation $\phi_{k_1 k_2}$. The coefficient of stretching k of the resultant transformation is equal to the product of the coefficients of the component transformations ϕ_{k_1} and ϕ_{k_2}:

$$k = k_1 \cdot k_2 .$$

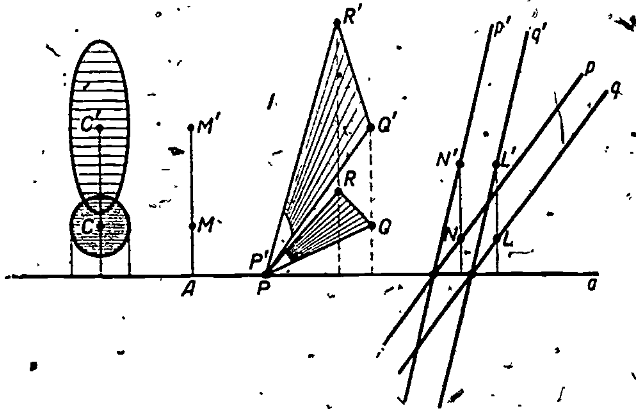


Fig. 126.

The transformation $\phi_{k_1 k_2}$ is called the sum of transformations ϕ_{k_1} and ϕ_{k_2}. We write:

$$\phi_{k_1} + \phi_{k_2} = \phi_{k_1 k_2} .$$

The sum of transformations ϕ_k and $\phi_{\frac{1}{k}}$ is equal to ϕ_1:

$$\phi_k + \phi_{\frac{1}{k}} = \phi_1 .$$

The transformations ϕ_k and $\phi_{\frac{1}{k}}$ are said to be inverse to each other; their sum is equal to the identity transformation ϕ_1.

We shall now give a general definition of a group of transformations:

Definition. A set G of transformations of a plane (or of space) into itself is called a group if:

1. the sum of two transformations of set G belongs to the set G (closure under addition),
2. the identity transformation belongs to the set G, (existence of a neutral element),
3. every transformation of the set G has an inverse transformation belonging to the set G (existence of an inverse element).

Thus, the set of all uniform stretchings ϕ of a plane π constitutes a group.

Hereinafter we shall encounter many groups of transformations.

20. THE GENERAL CONCEPT OF A GROUP

For the set of uniform stretchings there has been defined one algebraic operation -- that of addition. The operation was subject to prescribed conditions which characterize a group.

The general concept of a group belongs to algebra [4]. In the present section are presented those elements of general group theory indispensable for what is to follow.

It is sometimes important to exhibit some of the properties of sets in which there is defined an operation subject to certain conditions. These properties will hold irrespective of the elements of the set and regardless of the nature of the operation within the set.

To begin with we shall give a general definition of an algebraic operation.

Given some set G of elements of any kind, --

In set G an algebraic operation is said to be defined if to any two elements (different or identical) of the set G, taken in a definite order, there is assigned according to some law a fully determined third element belonging to the same set.

In our example, to any two elements in the set of stretching transformations there is assigned a third element in the same set of stretching transformations.

Arithmetic and algebra yield other examples of algebraic operations. Such, for instance, are addition and multiplication in number sets.

If the operation is such that for any two elements of the set G the result of the operation is independent of the order in which the elements are taken, we have a commutative operation. But if to some two elements (a, b) taken in a certain order the operation assigns a third element c, and to the same elements (b, a) but taken in the opposite order it assigns the element c' differing from c, the operation is non-commutative.

It is immaterial what sign we use for the operation, or what we call it.

We may use the signs of addition and multiplication,

$$c = a + b, \; c = a \cdot b, \; c = ab, \; c = a \times b$$

or any other signs, e.g.:

$$c = a \circ b, \; c = a * b, \; c = [ab],$$

We shall employ chiefly the additive notation $a + b$ and shall call the operation addition. On occasion we will also employ the multiplicative notation ab and call the operation multiplication.

A set G with one algebraic operation is called a group if it satisfies the following conditions:

I. The condition of associativity or the Associative Law.

For any three elements a, b, c of set G we have

$$a + (b + c) = (a + b) + c.$$

II. The condition of the existence of a neutral element.

Among the elements of set G there must exist a unique element called the neutral element and denoted by 0, such that

$$a + 0 = 0 + a = a$$

for every element a.

III. The condition of the existence of an inverse element for any given element.

For every given element a of a set G it must be possible to find in the same set G an element $-a$ such that

$$a + (-a) + (-a) + a = 0.$$

Let us consider some examples of groups.

1° In the set of uniform stretchings we had an example of a commutative group.

The neutral element was supplied by the identity transformation ϕ_1.

2° The set of all real numbers with the operation of ordinary addition is also a commutative group.

The neutral element of this group is the number 0.

3° The set of all real numbers excluding zero with the operation of ordinary multiplication is again a commutative group. The neutral element is the number 1. In this case it is more usual to write the group operation "multiplicatively". The group conditions will in this notation have the form:

I. $a(bc) = (ab)c$; II. $a \cdot 1 = 1 \cdot a = a$; III. $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

Here the neutral element is denoted by 1, and the element inverse to element a, by a^{-1}.

4° The set of all complex numbers is a group with respect to addition.

5° The set of integral complex numbers, that is, of numbers of the form $z = a + bi$, where a and b are integers, is also a group with respect to addition.

6° The set of complex numbers with zero excluded is a group with respect to multiplication.

The verification of this assertion is recommended as an exercise.

7° Let us now examine a set of transformations of an equilateral triangle ABC into itself.

The elements of this set may be written down most simply as follows. If for example a transformation carries over the vertices A, B, C respectively into the vertices B, C, A, this element is written in the form

$$\binom{ABC}{BCA}.$$

In all there are six elements:

$$a_0 = \binom{ABC}{ABC}; \quad a_1 = \binom{ABC}{ACB}; \quad a_2 = \binom{ABC}{BAC}; \quad a_3 = \binom{ABC}{BCA};$$

$$a_4 = \binom{ABC}{CAB}; \quad a_5 = \binom{ABC}{CBA}.$$

These are the elements of the set of all self-displacements or symmetries of the triangle ABC, i.e. the set of all displacements of the triangle ABC into itself. Transformation a_0 is the identity transformation, leaving every vertex of the triangle in its original position. Transformation a_5 carries vertex A into vertex C; B into B, i.e. leaves it in place; and A into C. Transformation a_5 is the "turning over" of triangle ABC about the altitude drawn from vertex B.

It is evident that two of these transformations, performed one after the other in a given order, can be replaced by one transformation belonging to the same set. Such a replacement of two transformations by one we shall call multiplication and, consequently, we shall use the "multiplicative language".

We have now defined in the set an operation which possesses the property of uniqueness and closure.

We shall agree to place in front that transformation which is to be performed first. For example

$$a_2a_4 = \binom{ABC}{BAC}\binom{ABC}{CAB} = \binom{ABC}{ACB}.$$

Here the first transformation takes vertex A into vertex B; the second, vertex B into vertex A; as a result, vertex A is taken by the transformation a_2a_4 into vertex A, i.e. it remains in place.

In the same manner we find that $B \longrightarrow A \longrightarrow C$; $C \longrightarrow C \longrightarrow B$.

We find that $a_2a_4 = a_1$. If we take the transformations in the opposite order, we shall have:

$$a_4a_2 = \binom{ABC}{CAB}\binom{ABC}{BAC} = \binom{ABC}{CBA} = a_5.$$

It is suggested that the student verify that the associative law is fulfilled here. The condition of the existence of a neutral element is also fulfilled. This element is a_o, since, for example, we have:

$$a_3a_o = \binom{ABC}{BCA}\binom{ABC}{ABC} = \binom{ABC}{BCA} = a_3,$$

and likewise $a_oa_3 = a_3$.

For every element there exists an inverse element.

If, say, $a_3 = \left(\begin{smallmatrix}ABC\\ \downarrow\downarrow\downarrow\\ BCA\end{smallmatrix}\right)$, then the inverse transformation must be such that $\left(\begin{smallmatrix}ABC\\ \uparrow\uparrow\uparrow\\ BCA\end{smallmatrix}\right)$, or in standard notation $\binom{ABC}{CAB}$, i.e., equal to a_4. It is easily verified that $a_3a_4 = \binom{ABC}{BCA}\binom{ABC}{CAB} = \binom{ABC}{ABC} = a_o$ and likewise $a_4a_3 = a_o$.

The set of symmetries of an equilateral triangle is a non-commutative group.

It is recommended that the student follow the carrying out of the transformations just considered, not only in symbolic notation, but also graphically in geometric terms.

Using the same example we shall explain the important concept of a subgroup.

We take from the six elements of our set just two, a_0 and a_1. This subset of the given set, with the same operation as was defined for the entire set, will also be a group, a subgroup of the given group.

Let us verify the closure of the operation (for the subset):

$$a_0a_1 = a_1; \quad a_1a_0 = a_1; \quad a_1a_1 = \begin{pmatrix} ABC \\ ACB \end{pmatrix}\begin{pmatrix} ABC \\ ACB \end{pmatrix} = \begin{pmatrix} ABC \\ ABC \end{pmatrix} = a_0.$$

The associative law does not have to be verified, since it is fulfilled for the entire set and hence is automatically fulfilled for a part of it.

The condition of the existence of a neutral element is also fulfilled, since the element a_0 belongs to the subset $[a_0, a_1]$.

Further, the condition of the existence of an inverse element is fulfilled. The element inverse to the neutral element is the latter itself; the element inverse to a_1 is a_1.

Furthermore, $a_0a_1 = a_1a_0$. Consequently, this subgroup of the given group is commutative.

It is easy to verify that subsets $[a_0, a_2]$ and $[a_0, a_5]$ are also subgroups of the given group. Other subsets of two elements of the given group will not be subgroups. The subset consisting of only the neutral element a_o also constitutes a subgroup, the identity subgroup.

Now let us take the subset $[a_o, a_3, a_4]$ consisting of three elements of the given set. The operation in this subset is that defined for the given group. Let us verify closure:

$$a_0a_3 = a_3a_0 = a_3; \quad a_0a_4 = a_4a_0 = a_4;$$

$$a_3a_4 = \binom{ABC}{BCA}\cdot\binom{ABC}{CAB} = \binom{ABC}{ABC} = a_o = a_4a_3.$$

Besides this, there exists a neutral element a_o; and there exists an inverse element for every element: $a_3^{-1} = a_4$; $a_4^{-1} = a_3$; $a_o^{-1} = a_o$.

Verify that the group of symmetries of triangle ABC has no subgroups of four or five elements.

Let us write out all the subgroups of the given group: --

$$[a_o], [a_o, a_1], [a_o, a_2], [a_o, a_5]$$

$$[a_o, a_3, a_4], [a_o, a_1, a_2, a_3, a_4, a_5].$$

The first and last of these subgroups of the given group are called its improper subgroups. In general, the identity subgroup of any group G and the group G itself are called improper subgroups of the group G.

We shall next establish the important concept of isomorphism in the sense in which it is applied to groups. This concept is familiar from algebra, where it was studied in connection with the concepts of ring and field.

Let there be given a one-to-one correspondence

$$a \longleftrightarrow a',$$

between the set of all elements of group G and the set of all elements of group G'. The correspondence is an isomorphism (or an isomorphic correspondence) provided the operation is preserved: that is, whatever relationship of the form $a + b = c$ exists among three elements of one group, for example G, upon replacing elements a, b, c of group G with the corresponding elements a', b', c' of group G' the relationship obtained remains valid: $a' + b' = c'$.

Two groups are called isomorphic if an isomorphic correspondence can be established between them.

The three subgroups $[a_0 a_1]$, $[a_0 a_2]$, $[a_0, a_5]$ of the group of symmetries of the triangle are isomorphic by virtue of the following isomorphic correspondences:

$$\left\{\begin{matrix} a_0, & a_1 \\ \updownarrow & \updownarrow \\ a_0, & a_2 \end{matrix}\right\}, \quad \left\{\begin{matrix} a_0, & a_1 \\ \updownarrow & \updownarrow \\ a_0, & a_5 \end{matrix}\right\}, \quad \left\{\begin{matrix} a_0, & a_2 \\ \updownarrow & \updownarrow \\ a_0, & a_5 \end{matrix}\right\}.$$

In general, all groups consisting of two elements are always isomorphic.

From the point of view of group theory, isomorphic groups do not differ from each other, whatever may be the nature of their elements and operations.

Let us examine another such example. We separate all the real non-negative numbers 0, 1, 2, 3, 4, 5,... into two classes: the class of even numbers

$$b_0 = [0, 2, 4, 6, \ldots]$$

and the class of odd numbers

$$b_1 = [1, 3, 5, 7, \ldots]$$

and consider the set of two elements b_0 and b_1. The elements of this set are classes.

Let us establish in this set of two elements the operation of addition, based on the following rule: to add two classes means to take one arbitrary number from each class and to add them in the ordinary way. The class to which the sum belongs we assign to the two given classes and call it the sum of the given classes. This operation uniquely assigns to the two classes a third one, and is closed.

From this rule it follows that:

$$b_0 + b_0 = b_0, \quad b_0 + b_1 = b_1 + b_0 = b_1, \quad b_1 + b_1 = b_0.$$

The isomorphism between group $[b_0, b_1]$ and the group $[a_0 a_1]$ will have the form:

$$\begin{array}{cc} a_0, & a_1 \\ \updownarrow & \updownarrow \\ b_0, & b_1 \end{array}$$

That the operation is preserved may be perceived from the table

$$a_0 a_0 = a_0, \quad a_0 a_1 = a_1 a_0 = a_1, \quad a_1 a_1 = a_0;$$
$$b_0 + b_0 = b_0, \quad b_0 + b_1 = b_1 + b_0 = b_1, \quad b_1 + b_1 = b_0 .$$

Theorem. Under the isomorphic mapping

$$a \longleftrightarrow a'$$

of a group G onto a group G', the neutral element of the second group corresponds to the neutral element of the first group, and to every pair of elements inverse to each other in the first group there corresponds a pair of elements inverse to each other in the second group.

To the neutral element a_o of group G let there correspond under the given isomorphism the element a'_o of group G':
$a_o \longleftrightarrow a'_o$.

We shall prove that a'_o is the neutral element of group G'.

For any arbitrary element of group G we have:

$$a + a_o = a, \quad a_o + a = a,$$

and by virtue of the isomorphism,

$$a' + a'_o = a', \quad a'_o + a' = a',$$

whence it follows that a'_o is the neutral element of group G'.

Let a and b be a pair of inverse elements of group G,

$$a + b = a_o, \quad b + a = a_o.$$

Then, by virtue of the isomorphism,

$$a' + b' = a'_o, \quad b' + a' = a'_o.$$

Since a'_o is the neutral element, a' and b' are inverse to each other. q.e.d.

ELEMENTARY GEOMETRICAL TRANSFORMATIONS

Chapter IV

PARALLEL TRANSLATIONS (TRANSLATIONS)

In chapter IV translations in the plane and in space are investigated. In connection with parallel translations we shall study the properties of parallel lines.

The theory of parallel lines will serve later as an introduction to the geometry of Lobachevskii.

In 26. we shall consider the application of the method of parallel translation to the solution of problems in construction.

21. PARALLEL LINES

Two distinct straight lines cannot intersect in more than one point, since if two lines had more than one common point they would coincide, i.e. they would have all their points in common.

After the proof of the existence of non-intersecting lines in a plane there is introduced the axiom of parallels.

IV (Euclid's axiom). Let *a* be an arbitrary line, and A a point lying outside of it; then there exists in the plane determined by the line *a* and the point A not more than one line passing through the point A and not intersecting the line *a*.

It should be emphasized that the assertion "not more than one" forms the content of the axiom while the assertion "one and only one" is the content of a theorem (See 2.). We shall call this theorem the first corollary of Euclid's axiom.

It should be noted that in the formulation of Euclid's axiom, as well as in the definition of a parallel, the parallel *b* is determined by the line *a* and a point A exterior to *a*. However the line *b* contains still other points. Suppose B is such another point. The first corollary tells us that there exists a line *c* parallel to line *a*, passing through B. But it does not tell us whether line *c* is the same as line *b*.

Observing, however, that line *b* lies in the plane determined by line *a* and point B, and furthermore, passes through the point B and does not intersect line *a*, we may utilize Euclid's axiom, which states that there is not more than one such line, to show that line *c* coincides with line *b*. Thus no matter what point B we choose on line *b* the parallel determined by line *a* and point B turns out the same line *b*. This is the second corollary of Euclid's axiom.

Only after establishing this second corollary may we speak about line *b* being parallel to line *a* without indicating any point A on line *b* with respect to which this parallelism holds.

Only now does it make sense to say: the line *b* is parallel to the line *a*.

Whether the line *a* is, in turn, parallel to the line *b* has yet to be proved. This fact constitutes the third corollary of Euclid's axiom.

If a line *b* is parallel to a line *a*, then line *a* is also parallel to line *b*: In fact, let us take a point C on the line *a* (fig. 127). The point C and the line *b* determine a plane in which the line *a* lies. But the line *a* does not have any common points with the line *b*, consequently, *a* is parallel to b, q.e.d.

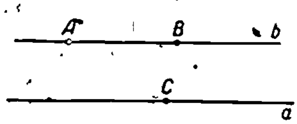


Fig. 127.

Only after the deduction of this third corollary is it possible to say: The lines a and b are parallel to each other. The property of parallel lines just established is called the property of symmetry for the parallelism of two straight lines.

Finally, we shall prove the fourth corollary of Euclid's axiom:

If a line a is parallel to a line b and the line b is parallel to a line c then the line a is also parallel to the line c.

The lines a and c are both parallel to the line b. Consequently, the lines a and b lie in one plane, and similarly, the lines c and b lie in one plane. These planes are either different or else coincide. If the planes are different, then they intersect in the line b.

Let us assume that the lines a and c intersect in the point S. In such a case, however, point S, being on the line a lies in the plane of the lines a and b; lying on the line c, the point S lies on the plane of the lines c and b. It follows that the point S lies on the line of intersection of these planes, i.e. on the line b. Thus the lines a and b have a common point S, which is impossible, since a and b are parallel.

The contradiction thus obtained shows that the lines a and c have no common point.

But a and c could be skew lines. We shall show that this cannot take place.

Let us take a point C on the line c and let us consider the line of intersection d of the planes determined, on the one hand by the lines c and b, and on the other by the line a and the point C.

The line of intersection d is parallel to the line b.

To prove this assertion: The line d lies in the plane of the lines c and b. Now, if the line d were to intersect the line b in a point Σ, then the point Σ, belonging to the line b, would lie in the plane of the lines a and b and at the same time, being on the line d, would lie in the plane of the lines d and a. But this would mean that the point Σ would be on the line of intersection(1) a of the indicated planes. In other words, the lines a and b would have a common point Σ, which is impossible. This shows that the line d is parallel to the line b.

Thus the line d, passing through the point C, and the line c also passing through the point C, are both parallel to the line b, and consequently coincide. But the line d lies in one plane with the line a. Thus the lines a and $c \equiv d$ cannot be skew, and consequently, the line a is parallel to the line c.

(1) Footnote: Unless lines a, b and d were all in the same plane. But then line b and point C would be in this plane (d passes through point C). Consequently line c, being the parallel to b passing through point C, would also be in this plane. Thus lines a and c would be coplanar and not skew. --Translators.

The first half of the fourth consequence is proved.

Let us now assume that the planes of lines (a, d) and (a, c) coincide, i.e. the lines a, b and c lie in one plane.

The lines a and c do not intersect the line b, since they are parallel to it. If the lines a and c were to intersect each other in a point S, then, through the point S, there would pass two lines parallel to the line b. But in accordance with Euclid's axiom this is impossible. Thus the lines a and c are parallel. The fourth corollary is completely proved.

The property of parallelism stated in the fourth corollary is called the property of transitivity.

This property holds for the plane as well as for space and allows us to introduce the notion of a pencil of parallel lines and a bundle of parallel lines.

The set of lines in a plane α parallel to a given line a is called a pencil of parallel lines.

Any two lines b and c of the pencil determined by a line a are parallel to each other. In fact, the line b is parallel to the line a, consequently, by the property of symmetry, the line a is parallel to the line b. Since the line c is also parallel to the line a, the transitivity property implies that line b is parallel to line c, q.e.d.

Thus a pencil of parallel lines can be determined by any line in it, i.e. we may speak of a pencil of mutually parallel lines.

Since the properties of symmetry and transitivity for parallelism of straight lines hold also in space, we have an analogous definition: The set of lines in space parallel to a given line a, including the line a itself is called a bundle of parallel lines.

In exactly the same manner we show that a bundle of parallel lines is determined by any line in it, i.e. one may speak of a bundle of mutually parallel lines.

22. TRANSLATIONS (PARALLEL TRANSLATIONS) IN THE PLANE

Let us examine a pencil of parallel lines belonging to a plane π (fig. 128).

We have seen that every line a completely determines a pencil, namely the aggregate of all lines parallel to a together with the line a itself. As a result of the fact that parallelism of lines is subject to the properties of symmetry and transitivity, and that the line a itself is in the pencil, the entire set of straight lines in the plane π is subdivided into classes of mutually parallel lines. Any line a of the plane represents a class of lines parallel to it, to which it itself belongs. Every line b belonging to the class determined by a line a, determines precisely that class to which it belongs.

To the set of lines in a class - the lines of a pencil of parallels - we ascribe <u>one direction</u>, and we shall say that all lines of the pencil have one and the same direction. Since each line belongs to only one pencil, a line has only one direction.

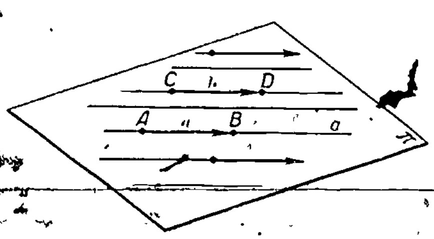


Fig. 128.

Let us take any two points A and B on a line a. Line a (and, consequently, a direction also) are fully determined by the two points A and B. However, for the two points, besides a direction it is also possible to indicate an orientation of the segment AB. If A is considered as the first point of the segment - its origin - and the point B its second point, or its end, then we thereby ascribe to the segment AB, besides a direction, also an orientation.

One and the same segment AB has one direction but two orientations $\overrightarrow{AB}$ and $\overrightarrow{BA}$. An oriented segment $\overline{AB}$ is called a vector.

Vectors are denoted by bold face lower case or upper case letters

$$a, b, c, A, B, C, \ldots$$

or by two upper case letters with a line above

$$\overline{AB}, \overline{CD}, \overline{LM}, \ldots$$

with the first letter denoting the origin of the vector, and the second denoting its end.

The line AB has two orientations, determined by the vectors $\overline{AB}$ and $\overline{BA}$. The orientations $\overline{AB}$ and $\overline{BA}$ are called opposite.

The length of the segment AB is called the length of the vector $\overline{AB}$.

A vector whose origin and end coincide is called a null vector. For such a vector neither the direction nor the orientation is determined, while the length is equal to zero.

Vectors are considered equal, $a = b$, if they have the same direction, the same length and the same orientation.

In the drawing the orientation of a vector is denoted by an arrow placed at its end (fig. 128).

Remark. If we agree to consider as equal only such vectors as have the same direction, the same length, the same orientation, and in addition, lie on the same line, then such vectors are called sliding vectors. The line on which they lie will be called the line of action of the sliding vector.

In mechanics, the forces applied to an absolutely rigid body are sliding vectors. A force applied to such a body may be translated along its line of action without changing the condition of motion or condition of rest of the body.

A force cannot, however, be displaced to a line parallel to its original line of action without changing the condition of motion or rest.

If we agree to consider as equal only such vectors as, in addition to the same direction, length and orientation, also have a common origin, then we arrive at the so-called fixed vectors.

The equality of vectors defined in the main text is the definition of equality of free vectors.

When we say that a plane is transformed into itself we mean that each point M of the plane is mapped into a point M' of the same plane. We say that under the given mapping the point M goes over into the point M', or that the point M is transformed into the point M', or that the point M is displaced into the point M'. The position of the image M' of the point M may readily be defined by means of the vector $\overline{MM'}$ (fig. 129).

Fig. 129.

Let us give the following definition: A vector having its origin at a point M and its end at the transformed point M', is called the displacement vector of the point M under the given transformation of the plane into itself.

We shall give an example.

Earlier (13, 14) we had a transformation of the plane into itself, under which the points of the triangle ABC inscribed in a circle passed into the points of the circle, and vice versa, while the remaining points of the plane remained where they were, i.e. went over into themselves (fig. 101, 111).

In fig. 130 certain displacement vectors under this transformation are drawn. The points other than those of the linear triangle and circle have null displacement vectors; the vertices of the triangle also have null displacement.

Since the point M passes into N (fig. 130), while N, in its turn passes into M, each segment MN carries two displacement vectors $\overline{MN}$ and $\overline{NM}$ with opposite orientations.

Let us give now a definition of translation (or parallel translation). A parallel translation (or translation) in the plane is a transformation of the plane into itself under which the displacement vectors of all points in the plane are equal to each other:

$$\overline{MM'} = \overline{NN'} = \overline{AA'} = \overline{BB'} = \dots = a \dots \text{ (fig. 131).}$$

A translation in the plane possesses a whole series of properties which we shall examine and formulate in the form of almost self-evident theorems.

<u>Theorem 1.</u> <u>A parallel translation in the plane is a one-to-one transformation of the plane into itself.</u>

In fact, a parallel translation f assigns to each point M of the plane the point

$$M' = f(M)$$

of the same plane where M' is the end of the given vector a, whose origin is at the point M.

The construction of the point M', the image of

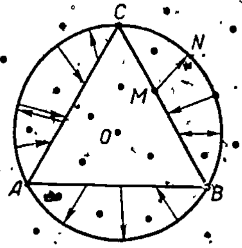


Fig. 130.

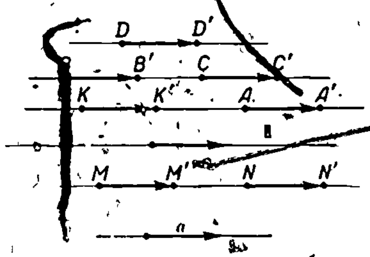


Fig. 131.

the point M, may be carried out as follows. Through the point M we pass a line of the pencil determined by the vector a, and we lay off on this line an oriented segment so that the vector $\overline{MM'} = a$.

The inverse image $f^{-1}(M')$ of the point M' consists obviously of the point M, so that

$$M = f^{-1}(M').$$

The transformation is one-to-one (14), q.e.d.

From here onward we shall not mention explicitly that we have to do with translations in the plane.

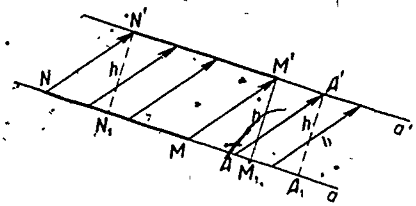


Fig. 132.

Theorem 2. A parallel translation transforms every line into a line which is parallel to it, i.e. it carries all points of the first line into points all lying on a line parallel to the first one.

To show this, let us draw from every point M', the image of a point M on the line a, a perpendicular $M'M_1$ to the line a (fig. 132). All these perpendiculars are equal to each other:

$$M'M_1 = N'N_1 = A'A_1 = \ldots = h.$$

This implies that the images M' of the points M of the line a, which of course lie on one side of the line a, are also at the same distance h from the line a, i.e. they lie on a line a' parallel to a, q.e.d.

The property expressed by this theorem gave this transformation its name.

Theorem 3. A parallel translation carries two parallel lines again into parallel lines, i.e. the image of two parallel lines consists of two parallel lines.

The theorem is an immediate consequence of the preceding one.

From theorem 3, there follows the following important theorem.

Theorem 4. A parallel translation carries every pencil of parallel lines into itself.

This is true because every line a of any pencil of parallels is transformed into a line a' parallel to the line a, i.e. the line a' belongs to the pencil of parallels determined by the line a, q.e.d.

A pencil of parallel lines remains unchanged or invariant, under every translation. Only the lines of the pencil are displaced, but the pencil as a whole goes over into itself.

The pencil of lines parallel to the vector a of a given translation also remains invariant under this translation, but in this case, every line of the pencil remains unchanged or invariant with respect to its position in the plane. Each such line goes over into itself; only its points are displaced, remaining however on the same line.

Lines whose position remains invariant under a transformation of the plane are called sliding lines. Under a translation the sliding lines are lines of a pencil of parallels.

A translation does not have any fixed points, unless it is the identity transformation.

Theorem 5. A parallel translation transforms a segment into an equal and parallel segment.

The points of the segment MN are transformed into the points of the segment M'N' (fig. 132). And since the figure MNN'M' is a parallelogram, we have MN = M'N', q.e.d.

It follows from this, that a parallel translation transforms every figure into an equal figure. For instance, a triangle, a square and a circle go over into the very same figures, being simply displaced with all their points by the given displacement vector.

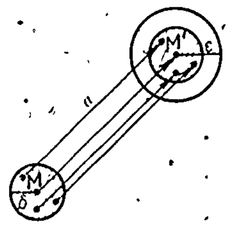


Fig. 133.

<u>Theorem 6.</u> <u>A parallel translation is a bicontinuous transformation.</u>

Let M' be the image of the point M. Let us choose an arbitrarily small ε-neighborhood of the point M' (fig. 133). For a δ-neighborhood of the point M we may select any open disk with the center at the point M and radius $\delta \leq \varepsilon$. In the same way we prove the continuity of the inverse transformation.

Thus a parallel transformation is a topological transformation of the plane into itself.

23. THE GROUP OF TRANSLATIONS IN THE PLANE WHICH ARE PARALLEL TO A LINE

Let us consider the set G of all translations of the plane π which leave invariant a given line a in this plane. Any translation in this set transforms also every line of the pencil parallel to a into itself.

Two translations of this set, performed one after the other, may be replaced by one translation belonging to the same set. If the first translation ϕ_1 carries over the point M of the line a into the point M', i.e. is determined by the displacement vector $\overline{MM'}$, while the second translation transforms the point M' of the line a into the point M" of the same line, i.e. is determined by the displacement vector $\overline{M'M''}$, then the translation ϕ which replaces ϕ_1 and ϕ_2 carries over the point M into the point M", and is characterized by the displacement vector $\overline{MM''}$. The resultant translation ϕ will be called the sum of the translations ϕ_1 and ϕ_2 and we shall write:

$$\phi = \phi_1 + \phi_2$$

The operation of assigning to two translations ϕ_1, ϕ_2 their sum ϕ will be called <u>addition</u> of translations.

The operation of addition of translations which leave invariant a line a is <u>closed</u>. This means that any two translations ϕ_1 and ϕ_2 belonging to a set G of <u>collinear</u> translations give, when added, a third translation ϕ, also belonging to the set G.

Thus we have a set G of translations in which there is defined one algebraic operation called <u>addition</u>.

It is easy to see that the sum of collinear translations obeys the associative law:

$$\phi_1 + (\phi_2 + \phi_3) = (\phi_1 + \phi_2) + \phi_3$$

Furthermore, <u>the identity transformation</u> ϕ_0 <u>belongs to the set of translations under consideration</u> and satisfies the condition

$$\phi + \phi_0 = \phi$$

for any translation ϕ belonging to the set.

The identity transformation is that transformation of the plane which carries every point of the plane into itself, i.e. leaves it in its original place.

Furthermore, for each translation ϕ in the given set of translations there exists a translation $(-\phi)$ in the same set such that their sum is the identity transformation ϕ_0:

$$\phi + (-\phi) = (-\phi) + \phi = \phi_0$$

This is so because if ϕ carries the point M into M', i.e. $\phi(M) = M'$, then the translation which takes M' into M is $(-\phi)$.

The translation $(-\phi)$ is called the <u>inverse</u> of ϕ.

The translation inverse to $(-\phi)$ is ϕ, i.e.

$$-(-\phi) = \phi.$$

Instead of

$$\phi + (-\phi) = \phi_0$$

we write

$$\phi - \phi = \phi_0$$

Every set of elements of any kind in which there is defined a one-valued operation assigning to any two elements a third element in the same set and obeying the conditions stated above is called, as we know, a group.

Let us remark that for sets of transformations some of the conditions for a group are fulfilled automatically. However we shall not discuss this here.

Let us also note that the commutative law holds for the group in question;

$$\phi_1 + \phi_2 = \phi_2 + \phi_1$$

Thus we may say that the set of collinear translations is a commutative group.

We shall establish a one-to-one correspondence between the elements of this group and the displacement vectors determining them.

We shall define the operation of addition of vectors by means of the addition of the corresponding transformations as follows: from two given vectors we pass to the corresponding translations and we add these translations. The vector corresponding to the translation-sum will be considered the sum of the two given vectors. The set of collinear vectors is, relative to the operation of addition, a group, isomorphic to the group of collinear translations.

Prove, as an exercise, that the set of all real numbers under the operation of ordinary addition is a group, isomorphic to the group of collinear vectors.

24. THE GROUP OF ALL TRANSLATIONS OF THE PLANE

In the set of all translations of the plane, we shall define an operation assigning to any two translations taken in a definite order a third translation in the following manner. A translation is completely determined by two points M and M', where M is any point in the plane and M' is the image of M. The displacement vector of this translation is $\overline{MM'}$.

Let the first translation ϕ_1 carry the point M into the point M', and the second translation ϕ_2 - the point M' into the point M''. Let us assign to the two translations ϕ_1 and ϕ_2 a third translation ϕ which carries the point M into the point M''. The translation ϕ is called the sum of the translations ϕ_1 and ϕ_2 and the operation of assigning ϕ to ϕ_1 and ϕ_2 is called addition. We write (1)

$$\phi = \phi_1 + \phi_2$$

For any three translations ϕ_1, ϕ_2, ϕ_3 we have:

$$\phi_1 + (\phi_2 + \phi_3) = (\phi_1 + \phi_2) + \phi_3$$

Let $\phi_1(M) = M'$, $\phi_2(M') = M''$, $\phi_3(M'') = M'''$

then in the left-hand member of (eq. 1), the transformation ϕ_1 carries the point M into M' while the transformation $\phi_2 + \phi_3$ takes M' into M'''. In the right-hand member of (eq. 1) the same transformation of the point M into the point M''' takes place differently. The transformation $\phi_1 + \phi_2$ carries the point M into the point M'', and then the transformation ϕ_3 takes the point M'' into the point M'''.

(1) It is easy to show that ϕ is the same no matter what point M is taken. --Translators.

The associative law holds.

The condition of the existence of a neutral element is also fulfilled, since the identity transformation ϕ_0 is a translation with a null displacement vector, and

$$\phi + \phi_0 = \phi$$

for any ϕ.

The requirement that an inverse element should exist for every given element is also filled.

Together with every translation ϕ, having the displacement vector $\overline{MM'}$, and taking the point M into the point M', there exists a translation $(-\phi)$, with displacement vector $\overline{M'M}$, which takes the point M' into the point M.

Thus $\phi + (-\phi) = (-\phi) + \phi = \phi_0$

All the group properties are thus verified for the set of all translations in the plane under the operation of addition which we have introduced. The set of translations in the plane is a group.

Since, furthermore,

$$\phi_1 + \phi_2 = \phi_2 + \phi_1 ,$$

the group of translations in the plane is commutative.

Together with the set of translations we also have the set of vectors in the plane. Moreover, between the translations and the vectors there has been established a one-to-one correspondence.

Let us introduce an operation of addition of vectors in such a way as to establish an isomorphism. Namely, to two vectors a and b we assign a third vector c in the following manner.

To the vector a there corresponds a translation ϕ_a, to the vector b - a translation ϕ_b. We shall call the vector c corresponding to the translation $\phi_a + \phi_b$ the sum of the vectors a and b and we write

$$c = a + b$$

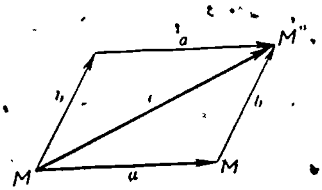


Fig. 134.

Alternatively, the addition of vectors may be defined by means of the triangle or parallelogram rule (fig. 134).

It is easy to see that the set of all vectors in the plane is, relative to addition, a commutative group, isomorphic to the group of parallel translations of the plane.

Recalling from algebra the well-known geometric interpretation of addition of complex numbers, we see that the set of all complex numbers is, relative to addition, a commutative group, isomorphic to the group of vectors and the group of translations of the plane.

We shall indicate certain subgroups of the group of translations of the plane.

Let us construct the following subgroup. Let us take an arbitrary but constant translation ϕ and add it to itself n times.

$$\underbrace{\phi + \phi + \ldots + \phi}_{n \text{ times}}$$

Further, let us introduce the notation

$$\underbrace{\phi + \phi + \ldots + \phi}_{n \text{ times}} = n\phi$$

We introduce an analogous notation for the inverse transformation

$$(-\phi) + (-\phi) + \ldots + (-\phi) = -n\phi$$

The translation $(-n\phi)$ is inverse to the translation $n\phi$, since, as we may easily verify,

$$n\phi + (-n\phi) = \phi_0 \quad ,$$

where ϕ_0 is the neutral element.

Further, let us define 0ϕ by the equation

$$0\phi = \phi_0 \; .$$

Thus we obtain a set of translations of the type $n\phi$. Here

$$n = 0, \pm 1, \pm 2, \pm 3, \ldots \pm n, \ldots$$

We shall show that this set of translations, under the same operation of addition as was defined for the set of all translations of the plane, is a group. It is easy to verify that for any integers p and q

$$p\phi + q\phi = (p+q)\phi$$

This implies, that the sum of two arbitrary translations of the type $n\phi$ is a translation of the same type; but this constitutes the condition of closure of addition.

It is not necessary to verify the condition of associativity; this condition holds true for all translations.

The condition of the existence of a neutral element is fulfilled, since for $n = 0$ we have $0\phi = \phi_0$.

The condition of the existence of an inverse element is also observed.

$$n\phi + (-n\phi) = \phi_0.$$

Moreover, addition is commutative.

We have constructed a subgroup of the group of all translations of the plane.

This subgroup is called the subgroup generated by a given translation ϕ.

If we should establish a one-to-one correspondence between the numbers belonging to the group, relative to addition, of all integers, and the translations of the subgroup just constructed

$$n \longleftrightarrow n\phi,$$

then this correspondence, as it is easy to see, will be an isomorphism, and the groups will be isomorphic to each other.

Groups, isomorphic to the group of integers are called infinite cyclic groups.

We shall call two points A_1 and A_2 equivalent relative to a subgroup of translations generated by a given translation ϕ, if there exists a translation in this subgroup which carries the point A into the point A_1. We shall denote equivalence by the sign $\sim$:

$$A \sim A_1$$

It is easy to verify that the following hold true:

1. Reflexiveness: $A \sim A$.
2. Symmetry: if $A \sim B$ then $B \sim A$.
3. Transitivity: if $A \sim B$ and $B \sim C$, then also $A \sim C$.

Thus a subgroup generated by a translation ϕ, divides all points of the plane into classes of equivalent points.

If we should regard all equivalent points as one ("paste them together"), we should obtain from the plane a new figure. This figure may be represented by any set of points in the plane consisting of one representative point from each class of equivalent points. No two points of this figure are equivalent.

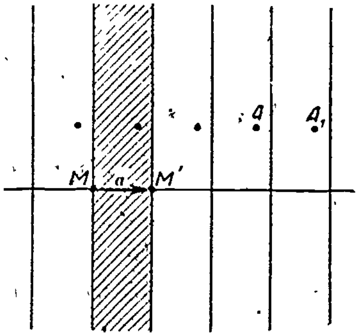


Fig. 135.

If we take the displacement vector $a = \overline{MM'}$ (fig. 135) of the translation ϕ which generates the subgroup and draw parallel lines through its ends M and M', we obtain a strip between the parallels in which there are no two points equivalent to each other. The boundary of this strip consists of pairs of equivalent points (such as M and M'), by identifying which we may obtain a cylinder.

If we add to the interior points of the strip only one of the lines which form its boundary, we obtain a region of the plane which is "swept out" entirely by every transformation of the subgroup, i.e. every point of the region is carried over into a point outside the region.

We shall now construct the group generated by two non-collinear translations ϕ_1 and ϕ_2.

Let us consider the set of all translations of the type

$$\phi = p\phi_1 + q\phi_2$$

where $p, q = 0, \pm 1, \pm 2, \ldots, \pm n$; the meaning of the notation $p\phi_1$ and $q\phi_2$ is clear from the foregoing.

It is easy to verify that the set of indicated translations forms a subgroup of the group of all translations with respect to the addition of translations.

As before, we shall call two points A and A_1 <u>equivalent</u> relative to the subgroup of translations generated by the given non-collinear translations ϕ_1 and ϕ_2 if there exists a translation in this subgroup which takes the point A into the point A_1. The relation of equivalence possesses the properties of reflexiveness, symmetry and transitivity.

The entire set of points in the plane is subdivided into classes of equivalent points.

If we regard all equivalent points as one ("paste them together"), we shall obtain from the plane a new figure. This figure consists of points of the plane such that among them there is no pair of equivalent points, and it is topologically equivalent to a torus.

If we take the displacement vectors $a = MM'$ and $b = MM''$ (fig. 136) of the translations ϕ_1 and ϕ_2 which generate the subgroup and examine the parallelogram with the sides MM' and MM'', we obtain a set of points, interior to the parallelogram,

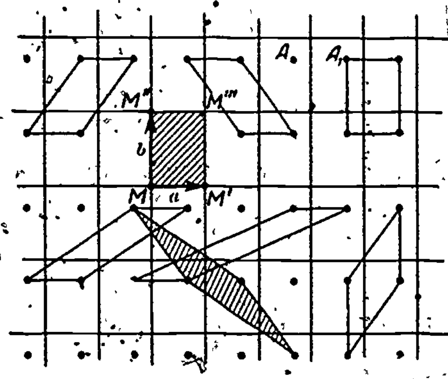


Fig. 136.

among which there are no two equivalent points.

Adding to these interior points the points of the segments MM' and MM", we obtain a set of points which it is impossible to enlarge without obtaining in the enlarged set equivalent points.

The region of the plane obtained, is entirely "swept out" by each translator of the subgroup, i.e. each point of the region goes over into an exterior point.

A set of all points equivalent to a given point forms a so-called lattice (fig. 136). Similar kinds of lattices play an important role in crystallography and in the geometrical theory of numbers.

If to the group of complex numbers, i.e. numbers of the type $p + qi$ where p and q are integers, we assign the set of translations of the type $p\phi_1 + q\phi_2$ where ϕ_1 and ϕ_2 are not collinear, then the correspondence obtained is an isomorphism

$$p + qi \leftrightarrow p\phi_1 + q\phi_2$$

This means that the group of integral complex numbers is isomorphic to the subgroup of translations generated by two non-collinear translations.

The same subgroup may be generated by different generating translations, for instance $\phi_2 - \phi_1$ and $\phi_2 - 2\phi_1$. The corresponding region of the plane, containing no equivalent points, will be a two-dimensional parallelogram with two non-parallel sides excluded (fig. 136).

If we use the notation

$$\phi_1^* = \phi_2 - \phi_1 \ ; \ \phi_2^* = \phi_2 - 2\phi_1 \ ,$$

then the set of translations

$$\rho\phi_1^* + g\phi_2^* \ ,$$

where ρ and g are integers, forms a group which coincides with the group of translations $\rho\phi_1 + g\phi_2$.

It is possible to set up a one-to-one correspondence between the translations of this group

$$\rho\phi_1 + g\phi_2 \longleftrightarrow \rho\phi_1^* + g\phi_2^* \qquad (*)$$

which is an isomorphism.

An isomorphism of a group with itself is called an <u>automorphism</u> of the group.

One may look upon the correspondence (*) as a transformation of one translation into another. Here the transformed elements are not points but translations.

Examine, as an exercise, the set of all automorphisms of the given group and, defining the addition of automorphisms as the addition of transformations of translations, show that the set of all automorphisms of the given group is a group.

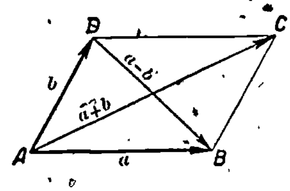


Fig. 137.

In conclusion we note that the parallel translations of a plane are closely connected with the properties of a parallelogram. To each parallelogram there corresponds a subgroup generated by translations whose displacement vectors a and b are determined by the sides of the parallelogram (fig. 137). The vector a represented by one side AB is the displacement vector which carries the second side AD into the position BC. This is connected with the fact that opposite sides of a parallelogram are equal and parallel. The opposite angles of a parallelogram are equal, while the angles on the same side add up to the sum of 180°. This is connected with the fact that translations transform any line into a line parallel to it.

It is possible to "pave" the entire plane with parallelograms. In fig. 136 this is shown for a rectangle; however, we may carry out a paving of the plane with each of the indicated parallelograms corresponding to various pairs of generating translations of the subgroup, and moreover, given two such "pavings," determined by two different parallelograms having one vertex A in common, then the lattice consisting of the vertices of the parallelograms will be the same for both "pavings." The parallelograms contain no points of the lattice (points equivalent to A) other than the vertices. The areas of such parallelograms are all equal (fig. 136). Every such parallelogram is cut by the lines equivalent to the lines MM' and MM" into parts out of which, by means of an appropriate parallel translation (belonging to the subgroup) of each part, it is possible to put together the rectangle MM'M"M'".

The verification of these statements is left as an exercise.

25. INVARIANTS AND INVARIANT PROPERTIES OF FIGURES

If a figure is subjected to any transformation, then some of its attributes, or characteristics, will change while others will remain unchanged. For example, under the topological transformation of a linear triangle into a circle (fig. 101) many attributes or properties of the triangle are changed. Among these are the straightness of the sides of the triangle, the length of the sides, the angles, the area etc. The characteristic of the triangle of being a closed curve without self-intersection remains unchanged.

General topological transformations (15) of figures are extraordinarily far-reaching. Very few properties of figures remain unchanged under topological transformations, but these most stable characteristics are also the deepest-lying properties of the figures.

A section of a right circular cylinder (fig. 27) and a section of a right circular cone (figs. 29, 31, 33) may be regarded as transformations of a circle by means of parallel or central projection. Each point of the circle, with the exception of not more than two, is transformed into a point on the conic section, lying on the same generator of the cylinder or cone.

Under the parallel projection of a circle into an ellipse and the concomitant projection of the entire plane of the circle into the plane of the ellipse, the straightness of lines is preserved; the parallelism of lines is preserved. The ratio of segments lying on the same line or on parallel lines also remains unchanged, in particular the image of the midpoint of a segment is the midpoint of the image of the segment.

The ratio of non-parallel segments is not preserved. Angles between lines also are not preserved, in particular, perpendicular lines are projected into lines which, generally speaking, are not perpendicular; the lengths of segments are, generally, changed.

Let us consider, for example, the perpendicularity of a diameter of a circle to a family of parallel chords and the fact that the diameter divides each of these chords into two equal parts (fig. 138).

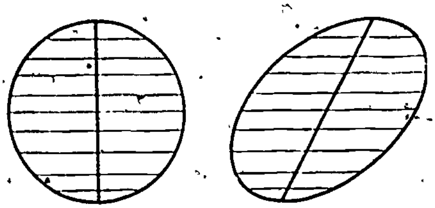


Fig. 138.

Under parallel projection the relation of perpendicularity between the diameter and the chords is not preserved, while the property of the diameter of being the locus of the midpoints of the parallel chords remains unchanged.

Having considered these examples, we shall give the following definitions:

The attributes or properties of figures which are not destroyed by any of the transformations of a system are called invariant properties relative to this system of transformations.

In particular if the system of transformations is a group, we speak of the invariant properties of a figure relative to the group of transformations.

If some magnitude, for instance a length, an angle or an area is associated with a figure in such a way that it remains unchanged under all transformations of a group, then this magnitude is called an invariant of the figure (1) relative to this group.

Thus, for instance, under all parallel projections the parallelism of straight lines is an invariant property, so is the ratio of parallel segments. Angles are invariant under stereographic projections and rotations of the sphere (page 174).

Let us consider in particular the invariant properties and invariants of the group of translations in the plane. The distance between two points is an invariant of the group of translations. As a consequence, we obtain from this the invariance of angles between lines and of the areas of figures. The attributes of a figure of being a point, being a line etc. are properties which are invariant under translations. The attribute of the sides of an angle consisting of their having given directions is also an invariant property relative to all translations. The attribute of a line consisting of the fact that it occupies a given place in the plane, i.e. the attribute of being fixed, is not an invariant property relative to the entire group of translations, but invariant relative to a collinear subgroup.

The position in space of each of the lines of a parallel pencil is an invariant property relative only to the group of translations collinear to this pencil.

(1) Footnote: An illustration of what is meant is the invariance of distance under translation proved later in the text. The distance is a magnitude associated with figures consisting of two points. --Translators.

It is namely such a pencil that we called invariant relative to a group of collinear translations. This pencil consists of the sliding lines of each translation of the group.

26. THE METHOD OF PARALLEL TRANSLATIONS IN THE SOLUTION OF CONSTRUCTION PROBLEMS

Frequently, in solving construction problems, it is helpful to carry over a figure or part of a figure by means of a translation into a position more convenient for the construction of a given plane locus, and after carrying out the construction to restore the previous position of the figure by means of the inverse transformation.

Problem. To place a segment AB, of given length l and parallel to a given line c, between two given lines a and b (fig. 139).

If we translate the line a (using the displacement vector $\pm l$ collinear with the line c and of length l), we obtain lines a_1 and a_2 whose intersections with the line b yield the points B and B'. The segments BA and $B'A_1$ parallel to the line C are the desired segments.

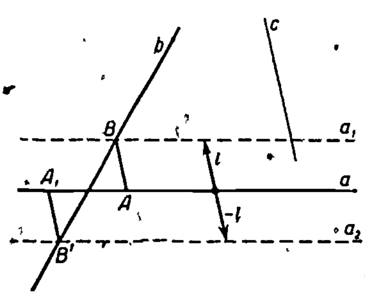


Fig. 139.

If the lines a and b intersect, the problem has two solutions. If the lines a and b are parallel then the problem either has no solution or an infinite number of solutions.(1)

Problem. Given two parallel lines a and b and two points A and B on different sides of the strip bounded by the parallels a and b. To find the shortest path from B to A if the part of the path lying between a and b is to be parallel to a given line c (fig. 140).

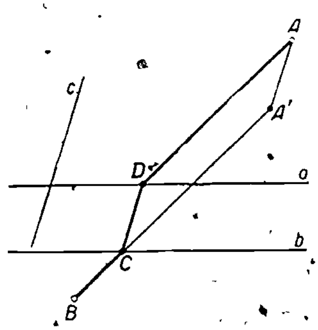


Fig. 140.

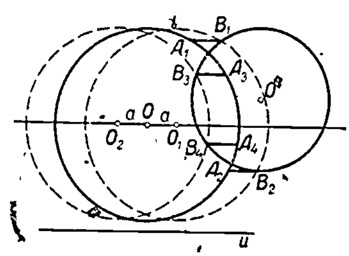


Fig. 141.

If we translate the point A and the line a collinearly with c so that line a coincides with line b, then the point A will occupy position A'. Joining the point B with the point A', we obtain the segment BA' equal to the desired path minus the part between the parallels. Let C be the point of intersection of the segment BA' with the line b. Carrying over by

(1) Footnote: This is a kind of "constructibility" where we assume that we can make a translation of any sort whenever we wish. This "ability to effect translations" is one of the basic tools of this kind of "constructibility", just as the ability to duplicate a distance by means of dividers is one "tool" in the familiar Euclidean "constructibility".
--Translators.

means of the inverse translation $\overline{(A'A)}$ the segment CA' into the position DA, we obtain the desired path BCDA from the point B to the point A.

Problem. To place, between two given circles O and O', a segment AB, of given length a and having the same direction as a given line u (fig. 141).

Let the end point A of the segment AB be on the circle O. To find the point B, we translate the circle O through the distance a parallel to the line u. For this, it is sufficient merely to translate the center O of the circle, and, using the same radius, draw the circles O_1 and O_2. The intersection of the circles O_1 and O_2 with the circle O, determines the end point B of the segment.

The problem has no more than four solutions provided the center of O' does not coincide with the centers of O_1 or O_2.

Problem. To construct a quadrilateral, knowing the four sides and the angle between two opposite sides.

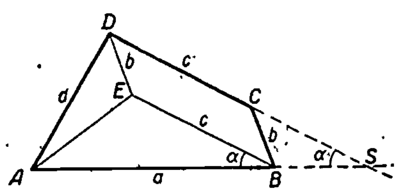


Fig. 142.

Suppose ABCD is the required quadrilateral (fig. 142), AB = a, BC = b, CD = c, DA = d and $\angle$ BSC = α.

Let us, by means of a parallel translation, bring the side CD into position BE. Let us find DE = b, BE = c, $\angle$ ABE = α. Thus it is possible to construct the triangle ABE having two of its sides equal to a and c and the angle ABE between them equal to α. We then construct upon the side AE the triangle AED, knowing its sides.

The investigation of the conditions under which a solution is possible is left as an exercise.

27. THE GROUP OF TRANSLATIONS IN SPACE

After the detailed study of translations in the plane, it is not difficult, following the same plan, to consider translations in space.

For translations in the plane it would have been helpful to think of two coinciding planes and to imagine one of these moving as a unit along the second. In exactly the same way one may think of two copies of space, a movable space and a fixed space, the first being free to move in any direction within the second.

The geometrical meaning of such displacements is merely that we assign, according to some definite rule or law, to each point M in space a corresponding point M'.

Let us consider the set of all lines in space parallel to a line a. This set of lines, including also the line a, is called, as we know, a bundle of parallel lines. The line a does not play any special role in the determination of the bundle. Because of the symmetry and transitivity of parallelsim, any line belonging to the bundle determines this bundle. The entire set of lines in space is subdivided, in this way, into classes. Each class consists of all the lines of a parallel bundle. All the lines of a bundle have one and the same direction. Since every line belongs to only one bundle, a line has only one direction.

Taking two points A and B on a line, we may establish in it an orientation determined by the vector $\overline{AB}$ or $\overline{BA}$. A line may be oriented in two ways. An oriented line is frequently called an <u>axis</u>.

Just as in the plane, we consider vectors as equal if the segments determining them are parallel and have equal length and the same orientation.

A translation or parallel translation of space is such a transformation of space into itself under which the displacement vectors of all points in space are equal to each other:

$$\overline{MM'} = \overline{NN'} = \overline{AA'} = \overline{BB'} = \ldots = a = \ldots$$

Theorems 1 through 6 (22) which express the basic properties of translations hold here also. Also, just as in the plane, we define, in the set of all translations, an algebraic operation of addition of translations. To the addition of translations there corresponds the addition of the corresponding displacement vectors:

$$\overline{MM''} = \overline{MM'} + \overline{MM''}$$

The identity transformation of space into itself is taken as the neutral translation.

It is not difficult to verify that all the conditions of a group are fulfilled (20) and thus establish that the set of all translations in space is a group.

The set of all translations leaving in place every line of a bundle of parallel lines, is a subgroup of the group of all translations. Such a subgroup is called a group of collinear parallel transformations in space.

A group of collinear translations is isomorphic to the group of real numbers relative to addition. This may be shown the same way as for collinear translations in the plane.

A bundle, each of whose lines remains unchanged under a subgroup of collinear translations is called an invariant bundle of lines relative to this group of collinear translations. The lines of this bundle, under the indicated translations only "slide upon themselves" and are called lines of sliding.

As an exercise, let the reader characterize the set of planes which also remain unchanged under all collinear translations of a collinear subgroup.

Cylindrical surfaces whose straight line generators belong to the invariant bundle also "slide upon themselves" under the indicated translations and are called sliding surfaces. The parallelism of planes as well as the parallelism of lines possesses the properties of symmetry and transitivity.

The entire set of planes in space is subdivided into classes of parallel planes. Every class consists of planes forming a parallel pencil. We say that all planes of a parallel pencil possess the same two-dimensional direction.

Any plane of a parallel pencil may serve as a representative of the class, and the plane α, by means of which the pencil is determined, plays no special role.

The set of all translations leaving in place every plane of a pencil of parallel planes, is a subgroup of the group of all translations. This subgroup is called a group of coplanar parallel translations in space.

It is easy to establish that a group of coplanar translations is isomorphic to the group of all translations in the plane considered earlier, and also, as we have seen, to the group of all complex numbers, relative to addition.

A pencil, every one of whose planes remains unchanged under each member of a group of coplanar translations is called an invariant pencil of planes relative to this subgroup of coplanar translations. The planes of such a pencil only "slide upon themselves" under the indicated translation and are sliding surfaces under each translation.

The distance between two points, the angle between two lines, the angle between two planes, the angle between a line and a plane are invariants of the group of translations.

We shall see below that areas and volumes of figures are also invariants of this group of transformations.

If we are given two oriented bundles of parallel lines, we say that any oriented line of the first bundle forms equal angles with any oriented line of the second bundle. The angles between two directions is frequently mentioned. However, there are four such angles, so that this angle is not uniquely determined. An analogous situation arises with the angle between planes; there are four angles, and without additional conditions there is no unique determination.

If we are given an oriented bundle of parallel lines and a pencil of parallel planes then each oriented line of the bundle forms equal angles with each plane of the pencil. The angle between an oriented line and a plane is uniquely defined as the angle between this line and its oriented projection.

Let us consider the subgroup G of translations, generated by three non-coplanar translations ϕ_a, ϕ_b and ϕ_c with displacement vectors a, b and c, i.e. translations of the form

$$\phi = \rho\phi_a + g\phi_b + r\phi_c$$

where ρ, g, r are integers.

Just as in the plane, we shall consider two points in space as equivalent relative to the subgroup G if one point is carried into the second by some translation of this subgroup.

Every point is equivalent to itself, $A \sim A$. If $A \sim B$, then $B \sim A$ - the property of symmetry. If $A \sim B$ and $B \sim C$, then $A \sim C$ - the property of transitivity. Thus the set of all points in space is subdivided into classes of equivalent points relative to the subgroup G.

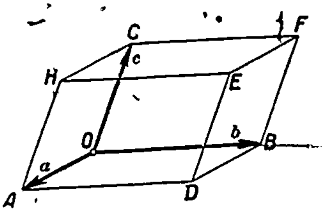


Fig. 143.

It is easy to see that the parallelepiped with the edges a, b and c (fig. 143) does not contain among its interior points even one pair of equivalent points. If we add to the interior points of the parallelepiped the points of the faces OADB, OBFC, OANC, we obtain a region of space which is completely "swept out" by each transformation of the subgroup and which can not be enlarged without introducing into the enlarged region equivalent points.

Opposite congruent faces of the parallelepiped are transformed one into the other by those translations of the subgroup G which have a, b, and c as displacement vectors. Thus, for example, the face OBFC is transformed by the translation ϕ_a into the face ADEN. All ~~the edges~~ between two parallel faces are transformed one into the other by translations of the subgroup G. For instance, the edge OC is transformed by the translation $\phi_a + \phi_b$ into the edge DE.

The entire space may be "paved" by such half-open parallelepipeds, each consisting of three faces and all interior points.

For any given pair of parallelepipeds, or cells, in this "paving" there exists a translation in the subgroup G which takes the first cell into the second.

The vertices of all parallelepipeds of a given subdivision of space form a space lattice.

The subgroup G may be generated not only by the translations ϕ_a, ϕ_b, ϕ_c, but also by other translations of this subgroup. To each system of three generating translations there corresponds a particular subdivision of space into cells. However a point set which forms a "lattice" under one subdivision also forms a "lattice" under any other.

As an exercise, it is suggested that the student find the entire set of systems of three translations of the subgroup G, which generate this subgroup.

Chapter V.

ROTATION.

In chapter V we take up the rotations of a plane around a point, the rotations of space around an axis and the rotations of space around a point.

In connection with the group of rotations of space around an axis we shall study the elementary properties of surfaces of revolution.

We shall define those transformations of the plane and of space which are known as motions.

The connection of motions with translations and rotations will be established.

In 30. we shall consider the application of the method of rotation to the solution of construction problems.(1)

28. THE TRANSFORMATION OF A PLANE INTO ITSELF BY MEANS OF A ROTATION AROUND A POINT.

Let O be a given point in the plane (fig. 144). By a rotation of the plane around the point O, through a given angle ϕ, we mean a transformation of the plane into itself under which to every point M there is assigned a point M' such that the conditions

$$OM' = OM, \angle MOM' = \phi.$$

are fulfilled. Rotations through angles which are multiples of 2π are considered to be identical. The point O goes over into itself, i.e. remains fixed.

Measuring the angle ϕ in radians, we obtain a real number x. To each rotation through an angle ϕ we assign the set of numbers

(1) That is, we shall consider construction problems where the possibility of effecting an arbitrary rotation is one of the admissable "tools" of constructibility.
--Translators.

$x+2k\pi$, where k is any integer. We introduce the usual convention with regard to the sign of angles: we consider counterclockwise rotations as positive.

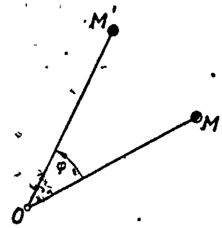


Fig. 144.

The addition of rotations having the same center is carried out in the same manner as for transformations in general. If a rotation through the angle ϕ_1 carries any point M into M', and a rotation through the angle ϕ_2 takes the point M' into the point M", then the sum of these rotations takes the point M directly into the point M". This sum of rotations is again a rotation through the angle $\phi_1 + \phi_2$.

It is easily seen that the set of rotations of the plane around a point O is a group with respect to addition of rotations.

Among these rotations there is a neutral element - rest, or the rotation through an angle equal to zero. For each rotation through an angle ϕ there is an inverse rotation through an angle $(-\phi)$. The associative law, as may be easily verified, holds for transformations of any kind. Furthermore, the group of rotations of the plane around a point O is commutative.

The set of real numbers can be mapped into the set of rotations, assigning to each real number x a rotation through an angle whose radian measure is x.

The inverse image of the angle ϕ under this mapping is the set of numbers of the type:

$$x + 2K\pi$$

where K is any integer.

Each such inverse image forms a class of numbers which are assigned to the given angle ϕ.

The set of such classes can be mapped one-to-one onto the set of rotations; to each class of numbers there is assigned one rotation, and each rotation has only one class as its inverse image.

The set of the above mentioned classes of real numbers is a group, isomorphic to the group of rotations.

The addition of classes is carried out according to the following rule: We take an arbitrary number from each of the classes to be added; their sum is a number which belongs to some third class which we now regard as the sum of the two classes. This rule determines uniquely the class which is the sum.

The set of real numbers of the form $2K\pi$, where K is any integer, is the neutral element in the group of classes and is itself a group with respect to addition:

$$2K_1\pi + 2K_2\pi = 2(K_1 + K_2)\pi .$$

Let us consider now some properties of a figure which are invariant under rotation. A point is transformed into a point. The property of being a point is invariant. However, the position of each point, with the exception of the center of rotation 0, changes. The unchanged center of rotation O will be called a point invariant with respect to position or simply an invariant point.

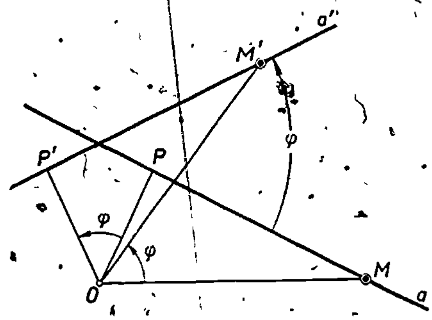


Fig. 145.

Points lying on a line a (fig. 145) go over into points lying on a line a'. In other words, a rotation transforms a line into a line. Proof: Let OP be perpendicular to the line a. All points of the segment OP are transformed into points of the segment OP'. Let us pass through the point P' a line a' perpendicular to the line OP'. The line a' is the image of the line a, since any point M of the line a goes over, under this rotation, into a point M' on the line a' such that P'M' = MP. The last assertion follows from the equality of the triangles OPM and OP'M'. (Since the triangles OPM and OP'M' are right triangles, and OP = OP', PM = P'M'.) It also follows from this that the angle ϕ between the line a and its image a' is the angle of rotation.

<u>The distance between two points is an invariant of the group of rotations of a plane around a point.</u>

The segment AB passes, under a rotation through an angle ϕ, into the segment A'B' (fig. 146). The equality of the segment AB and its image A'B' follows from the congruence of the triangles OAB and OA'B', which, in turn, follows from the equalities:

$$OA = OA', \ OB = OB' \text{ and } \angle AOB = A'OB'.$$

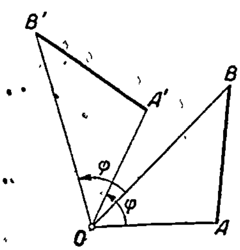


Fig. 146.

From the invariance of the lengths of segments there immediately follows the invariance of angles under rotations around a point.

It is now easy to verify the congruence of a figure and its image.

A circle, for instance, is transformed into a circle of the same radius, where the center of the image is the image of the center. Circles whose centers do not coincide with the center of rotation O change their position in the plane. Circles with center at the point O preserve their position in the plane under any rotation of the group, and will only "slide along themselves." The concentric circles with center at the point O are called invariant circles of the rotation group, or curves of sliding under the rotations of the plane around the point O.

Every line passing through the center of rotation O goes over into a line also passing through the center O. However, the entire pencil of lines having its center at the center of rotation

O, is preserved under any rotation of the group. The pencil of lines whose center coincides with the center of rotation O is called the <u>invariant pencil</u> of the group of rotations of the plane around the point O.

Intuitively it is possible to imagine the transformations of the plane into itself which are called rotations as taking place in the following manner. Suppose we have two planes coinciding at all their points; one plane is fixed while the other may slide upon the first. The center of rotation O remains in the same place and does not change its position as do the other points of the movable plane. The movable plane rotates around the point O while maintaining its identity as an absolutely unchanged system of points.

29. THE GROUP OF MOTIONS IN THE PLANE

If, in addition to the point O, we require that one more point O_1 should remain fixed, then it is no longer possible to move the second plane relative to the first and still have the planes touching at all points. We exclude the possibility of "turning over" the movable plane onto its other side around the line OO_1.

If we agree to exclude all such "turning over", then we may say that the position of the movable plane relative to the fixed plane is fully determined by the position of any segment AB of the movable plane.

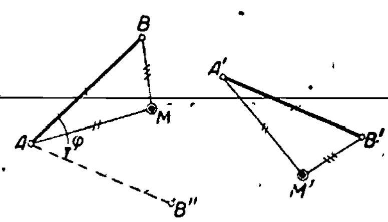


Fig. 147.

With this in mind, we may construct the following mapping of the plane onto itself. Let us take in the plane any two equal segments AB and A'B' (fig. 147) and let us assign the point A' to the point A and the point B' to the point B. To an arbitrary point M in the plane we assign a point M' also in the plane, such that

$$AM = A'M', \quad BM = B'M'$$

and in addition, the triangles AMB and A'M'B' are similarly oriented. By the latter we mean that the vertices A, M, B and

A', M', B' are either arranged counterclockwise in both triangles or clockwise in both triangles. If the point M happens to lie on the line AB then its image M' is uniquely determined without any specifications concerning orientation. It is clear that under this mapping there need not be any fixed points.

Given a mapping such as we have described, constructed by means of segments AB and A'B'. Let the segment C'D' be the image, under this mapping, of the segment CD. Then the mapping constructed by means of the segments CD and C'D' is identical with the mapping constructed by means of segments AB and A'B'. Intuitively it is easy to imagine the indicated mapping as sliding the movable plane over the fixed plane until the segment AB coincides with the segment A'B' and consequently every point M coincides with its image M'.

The mapping of the plane onto itself just considered is called a motion. It is easy to see that a motion is a one-to-one bi-continuous transformation. It is also easy to verify that the set of all motions in the plane forms a group.

The sum of two motions, i.e. the transformation consisting of carrying through one motion and then carrying through the second, is again a motion.

The algebraic operation of addition of motions is closed.

The associative law holds, as it does for all mappings of the plane onto itself.

The requirement of the existence of a neutral element is fulfilled, since the identity transformation is a motion (where the segment A'B' coincides with the segment AB).

The fulfillment of the requirement of the existence, for each motion, of an inverse motion follows from the fact that there exists a motion which carries the segment A'B' back into the segment AB (fig. 147).

The translations in the plane and the rotations of the plane around a point are motions. The group of all translations in the plane is a subgroup of the group of all motions in the plane. The group of rotations of the plane around a given point is also a subgroup of the group of all motions.

Any motion of the plane may be obtained by performing successively one rotation and one translation.

As proof, let us turn the plane around the point A of the segment AB (fig. 147) through an angle ϕ until the segment AB occupies the position AB", parallel to the segment A'B', and afterwards perform the translation whose displacement vector is $\overrightarrow{AA'}$. The sum of the indicated rotation and translation is equal to the given motion. Of course one, and even both, of these transformations may consist of the identity transformation.

Having shown that a motion may be decomposed into a rotation and a translation, it is easy to establish the invariants and the invariant properties of figures relative to the group of all motions. The basic invariant is an invariant of two points, namely the distance between two points.

An assertion even stronger than the one previously made is true. Namely: Every motion in the plane is either a translation or a rotation around a point.

Let the motion carry the plane together with the segment AB into the position A'B' (fig. 148).

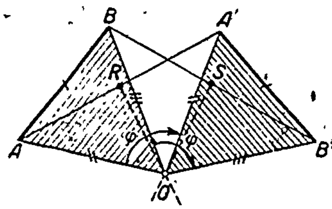


Fig. 148.

If there exists a center of rotation O, then we must have OA = OA' and OB = OB'. The first equation shows that the point O must belong to the locus of points equidistant from the points A and A', i.e. the point O must lie on the perpendicular bisector RO of the segment AA'. Similarly, from the second equality it follows that the point O lies on the perpendicular bisector SO of the segment BB'.

The perpendicular bisectors RO and SO either intersect at the point O, coincide or are parallel.

Let us examine these three possibilities.

1. The perpendicular bisectors RO and SO intersect at the point O (fig. 148). The motion is a rotation through the angle

$$\phi = \angle AOA'.$$

That the angle AOA' is equal to the angle BOB', follows from the fact that in the triangles ABO and A'B'O' the three corresponding sides are equal.

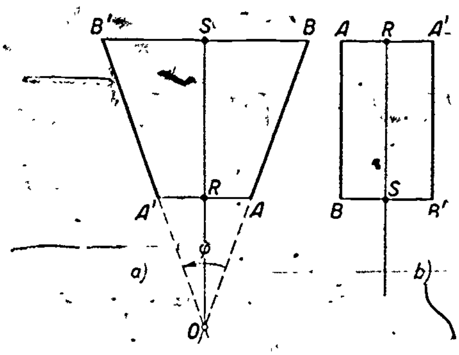


Fig. 149

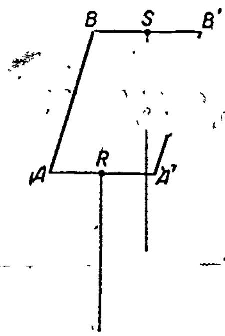


Fig. 150.

2. If the perpendicular bisectors RO and SO coincide, then the figure ABB'A' is a trapezoid with parallel sides AA' and BB' (fig. 149a, b). In this case we have a rotation through the angle $\phi = \angle AOA' = \angle BOB'$ around the point of intersection O of the lines AB and A'B' provided AB and A'B' are not parallel. Otherwise, we have a translation (fig. 149b).

3. The perpendicular bisectors RO and SO are parallel, and a center of rotation O does not exist. However, this is only possible in case the directed segments $\overrightarrow{AB}$ and $\overrightarrow{A'B'}$ (fig. 150) are parallel and have the same orientation; the motion is therefore a translation with the displacement vector $\overrightarrow{AA'}$.

The truth of our assertion is established. It follows that the set of all translations and all rotations of the plane around every one of its points is the group of motions.

The set of all translations of the plane is, as we have emphasized, a subgroup of the group of motions; however the set of all rotations of the plane around every one of its points is not a group.

One can easily prove this by means of the following example. The segment O_1A, together with the movable plane "attached" to it, is rotated around its end point O_1 through the angle ϕ (fig. 151). The segment O_1A will then occupy the position O_1O_2. A second rotation is performed around the center O_2 through the angle $(-\phi)$; the segment O_1O_2 thus finally occupies the position BO_2.

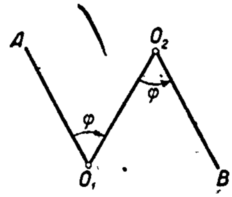


Fig. 151.

The result of performing the two rotations is therefore a translation with the displacement vector $\overrightarrow{AO_2} = \overrightarrow{O_1B}$, and not a rotation around a point.

The operation of addition of rotations is not closed in the set of all rotations of the plane, and consequently the set of all rotations with all possible centers does not turn out to be a group. As an exercise, it is suggested to prove that the group of motions is not commutative.

30. THE METHOD OF ROTATION IN THE SOLUTION OF CONSTRUCTION PROBLEMS

In solving construction problems, it is often convenient to rotate a figure or part of a figure around some point through some angle, thus making possible the construction of the figure or its part. By the inverse rotation it is then possible to return the constructed figure to the original position.

Other methods of applying rotations are also possible.

The method of rotation is best explained by means of examples.

Problem. To construct a triangle, given two sides $AB = c$, $AC = b$ and the median $AM = m$.

Rotating the triangle AMB through an angle equal to π around the point M, we bring it into position DMC (fig. 152), where the segment MD is the continuation of the segment AM. It is easy to construct the triangle ADC from its three known sides ($AD = 2m$). The solution is possible under the condition $2m < b + c$.

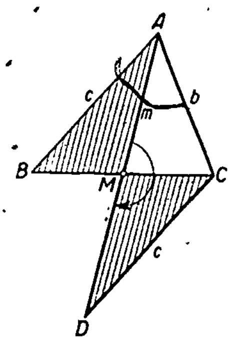


Fig. 152.

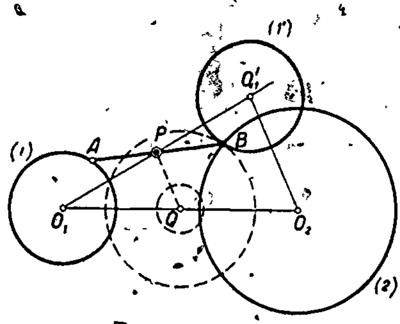


Fig. 153

<u>Problem</u>. <u>To draw a line through a given point P so that its segment AB between two given circles is divided by the point P into two equal parts.</u>

Let us turn circle (1) through an angle equal to π around the point P so that it falls in position (1') (fig. 153). Since it is required that AP = PB, the desired point A on circle (1) goes over into point B on circle (1'). But point B must lie on circle (2), consequently, the desired point B is the point of intersection of circles (1') and (2).

Circles (1') and (2) will intersect provided

$$R_2 - R_1 \leq O'_1O_2 \leq R_2 + R_1$$

where R_1 and R_2 are the radii of circles (1') and (2) respectively.

Dividing the members of the inequality by 2, we get:

$$\frac{R_2 - R_1}{2} \leq PQ \leq \frac{R_2 + R_1}{2}$$

Here Q is the midpoint of the segment O_1O_2 and hence $PQ = \frac{O'_1O_2}{2}$. Describing two circles with center at Q and radii $\frac{R_2 - R_1}{2}$ and $\frac{R_2 + R_1}{2}$, we find that if the point P lies inside the ring between these two concentric circles, we have two solutions; if the point P lies on one of these circles - one solution; if the point P lies outside of the indicated ring and not on its boundary then there is no solution.

<u>Problem</u>. <u>To construct an equilateral triangle so that its vertices lie on three given parallel lines.</u>

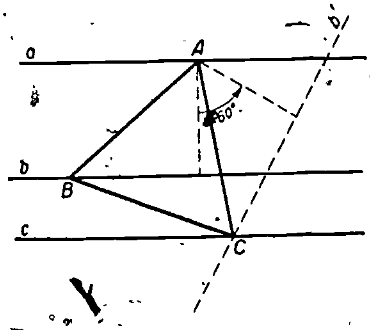


Fig. 154.

Let a, b, c be the three given parallel lines (fig. 154). The vertex A on line a may be taken arbitrarily, since if the triangle ABC is constructed, any translation collinear with the given parallels, will also yield a triangle satisfying the conditions of the problem.

Let ABC be the desired triangle. If the plane is turned around the point A through 60°, so that the side AB coincides with the side AC, the point B on line b will go over into point C.

Let us rotate line b around point A through 60°. Its intersection, after the rotation, with line c will give the desired vertex C of the triangle.

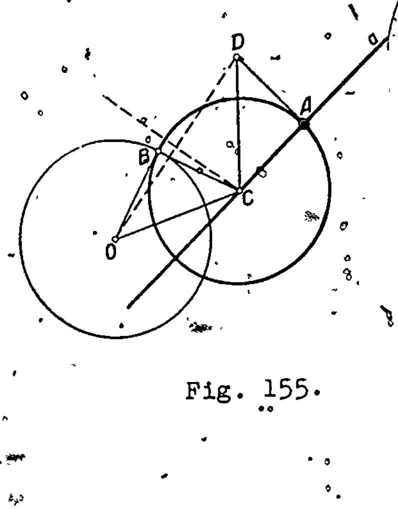


Fig. 155.

Problem. To draw a circle which passes through the point A of a given line a, has its center on the line a and intersects orthogonally a given circle.

Let c be the center of the desired circle (fig. 155). The triangle OBC is right-angled. Let us rotate it around c through the angle BCA. BO assumes position AD, perpendicular to line a at the point A. The desired center c is equidistant from the points O and D. Constructing point D, it is easy to carry the solution of the problem to completion (fig. 155).

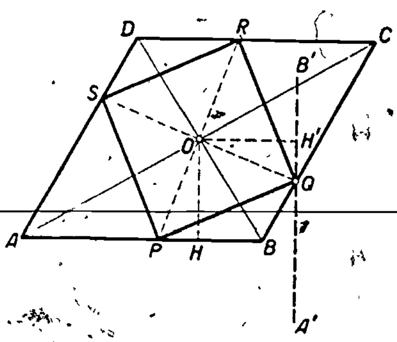


Fig. 156.

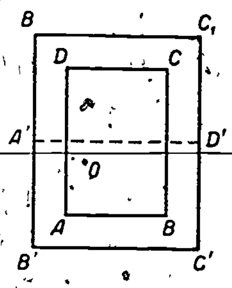


Fig. 157.

Problem. To inscribe a square in a parallelogram.

Let PQRS be the desired square (fig. 156). It is clear, first of all, that the center of the desired square coincides with the center of the given parallelogram. In proof, the midpoint O of the diagonal RP lies on the line connecting the midpoints of the sides AD and BC of the parallelogram; the same midpoint O of the diagonal QS lies on the line connecting the midpoints of the sides AB and DC.

It follows, that the centers of the parallelogram and the desired square coincide.

Let us turn the right-angled isosceles triangle POQ through a right angle around the point O. The point P on line AB goes over into point Q on line BC. This leads us to the construction.

Let us turn line AB through a right angle around the center O of the given parallelogram and we shall find the desired point Q of the square as the point of intersection of the image A'B' of line AB with the side BC (fig. 156). We then construct the diagonal QS of the square and, finally, the entire square.

It is suggested that the investigation of the conditions under which the problem has a solution be carried out as an exercise. Prove, in particular, that every rectangle circumscribed around a square is a square. Also solve the following problem.

Problem. The rectangular top of a folded card table is brought from position ABCD into position A'B'C'D' by a rotation around a point O within itself and then unfolded to form the rectangle $B'C'C_1B_1$ (fig. 157). Construct the position of the center of rotation.

31. THE ROTATION OF SPACE AROUND AN AXIS

Let there be given an axis in space, i.e. an oriented line. By a <u>rotation</u> of space through an angle ϕ around the axis a we mean a transformation of the space into itself such that:

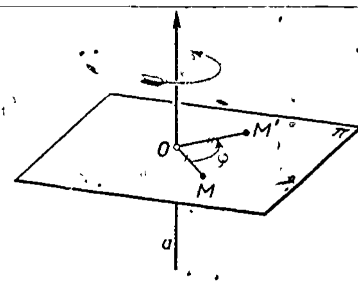


Fig. 158.

1) The image M' of each point M lies, together with M, in a plane π perpendicular to the axis of rotation a;

2) the distances from the points M and M' to the axis of rotation are equal, OM' = OM (fig. 158);

3) the angle MOM' is equal to the given angle ϕ; the direction of the rotation from OM to OM' corresponds to the turning of a right handed screw pointed in the direction of the oriented axis a when $\phi > 0$, and in the opposite direction when $\phi < 0$.

It is not difficult to verify that a rotation of space around an axis is a one-to-one transformation.

It follows from our definition that a rotation of space around an axis a through an angle ϕ induces a rotation, through the same angle ϕ, in each plane π perpendicular to the axis a around the point of intersection O of the plane π with the axis a.

The set of all rotations of space around an axis forms a group relative to addition of angles. Rotations through angles which differ by integral multiples of 2π are considered identical. This group, clearly, is isomorphic to the group of rotations in a plane π around the point O.

We shall now show that the distance AB of two points A and B is an invariant of the rotation of space around the axis a:

$$A'B' = AB$$

where A' and B' are the images of A and B.

In proof, if points A and B lie in one plane π perpendicular to the axis of rotation a then, by the preceding (28), the distance AB is invariant under the rotation of the plane π around the point O in which the axis a intersects the plane π.

Suppose now that the points A and B do not lie in one plane perpendicular to the axis a (fig. 159).

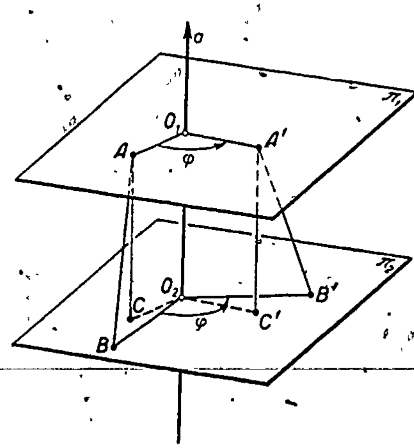


Fig. 159.

The point A and its image A' lie in a plane π_1, perpendicular to the axis a and intersecting it in the point O_1.

The point B and its image B' lie in a plane π_2, perpendicular to the axis a and intersecting this axis in the point O_2.

Dropping perpendiculars AC and A'C' on plane π_2, we find:

$$AC = A'C', \ O_2C = O_1A, \ O_2C' = O_1A', \ \angle CO_2C' = \phi .$$

Consequently, the point C' is the image of point C under the rotation of space around the axis a and B'C' = BC.

From the equality of triangles ABC and A'B'C' it follows that their hypotenuses are equal, q.e.d.

Under a rotation of space around an axis, planes are transformed into planes.

This is because any plane may be regarded as the locus of points equidistant from two appropriately chosen points A and B, and distances are invariant under rotation.

The property of being a plane is one of the properties of figures which are invariant under a rotation of space around an axis.

Under rotations of space around an axis, a straight line is transformed into a straight line.

This is because any straight line may be regarded as the locus of points equidistant from three appropriately chosen points A, B and C and distances are invariant under rotation.

The property of being a line is a property of figures invariant relative to rotation of space around an axis.

A rotation of space around an axis through an angle ϕ is a one-to-one transformation of space into itself. This leads us to the conclusion that parallel planes and parallel lines are transformed by a rotation of space around an axis into parallel planes and lines respectively.

Every plane, since it is transformed ~~by a~~ rotation again into a plane, is an invariant figure with respect to the group of rotations of space around a given axis; the same may be said of a straight line.

There exist, however, planes whose position in space remains unchanged under all rotations of the indicated group. Planes which do not change their positions in space under all rotations of space around a given axis are called invariant planes of the group of rotations of space around the given axis. The invariant planes are, clearly, the planes which are perpendicular to the axis of rotation. Every such plane only slides upon itself under all transformations of the group, rotating around its point of intersection with the axis of rotation. Such a plane is a sliding surface under every rotation of space around the given axis.

The pencil of planes passing through the axis of rotation (fig. 160) is also an invariant in regard to its position in space. Under each rotation of the group any plane of this pencil is transformed into a plane of the same pencil. The pencil of planes as a whole remains unchanged with regard to its position in space.

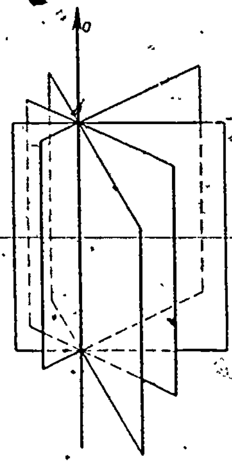


Fig. 160.

We have two invariant pencils of planes relative to the group of rotations of space around an axis: the pencil of parallel planes, each of which is perpendicular to the axis of rotation, and the pencil of planes which pass through the axis of rotation.

Every plane of the first pencil is invariant in regard to its position in space under all rotations of the group, i.e. these planes are sliding surfaces; any plane of the second pencil changes its position in space under every rotation of the group, except the neutral one.

Every invariant plane intersects the pencil of planes passing through the axis of rotation in a pencil of lines.

A pencil of lines lying on an invariant plane and with its center on the axis of rotation is invariant with regard to its position in space under all rotations of the group. This is an invariant pencil of lines on the indicated plane relative to the

the group of rotations of the plane around the center of the pencil.

By a <u>bundle of lines</u> is meant a set of lines in space, passing through a given point S - the center of the bundle. The set of lines in space parallel to a given line, including the line itself, is called, as we already know, a <u>bundle of parallel lines</u>.

It is clear that each bundle of lines with its center on the axis of rotation and the bundle of lines parallel to the axis of rotation is invariant with regard to its position in space, or simply, is an invariant bundle of the group of rotations around the given axis.

32. SURFACES OF REVOLUTION

In 3. we gave a general definition of figures of revolution. It is easy to see that any figure of revolution is transformed into itself, or, more precisely, mapped onto itself, by every rotation around its axis. A figure of revolution is invariant with respect to its position in space or is simply an invariant figure relative to the group of rotations around the axis of the figure.

When the figure ϕ generating the figure of revolution (3.) is a curve lying in a plane π, passing through the axis of revolution ℓ, then the figure of revolution is called a surface of revolution (fig. 161).

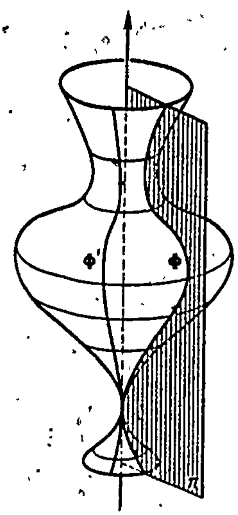


Fig. 161.

Remark. It is possible to obtain a surface of revolution by taking as the generating figure ϕ any curve in space not necessarily lying in a plane passing through the axis of rotation (see for example, fig. 18). It is easy to see by examining a cross-section of the surface by such a plane, that the definition given above does not give a narrower class of surfaces of revolution.

Intuitively one can see that under any rotation of space around the axis ℓ a surface of revolution slides along itself, i.e. is a <u>sliding surface</u>.

Every rotation around the axis ℓ carries the generating figure ϕ into a position ϕ'. The images ϕ' of the generating curve ϕ under the rotations of the group are called <u>meridians</u> of the surface of revolution. Each meridian may serve as the generating curve of the surface of revolution. Thus the intersection of the surface of revolution with the invariant pencil of planes whose axis is the axis of revolution is the family of all meridians of the surface.

The sections of the surface of revolution by planes perpendicular to the axis of revolution, are called <u>parallels</u> of the surface of revolution. The intersection of the surface of revolution with the invariant bundle of parallel planes is the family of all parallels of the surface.

Each parallel is, clearly, a circle with its center on the axis of rotation. Since every parallel intersects the plane of any meridian at a right angle, each meridian of a surface of revolution intersects every parallel also at a right angle. We say that the meridians and parallels form an orthogonal net on the surface of revolution (fig. 162).

A figure of rotation may be regarded as a locus of the meridians or as a locus of the parallels.

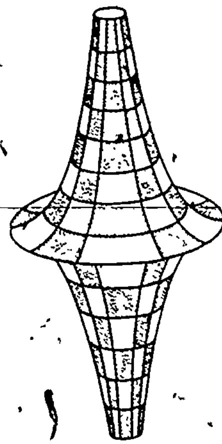


Fig. 162.

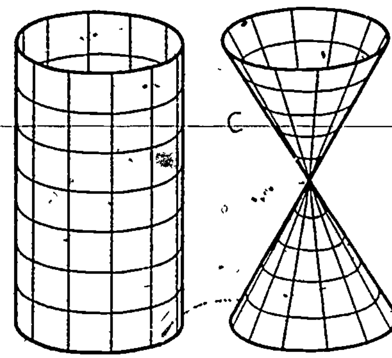


Fig. 163.

All circles in space with centers on the axis of rotation ℓ and lying in planes perpendicular to the axis are invariant in regard to their position in space, or simply, are <u>invariant circles</u> relative to the group of rotations with the given axis ℓ. Each invariant circle "slides" along itself under every rotation ~~of~~ the group and is a <u>sliding curve</u> under every rotation around the axis. Through each point in space there passes a unique invariant circle relative to the group of rotations around a given axis.

Choosing any curve intersecting each plane perpendicular to the axis of revolution in a single point and taking the union of all invariant circles passing through its points, we thus construct a surface of revolution from its parallels.

In 3. we considered the simplest surfaces of revolution generated by straight lines. These were cylinders (fig. 16) and cones (fig. 17). The meridians of the cylinder and the cone turn out to be their straight line generators; the parallels - circles intersecting the meridians at right angles (fig. 163).

We also examined surfaces of revolution generated by circles. The first of these is the torus (fig. 20) - a surface obtained by revolving a circle around an axis lying in the plane of the circle and not intersecting it.

The meridians of the torus are circles congruent to the generating circles, and the parallels are invariant circles determined by the points of any meridian (fig. 164).

Let us consider the surface generated by revolving a circle around its diameter.

This surface is, as we know, a sphere. To obtain a sphere, it is sufficient to rotate only a semi-circle around its diameter.

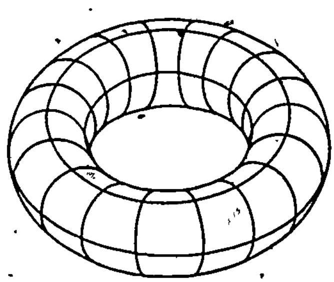


Fig. 164.

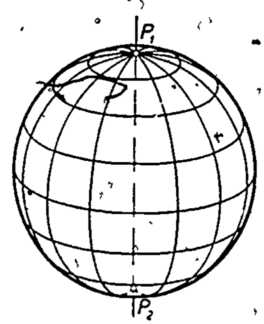


Fig. 165.

The images of the generating semi-circle are the meridians of the sphere. Each point on the semi-circle describes one of the parallels of the sphere. The largest parallel is the equator of the sphere. The family of parallels together with the family of meridians form an orthogonal net on the sphere (fig. 165).

The diameter of the sphere lying on the axis of revolution of the generating semi-circle and perpendicular to the plane of the equator has as its endpoints the poles of the orthogonal net. The poles are singular points of the net: all the meridians of the sphere pass through each pole, while through each ordinary point, i.e. other than the poles, there passes only one meridian and one parallel.

A sphere is a surface of revolution relative to each of its diameters; any diameter of the sphere may be taken as the axis of rotation. Revolving the sphere around any one of its diameters causes the sphere only to slide upon itself. The invariant circles under this group of rotations are the intersections of the sphere with the invariant planes.

All these propositions follow from the familiar fact that the intersection of a sphere by any plane is a circle, provided that the cutting plane is at a distance from the center of the sphere not greater than the length of the radius of the sphere.

Let us especially note this fact: Through each two points A and B on a sphere, not forming the end points of a diameter, there passes only one great circle of the sphere. The points A and B together with the center O of the sphere determine a unique plane, which intersects the sphere in a great circle passing through the given points A and B.

Through any two points A and B on the sphere which are not endpoints of a diameter, it is possible to pass a pencil of planes.

Each plane of this pencil, except the one passing through the center O of the sphere, intersects the sphere in a small circle of the sphere.

Among all circular arcs on a sphere joining two non-diametrically-opposite points A and B, the arc of the great circle is the smallest.

In proof, let us, by means of rotation around the axis AB, bring each plane of the pencil into the same position as the plane of the great circle, and, furthermore, let the smaller of the two arcs AB of each circle lie on one side of the line AB. Of two such arcs AB the one with the larger radius will be the smaller, since the encompassed arc A_mB is shorter than the encompassing arc A_nB (fig. 166). The great circle, however, has the largest radius, consequently, the arc of the great circle is the shortest, q.e.d.

Let us mention without proof that of all curves on the sphere joining two non-diametrically-opposite points A and B, the shortest is an arc of a great circle of the sphere.

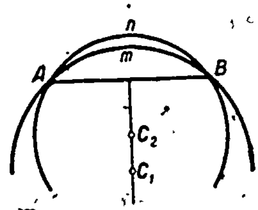


Fig. 166.

The length of the shortest curve joining two given points A and B on a sphere is called the spherical distance between the two points or, if only measurements on the sphere are considered, simply the distance between the two points A and B.

Making use of the notion of spherical distance, it is easy to establish a series of properties of figures on a sphere. Thus, for instance, the locus of points on a sphere equidistant in the sense of spherical distance - from a given point on the sphere, is a circle.

All the points on a parallel are equidistant from each of the poles P_1 and P_2 (fig. 165). Thus each circle on a sphere has two spherical centers.

Similarly, any parallel - and any circle on a sphere may be considered a member of a family of parallels having the diameter joining its spherical centers as its axis of revolution - is the locus of points equidistant from the equator and lying on one side of it.

In the plane the shortest curve joining two points is a straight line; on the sphere it is a great circle. The locus of points in the plane at a given distance from a given line, is a pair of straight lines parallel to the given one.

The locus of points on a sphere which are at a given spherical distance from a given great circle is a pair of parallels (or a pair of points).

However, parallels other than the equator are not shortest curves on the sphere. Here we have no analogy with the plane.

It is suggested as an exercise to prove that the locus of points on a sphere at equal spherical distances from two given points on the sphere is a great circle.

In conclusion, let us mention that the figure consisting of the arcs AB, BC and CA of great circles is called the spherical triangle ABC.

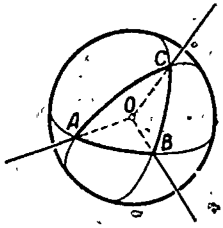


Fig. 167.

A spherical triangle is called Eulerian if its sides AB, BC, CA are each less than πR, where R is the radius of the sphere (fig. 167).

In a spherical triangle, all three angles may be right angles or obtuse angles. Examples of such triangles are easy to find.

The sum of the angles of a spherical triangle is more than two right angles. This follows from the theorem about the sum of the dihedral angles of a trihedral angle. The edges of this angle are the rays joining the center O of the sphere with the vertices A, B, C of the spherical triangle (fig. 167); the faces of the angle are the planes passing through the center O and containing the sides AB, BC, CA of the spherical triangle.

The sum of the angles of a spherical triangle is a variable magnitude changing from triangle to triangle.

Similar but unequal spherical triangles on the same sphere do not exist.

The following theorem holds: If three angles of one spherical triangle are equal to three angles of another, then the triangles are either congruent or symmetric.

Symmetry plays a special role on the sphere, since, for example, one cannot turn a spherical triangle over onto its other side and replace it on the sphere.

Constructions on a sphere are extremely valuable exercises. It is convenient to use a wooden model of a globe and a compass with bowed legs for drawing circles on the sphere.

33. ROTATION OF SPACE AROUND A POINT

With the rotations of a sphere around its different diameters are closely connected the transformations of space called the rotations of space around a given point.

Under all rotations of space around axes passing through a given point O, each sphere with center O is transformed into itself, with the spherical distance between any two of its points left unchanged.

Suppose we have two spheres coinciding at all their points and having therefore a common center O. Let us imagine that one sphere is fixed, but the second one is able to slide over the first, with the preservation of the spherical distance between any two points.

If we require that one point A of the movable sphere maintain its position unchanged, then the sliding of the movable sphere over the fixed one will be a rotation around the axis OA.

If we require that, in addition to the point A, another point B on the movable sphere, not diametrically opposite to the point A, remain fixed, then it will become impossible to rotate the movable sphere.

In other words, the position of the movable sphere relative to the fixed one is determined by the position of any spherical segment ($AB \neq \pi R$) of the movable sphere.

Noting this, we may construct the following mapping of the sphere onto itself. Let us take on the sphere any two equal spherical segments AB and A'B' not equal to πR (fig. 168), and let us assign point A' to point A and point B' to point B.

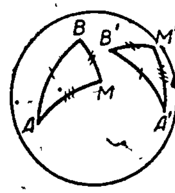


Fig. 168.

To any arbitrary point M of the sphere we shall assign a point M' on the same sphere such that we have equality of the spherical distances

$$AM = A'M'; \quad BM = B'M'$$

and in addition, as for motions in the plane, the spherical triangles AMB and A'M'B' are oriented in the same way (the arrangement of the vertices of the triangles, looking from the outside of the sphere, is the same).

If the point M lies on the great circle AB, then its image is uniquely determined without additional specifications as to orientation.

Here, as in the case of motions in the plane, given a mapping of the sphere onto itself determined by the equal spherical segments AB and A'B'. Let CD be a third spherical segment (not equal to πR) and let C'D' be its image under this mapping. Then the mapping determined by the segments CD and C'D' is identical with the one determined by AB and A'B'.

Intuitively, one can easily visualize the indicated mapping as a sliding of the movable sphere over the fixed one until the spherical segment AB coincides with the segment A'B' and, as a result, any point M will coincide with its image M'.

The mapping of the sphere onto itself just described, is called a rotation of the sphere around its center.

We shall show that every rotation of a sphere around its center is a rotation of the sphere around one of its diameters.

Suppose that the rotation of the sphere around its center is determined by a pair of spherical segments AB and $A'B'$ ($AB = A'B' \neq \pi R$) (fig. 169). Let us repeat on the sphere the same constructions as we performed in the plane for locating the center of rotation (29, fig. 148).

We join points A and A' by the arc AA' of a great circle and we pass through the midpoint R of this arc a spherical perpendicular RO; similarly, we join the points B and B' and construct a spherical perpendicular SO at the center of the spherical segment BB'.

The perpendiculars RO and SO either intersect at a point O or else they coincide.

In case of intersection, the given rotation of the sphere around its center O_1 (not shown in the diagram) is a rotation of the sphere around the diameter determined by the line OO_1 through an angle

$$\phi = \angle AOA' = \angle BOB'.$$

If the perpendiculars RO and SO coincide then the given rotation of the sphere around its center is a rotation of the sphere around the line OO_1, where O is the point of intersection of the great circles AB and $A'B'$. It is suggested to make a diagram analogous to fig. 149 and carry through the proof.

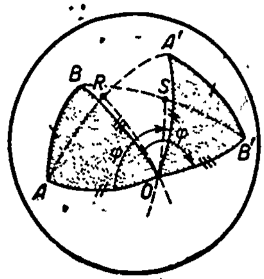


Fig. 169.

Remark. Since two great circles of a sphere always intersect, a sliding of a sphere upon itself analogous to parallel translations in the plane does not exist. If the points A and A' (or B and B') are diametrically opposite, the proof still remains valid.

The theorem just proved permits us to conclude that the set of all rotations of a sphere about its center is a group. In other words, the set of all rotations of a sphere about its diameters is a group.

Here the group operation is the addition of rotations, that is, the substitution of a single rotation of the sphere for two successive rotations.

It is recommended as an exercise to verify the closure of the operation of addition of rotations, as well as the fulfillment of the remaining requirements for a group. Show that the group of rotations of the sphere is not commutative.

A rotation of a sphere about a diameter generates a rotation of space around an axis, and vice versa. Thus we arrive at the concept of the group of rotations of space about all possible axes passing through a given point.

As an exercise, show that successive rotations of space through two right angles about three mutually perpendicular axes intersecting in one point will return space to its original position.

34. THE GROUP OF MOTIONS OF SPACE

Let us consider the following mapping of space onto itself, called a motion.

Let ABC and A'B'C' be two congruent triangles occupying any fixed positions in space. We assign the point A' to the point A, the point B' to the point B, and the point C' to the point C. To any point M in space we assign a point M' such that the equalities

$$A'M' = AM, \quad B'M' = BM, \quad C'M' = CM$$

hold, and the orientation of the tetrahedrons M'A'B'C' and MABC is the same (fig. 170).

By the same orientation of the tetrahedrons we mean such a disposition of their vertices that if one were to look from the vertices M and M' on the triangles ABC and A'B'C' then the sets of vertices A, B, C and A'B'C' are either both arranged counterclockwise or both arranged clockwise.

If the point M lies in the plane of the triangle ABC then the specification about the orientation is unnecessary.

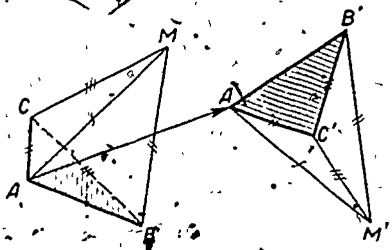


Fig. 170.

Intuitively, the given motion may be thought of as a displacement of the triangle ABC together with the whole of space, rigidly attached to it, into the position A'B'C'; then any point M is brought into coincidence with its image M'.

It is easy to establish that a motion is a one-to-one, bicontinuous transformation of space onto itself, i.e. a topological transformation. The set of all motions in space forms a group.

In effect, a motion ϕ_1 in space, carrying the triangle ABC into the triangle A'B'C' followed by the mapping ϕ_2, of space mapping the triangle A'B'C' into the triangle A"B"C" may be replaced by one motion

$$\phi = \phi_1 + \phi_2$$

carrying ABC into A"B"C".

This implies that the operation of addition of motions is closed in the set of all motions.

The verification that the other requirements for a group are fulfilled is suggested as an exercise. It is also easy to verify that any triangle PQR and triangle P'Q'R', corresponding to it under a given motion, may serve to define this motion. Thus the triangles ABC and A'B'C' do not play any special rôle.

The group of all translations in space is, clearly, a subgroup of the group of motions. The group of rotations of space around a given point is also a subgroup of the group of motions.

It is easy to show that each motion in space may be obtained by performing successively one parallel translation and one rotation around some axis.

In proof, let us choose a displacement vector, $\overrightarrow{AA'}$ where A is any point and A' is its image under the given motion (fig. 170).

After carrying out the translation whose displacement vector is $\overrightarrow{AA'}$. there still remains to bring the space into the desired position by means of a motion which leaves the point A' unchanged. But the last motion is a rotation of space around the point A' and, consequently, is a rotation around an axis passing through A', q.e.d.

Since we may take as the displacement vector, a vector whose origin is any point A of space, and its endpoint the image A' of the point A under the given motion, the decomposition of the given motion into a translation and a rotation around an axis is not determined uniquely.

Any motion in a plane generates a motion in space, which is easy to see if we take for the triangles ABC and A'B'C' (fig. 170) two triangles in the plane which correspond to each other under the plane motion.

The motions of space, generated by the motions in a given plane form a subgroup of the group of all motions of space. The motions of this subgroup are called coplanar motions.

The subgroup of coplanar motions considered here leaves invariant, with regard to its position in space, every plane of the pencil of parallel planes containing the plane which determines the subgroup.

In view of the isomorphism of the group of coplanar motions in space and the group of motions in the plane, there corresponds to each proposition concerning motions in the plane, a proposition regarding motions in space.

Thus, we know that every motion in the plane which is not a translation is a rotation.

As a result of the indicated isomorphism, we have the proposition: every coplanar motion of space which is not a translation, is a rotation around an axis perpendicular to the planes of the invariant pencil.

In particular, the sum of a translation determined by the displacement vector $\overrightarrow{AA'}$ and a rotation through an angle $\phi \neq 0$ around an axis perpendicular to the displacement vector $\overrightarrow{AA'}$ is a rotation around an axis parallel to the first.

Furthermore, as it is easy to verify, the resulting rotation is a rotation through the same angle ϕ.

One of the most important theorems concerning motions in space is the following:

Theorem. Any transformation in space is the sum of a translation and a rotation around an axis collinear with the translation.

In proof, let the given motion be composed of a translation with the displacement vector $\overrightarrow{AA'}$ and a rotation around the axis ℓ passing through the point A'.

We decompose the given translation into a sum of two translations, one parallel to the axis of rotation ℓ, and the other perpendicular to the axis ℓ (fig. 171):

$$\overrightarrow{AA'} = \overrightarrow{AB} + \overrightarrow{BA'}$$

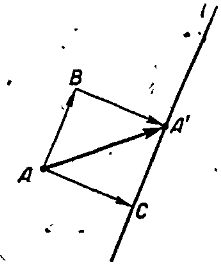


Fig. 171.

By the preceding proposition, we may replace the sum of the translation $\overrightarrow{BA'}$ perpendicular to the axis ℓ', and the rotation around the axis ℓ_1, by a rotation around an axis parallel to the axis ℓ. Consequently, the given motion is the sum of the translation $\overrightarrow{AB}$ and a rotation around an axis parallel to $\overrightarrow{AB}$, q.e.d.

A motion which is the sum of a translation and a rotation around an axis parallel to this translation is called a <u>screw-motion</u>.

Thus every motion in space is a screw-motion.

Here we do not exclude the possibility that either the rotation or the translation, or both, may consist of the identity transformation.

Clearl[illegible] a screw motion, the order in which the translation and rotation are taken does not matter. The pair consisting of a translation and a rotation around an axis parallel to the translation, is called a <u>screw</u>.

The ratio of the length of the displacement vector to the angle of rotation is called the <u>parameter of the screw</u>.

It must be emphasized, that the concept of motion in geometry does not require the actual displacement of space or a part of space from one position into another through all intermediate positions. What is essential for the geometric concept of motion is only the initial and final position of the figure. We may say that under a geometrical motion of a figure its points do not have trajectories.

Under motions in space as well as motions in a plane any two points have an invariant, the distance between them.

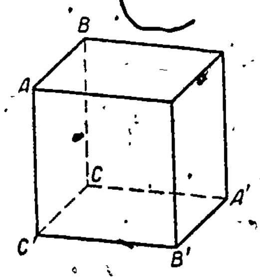


Fig. 172.

As an exercise, let the student show that successive rotations through an angle of 180° around two axes which are skew and inclined toward each other at a right angle add up to a screw motion collinear with the vector of the shortest distance between the axes; the displacement is equal to twice the distance between the axes, while the angle is 180°.

Prove that a segment AB may be brought into any non-parallel position A'B' by means of a rotation around some axis and that consequently any motion may be obtained as a sum of two rotations around appropriate axes.

A cube is moved so that three of its vertices A, B, C go over into A', B', C' (fig. 172). Find the screw.

The properties of motions as geometric transformations which we have just studied play a major role in the mechanics of rigid bodies.

35. THE GROUPS OF ROTATIONS OF A REGULAR PYRAMID

If we rotate the right pyramid $SA_1A_2 \ldots, A_n$ around its altitude SO (fig. 173) then the set of those motions which carry the pyramid over into itself forms a group.

This group contains n rotations through the angles $\phi_0 = 0, \phi_1 = \frac{2\pi}{n}; \phi_2 = 2\frac{2\pi}{n}, \ldots \phi_{n-1} = (n-1)\frac{2\pi}{n}$.

The general expression for the angle of rotation is given by:

$$\phi_k = k\cdot\frac{2\pi}{n}(k = 0,1,2,\ldots, n-1).$$

or

$$\phi_k = k\,\phi_1 .$$

Rotations which differ by an integral number of revolutions are considered identical. It is easy to see that rotations other than those enumerated will not carry the pyramid $SA_1A_2,\ldots, A_n$ into itself.

The sum of two rotations through angles $p\,\phi_1$, and $q\,\phi_1$ is again a rotation through an angle $r\,\phi_1$ where r is the remainder obtained by dividing $p + q$ by n.

The group of rotations of a regular pyramid is isomorphic to the group of rotations of its base $A_1A_2,\ldots,A_n$ around the center O. It is understood that the rotations under consideration here are those which carry the regular polygon into itself without taking the polygon out of its plane (do not "turn it over").

Groups isomorphic to the groups of rotations of regular polygons are called finite cyclic groups. A finite cyclic group is generated by one element, for example, in our case, by the element ϕ_1.

In 24. we had an example of a group isomorphic to the group of integers; such groups are called <u>infinite cyclic groups</u>.

A group generated by one of its elements is called a <u>cyclic group</u>.

We know from algebra that the roots of the equation $W^n = z$ in the field of complex numbers are:

$$W_k = \sqrt[n]{p}\left[\cos\left(\frac{\phi}{n} + \frac{2k\pi}{n}\right) + i \sin\left(\frac{\phi}{n} + \frac{2k\pi}{n}\right)\right]$$

where

$$k = 0, \pm 1, \pm 2, \ldots,$$

ϕ is the argument of z and $p = |z|$.

It is also known that among the numbers W_k there are only n different ones. Since arguments of complex numbers differing by multiples of 2π may be considered as identical and since in multiplying complex numbers the arguments are added, it follows that the set of n^{th} roots of a number z is a group with respect to multiplication. This group is isomorphic to the group of rotations of a regular pyramid whose base is a regular polygon.

As an exercise, let the reader find all the elements of a cyclic group of order n which generate the group. Compare with the notion of a primitive root of a binomial equation.

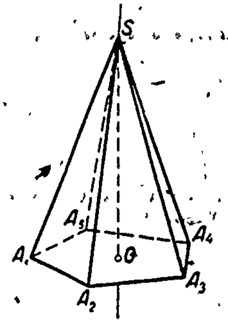


Fig. 173.

36. THE DEFINITION AND PROPERTIES OF ORTHOGONAL TRANSFORMATIONS

The group of translations as well as the group of rotations of space around an axis and around a point turned out to be subgroups of the group of motions in space. The group of motions is, in turn, a subgroup of a wider group of transformations.

Definition. A transformation of a plane onto a plane or of space onto itself is called orthogonal if under this mapping the distance between points remains unchanged, i.e. the distance $A'B'$ between the images A', B' of the points A and B is the same as the distance between A and B:

$$A'B' = AB.$$

From this definition there follows an entire series of simple theorems describing the properties of orthogonal transformations.

Theorem 1. An orthogonal transformation is one-to-one.

Proof: If two distinct points A and B had one and the same image A', then the distance AB between them would have been, by the definition of orthogonal transformations, equal to zero and, hence, the points A and B could not be distinct.

Theorem 2. The set of all orthogonal transformations in space forms a group with respect to the addition of transformations.

Proof: 1° The sum of two orthogonal transformations - i.e. a transformation obtained by performing successively two orthogonal transformations - is an orthogonal transformation, since the sum-transformation leaves unchanged the distance between points.

Thus we have closure for the algebraic operation of addition of orthogonal transformations.

2° The associative law holds for all transformations of space into itself.

3° The condition of the existence of a neutral element is fulfilled since the identity transformation is orthogonal.

The condition that every element in a group should have an inverse is fulfilled here, since the inverse transformation of an orthogonal transformation is orthogonal. The theorem is thus established.

Since motions preserve the distance between points, it follows that the group of motions of space is a subgroup of the group of orthogonal transformations.

The properties of orthogonal transformations follow from the following simple properties of the distance between two points in space:

A. If A, B, and C are three non-collinear points then the sum of any two segments determined by these points is larger than the third segment.

B. If A, B, and C are three collinear points then

$$AB + BC = AC$$

provided that the point B lies between the points A and C.

From these two propositions there follows the converse of proposition B.

C. If the segments determined by three points A, B, and C are connected by the relation

$$AB + BC = AC,$$

then these three points lie on one line and, furthermore, the point B lies between A and C.

In proof: if the points A, B, C did not lie on the same line, then by property A we would have $AB\ BC > AC$.

Furthermore, B lies between A and C since if, for example, C were to lie between A and B, we would have

$$AC + CB = AB,$$

which contradicts the relation $AB + BC = AC$.

In view of these properties of distance between points the collinearity or non-collinearity of three points is completely determined by the distance between these points. Since orthogonal transformations leave distances unchanged, the following theorems hold.

Theorem 3. Under an orthogonal mapping, every collinear triplet of points is transformed into a collinear triplet of points. Every non-collinear triplet of points is transformed into a non-collinear triplet.

Theorem 4. Under an orthogonal mapping the image of a straight line is a straight line: the inverse image of a straight line is a straight line.

Theorem 5. Under an orthogonal transformation a segment is transformed into a segment.

Theorem 6. Under an orthogonal transformation a point dividing a segment in a given ratio is transformed into a point which divides the image of the given segment in the same ratio.

In other words the simple ratio of three points on a line is an invariant of an orthogonal transformation.

Theorems 3-6 are an immediate consequence of the properties of the distance between two points and the definition of orthogonal transformations. The proof of these theorems is left to the student.

Theorem 7. Under an orthogonal transformation of space the image of any plane is a plane and, conversely, the inverse image of any plane is a plane.

In proof, let us take in the given plane π two lines AB and AC intersecting at the point A. Let M be an arbitrary point other than A in the plane π. Through the point M draw a line m intersecting lines AB and AC in two distinct points P and Q.

The image of the point A under the given orthogonal transformation is a point A'; the images of the lines AB and AC will be the distinct lines A'B' and A'C'. The images of line m will be the line m' intersecting lines A'B' and A'C' in the points P' and Q' — the images of the points P and Q. Hence, the image M' of the point M will lie in the plane π' passing through the lines A'B' and A'C'.

Thus an arbitrary point M in the plane π (AB, AC) is transformed into M' in the plane π'(A'B', A'C') and, consequently, the points of the plane π are transformed into the points of the plane π'.

Since the inverse transformation is orthogonal, the inverse image of a plane is a plane. The theorem is now proved.

Theorem 8. Under an orthogonal transformation parallel lines go over into parallel lines, and parallel planes into parallel planes.

For otherwise, the orthogonal transformation would not be one-to-one.

Theorem 9. Under an orthogonal transformation, angles are preserved.

This theorem follows immediately from the preservation of segments.

If we are given three non-coplanar vectors e_1, e_2, e_3, originating at the same point O, then, as we know, every vector $\overrightarrow{OM}$ may be decomposed in only one way with respect to the three vectors e_1, e_2, e_3:

$$\overrightarrow{OM} = xe_1 + ye_2 + ze_3 .$$

The numbers x, y and z are called coordinates of the point M in the coordinate system determined by the triplet of vectors e_1, e_2, e_3. Furthermore, x, y, z are the respective ratios to the segments e_1, e_2, e_3 of the projections--generally non-perpendicular - of the vector $\overrightarrow{OM}$ on the axes of e_1, e_2, and e_3.

We shall denote this system of coordinates by the symbol $O\{e_1, e_2, e_3\}$.

Theorem 10. (First fundamental theorem on orthogonal mappings). Under an orthogonal transformation every triplet of non-coplanar vectors e_1, e_2, e_3 originating at a point O goes over into a triplet of non-coplanar vectors e'_1, e'_2, e'_3 with the origin at O' and with their lengths and angles unchanged. Moreover, every point M goes over into a point M' having the same coordinates relative to (e'_1, e'_2, e'_3) as the point M had relative to (e_1, e_2, e_3). Conversely, every transformation having these properties is orthogonal.

Proof: Under an orthogonal transformation the point O passes into O', the vectors e_1, e_2, e_3 into vectors e'_1, e'_2, e'_3 with their lengths and the angle between them unchanged. In addition, the respective parallelism of the planes projecting the point M on the axes e_1, e_2, e_3 to the planes (e_2, e_3), (e_3, e_1), (e_1, e_2) is preserved. The ratios, or coordinates, are also left unchanged.

Since the length of a segment is completely determined by the coordinates of its end points, the converse of the theorem also follows and the theorem is completely proved.

This theorem has the following corollary. An orthogonal transformation of space is completely determined by indicating the images of any four points not lying in the same plane.

In proof, taking any of the four points as the origin O, and the remaining three as the end points of the vectors e_1, e_2, and e_3, their images - the point O' and the vectors e'_1, e'_2, and e'_3 - are determined. Consequently, we also know the point M' (x, y, z), the image of the point M (x, y, z). The point M' has the same coordinates in the system O' (e'_1, e'_2, e'_3) as the point M in the system O' (e'_1, e'_2, e'_3). This proves the corollary.

It is suggested that the student obtain by himself an analogous result for orthogonal mappings of the plane onto itself.

A transformation of space is called a reflexion in the plane π, if to each point M there is assigned a point M' such that

$$AM' = AM$$

where the line MM' is perpendicular to the plane π, A is the foot of the perpendicular from the point M to the plane π and the points M and M' lie on different sides of the plane π. The points of the plane π are the fixed points of the reflexion.

A reflexion in a plane π preserves the distance between points, and is consequently an orthogonal transformation.

A transformation in a plane is called a reflexion in a line a, if to every point M in the plane, we assign a point M' such that

$$AM' = AM$$

where A is the foot of the perpendicular from the point M to the line a, the line MM' is perpendicular to the line a and the points M and M' lie on different sides of the line a.

It is easy to see that the reflexion of the plane in a line is an orthogonal transformation of the plane.

The properties of reflexions will be taken up in detail in the next chapter.

Theorem 11'. (Second fundamental theorem on orthogonal transformations). Each orthogonal transformation of space is either a motion, or the sum of a motion and a reflexion in a plane.

Proof. Suppose the orthogonal transformation takes the point O into the point O' and the non-coplanar vectors e_1, e_2, e_3 with origin at O into the vectors e'_1, e'_2, e'_3 with the origin at O'. Let us consider the motion which carries O into O', the vectors e_1 and e_2 into the vectors e'_1 and e'_2. Such a motion, as we know, is unique. Because of the equality of distances and angles, this motion either takes the vector e_3 into the vector e'_3 - and then the given orthogonal transformation coincides with the motion - or it takes the vector e_3 into the vector e''_3, which is the reflexion of the vector e'_3 in the plane (e'_1, e'_2).

In the latter case, the orthogonal transformation is a sum of a motion and a reflexion in a plane. The proof is completed.

We have an analogous result for orthogonal transformations in the plane.

Those orthogonal transformations which are also motions are called orthogonal transformations of the first kind. All others are called orthogonal transformations of the second kind.

Frequently any orthogonal transformation is called a motion or a motion in a broad sense. Particular care should be taken when using this terminology.

Chapter VI

SYMMETRY

In Chapter VI we take up the symmetry of a plane with respect to an axis and with respect to a point, and the symmetry of space with respect to a plane and with respect to a point. The performance of translations and rotations by means of successive application of reflections is examined. The groups of symmetries of a cube and of a tetrahedron are studied. Feodorov's general definition of the symmetry of figures is given. In 39 we examine the application of the method of reflections to the solution of construction problems.

37. SYMMETRY IN A PLANE WITH RESPECT TO AN AXIS

Point A' is said to be symmetrical to point A with respect to line ℓ if points A and A' lie: 1) on opposite sides of this line; 2) at equal distances from it; and 3) on a line AA' perpendicular to line ℓ (fig. 174).

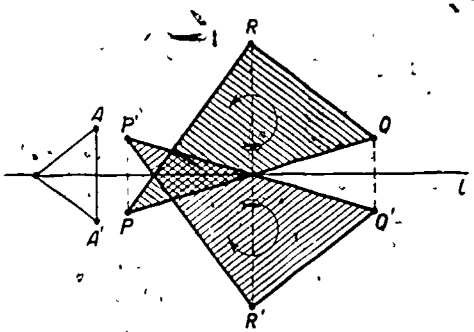


Fig. 174

Line ℓ is called the axis of symmetry of points A and A'. Every point of the axis of symmetry is symmetrical with itself.

Two points A and A' of a plane have only one axis of symmetry. This is the locus of points M in the plane which are equidistant from points A and A' (fig. 174).

A transformation of the plane into itself is said to be a *symmetry about an axis* ℓ, or an axial symmetry if to each point A is assigned its symmetrical point A' with respect to the axis ℓ.

A symmetry of a plane about a line is also called a *reflexion of the plane in the line*.

It can readily be verified that a symmetry about an axis is a one-to-one and bicontinuous transformation of the plane into itself. A point A is transformed into the point A', symmetrical with A. This transformation of the plane generates a transformation of any figure into a figure symmetrical with the first with respect to the axis. Thus, triangle PQR is transformed into triangle P'Q'R', symmetrical with the first, the orientation of the image being, however, opposite to that of the antecedent (Fig. 174).

Speaking in terms of motions in space we may say that a symmetry of a plane with respect to an axis ℓ is a half-turn of this plane about the axis ℓ.

Precisely this property of a symmetry is used in defining symmetrical points as points which will coincide when the plane of the drawing "is folded" along the axis of symmetry.

From the above follow all the properties of a symmetry about an axis. An axial symmetry preserves distances between points, takes a line into a line and parallel lines into parallel lines, preserves angles between lines, and so on.

Since an axial symmetry "turns the plane over" onto its other side, the orientation, say, of a triangle is changed by a symmetry into the opposite orientation.

The properties of a symmetry can be obtained without reference to motions in space. Thus, for example, it can easily be proved that if any two points on one line are symmetrical with two points on a second line with respect to a given axis, then all points of the first line are symmetrical with the points of the second line.

Given lines AB and A'B' (fig. 175), where points A and B are respectively symmetrical with points A' and B' with respect to an axis ℓ.

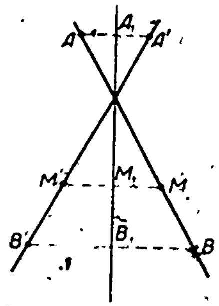


Fig. 175.

It is required to prove that for every point M of line AB the symmetrical point M' lies on line A'B'.

Let A_1 and B_1 be the points of intersection of lines AA' and BB' with line ℓ. Since $BB_1 = B'B_1$ and $AA_1 = A'A_1$, points A_1, and B_1 belong to the bisector of the angle between AB and A'B', if these lines intersect, or to the parallel line midway between them if they are parallel. In either case, the line A_1B_1 coincides with the axis of symmetry ℓ. Letting M be an arbitrary point on line AB, drawing line MM' parallel to lines AA' and BB' and marking its point of intersection M' with line A'B', we find by the property of the line ℓ that $MM_1 = M'M_1$, where M_1 is the point common to lines MM' and ℓ; q.e.d.

Two successive symmetries about the same axis ℓ are equivalent to the identity transformation of the plane.

Proof: if M' is the point symmetrical with an arbitrary point M, then the point symmetrical with M' is the original point M.

The set of symmetry transformations about all possible lines of a plane does not possess the property of closure. More precisely, the sum of two symmetries will be some transformation of the plane, but not a symmetry.

Proof: two successive symmetries leave unchanged the orientation of an arbitrary triangle in the plane, consequently, the resulting transformation is not a symmetry about any axis, q.e.d.

From this it follows that the set of all axial symmetries of a plane is not a group.

Theorem. The sum of two symmetries about parallel axes is a translation of the plane.

Let us first note that whatever the arrangement of three points A, A' and A" on an oriented line, the vector equation

$$\overrightarrow{AA''} = \overrightarrow{AA'} + \overrightarrow{A'A''}$$

always holds good.

If to collinear segments we prefix signs determined by the orientation of the axis on which the segments lie, the foregoing equation is equivalent to:

$$AA' + A'A'' + A''A = 0 .$$

These equations can be verified without difficulty for all possible arrangements of three points A, A' and A'' on an oriented line.

Now let ℓ_1 and ℓ_2 be the two parallel lines, taken as axes of symmetry; let ℓ_1 be the axis of the first symmetry and ℓ_2 the axis of the second. Let $\underline{a}$ be the displacement vector for the translation of line ℓ_1 into the position ℓ_2.

For an arbitrary point A we have (fig. 176):

$$\overrightarrow{AA''} = \overrightarrow{AA'} + \overrightarrow{A'A''} ,$$

where A' is the point symmetrical with A about axis ℓ_1 and A'' the point symmetrical with A' about axis ℓ_2.

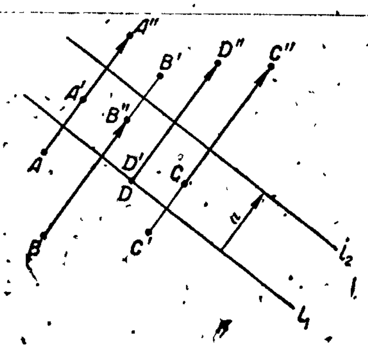


Fig. 176

But, by virtue of the properties of a symmetry we have:

$$\overrightarrow{AA'} = 2\overrightarrow{A_1A'} , \quad \overrightarrow{A'A''} = 2\overrightarrow{A'A_2}$$

where A_1 and A_2 denote the intersections (not lettered in the drawing) of line AA'' with the axes ℓ_1 and ℓ_2.

Consequently,

$$\overrightarrow{AA''} = 2(\overrightarrow{A_1A'} + \overrightarrow{A'A_2}) = 2\overrightarrow{A_1A_2} = 2\underline{a} .$$

Thus, the addition of two successive symmetries about parallel axes carries an arbitrary point A over into a point A''

with a displacement vector of 2a , which is constant for all points A , that is, the sum of these symmetries is a translation, q.e.d.

In figure 176 are shown several pairs of corresponding points in various positions relative to the axes of symmetry.

If we reverse the order of addition of the symmetries, we obtain a translation inverse to the first one.

The converse theorem also holds good: every translation in a plane can be represented as the sum of two symmetries about parallel axes.

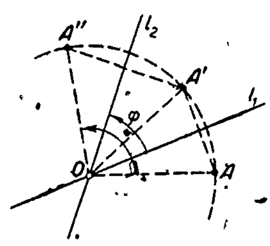
The proof of this theorem and the choice of the axes of symmetry is suggested as an exercise.

As we already know, the distances of any point on the axis of symmetry from two symmetrical points A and A' are equal.

Hence, for the point of intersection O of the axes of symmetry ℓ_1 and ℓ_2 (fig. 177) we have:

$$OA = OA' = OA'' ,$$

where A is an arbitrary point in the plane, A' is its symmetrical point with respect to axis ℓ_1 and A" the point symmetrical with A' with respect to axis ℓ_2 .



Furthermore, two successive symmetries "do not turn the plane over". Thus, the sum of the two symmetries here considered is a rotation.

The angle of rotation AOA" is twice the angle between the axes:

$$\angle AOA'' = 2\phi .$$

The converse theorem is likewise true: Every rotation of a plane about a point can be represented as the sum of two symmetries about axes passing thru the center of the rotation.

The proof of this theorem is recommended as an exercise.

From the properties of axial symmetry which we have demonstrated there follows the important conclusion: the set of all axial symmetries generates a group which includes all rotations and translations. This group is the group of orthogonal transformations of the plane. The motions in the plane considered earlier are called motions or orthogonal transformations of the first kind.

The group of motions in the plane is a subgroup of the group of orthogonal transformations.

38. SYMMETRY IN A PLANE WITH RESPECT TO A POINT

Point A' is said to be symmetrical to point A with respect to point O if O is the midpoint of segment AA'.

Point O is called the <u>center of symmetry</u>.
The center of symmetry is symmetrical with itself.

Two points A and A' of a plane have only one center of symmetry - the midpoint of segment AA'.

A transformation of the plane into itself is called a symmetry about a center, or a <u>central symmetry</u>, if to each point A is assigned its symmetrical point A' with reference to a given center O.

As may readily be verified, symmetry about a point is also a one-to-one and bi-continuous transformation.

A central symmetry is, of course, a rotation of the plane through 180° about the center of symmetry [14].

Thus, central symmetry does not confront us with any new transformation of the plane; in this it differs from the case of axial symmetry.

A central symmetry, like every rotation of the plane about a point, is the sum of two axial symmetries whose axes pass through the center. In this case the axes of symmetry are perpendicular to each other.

<u>Theorem</u>. <u>The sum of two central symmetries of the plane with different centers is a translation in the plane.</u>

<u>Proof</u>: We know that the sum of two rotations of the plane about two different centers is a motion of the first kind in the plane (29). Every motion of the first kind is completely deter-

mined by giving two pairs of corresponding points, or -- what is the same thing -- by giving a pair of corresponding segments. A rotation of 180° about center O_1 transforms the given segment AB into segment A'B' parallel to AB but oppositely oriented (fig. 178).

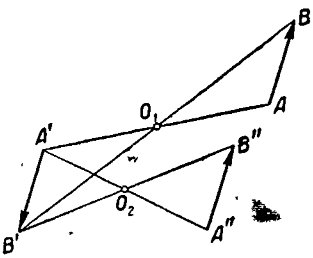


Fig. 178

A second rotation through 180° about center O_2 transforms segment A'B' into segment A"B" parallel to AB and having the same orientation as the latter. But such a transformation of AB into A"B" is a translation, q.e.d.

Conversely, every translation of the plane is the sum of two central symmetries, one of whose centers can be arbitrarily chosen.

The proof of this assertion is suggested as an exercise.

39. THE METHOD OF SYMMETRY IN THE SOLUTION OF CONSTRUCTION PROBLEMS

In the solution of construction problems it is often advantageous to subject a figure or part of one to a symmetry transformation which will enable the figure or part to be constructed. By a second symmetry the figure can be returned to its original position.

Other applications of symmetry are also possible.

The method of symmetry is best elucidated by examples.

Problem. Given two points A and B on the same side of the line a . (fig. 179) To find on line a a point C such that ∠ACP = ∠BCP, where CP is the perpendicular to line a at point C .

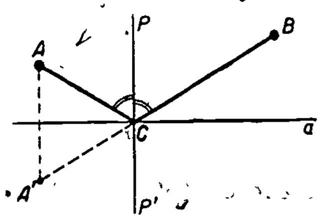


Fig. 179

Taking point A', symmetrical with point A about line a we find ∠A'CP' = ∠ACP = ∠PCB .

Consequently, the segments BC and CA' lie on the straight line BCA'. The required point C is obtained by joining B with A', the point symmetrical with point A about line a .

If A is a luminous point, all rays issuing from it are reflected from the mirror a in accord with the foregoing law, and it will appear that they issue from point A', the image of point A .

In the same way exactly, if A is a billiard ball and a the cushion, in order that ball A, upon its rebound from the cushion, shall strike ball B - it is necessary to shoot ball A in the direction of point C (fig. 179).

Furthermore, it is clear that the length AC + CB of the broken line ACB is the shortest distance from A to B by way of any point on line a.

Problem. To find the path from A to B of a billiard ball which shall rebound from two cushions during its travel between these points (fig. 180).

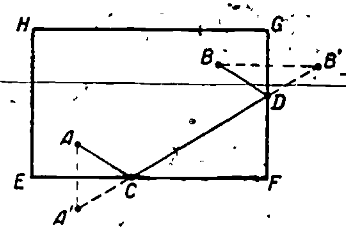


Fig. 180

On the basis of the preceding problem, we join points A' and B', respectively symmetrical with A and B about the axes EF and FG. The broken line ACDB is the required path.

The investigation as to what the position of the balls A and B must be in order that a solution of the problem be possible is recommended as an exercise.

Problem. To construct a pentagon, given the midpoints O_1, O_2, O_3, O_4, O_5 of the five sides.

Let the required pentagon be ABCDE (fig. 181).

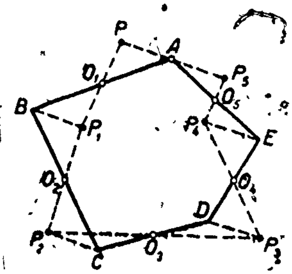


Fig. 181

If point A were known, then by taking the point symmetrical to A with respect to O_1 we would have B; vertex C would be symmetrical with B about O_2, and so on; the fifth symmetrical point, relative to center O_5, would be the starting point A.

Let us take an arbitrary point P and find successively the symmetrical points P_1, P_2, P_3, P_4, P_5 relative to the centers O_1, O_2, O_3, O_4, O_5. By the properties of central symmetry we find:

$$AP = BP_1 = CP_2 = DP_3 = EP_4 = AP_5 ;$$

furthermore, all the segments entering into this equation are parallel to each other and each has an orientation opposite to the preceding one. From this it follows that the required vertex A is the midpoint of segment PP_5. Thus, starting from an arbitrary point P we can construct vertex A and then all the remaining vertices of the required pentagon.

Prove as an exercise that the foregoing construction is applicable only to polygons with an odd number of sides.

Further examples, as well as indispensable problems for independent work will be found in I. I. Aleksandrov: "Collection of Geometrical Construction Problems."

40. SYMMETRY OF SPACE ABOUT A PLANE

Points A and A' are said to be symmetrical with respect to the plane α if they: 1) lie on opposite sides of plane α; 2) lie on the same perpendicular to plane α; and 3) are equidistant from plane α.

Plane α is called the plane of symmetry of points A and A'.

The points in the plane of symmetry are symmetrical with themselves.

A transformation of space into itself which assigns to each point A the point A' symmetrical to A about some plane α is called a symmetry about a plane. A symmetry about a plane is also called a reflexion in the plane.

The following properties of symmetry about a plane are familiar from secondary - school geometry [17].

1. Two points have only one plane of symmetry.

2. Every point lying in the plane of symmetry of two points is equidistant from these points.

3. If a point is equidistant from two given points, it lies in the plane of symmetry of these points.

4. If any two points on one line are symmetrical about a given plane with two points on a second line, then all points of the first line are symmetrical with the points of the second line.

This property may also be formulated as follows: a symmetry about a plane transforms lines into lines.

This property is most simply verified by passing through the line to be transformed a plane perpendicular to the plane of symmetry, and taking the line in the constructed plane which is symmetrical to the given line with respect to the line of intersection of the two planes.

5. If any three non-collinear points A, B and C in one plane are, respectively, symmetrical to three points A', B' and C' in a second plane with respect to some third plane, then all points of the first plane are symmetrical with the points of the second plane.

This property can readily be established with the aid of the preceding one.

Proof: lines AB, BC, CA are transformed, by the preceding property, respectively into lines A'B', B'C' and C'A'. We draw a line through an arbitrary point M in the first plane and mark its points of intersection D and E with any two of the three lines AB, BC, CA. Points D' and E', symmetrical with respect to the third plane to points D and E, lie on those two of the three lines A'B', B'C', C'A' which are the images of the first two. Consequently, line D'E', symmetrical to line DE, belongs to the second plane; but this means that point M', symmetrical with point M, lies in the second plane, q.e.d.

In brief, we may say that a symmetry about a plane transforms a plane into a plane.

Analogously to the case of the plane, it can be shown that two successive symmetries of space about parallel planes are equivalent to one translation; and conversely.

It can also be shown that the sum of two symmetries about intersecting planes is equal to a rotation of space about the line of intersection of these planes through twice the angle between the planes, with the orientation taken from the first plane of symmetry to the second.

The distance between two points, the angle between two planes, the angle between two lines, the angle between a line and a plane all are invariants under a symmetry about a plane.

It is also easy to demonstrate that the set of symmetries about all planes in space is not a group, and that a symmetry about a plane is a one-to-one and bi-continuous transformation, that is to say, it is topological.

Let us consider the tetrahedron SABC and its symmetrical image S'ABC relative to the plane ABC. These two tetrahedrons are, as we know, oppositely oriented. It is not possible to make their corresponding vertices coincide by a motion of space.

Two figures are said to be symmetrical about a plane if a symmetry about this plane transforms one figure into the other.

Figures symmetrical about a plane are in general not congruent. By congruent figures we understand those which can be made to coincide by a motion of the first kind in space; for example, by a helical motion such as we have previously studied. Thus, one's left hand is symmetrical with his right relative to a plane, but not congruent to it.

41. SYMMETRIES OF SPACE ABOUT A POINT

Points A and A' are said to be symmetrical with respect to a point O if O is the midpoint of segment AA'. Point O is called the center of symmetry of points A and A'.

A transformation of space into itself is called a symmetry about a center, or a central symmetry, if it assigns to every point A the point A' symmetrical with A about the given center O.

The following properties of a central symmetry should be recalled from secondary - school geometry [17].

1. Two points A and A' have only one centre of symmetry.

2. The point symmetrical with the center of symmetry is the center itself.

3. If two points of one line are symmetrical with respect to some center with two points of a second line, then all points of the first line are symmetrical with respect to the same center with points of the second line.

Every line passing through the center of symmetry is symmetrical with itself.

Let a be a line not passing through the center of symmetry O. We pass a plane through line a and point O and construct in this plane line a' symmetrical with line a. Line a' is also the image of line a under a symmetry of space about center O.

In brief, a central symmetry transforms a line into a line.

4. If three non-collinear points A, B and C of one plane are symmetrical with respect to some center with three points

A', B', C' of a second plane, then all points of the first plane are symmetrical with respect to the same center with points of the second plane.

The proof is a formal repetition of the proof of proposition 5 of the preceding section, and is left to the student.

In brief, a central symmetry of space transforms a plane into a plane.

It is recommended that the student also prove that two planes which are symmetrical with respect to a center are parallel.

In contrast to a central symmetry in a plane, a central symmetry of space cannot be reduced to a simple rotation of space. On the other hand a symmetry of space about an axis is nothing other than a rotation of space through 180° about this axis.

Two figures symmetrical about a point are in general not congruent, that is, the one cannot be transformed into the other by a motion in space.

For example, a spherical triangle is in general not congruent to the triangle which is symmetrical with it about the center of the sphere.

It is suggested that the student prove that the spherical triangle A'B'C' symmetrical about the center of the sphere with the given spherical triangle ABC, can by a rotation of the sphere about its center be brought into the position ABC", where C" is the point symmetrical with point C relative to the plane of the circle AB, or as we say, spherically symmetrical about the great circle AB.

Due to the fact that it is impossible to make the surface of a sphere coincide with itself "after turning it over" (as can be done with a plane), the distinction between symmetrical and congruent figures on a sphere is an essential one.

It is easy to verify that the set of central symmetries of space about all points is not a group.

42. SYMMETRY OF FIGURES IN A PLANE

Let us consider symmetrical figures in a plane.

A figure is said to be symmetrical about some axis if by a symmetry about this axis it is transformed into itself.

We shall consider some examples of plane figures which possess axial symmetry.

1°. An isosceles triangle is symmetrical about its altitude, which is at the same time also a median and an angular bisector.

In proof, let CD be the altitude of isosceles triangle ABC with base AB (fig. 182).

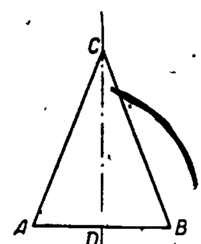


Fig. 182

Since segment AB is perpendicular to CD and AD = BD, point B is symmetrical with point A with respect to the axis CD; point C is symmetrical with itself. Consequently, the symmetry about the axis CD transforms the zero-dimensional triangle ABC into itself.

Further, line AC is transformed by this symmetry into line BC and line BC into AC, while AB is transformed into itself. It follows that the symmetry about CD transforms the linear isosceles triangle ABC into itself.

It can be proved analogously that the two-dimensional isosceles triangle ABC is also transformed into itself by the symmetry about CD.

2°. An equilateral triangle has three axes of symmetry.

This assertion is a corollary of the preceding one.

The set of all symmetries of an equilateral triangle, that is, of all motions (in space or in the plane) which cause this triangle to coincide with itself, constitutes a group.

This group consists of a third-order cyclical subgroup of rotations about the center of the triangle and of three turnings-over (symmetries) about the three axes of symmetry.

This group of six elements is designated as <u>the group of the triangle</u> (20).

3°. A rectangle is symmetrical about a line joining the midpoints of opposite sides.

The proof is suggested as an exercise. The student should also prove that:

4°. A rhombus is symmetrical about each of its diagonals.

As a direct corollary to the two foregoing assertions we have:

5°. A square is symmetrical about the lines joining the midpoints of opposite sides and also about its diagonals. Consequently, a square has four axes of symmetry. It is recommended that the student prove that a square has no other axes of symmetry lying in its own plane.

The set of all symmetries of the square, that is, of motions in the plane or in space which cause the square to coincide with itself, consists of the following transformations: rotations of the square around its center which cause it to coincide with itself (these rotations form, as we know, a fourth-order cyclical subgroup of the group of all symmetries of the square; this subgroup consists of four rotations, including the identity rotation) and of four turnings-over about the four axes of symmetry. These transformations form an eighth-order group, the group of the square.

6°. A regular polygon of n sides has n axes of symmetry. In the case in which n is even, the axes of symmetry are $\frac{n}{2}$ lines joining pairs of opposite vertices of the polygon and $\frac{n}{2}$ lines joining the midpoints of pairs of opposite sides. In the case in which n is odd, the axes of symmetry are lines joining the vertices with the midpoints of the opposite sides.

Before examining the set of all symmetries of a regular polygon of n sides we shall make the following general remark, to which we shall have frequent occasion to refer.

Remark. It is easily seen that all the symmetries of a figure, that is, all motions of the first or second kind which cause the figure to coincide with itself, constitute a group.

Proof: the sum of two motions which cause the figure to coincide with itself is a motion which is likewise a displacement of the figure into itself. We thus have closure of the operation of addition of symmetries. Furthermore, the addition of symmetries, like that of motions in general, possesses the property of associativity. There exists in the aggregate of symmetries of a figure a neutral element, namely the motion which leaves every point of the figure in its original place. Finally, for each symmetry of the figure there exists also an inverse symmetry.

We have already considered the examples of the groups of symmetries of the equilateral triangle and of the square.

Let us study in more detail the group of symmetries of the regular polygon of n sides. We shall show, first, that the n rotations of the n-sided polygon about its center and n symmetries about the axes of symmetry, or turnings-over about the axes, exhaust the entire set of displacements of the regular n-sided polygon into itself. Every rotation of the n-sided polygon about its center is a motion in the plane of the first kind. Every turning-over of the plane is a motion of the second kind.

The addition of symmetries of the first kind yields a symmetry of the first kind; the addition of symmetries of the first kind to symmetries of the second kind yields a symmetry of the second kind; and finally, the sum of two symmetries of the second kind is a symmetry of the first kind.

The symmetries of the first kind form, as we have noted, a cyclical group of rotations about the center of the n-sided polygon through the angles

$$0 ; \frac{2\pi}{n} ; 2\frac{2\pi}{n} ; 3\frac{2\pi}{n} ; \ldots ; (n-1)\frac{2\pi}{n} .$$

There exist n symmetries of the first kind, and they constitute a subgroup of all symmetries of the n-sided polygon.

We shall examine an arbitrary symmetry of the second kind and show that it coincides with one of the previously mentioned turnings-over of the n-sided polygon about one of its axes of symmetry.

After the performance of a given symmetry of the second kind, we carry out some turning-over, chosen once and for all, of the n-sided polygon about one of its axes of symmetry.

The sum of the two symmetries in question will be a symmetry of the first kind, that is, a rotation of the n-sided polygon about its center.

Thus, any symmetry of the second kind when followed by our fixed turning-over results in some symmetry of the first kind:

$$x_2 + a_2 = x_1 ,$$

Where x_2 is any symmetry of the second kind, a_2 is a turning-over about some fixed axis of symmetry of the polygon, and x_1 -- the sum -- is a symmetry of the first kind.

From the above it follows that

$$x_2 = x_1 + (-a_2) .$$

Consequently, every symmetry of the second kind is the sum of one of the rotations x_1 of the n-sided polygon about its center and a completely determined turning-over $(-a_2)$ about a fixed axis of symmetry.

But there exist exactly n rotation symmetries; consequently, there also exist not more than n symmetries of the second kind. But since n symmetries of the second kind -- i.e. turnings-over about axes of symmetry -- are known to exist, these n turnings-over exhaust all the symmetries of the second kind, q.e.d.

At the same time we have established that a regular n-sided polygon has no further axes of symmetry than those already indicated. From the foregoing it also follows that the group of symmetries of a regular n-sided polygon is generated by two elements: a rotation through the angle $\frac{2\pi}{n}$, and any one turning-over.

The group of symmetries of the regular n-sided polygon is of order 2n .

7°. Every diameter of a circle is an axis of symmetry of the circle.

8°. Two intersecting lines have two axes of symmetry, namely, the mutually perpendicular bisectors of the angles formed by the given lines.

9°. Two parallel lines have an infinite set of axes of symmetry.

With this we conclude our consideration of axial symmetry of figures in the plane and turn our attention to central symmetry of figures in the plane.

A figure is said to be symmetrical with respect to a point if by a symmetry about this point it is transformed into itself.

Since a central symmetry in the plane is equivalent to a rotation through the angle π about the center of symmetry, this rotation transforms a figure symmetrical about a point into itself.

Since a rotation about a point through the angle π is equal to the sum of two symmetries about mutually perpendicular axes, a figure symmetrical with respect to each of two mutually perpendicular axes is symmetrical with respect to the point of intersection of these axes.

The converse proposition is not true. A figure can be symmetrical about a point but have no axis of symmetry. As an example of this we may take a parallelogram other than a rectangle or a rhombus.

A regular n-sided polygon has a center of symmetry only when n is even. The center of a circle or of a disk is its center of symmetry.

43. THE SYMMETRY OF FIGURES IN SPACE

A figure is said to be symmetrical with respect to a plane π if it is transformed into itself by a symmetry of space about this plane.

Plane π is called the plane of symmetry of the figure.

A figure is said to be symmetrical with respect to a point O if it is transformed into itself by a symmetry about this point.

Point O is called the center of symmetry of the figure.

A figure is said to be symmetrical about the axis ℓ if it is transformed into itself by a symmetry about this axis.

Axis ℓ is called the axis of symmetry of the figure. Upon a rotation of 180° about this axis the given figure coincides with itself.

This axis is also called an axis of symmetry of the second order because under a complete rotation of the given figure about this axis the figure will in the process of rotation twice coincide with itself.

If under a complete rotation of a figure about some axis the figure coincides with itself several times during the process of rotation, such an axis is said to be an axis of symmetry of higher degree. The order of an axis of symmetry is the number of displacements of the figure into itself which occur during a complete rotation of the figure about the axis.

In defining an axis of symmetry of higher degree the motion of the figure may easily be stripped of its kinematic character and use made only of the geometric concept of a motion.

For this purpose we note that the set of all displacements of a figure into itself occurring during its rotations about an axis of symmetry is a group. This group is called the group of the axial symmetry of the figure. The order of this group is also the order of the axis of symmetry. For example, the perpendicular to the plane of a regular n-sided polygon passing through the center of the polygon is an axis of symmetry of order n of the n-sided polygon.

Let us examine the axes of symmetry of a regular tetrahedron. About the tetrahedron may be circumscribed a sphere, whose center is called the center of the tetrahedron. The center of a tetrahedron is equidistant from its vertices.

Since under all displacements of the tetrahedron into itself its center preserves the same position, all the symmetries of the tetrahedron are rotations about a point. But every rotation about a point is a rotation about an axis (33).

Thus, any symmetry of a tetrahedron is a rotation about some axis passing through its center. This axis is an axis of symmetry of the tetrahedron.

It is thus easy to find the total number of symmetries of a regular tetrahedron.

We drop from any vertex of the tetrahedron a perpendicular to the base. This perpendicular is seen to be an axis of symmetry of the third order. There will be four third-order axes of symmetry, one for each vertex of the tetrahedron (fig. 183).

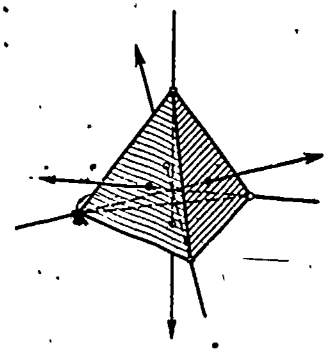


Fig. 183

These axes of symmetry are also axes of symmetry for the respective sides of the tetrahedron.

Edges of the tetrahedron not having a common vertex are called opposite edges. By joining the midpoints of opposite edges we obtain a further three axes of symmetry of the tetrahedron which are of the second order (fig. 184).

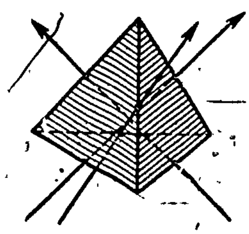


Fig. 184

The tetrahedron has no other axes, since an axis of symmetry must pass either through a vertex or through the center of a side or through the midpoint of an edge of the tetrahedron.

The full number of symmetries of the tetrahedron may now be easily calculated.

Each third-order axis of symmetry furnishes two symmetries which are not the identity transformation, and each second-order axis of symmetry furnishes one. Consequently the full number of the symmetries of the tetrahedron, including the identity, is

2.4 + 1.3 + 1 = 12. We note that the total number of symmetries of the regular tetrahedron is equal to twice the number of its edges.

The set of all symmetries of the tetrahedron, or what is the same, the set of rotations about its seven axes of symmetry, is a group. This group is called the group of the tetrahedron. In this group there are eight proper subgroups. Of these, seven are cyclical subgroups corresponding to rotations of the tetrahedron about each of the axes of symmetry of the tetrahedron. The eighth proper subgroup consists of all the rotations about the three second order axes of symmetry and the identity rotation.

It is suggested as an exercise that the student find all the planes of symmetry of the tetrahedron.

We turn to the determination of the axes of symmetry of the cube. We observe that, as in the case of the tetrahedron, under all symmetries of the cube its center preserves its position unchanged. It follows from this that all symmetries of the cube are rotations of the cube about axes passing through its center. These axes are the axes of symmetry of the cube.

We shall begin with the axes passing through vertices of the cube. Through each vertex A pass three distinct edges of the cube, AB , AD and AE (fig. 185).

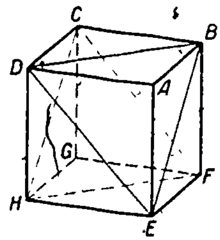


Fig. 185

The other ends of these edges determine the equilateral triangle BDE , which forms the base of a regular pyramid with vertex at A . The axis passing through vertex A and perpendicular to the plane BDE is an axis of symmetry of the cube, since a rotation of the regular pyramid ABDE about its altitude which displaces the pyramid into itself is at the same time a symmetry of the cube.

There exist four distinct axes of symmetry of the cube which pass through its vertices. This is because, for instance, the axis of symmetry passing through vertex G of the cube will be perpendicular to plane CFH which is parallel to plane BDE and, like the first axis will pass through the center of the cube. Consequently, the axes of symmetry passing through opposite vertices of a cube are identical.

In the same way we establish that there are three axes of symmetry each perpendicular to a pair of opposite sides of the cube and passing through their centers; these are fourth-order axes of symmetry. There exist six axes of symmetry passing through the midpoints of pairs of opposite edges of the cube; these are second-order axes of symmetry.

Thus, the cube has thirteen axes of symmetry: four diagonals, three lines joining the midpoints of opposite sides of the cube, and six lines joining the midpoints of opposite edges of the cube.

The set of all symmetries of the cube constitutes a group, which is called the group of the cube. Let us calculate the number of the elements of this group.

Aside from the identity symmetry we have: two rotations about each axis passing through a vertex; three rotations about each axis passing through the center of a side; one rotation about each axis joining midpoints of opposite edges. Including the identity symmetry we have in all:

$$2 \cdot 4 + 3 \cdot 3 + 1 \cdot 6 + 1 = 24.$$

The group of the cube is a group of order 24. The number of all symmetries of the cube is twice the number of its edges.

As an exercise let the student find all the proper subgroups of the group of the cube. Find also all the planes of symmetry of the cube. It will be a useful exercise to find the group of symmetries of each of the regular polyhedrons and their subgroups. [4], [17].

To conclude this section we shall give a general definition of symmetrical figures which is due to the eminent Russian crystallographer and mathematician, Academician Y. S. Feodorov (1853-1919).

"A symmetrical figure is one which can be made directly to coincide with itself in various positions or else such that coincidence in various positions can be produced if we replace it by a second figure, related to the first as a mirror image to the object mirrored ... that form of symmetry under which in order to make the figure coincide with its original position it is necessary simultaneously to rotate it about some axis through a determined angle and to replace it with its mirror image, I call complex symmetry ... the most general motion is a helical motion ... In general, therefore, if a figure can be made directly to coincide with itself in various positions, we can always find this helical axis of symmetry." [45]

Feodorov shows that for finite figures the parameter of the helical motion (see 34) is zero, and furthermore, for finite figures the axes of symmetry always intersect. For infinite figures an axis of helical motion with helical parameter differing from zero is possible, as are non-intersecting axes of symmetry.

Feodorov introduces the concept of regular systems of figures. "By a regular system of figures I mean an aggregate of finite figures extending infinitely in every direction such that if in accord with the laws of symmetry we bring into coincidence two of the component figures of the system, all the figures of the system are made to coincide simultaneously [45].

Feodorov proves that exactly 230 regular systems of figures exist, and determines all of them, and at the same time solves the problem of the possible structure of crystals. E. S. Feodorov's results constitute one of the great achievements of science.

Chapter VII

SIMILARITY

In this chapter we shall study the transformations of the plane and of space known as central similitudes.(*) In connection with central similitudes the general group of similitude transformations and its invariants will be considered. A theorem is proved which asserts that a general similitude, distinct from a motion, is the sum of a rotation and a central similitude. A definition of similar figures is given and their properties are studied.

In 48 we present the method of similarity in the solution of construction problems.

44. CENTRAL SIMILITUDES OF THE PLANE

A point A' of a plane is said to be centrally similar to a point A with respect to a center O if the following conditions are fulfilled:

1) Point A' lies on the line OA

2) the equation OA' = k·OA holds good, where the given number $k \neq 0$;

3) if $k > 0$, point O lies outside the segment AA', while if $k < 0$, point O lies between points A and A'.

A transformation under which every point A of a plane is mapped into a point A' which is centrally similar to point A with respect to a center O, with the number k constant for all points, is called a central similitude of the plane.

Point O is the center of similarity. The number k is the coefficient of similarity.

For $k > 0$ a central similitude is said to be direct; for

(*) Also known as Homothetic transformations -- translators.

$k < 0$, inverse. The center of similarity 0 in the first case is said to be external; in the second case, internal (figs. 186, 187).

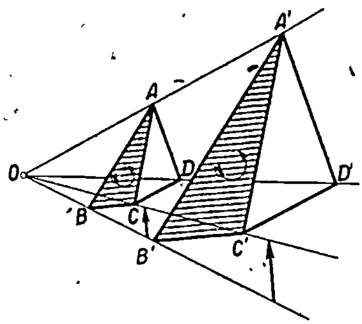


Fig. 186.

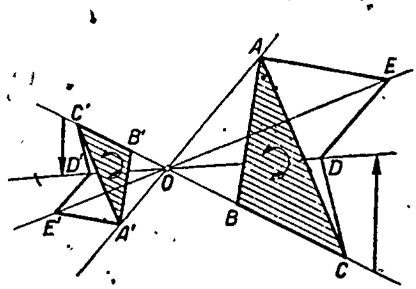


Fig. 187.

The center of similarity 0 is the only point which goes over into itself under a central similitude (if $k \neq 1$).

It follows from the definition of a central similitude that it is a one-to-one transformation.

The following properties of central similitudes are familiar from secondary - school geometry [17].

1°. A segment AB not lying on a line passing through the center 0 is transformed into the segment A'B' parallel to the given segment, and

$$A'B' = k \cdot AB .$$

This equation follows directly from the similarity of triangles OAB and OA'B' (figs. 186 and 187).

2°. Under a central similitude a straight line is transformed into a straight line parallel to the given line.

3°. Angles between lines preserve their magnitudes, that is, the angle is an invariant of central similitudes.

This property is a direct corollary of the preceding one.

4°. In every pair of centrally similar figures, corresponding segments of straight lines are proportional and corresponding angles are equal.

5°. Under a central similitude a circle is transformed into a circle, the image of the center being the center of the transformed circle and the ratio of the radii of the given and the transformed circles equal to the coefficient of similarity.

Making use of this property, it is not difficult to prove that a central similitude is not only one-to-one but also bicontinuous, i.e., it is a topological transformation.

6°. The ratio of the area of a figure obtained by a central similitude to the area of the given figure is equal to the square of the coefficient of similarity.

If S is the area of the given figure and S' the area of its image, then $S' = k^2 \cdot S$.

The following further properties of central similitudes are immediately obvious.

7°. Any two centrally similar triangles have the same orientation; any two corresponding angles in centrally similar figures have the same direction (figs. 186 and 187).

8°. Any two centrally similar vectors have the same orientation under a direct similitude (with $k > 0$) and opposite orientations under an inverse similitude (with $k < 0$) (figs 186, 187).

In contrast to the elementary transformations already studied, under central similitudes (with $k \neq 1$) the distance between two points is not preserved.

The length of a segment is not an invariant of similitudes (if $k \neq 1$), whereas the property of being a segment is an invariant property.

Three points A B C have an invariant namely, the ratio of the distances of any one of the three points to the two others:

$$\frac{AB}{BC} \text{ is invariant.}$$

Proof: the transformation takes points A, B and C into points A' B' C' where $A'B' = k \cdot AB$ and $B'C' = k \cdot BC$.

The invariance of the ratio in question follows from the above, that is,

$$\frac{A'B'}{B'C'} = \frac{AB}{BC}$$

In particular, if the three points are collinear there exists invariance of the ratio in which one point divides the segment bounded by the two others.

Under a central similitude the shape of a figure is invariant, but the dimensions of the figure (if $k \neq 1$) are changed.

In conclusion we shall prove the theorem on three centers of similarity:

<u>The centers of similarity of three figures which are pairwise similar lie on a straight line.</u>

Let O_{12}, O_{13}, O_{23} be the centers of similarity of the figures whose numbers appear in the subscripts (fig. 188). We shall denote by s the line $O_{23} O_{13}$ and show that line s passes through point O_{12}.

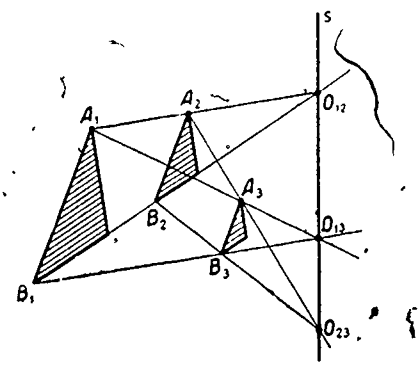


Fig. 188.

Let us consider line s as a part of figure one. Since line s passes through the center O_{13}, under the similitude having this center and taking figure one into figure three the line s goes over into itself.

Now let us consider line s as a part of figure three. The similitude having the center O_{23} and taking figure three into figure two also takes line s into itself. From this it follows that the line s which is part of figure two corresponds to the same line s as a part of figure one. Thus, the similitude having center O_{12} and taking figure one into figure two takes line s into itself. This is possible only in the case that line s passes through center O_{12}, q.e.d. [1]

The line s is called the <u>axis of similarity</u> of the three pairwise centrally similar figures.

As an exercise, define the central similitudes of space and establish their properties, analogous to those of the central similitudes of the plane.

---

1) If we denote by S_{ij} the similitude with center O_{ij} and taking figure i into figure j (i, j - 1, 2, 3) then an implicit assumption in the proof is that if S_{13} carries <u>any</u> point P of the plane into a point P' and S_{32} carries P' into P", then S_{12} will carry P into P". This may either be considered as a part of the hypothesis of the theorem. However, we may consider the hypothesis to be merely that S_{12} carries P into P" when P is a part in figure 1. That this is true for <u>any</u> point P in the plane will then follow easily from the proposition that the sum of two central similitudes is either a translation or a central similitude (end of 46). -- Translators

45. SIMILARITY OF CIRCLES

Any two unequal and non-concentric circles can be regarded as centrally similar; and this can be done in four different ways. That is to say that there exist four central similitudes of the plane under which one circle is transformed into the other.

In order to find a central similitude it is necessary to discover its center and its coefficient of similarity; for this it is sufficient to know the end-points of two parallel corresponding segments having the same or opposite orientations.

Let S_1 and S_2 be the centers of two unequal circles, the radii of which are R_1 and R_2 (fig. 189).

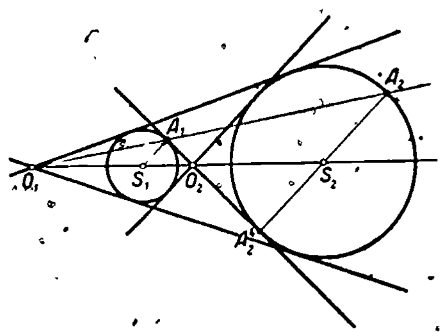


Fig. 189.

If the required transformations are to transform the first circle into the second, then the coefficient of similarity must be

$$k = \frac{R_2}{R_1}.$$

In order to find the centers of the required similitudes, we draw the diameter $A_2A'_2$ of the second circle, and parallel to it the radius S_1A_1 of the first circle. Regarding A_1 and A_2 as corresponding points under the required similitude we find the center O_1 of the similitude as the intersection of the lines A_1A_2 and S_1S_2. If A_1 and A'_2 are regarded as corresponding points, the center of similiarity will be point O_2.

We have thus found four central similitudes which transform one of the given circles into the other. The first transformation with center O_1 transforms circle (S_1) into circle (S_2); its coefficient of similarity is $k_1 = \frac{R_2}{R_1}$. The second transformation, with the same center O_1 but with the coefficient of similarity $k_2 = \frac{R_1}{R_2}$, transforms circle (S_2) into circle (S_1). In exactly the same way there are two transformations with center O_2 transforming the first circle into the second and vice versa.

Point O_1 is the exterior center of similarity of the two circles; point O_2 is the interior center of similarity.

If the centers O_1 and O_2 of the circles coincide, the common and unique center of the two central similitudes, inverse to each other, which transform these concentric circles one into the other will coincide with the common center of the two circles and the coefficients of similarity will be:

$$k_1 = \frac{R_2}{R_1}, \quad k_2 = \frac{R_1}{R_2}.$$

It follows from the one-to-one property of central similitude transformations that tangents drawn from the exterior center of similarity of two circles to one of the circles are also tangent to the other circle. The same statement holds good relative to the interior center of similarity, providing that this center does not lie within either of the given circles (fig. 189).

From the theorem on the three centers of similarity presented in 46 there follows the theorem on the six centers of similarity of three circles.

If the centers of three circles no two of which are equal are not collinear, then the six centers of similarity of pairs of these circles lie in sets of three on each of four straight lines. (fig. 190). The proof is left as an exercise for the student.

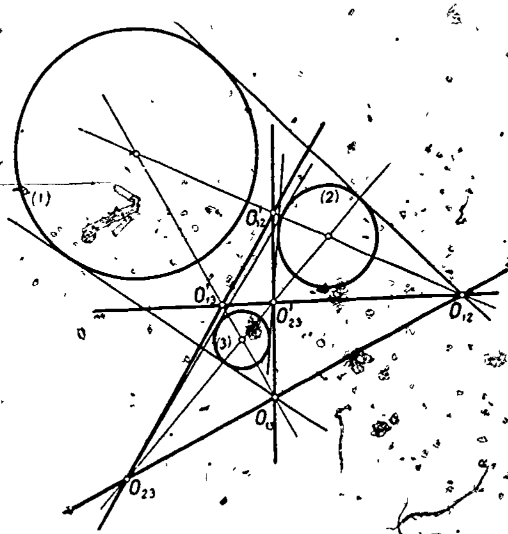


Fig. 190.

In the diagram the six centers of similarity are designated by letters with appropriate subscripts: O_{12}, O_{13}, O_{23}, O'_{12}, O'_{13}, O'_{23}.

The four lines on each of which lie three of the centers of similarity of the three circles are called the axes of similarity of the three circles.

46. THE GROUP OF SIMILITUDE TRANSFORMATIONS WITH A COMMON CENTER

The set of all central similitudes of the plane which have a common center O constitutes a group.

Proof: the sum of two such transformations with coefficients k_1 and k_2 and common center O will be a similitude with the coefficient $k = k_1 \cdot k_2$ and the same center, since

$$OM' = k_1 \cdot OM \text{ ; } OM'' = k_2 \cdot OM' = k_2 \cdot k_1 \cdot OM$$

and the points O , M , M' and M'' are collinear.

The requirement of the existence of a neutral element is satisfied: the identity transformation is the similitude with coefficient $k = 1$. The requirement of the existence of an inverse for each element is also satisfied. The central similitude with coefficient $k_2 = \frac{1}{k_1}$ is inverse to the similitude with coefficient k_1 and the same center O. Thus, all the requirements for a group are satisfied and the assertion is proved.

The group of similitudes having a common center is isomorphic to the group of all non-zero real numbers, with respect to the operation of multiplication.

If to every similitude with center O and coefficient $k \neq 0$ is assigned the number k , a sum of similitudes will correspond to the product of the numbers assigned to each of the similitudes added. This isomorphism shows that for the group of similitudes with a common center it is more convenient to employ the terminology of multiplication (20).

The set of all possible central similitude transformations in a plane, with all points of the plane as centers, does not constitute a group. In order to verify this it is sufficient to find even one example of a sum of two transformations, elements of this set, which is not a central similitude, since this would mean that the requirement of closure of the operation of addition was not satisfied. It is in fact easy to find two central similitudes, elements of this set, whose sum is a translation and therefore not a central similitude.

Let there be given an arbitrary translation in a plane, transforming points A and B respectively into points A" and B".

Let us select the center O_1 of a similitude which, as indicated in figure 191, will take points A and B into points A'B'.

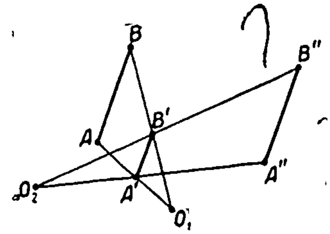


Fig. 191.

Point O_2, the intersection of lines A'A" and B'B", will be the center of a second similitude which takes points A' and B' respectively into points A" and B". The required similitudes have been constructed, q.e.d.

Proposition. The sum of two similitudes with coefficients k_1 and k_2 is a translation when $k_1k_2 = 1$, and a similitude transformation otherwise.

Proof: Let S_1, S_2 be similitudes with coefficients k_1, k_2, and let $S_3 = S_1 + S_2$ be their sum. Let A, B be any two points, A', B', their images under S_1 and A'', B'' the images of A', B' under S_2. Then the segments AB, $A'B'$ and $A''B''$ are parallel and $A''B'' = k_1k_2\, AB$. If $k_1k_2 = 1$ then $\overrightarrow{AB} = \overrightarrow{A''B''}$; hence $\overrightarrow{AA''} = \overrightarrow{BB''}$, and S_3 is a translation. If $k_1k_2 \neq 1$ then lines AA'' and BB'' intersect at a point O. We shall show that O is fixed under the transformation S_3. For, if O'' is the image of O then the segment $O''A''$ is parallel to OA and since A'' lies on line OA so must O''. Similarly O'' lies on line OB. Therefore, $O = O''$. We shall now show that any line passing through O is transformed into itself by S_3. For let P be any part and P'' its image then the image of the segment OP is the segment OP'' which is parallel to it and passes through O. Hence P'' is collinear with O and P. Furthermore, $OP'' = k_1k_2 \cdot OP$. This shows that S_3 is a central similitude.

If we describe the group operation in the multiplicative instead of additive terminology, then instead of the sum we have to speak of the product of central similitude transformations.

It is suggested as an exercise that the student examine the properties of central similitude transformations of space, which are analogous to the properties of central similitude transformations in the plane just studied.

47. THE GENERAL CASE OF THE SIMILARITY OF FIGURES IN THE PLANE

Let us consider the one-to-one transformation of the plane into itself possessing the property that, if A' and B' are the images of two points A and B, the ratio of segment A'B' to segment AB is a constant for all pairs of points A and B in the plane:

$$\frac{A'B'}{AB} = k = \text{const.}$$

The constant ratio k is called the coefficient of similarity. The mapping itself is called a similitude transformation, or simply a similitude.

That such transformations exist may be shown by means of examples. Any motion in the plane is seen to be a similitude transformation: its coefficient $k = 1$. Any central similitude with $k > 0$ is likewise a similitude. The identity transformation of the plane is a similitude.

We shall consider some properties of similitudes.

1. Under a similitude transformation a line in the plane goes over into a line, that is, collinear points are transformed into points which are likewise collinear.

In proof, let point C lie on the segment AB and let the images of points A, B and C be A'B' and C'. From the definition of similitude we have:

$$A'C' : AC = C'B' : CB = A'B' : AB,$$

whence

$$(A'C' + C'B') : (AC + CB) = A'B' : AB.$$

Thus, from the equation

$$AC + CB = AB ,$$

it follows that

$$A'C' + C'B' = A'B' .$$

But the last equation is true only on the condition that point C' lies on the segment $A'B'$; q.e.d.

It follows from the foregoing that:

2. A segment is transformed by a similitude into a segment; a ray, into a ray; a half-plane into a half-plane; an angle into an angle.

3. The angle between two lines is an invariant under a similitude.

This follows from the fact that under a similitude any triangle ABC is transformed into a triangle $A'B'C'$ similar to the first since by the definition of similitude transformations we have

$$A'B' : B'C' : C'A' = AB : BC : CA .$$

The following property pertains to the construction of similitudes.

4. A similitude is completely determined by giving two pairs of corresponding points and the orientation of the image of any triangle. 1)

1) In the following, we take for granted the following property of similitudes: a similitude of the plane either preserves the orientation of all triangle or reverses the orientation of all triangles. -- Translators.

Let A' and B' be the images of points A and B (fig. 192).

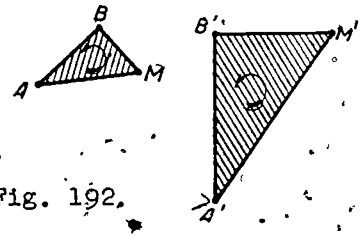


Fig. 192.

The coefficient of similarity is determined by the relationship

$$k = \frac{A'B'}{AB}$$

Let us construct the image of an arbitrary point M' as the vertex M' of a triangle A'B'M', which is similar to triangle ABM if M does not lie on line AB. If triangles AMB and A'M'B' have the same orientation we have a proper similitude; with opposite orientations we have a mirror similitude.

The relationship

$$\frac{A'M'}{AM} = \frac{B'M'}{BM} = k$$

shows also how to construct the image M' of point M in the case in which M lies on line AB.

The construction of corresponding points under a similitude has thus been indicated. From the construction it follows also that a similitude is uniquely determined by giving two pairs of corresponding points and the preservation or reversal of the orientation. Furthermore, it is clear that any two pairs of corresponding points under a similitude completely determine this transformation if orientation is preserved.

Two figures in a plane are similar if one can be transformed into the other by a similitude.

Corresponding points under a similitude are called analogous points.

Corresponding segments, lines, etc. of two similar figures are called similiar lines, similar segments, etc.

5. Every proper similitude which is not a motion or a central similitude is the sum of a rotation of the plane about some point O and a central similitude with its center at that point.

Note that in such a case the point O will be a fixed point under the transformation, that is, it will coincide with its image. We shall give a method of finding the fixed point which will at the same time enable us to find the rotation and the central similitude, the sum of which will be the given proper similitude.

Let us suppose that O (fig. 193) is the required fixed point of the similitude determined by the two pairs of corresponding points

$A \rightarrow A'$ and $B \rightarrow B'$.

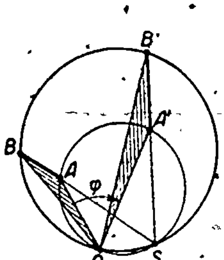


Fig. 193.

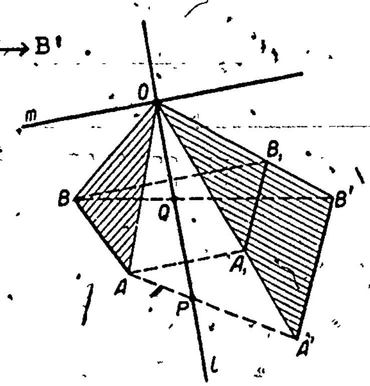


Fig. 194.

Since by assumption the given sillilitude is not a motion and not a central similitude, $A'B' \neq AB$ and the segment $A'B'$ is not parallel to the segment AB .

Let S be the point of intersection of lines $A'B'$ and AB . Triangle $OA'B'$ is similar to triangle OAB. Consequently,

$\angle OAS = \angle OA'S$; but this means that the four points O , S , A' and A lie on a circle. Similarly, the points O , S , B and B' also lie on a circle.

Thus, the fixed point O is a point of intersection of the circle passing through points S , A and A' with the circle passing through points S , B and B' . The given imilitude is the sum of a rotation about center O through the angle $\phi = \angle AOA'$ plus a central similitude having the same center O and the coefficient $k = \frac{A'B'}{AB}$.

The analysis of the case in which the fixed point S and the point of intersection O coincide is suggested as an exercise. In order to avoid such a coincidence it is sufficient to take any other corresponding points distinct from say, B and B' . The theorem is proved.

At the same time it has been shown [1] that under every proper similitude of the plane which is not a translation, there exists a unique double point O .

---

1) Although what the author seems to have given is a procedure for finding the fixed point O , on the assumption that it exists, the student will find after a little reflexion that we have here also a proof that a fixed point does exist. Thus if the circles determined by S , A , A' and S , B , B' intersect in a second point O , it is easy to see that the triangles OAA' and OBB' have corresponding angles equal and the same orientation. Hence the point O coincides with its own image. If the circles determined by S , A , A' and S , B , B' intersect only at S , it is easy to show that

$$\frac{SA'}{SA} = \frac{A'B'}{AB}$$

which, in turn, may be used to derive that S coincides with its image. Thus, in all cases, a fixed point exists under a proper similitude. -- the Translators.)

6. Every mirror similitude transformation which is not a motion of the second kind is the sum of a central similitude about some point O and a reflexion about a line passing through that point.

Let O be the center of the required similitude (fig. 194). Since line OA' must be symmetrical with line OA and line OB' with OB, the axis of symmetry ℓ must bisect the angle O in triangles AOA' and BOB', and consequently must pass through points P and Q which divide the corresponding segments AA' and BB' in a ratio equal to the coefficient of similarity k :

$$\frac{A'P}{PA} = \frac{OA'}{OA} = \frac{A'B'}{AB} = k\ ;\ \frac{B'Q}{QB} = \frac{OB'}{OB} = \frac{A'B'}{AB} = k\ .$$

Taking segment A_1B_1 symmetrical with AB about axis ℓ, we find center O as the point of intersection of axis ℓ with line A_1A'.

The investigation of the case when lines ℓ and A_1A' are parallel is suggested as an exercise. 1)

It follows from property 6 that under every mirror similitude of the plane which is not a motion there exist a double point O and two mutually perpendicular double lines ℓ and m passing through point O (fig. 194).

7. The set of all proper similitudes of the plane constitutes a group.

1) Here also, the authors seem to begin by assuming that a fixed point and an axis of symmetry exist and merely gives a procedure for finding them. The student should prove that the point O determined by this procedure is a fixed point, and the line ℓ is an axis of symmetry. -- the Translators.

Motions of the second kind constitute a subgroup of this group.

8. <u>The set of all similitudes of the plane constitutes a group.</u>

The proper similitudes form a subgroup of this group. The set of mirror similitudes is not a group.

The set of all motions of both the first and second kinds, i.e., the set of all orthogonal transformations, is also a subgroup of the group of similitude transformations.

The detailed proof of these last assertions is suggested as an exercise.

The group of similitudes is precisely that group the invariants of which -- and the invariants of whose subgroups -- are studied in elementary geometry. We note, in particular, that two points have no invariant under similitude transformations. Three points A, B and C have an invariant. This invariant is the <u>ratio</u> of the distances of one of the three points from the other two : $\frac{AB}{AC}$ <u>is invariant</u> = const.

If three points are collinear, this invariant of the three points is the ratio in which one point divides the segment bounded by the other two.

Analogously to the case of the plane just considered, it is easy to establish the properties of similitude transformations of space into itself.

In conclusion, we shall point out that the group of similitudes is only a subgroup of the broader group of <u>affine</u> transformations. Orthogonal transformations and similitudes are affine. The stretchings considered in 19 are also affine transformations.

The general definition of an affine transformation of space is as follows: a one-to-one mapping of space onto itself is said to be affine if every collinear triad of points is mapped into a collinear triad of points. That such transformations exist is shown by the examples given. Starting from this definition, we can obtain all the properties of affine transformations.

The set of the affine transformations of space is easily seen to constitute a group, known as the group of affine transformations. The study of the invariants of this group of transformations and of the properties of the group itself forms the content of affine geometry. [21]

In this geometry there is no notion of angle, of length, no ratio of non-parallel segments, no notion of area, since all these magnitudes are not invariants of the affine group of transformations. The ratio of parallel segments and the ratio of areas are invariants. The study of affine geometry is outside the scope of the present text.

48. THE METHOD OF SIMILITUDE IN THE SOLUTION OF CONSTRUCTION PROBLEMS

The method of similitude, like the methods involving the other elementary transformations, is often useful in solving many construction problems. Namely, it is sometimes easier to construct a figure, or part of one, which is similar to the required figure. This may be the case, for example, when one part of the conditions of the problem determines the shape of the required figure and another part determines its size. Setting aside the second part of the conditions, we proceed to construct a figure similar to the required one. The construction of similar figures amounts in practice to the construction of centrally similar figures (44 and 47).

Problem. To inscribe a square in a triangle.

We construct the square BCDE on the side BC so that vertex A and the constructed square lie on opposite sides of the line BC (fig. 195).

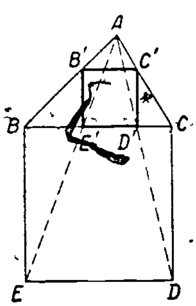


Fig. 195.

The required square B'C'D'E' is centrally similar to square BCDA, the center of similarity being point A. From this the construction of the required square becomes clear.

Problem. In a given triangle ABC to inscribe a second triangle whose sides are parallel to the given lines u, v and w (fig. 196).

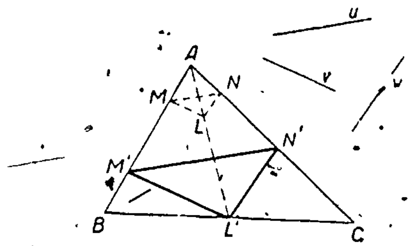


Fig. 196.

We draw arbitrarily a line MN parallel to line u and intersecting sides AB and AC of the given triangle at M and N. Through points M and N we draw lines ML and NL, respectively parallel to lines v and w. We mark point L', the intersection of line AL with side BC. The required triangle M'L'N' is similar to triangle MLN with the center of similarity at point A.

Problem. Given three lines, a, b and c, passing through point O (fig. 197).

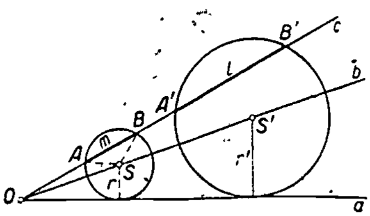


Fig. 197.

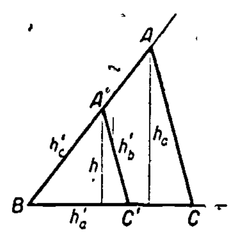


Fig. 198.

To describe a circle tangent to line a, having its center on line b and cutting off on line c a chord of a given length ℓ.

From an arbitrary point S on line b as center, we draw a circle of radius r tangent to line a. Let AB = m be the segment cut off by this circle on line c. The required circle, with center S' and radius r', is one similar to the first circle, the center of similarity being at point O and the coefficient $k = \frac{l}{m}$; the radius $r' = k \cdot r$.

Problem. To construct a triangle, being given its altitudes h_a, h_b, h_c.

In the first place we note that the altitudes of any triangle are inversly proportional to the corresponding sides:

$$h_a : h_b : h_c = \frac{1}{a} : \frac{1}{b} : \frac{1}{c} .$$

In proof of this, we have the following expression for twice the area of a triangle:

$$ah_a = bh_b = ch_c = 2S ,$$

whence

$$h_a : h_b : h_c = \frac{2S}{a} : \frac{2S}{b} : \frac{2S}{c} = \frac{1}{a} : \frac{1}{b} : \frac{1}{c} .$$

In exactly the same way, the equation

$$a : b : c = \frac{1}{h_a} : \frac{1}{h_b} : \frac{1}{h_c}$$

holds good.

With this in mind, we construct a triangle the sides of which are equal to the altitudes of the required triangle: by what we have just proved we have for the altitudes of the constructed triangle the expression

$$h'_a : h'_b : h'_c = \frac{1}{h_a} : \frac{1}{h_b} : \frac{1}{h_c} = a : b : c .$$

This means that the required triangle is similar to the triangle whose sides are the altitudes of the triangle with sides h_a, h_b and h_c.

The construction of the required triangle follows from this. We construct a triangle with sides h_a, h_b, h_c and having found its altitudes h'_a, h'_b and h'_c we construct triangle A'BC' with sides equal to these altitudes and then enlarge its dimensions by means of a similarity with center at point B, as shown in fig. 198. The construction also makes evident the condition for the possibility of a solution to the problem:

$$h_a + h_b > h_c .$$

Further examples and essential problems for independent work are to be found in the book by I. I. Aleksandrov already mentioned.

Chapter VIII

INVERSION

In Chapter VIII we study inversions of the plane and of space. In 51. the application of inversion transformations to the solution of geometrical construction problems is considered, and in connection with that the question is answered as to the possibility of solving with the aid only of compasses which can be set without restriction as to radii all construction problems solvable with compasses and straight-edge. In the selection of examples those problems are examined which will be of basic importance for the interpretation of the geometry of Lobachevskii, to be studied in Part IV.

Among the invariant properties of figures under inversion transformations, emphasis is given to those properties having the greatest importance for the study of Lobachevskian geometry.

49. INVERSIONS IN THE PLANE

Let there be given a circle with center O and radius R. Point M' is said to be <u>inverse</u> to point M with respect to the given circle if:

1) points O, M and M' lie on the same ray issuing from the center O;

2) the following equation holds:

$$OM' \cdot OM = R^2.$$

The point inverse to point M' will in turn be point M. Hence points M and M' are said to be inverse to each other with respect to the given circle.

A transformation of the plane under which to each point M there is assigned its inverse point M', is called <u>an inversion of the plane with respect to the given circle</u>, or simply <u>an inversion</u>.

The given circle is called the circle of inversion; its center, the center or pole of inversion; its radius, the radius of inversion; the square of the radius, the degree of inversion.

Two figures which correspond to each other under an inversion are said to be inverse to each other.

The pole of inversion has no inverse point, and for this reason an inversion is not a one-to-one transformation of the plane into itself.

We have already encountered such a phenomenon in studying the stereographic projection of the sphere onto a plane (17.).

In what follows we shall be dealing with a plane "punctured" at the pole, that is, a plane with point O removed.

An inversion is a one-to-one mapping onto itself of the plane punctured at the pole of inversion.

One considers also the transformation known as an inversion of negative degree $-R^2$, by which is understood the sum of an inversion of degree R^2 and a symmetry about the pole of the inversion.

Prefixing the signs plus or minus to oriented segments of an axis passing through the center of inversion O, we can write for points M and M' under an inversion of negative degree the equation

$$OM' \cdot OM = -R^2.$$

An inversion of negative degree does not have a circle consisting of double points. The circle of inversion remains invariant, but its points do not preserve their original position. Hereinafter we shall be speaking of inversions of positive degree unless the contrary is stated.

As an exercise it is suggested that for each property of inversions studied hereinafter the student should formulate and prove the corresponding property for inversions of negative degree.

We shall give methods for the construction, of the inverse point M' of any point M once the pole and circle of the inversion are known.

1. The following construction takes the definition of reciprocally inverse points as its point of departure:

a) Suppose point M to lie inside the circle of inversion. We draw through M, perpendicular to the ray OM, a chord, T_1T_2, of the circle of inversion (fig. 199). We construct the tangents T_1M' and T_2M' to the circle at the endpoints of this chord. The point M' at which these tangents intersect will be the point inverse to point M.

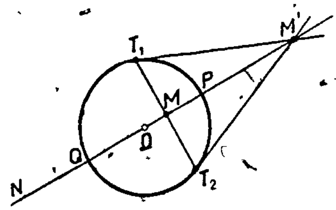


Fig. 199.

The closer point M is to the pole O, the farther away will be its inverse point M'; if the center O is taken as point M, the tangents at the ends of a diameter will be parallel and have no point of intersection. The closer point M is to the circle of inversion, the closer to this circle will the inverse point M' be also.

Every point P of the circumference (fig. 199) is seen to be fixed under the inversion of the plane, since $OP \cdot OP = R^2$.

b) Suppose point M to lie outside the circle of inversion. Since the image of point M' is point M, the same construction, performed in the opposite order, will yield the required point.

We draw from the given point M' (fig. 199) tangents MT_1 and MT_2 to the circle of inversion and, joining the points of tangency T_1 and T_2, we find the required point M as the midpoint of this chord.

Before presenting a second method of constructing inverse points we shall prove the following lemmas:

α) If A, A' and B, B' are two pairs of reciprocally inverse points and the points O, A and B are not collinear, the triangles OAB and OB'A' (O being the pole of inversion) are mirror-similar.

By the definition of inverse points, O, A and A' lie on one ray, as do O, B and B' (fig. 200), and

$$OA' \cdot OA = R^2 \; ; \; OB' \cdot OB = R^2.$$

Consequently, in the triangles OAB and OA'B', having a common angle at O,

$$\frac{OA'}{OB'} = \frac{OB}{OA},$$

which proves the proposition.

The converse proposition is also true, as can be verified without difficulty.

β) If triangles OAB and OB'A', with the angle O in common, are mirror-similar, then the pairs of points A, A' and B, B' are pairs of inverse points under an inversion having O as its center.

From the foregoing it follows also that a circle may be circumscribed about the quadrilateral AA'B'B.

2. If the pole of inversion O and one pair of reciprocally inverse points A and A' are given, the construction of a point B' inverse to a point B is performed as follows.

We draw ray OB and lay off on it the segment OA_1 equal to segment OA (fig. 200). On ray OA we lay off segment OB_1 equal to segment OB. Drawing from point A' a line A'B' parallel to line B_1A_1, we find the required point B' as the intersection of lines A'B' and OB.

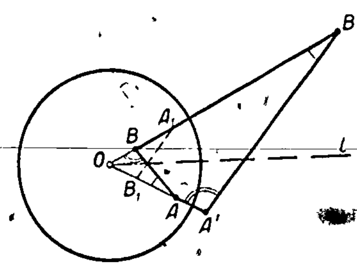


Fig. 200.

The construction is based on the propositions just proved. This construction is no other than the decomposition of a mirror-similarity transformation into the sum of a symmetry about the axis l and a similarity with O as center (fig. 200).

3. The construction of an inverse point using compasses only.

We know (11.) that every construction possible with the free use of compasses and straight-edge can also be carried out with the use of compasses only. This proposition was established by carrying out successively the five basic constructions (10) using only a pair of compasses. It is possible, however, to seek a desired construction with compasses directly, without reducing it to a chain of the basic constructions.

Let us find the point A' inverse to a given point A with respect to a given pole O and circle of inversion Γ (fig. 201). With radius AO and center at A we draw a circle and denote by U_1 and U_2 the intersections of this circle with the circle of inversion Γ. (In fig. 201, point A -- the center of circle U_1OU_2 -- is not shown.)

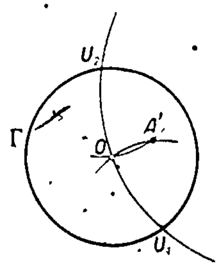


Fig. 201.

With radius $U_1O = U_2O$ and centers at U_1 and U_2 we draw circular arcs. Their point of intersection distinct from pole O will be the required point A' inverse to point A.

<u>Proof:</u> Points A and A' lie on one ray OA, and from the similarity of the isosceles triangles OAU_1 and OU_1A' it follows that

$$\frac{OA'}{OU_1} = \frac{OU_1}{OA}, \text{ or } \frac{OA'}{R} = \frac{R}{OA},$$

that is to say, $OA' \cdot OA = R^2$, q.e.d.

This construction can always be carried out for points A which are exterior to the circle of inversion. If A is a point interior to this circle, the circle drawn from A with radius OA will intersect the circle of inversion only if $AO \geq \frac{1}{2} R$.

Thus, the above construction does not possess the necessary generality. It is suggested that the student develop a generally applicable construction of the point inverse to point A, using only the pair of compasses.

Let us examine some properties of inversion.

1. <u>A line passing through the pole of inversion is transformed under the inversion into itself.</u>

This follows from the first clause of the definition of inverse points.

Points M on the radius OP in fig. 199 are transformed into points M' of ray PM', while points on radius OQ are transformed into points of the ray QN. Point O has no inverse point.

An inversion, like a symmetry, turns around every ray OP about point P (about the tangent to the circle of inversion at point P), the radius PO being at the same time stretched to infinity while the infinite ray PM' is shrunk to the dimensions of radius PO.

A symmetry is a half turn of the plane about an axis of symmetry; in an inversion, the role of the axis of symmetry is played by the circle of inversion.

If the radius of the circle of inversion is "large", then the circle of inversion, "differs little from a straight line", and "in the neighborhood" of the circle, the inversion transformation is "approximately" a symmetry.

2. A circle having its center at the pole of inversion is transformed into a concentric circle.

From the definition of inversion it follows that a circle of radius OM is transformed into a circle of radius OM', where M and M' are reciprocally inverse points.

A circle concentric with and interior to the circle of inversion is transformed into a circle exterior to the latter, and conversely. The smaller the radius of such an interior circle, the larger will be the radius of its image. An inversion casts all the points interior to the circle of inversion out into the exterior region of the plane, and gathers all of the exterior region into the interior. At the same time the circle of inversion itself remains invariant.

3. A circle which intersects orthogonally the circle of inversion is transformed into itself.

Let a circle intersect orthogonally the circle of inversion at points P and Q. The radii OP and OQ are tangent to the orthogonal circle at points P and Q (fig. 202).

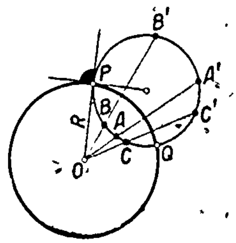


Fig. 202.

We draw rays OAA', OBB', OCC', ..., where A and A', B and B', C and C', ... are the points of intersection of these rays with the orthogonal circle.

By the theorem on the product of the segments of secants we have:

$$O'A \cdot OA = OB' \cdot OB = OC' \cdot OC = \ldots = OP^2 = R^2.$$

These equalities prove that an orthogonal circle is transformed under the inversion transformed into itself, that is, it is an invariant circle, with the exterior arc PA'Q transformed into the interior arc PAQ, and vice versa.

Here also we observe the analogy with a symmetry.

The converse proposition is also easily proved:

4. A circle, distinct from the circle of inversion, which under the inversion is transformed into itself, intersects the circle of inversion orthogonally.

5. A circle passing through the pole of inversion is transformed into a line.

Let L be a circle passing through the pole O of inversion (fig. 203). We draw the diameter OA of circle L and construct the point A' inverse to endpoint A of the diameter.

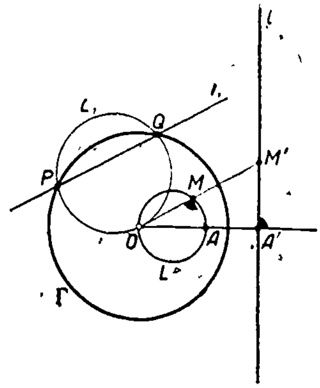


Fig. 203.

The line ℓ passing through A' and perpendicular to the ray OA will be the image of circle L. Proof: let M and M' be points respectively of circle L and line ℓ, lying on a ray OMM'. The right triangles OMA and OA'M' with a common angle at O are mirror-similar and consequently the points A, A' and M, M' are reciprocally inverse pairs of points under some inversion with center at O. Let us find the radius of this inversion. From the similarity of the triangles just mentioned we have:

$$\frac{OA}{OM} = \frac{OM'}{OA'}, \text{ or } OM' \cdot OM = OA' \cdot OA = R^2.$$

Thus the inversion transforming circle L into line ℓ is identical with the given inversion, q.e.d.

From the property of reciprocity of an inversion we have:

6. <u>A line ℓ not passing through the pole is transformed by the inversion into a circle L passing through the pole O.</u>

It is particularly simple to construct the image of a line, or of a circle passing through the pole, when this line or circle intersects the circle of inversion (fig. 203), since we then know two double points P and Q of the given figure.

From lemmas α and β there follows the property:

7. A circle passing through two reciprocally inverse points A and A' is transformed into itself, and consequently intersects orthogonally the circle of inversion Γ (fig. 204).

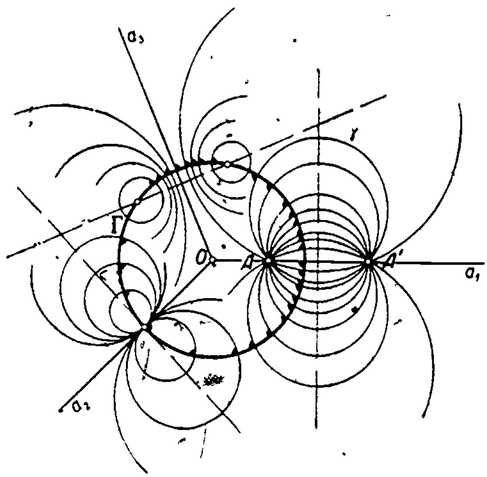


Fig. 204.

Line OAA' is the radical axis of all circles passing through the inverse points A and A' (fig. 204). The set of circles having a common radical axis is called a pencil of circles.

Every circle of the pencil in question is transformed into itself under the inversion and intersects orthogonally the circle of inversion Γ. For all the circles of the pencil we have

$$OA' \cdot OA = R^2 ;$$

but this means that the lengths of the tangents from the pole of inversion O to all the circles of the pencil are equal and the locus of the points of tangency is the circle of inversion with the exclusion of its point of intersection with line AA'.

Recalling the definition of a radical axis we find:

8. Given a pencil of circles and let P be any point on the radical axis (exterior to the circles of the pencil). Then there exists an inversion whose pole is at P and such that every circle of the pencil is orthogonal to the circle of inversion and, consequently, invariant under the inversion.

The pencil of circles with radical axis a_1 (fig. 204) is said to be elliptical; the pencil with radical axis a_2, parabolic; and the pencil with radical axis a_3, hyperbolic.

The circles of the elliptical pencil have two common points; those of the parabolic pencil, one common point; and those of the hyperbolic pencil have no points in common.

The set of circles orthogonally intersecting a given circle Γ is called a sheaf of circles. The inversion with respect to Γ transforms every circle of such a sheaf into itself.

The best method of studying the geometry of circles is by the independent solution of a series of appropriately ordered problems. (1) The solution of each problem does not offer any great difficulty, and the aggregate of such problems makes the student well acquainted with an important part of elementary geometry having broad

(1) A set of well chosen problems are found in the book by B. Delone and O. Zhitomirski: "Exercises in Geometry" (Zadachnick po geometrii) Gostechizdat 1949 [20].

application in the theory of electromagnetic fields, in aerodynamics, in heat theory and in many other disciplines. Of no less importance is the problem-solving skill gained from such exercises.

9. A circle not passing through the pole of an inversion is transformed by the inversion into a circle.

The case where the given circle contains one of its image point, is already covered in proposition 7 above. We may, therefore, confine our proof to the case where the given circle and its image do not intersect.

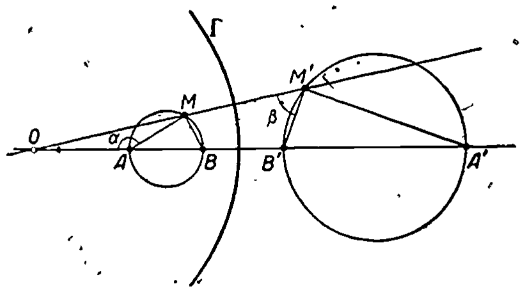


Fig. 205.

Let AB be the diameter of the given circle which passes through the pole O, and A' and B' the images of points A and B (fig. 205). If M' is the image of an arbitrary point M of the given circle, the triangles OAM and OM'A' as well as the triangles OBM and OM'B' are mirror-similar, and consequently

$$\angle OAM = \angle OM'A' = \alpha \text{, and } \angle OBM = \angle OM'B' = \beta .$$

But $$\angle OAM - \angle OB'M' = \alpha - \beta = \angle AMB = \frac{\pi}{2} ,$$

whence $$\angle A'M'B' = \angle OM'A' - OM'B' = \alpha - \beta = \frac{\pi}{2} .$$

Point M' possesses this property, that segment A'B' subtends at it a right angle. Thus, the locus of the points M' inverse to the points of the given circle is a circle, q.e.d.

It must be emphasized that the image of the center of the given circle is not the center of the transformed circle.

The propositions proved enable us to say that the set consisting of all straight lines and circles is transformed by an inversion into itself.

10. (Invariance of angles under inversion.) The magnitude of the angle between two intersecting curves is not changed by an inversion.

Let curves γ and δ, intersecting in point A, be transformed by an inversion into curves γ' and δ' (fig. 206). The point of intersection A' of curves γ' and δ' is the image of point A.

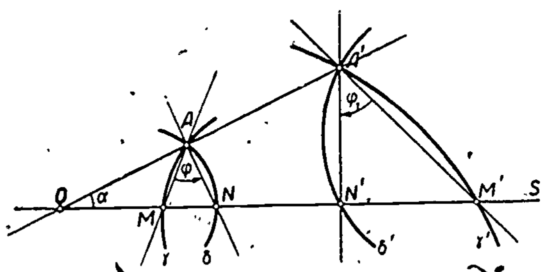


Fig. 206.

We draw through the pole O a line OS and mark the points M and N, and their images M' and N', at which the line OS intersects the four curves.(1)

(1) If the line OS intersects one of the curves, say γ, in more than one point, then we take the point of intersection M closest to A (i.e. there is no other point on γ between A and M which also lies on OS.) Since γ' is the image of γ, M' the image of M will be the point of intersection of OS and γ' closest to A'. --Translators.

The triangles OMA and OA'M' are mirror-similar; consequently

$$\angle OMA = \angle OA'M'.$$

It likewise follows from the mirror-similarity of triangles ONA and OA'N' that

$$\angle ONA = \angle OA'N'.$$

From these equations we obtain:

$$\angle MAN = \angle OMA - \angle ONA = \angle OA'M' - \angle OA'N' = \angle M'A'N'.$$

This signifies that the angle ϕ between the secants AM and AN of the curves γ and δ is equal to the angle ϕ_1 between the secants A'M' and A'N' of the curves γ' and δ'.

Passing to the limit as $\alpha \to 0$, we note that points M and N will tend toward point A, while points M' and N' will tend toward point A'; the secants AM, AN, A'M' and A'N' will tend to take the position of tangents to their respective curves. Thus we have

$$\lim_{\alpha \to 0} \phi = \lim_{\alpha \to 0} \phi_1.$$

What this means is that the angle of intersection of curves γ and δ is equal to the angle of intersection of curves γ' and δ', since the angle between curves is defined as the angle between their tangents at the point of intersection; q.e.d.

Inversion is thus a conformal transformation.

Note that an inversion, while preserving the magnitude of angles, changes their orientation (in complete analogy with a symmetry).

From the property of invariance of angles it follows that if two curves are tangent at point A, their images under an inversion are also tangent at point A', the image of point A.

Let us note two important propositions which will be very useful to us later.

A. If we take as the circle of inversion any circle γ which is orthogonal to a given circle Γ (fig. 204), then this inversion maps the closed disk Γ onto itself.

In this mapping the part of disk Γ external to circle γ is transformed into the part lying inside circle γ, and vice versa.

Disk Γ is as it were turned over onto its other side about the arc of circle γ lying within disk Γ, which is completely analogous with a symmetry, the latter being indeed the limiting case of an inversion with respect to an orthogonal "arc". If the radius of circle γ increases without limit while the orthogonality of this circle to circle Γ is preserved, the arc tends toward a diameter of circle Γ, and the inversion tends to become an ordinary symmetry.

The second proposition consists in the following. Evidently, two arbitrary points have no invariant under an inversion. Likewise, three points do not possess any invariant.

However, four points have an invariant under inversion. In order to describe this invariant of the four points A, B, C, D we introduce the following definition:

By the cross ratio of four points A, B, C, D of a plane, taken in that order, we mean the ratio of two ratios

$$\frac{AC}{AD} : \frac{BC}{BD} .$$

The cross ratio of four points is denoted by the symbol (ABCD), so that $(ABCD) = \frac{AC}{AD} : \frac{BC}{BD}$.

B. <u>An invariant of four points ABCD under an inversion is the cross ratio (ABCD).</u>

In proof, we note that from the mirror-similarity of triangles OAC and OC'A' it follows that

$$\frac{A'C'}{AC} = \frac{OC'}{OA} = \frac{OC'\cdot OC}{OA\cdot OC} = \frac{R^2}{OA\cdot OC}. \quad (1)$$

Here A' and C' are the points inverse to A and C, R is the radius and O the pole of the inversion. In the same way we have:

$$\frac{B'C'}{BC} = \frac{R^2}{OB\cdot OC}; \quad (2)$$

$$\frac{A'D'}{AD} = \frac{R^2}{OA\cdot OD}; \quad (3)$$

and

$$\frac{B'D'}{BD} = \frac{R^2}{OB\cdot OD}. \quad (4)$$

From these four equations we obtain:

$$\frac{A'C'}{A'D'} : \frac{B'C'}{B'D'} = \frac{AC}{AD} : \frac{BC}{BD},$$

or $(A'B'C'D') = (ABCD)$.

This means that the cross ratio of four points is not changed by an inversion, q.e.d.

<u>Remarks</u>: 1. The four points whose cross ratio we have defined need not be collinear; they can be any four points of the plane, excluding the pole of the inversion.

2. An inversion does not transform straight-line segments into straight-line segements; the straight-line segments referred to in the cross ratio (A'B'C'D') are not the images of the segments referred to in the cross ratio (ABCD).

50. INVERSIONS IN SPACE

Two points M and M' in space are said to be inverse to each other under an inversion with pole O and radius of inversion R if:

1) points M and M' lie on one ray issuing from the pole O;

2) the equation

$$OM' \cdot OM = R^2$$

holds good.

A mapping of space (with the pole O removed) onto itself under which to every point M there is assigned its inverse point M' is called an <u>inversion</u>.

All points of the sphere with center at the pole O and radius R are transformed into themselves, i.e. are <u>double points</u> of the inversion. This sphere consisting of double points is called <u>the sphere of the inversion</u>.

From this definition it follows that every plane passing through the pole of inversion and perforated at this point O is mapped onto itself by the inversion of space. Moreover, this mapping of the plane onto itself is an inversion in the plane with pole O, and radius R.

From this remark and the properties of an inversion in the plane it follows that:

1) lines passing through the pole of inversion O are (with the exception of the point O) transformed into themselves;

2) a sphere with center at the pole of inversion O is transformed into a concentric sphere;

3) a sphere orthogonally intersecting the sphere of inversion is transformed into itself;

3*) a circle orthogonally intersecting the sphere of inversion is transformed into itself;

4) a sphere other than the sphere of inversion which is transformed into itself under the inversion intersects orthogonally the sphere of inversion;

4*) a circle not belonging to the sphere of inversion which is transformed into itself under the inversion intersects orthogonally the sphere of inversion;

5) a sphere passing through the pole of inversion is transformed into a plane.

In particular, if such a sphere is also tangent to the sphere of inversion at a point P, then the plane π into which the sphere is transformed is tangent to the sphere of inversion at the same point P.

Stereographic projection is seen to be an inversion transformation of the sphere into the plane π (figs. 117, 118, 122, 124). The point S serves as the pole of the inversion, and the radius of the inversion is equal to SP.

As an exercise the student should derive the properties of stereographic projection considered in 17. from the properties of inversions.

5*) a circle passing through the pole of inversion is transformed into a line;

6) a plane not passing through the pole of inversion is transformed into a sphere which passes through the pole of inversion;

6*) a line not passing through the pole of an inversion is transformed into a circle which passes through the pole of inversion;

7) a sphere passing through two reciprocally inverse points A and A' is transformed into itself and consequently it intersects orthogonally the sphere of inversion;

8) an inversion of space preserves the angles between surfaces, between curves, and between curves and surfaces.

In concluding this section we shall note an important proposition which will be of use to us later.

If as the sphere of inversion we take any sphere γ orthogonal to a given sphere Γ, then this inversion maps the closed solid sphere Γ into itself.

The detailed carrying out of simple proofs of the enumerated properties of inversions of space is left as an exercise.

51. THE APPLICATION OF INVERSIONS TO GEOMETRICAL CONSTRUCTIONS

Inversion transformations can be applied very effectively to the solution of construction problems and in proofs.

We shall study the application of the method of inversion transformations by means of examples. The choice of examples will be dictated in part by their usefulness later on.

<u>Problem.</u> <u>Through two given points A and B lying within a given circle Γ, to draw a circle orthogonal to the given circle.</u>

The required circle is invariant under inversion with respect to circle Γ. The construction follows from this. Taking Γ as the circle of inversion we construct point A' inverse to point A. The circle circumscribed around the triangle ABA' will be the required circle (fig. 207).

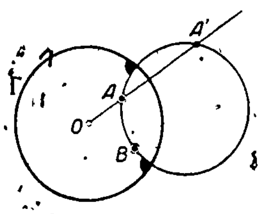


Fig. 207.

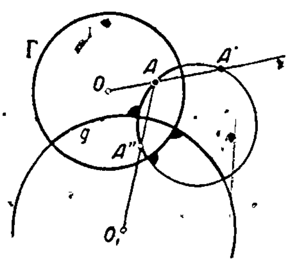


Fig. 208.

If the given points A and B are collinear with the center O of circle Γ, the required circle degenerates into a line. The problem has a unique solution.

<u>Problem.</u> <u>Through a point A to draw a circle orthogonal to two given circles which are orthogonal to each other. Point A lies inside one of the given circles.</u>

Let Γ and g of fig. 208 be the given orthogonal circles.

The required circle will be invariant under each of two inversions having Γ and g as their respective circles of inversion. This determines the construction. Taking Γ as a circle of inversion we construct point A' inverse to A. Next, taking g as a circle of inversion we construct point A" inverse to A.

The circle circumscribed about triangle AA'A" will be the required circle.

It may happen that the three points A, A' and A" will be collinear. In such a case the line of centers of the given circles will take the place of the required circle.

The problem has a unique solution.

Remark. The indicated construction does not require that the given circles Γ and g be orthogonal. We chose such an arrangement of these circles with later needs in mind.

Problem. Through a given point A to draw a circle tangent to one given circle g and orthogonally intersecting a second given circle Γ, the circles g and Γ being orthogonal to each other (fig. 209).

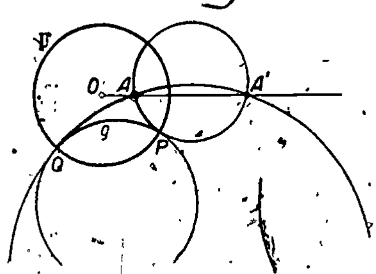


Fig. 209.

The required circle, being orthogonal to circle Γ, must be transformed into itself under inversion with respect to circle Γ. Circle q is invariant under this same inversion.

Since the point of tangency P of the required circle with the given circle q is transformed into the point of tangency of these same circles, point P is a double point under the inversion. This means that P lies on the circle of inversion Γ.

From this follows the construction of the required circle. We mark a point of intersection P of the circles Γ and q and after finding point A' inverse to the given point A with respect to Γ we pass a circle through the three points A, P and A'. The circle thus constructed is a solution of the problem. Since the given circles Γ and q intersect in two points P and Q, the problem has two solutions.

If the points O, A and P are collinear, the required circle degenerates into a line passing through the center O of circle Γ.

What form does the construction take if point A lies on or outside of circle Γ?

<u>Problem.</u> <u>Given circle Γ and two circles q_1 and q_2 orthogonal to Γ and passing through point A (fig. 210), to construct a circle orthogonal to circles Γ and q_1 and tangent to circle q_2.</u>

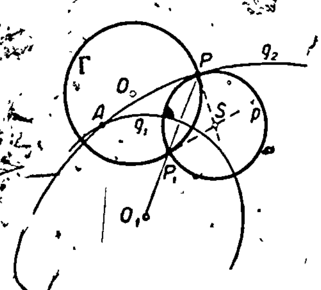


Fig. 210.

As in the preceding problem it is evident that the point of tangency of the required circle with circle g_2 is a double point under inversion with respect to circle Γ and consequently must be the point of intersection of circles Γ and g_2. Thus, the problem is reduced to that of drawing through a given point P a circle orthogonal to two given circles Γ and g_1. The center S of the required circle p lies at the intersection of the tangents PS and P_1S to circle Γ at the points P and P_1 (fig. 210). The problem has not more than four solutions. Its complete investigation is recommended as an exercise.

<u>Problem.</u> <u>Given a circle Γ and the two circles g_1 and g_2 having no points in common and both orthogonal to circle Γ (fig. 211), to construct a circle orthogonal to all three of the given circles.</u>

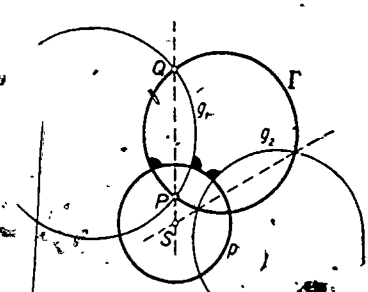


Fig. 211.

We have already (in 7.) solved this problem in a more general form with the aid of the concepts of the radical axis and radical center.

We construct the radical center S and from S as center draw a circle with radius equal to the common length of the tangents from S to the three given circles. The problem has a unique solution.

We shall describe a solution of this problem, more complicated in construction but very instructive, which is based on the use of an inversion transformation.

We transform the given figure, that is, the three given circles Γ, g_1 and g_2, by means of an inversion having its pole at Q, the intersection of circles Γ and g_1, and an arbitrary radius of inversion R. We shall describe this inversion without actually carrying it out.

The circles Γ and g_1, since they pass through the pole Q, are transformed into the lines Γ' and g_1', with the preservation of the angle between them at their point of intersection P', which is the image of P (fig. 212). Point Q has no image.

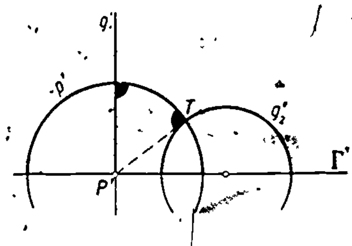


Fig. 212.

Circle g_2, not passing through the pole Q, is transformed into circle g_2', orthogonal to line Γ', that is, with its center on line Γ'. (It must be remembered that the center of the image is not the image of the original center.)

With the aid of the inversion the problem has been reduced to the following:

Given two perpendicular lines Γ' and g_1' and the circle g_2' with its center on line Γ' and not intersecting line g_1', to construct a circle p' orthogonal to the given lines Γ' and g_1' and to the given circle g_2'.

The solution is extremely simple. Since the required circle p' is orthogonal to lines Γ' and g_1', its center must be at P', the intersection of these lines, and its radius must be $P'T$ equal to the length of the tangent from P' to the circle g_2' (fig. 212). This problem is solved.

The same inversion with pole at Q transforms, in the other direction,

$$\Gamma' \rightarrow \Gamma, \; g_1' \rightarrow g_1, \; g_2' \rightarrow g_2, \; P' \rightarrow P$$

and the circle p' into the circle p required in the original problem.

Since the images of points, lines and circles under an inversion can be constructed easily (even with a single pair of compasses), we may find it helpful to solve a problem by solving the "problem transformed by inversion" instead of the original one. The choice of pole and circle of the inversion is at our disposal and often can be a means of simplifying the problem.

As an exercise the student should rework all the problems solved in this section, inverting the circle Γ into a line Γ' and formulating and solving "the inverted problem".

Problem. Given only the pairs of points A_1, B_1 and A_2B_2, to construct the point of intersection of the lines A_1B_1 and A_2B_2, using compasses only.

As a preliminary we shall learn how to construct with compasses only a circle inverse to a line which is given by means of two points A and B. The pole O and circle Γ of the inversion are assumed given (fig. 213).

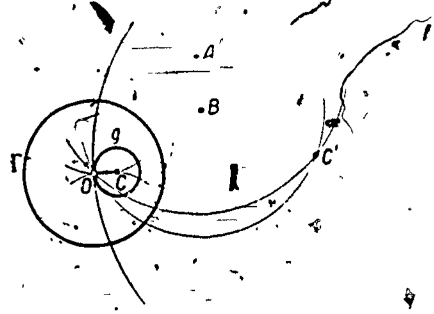


Fig. 213.

The method of construction of the image of the line consists in the following.

We construct point C', symmetrical to the pole O about the given line AB, which is easily done by drawing circles with centers A and B and radii AO and BO respectively.

Point C,(1) inverse to point C', is the center of the required circle g; knowing this point we construct this circle with center C and radius CO.

Proof: Let P and P', inverse to each other, be the points of intersection of line OC with line AB and with the required circle g. (These points are not shown in the diagram.)

In virtue of the construction we have:

$$OC' = 2\cdot OP, \quad OP'\cdot OP = R^2, \quad OC'\cdot OC = R^2,$$

and consequently

$$OP'\cdot OP = OC'\cdot OC = 2\cdot OP\cdot OC,$$

whence $OP' = 2\cdot OC$, q.e.d.(2)

(1) The construction, with compasses only, of a point inverse to a given point, has already been studied in 53.
(2) By referring to fig. 203, the student will see that OP' is the diameter of the required circle g. It follows then from $OP' = 2\cdot OC$ that C is the center. --Translators.

We have now the means of solving easily the original problem. We invert the lines determined by the given points A_1B_1 and A_2B_2 with respect to an arbitrary pole and radius of inversion. We find the point of intersection S of the circles inverse to these lines and construct point S' inverse to S under the same inversion. S' is the required point of intersection of lines A_1B_1 and A_2B_2. The construction has been accomplished using only a pair of compasses.

By an analogous construction we can find the points of intersection of a circle with a line given only by two of its points. It is precisely this method by which it is most simply proved that all problems solvable with compasses and straight-edge can also be solved using compasses only.

52. THE PROBLEM OF APOLLONIUS

The problem of Apollonius consists in constructing with compasses and straight-edge a circle tangent to three given circles.

This problem, of great importance in the elementary development of the theory of conic sections, is interesting with regard both to methods of solution and to the large number of limiting cases dealt with in the secondary school course.

The limiting cases of this problem arise when any of the given circles become points or lines.

Let us examine some limiting cases.

1. If the three circles degenerate into points we have the familiar problem: to circumscribe a circle about a triangle.

In this formulation we exclude the case in which the three points are collinear. If one given point is left undetermined, then by letting this third point vary we obtain the locus of centers of circles passing through the other two points. This will be the perpendicular bisector of the segment joining the two points.

2. If the three circles degenerate into lines, we again have a familiar problem: to construct a circle tangent to three given lines.

Let two of the three given lines intersect, and the third line be undetermined; then letting the third line vary, we obtain the locus of centers of the circles tangent to the two fixed lines. This locus is the pair of bisectors of the angles between these two given lines. If two of the given lines are parallel, and the third line, assumed to intersect them, is undetermined, we obtain the locus of the centers of the circles tangent to the two parallels which is the parallel lying midway between them.

3. If two of the given circles degenerate into points and the third into a line, we have the problem: to construct a circle passing through two given points and tangent to a given line.

For this construction to be possible the given points A and B must not lie on opposite sides of the given line a (fig. 214).

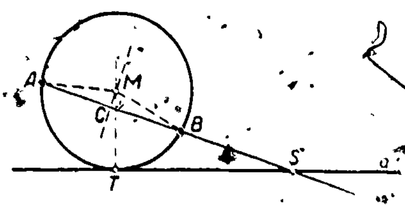


Fig. 214.

The solution presented in secondary school geometry courses consists in finding the point of tangency T of the required circle with the given line a, and is based upon the equation $ST^2 = SA \cdot SB$, where S is the point of intersection of the lines AB and a.

Having found the segment ST we lay it off on line a in both directions from [illegible] and obtain the two solutions of the problem. Only one of the circles has been constructed in fig. 214.

If lines AB and a are parallel, the point T of the required circle lies on the line perpendicular to both and bisecting segment AB. The student should carry out the complete investigation of this problem as an exercise.

With the aid of inversion the problem can be solved as follows: making one of the given points, A, the pole of an inversion, we invert line a into circle a'; the required circle will now be a line. The problem is reduced to that of drawing, from point B', the image of point B, a tangent to circle a'.

Repeating the same inversion we obtain the solution of the original problem.

As an exercise the student should actually carry out this construction.

If one of the given points, let us say point B, is undetermined, the locus of the centers M of the circles passing through the given point A and tangent to line b is a parabola.

Proof: Since MA = MT, the point M is equidistant from the given point A and the given line a, point A is the focus of the parabola and line a is its directrix.

4. If two of the given circles degenerate into lines and the third into a point, we have the problem: to construct a circle tangent to two given lines and passing through a given point.

With the aid of an inversion having the given point as its pole the problem is reduced to that of drawing a line tangent to two circles. The complete investigation of the problem and the construction are suggested as an exercise.

5. If two of the given circles degenerate into points we have the problem: to construct a circle passing through two given points and tangent to a given circle.

By an inversion having its pole at one of the given points we reach the "inverted problem": through a given point to draw a tangent to a given circle.

The construction and investigation of this problem is suggested as an exercise.

If one of the given points, let us say B, is undetermined, then by letting B vary, we obtain the locus of the centers of the circles passing through point A and tangent to the given circle Γ. If point A lies outside circle Γ, the locus in question is a branch of an hyperbola; if point A lies inside circle Γ it is an ellipse.

The correctness of these assertions follows from the equations

$$OM - MA = OP = \text{const. (fig. 215a)}$$

$$OM + MA = OP = \text{const, (fig. 215b).}$$

6: If one of the given circles degenerates into a point, we have the problem: to construct a circle passing through a given point and tangent to two given circles Γ_1 and Γ_2.

Performing an inversion, of any radius, with point A as its pole, we obtain the familiar problem: to construct a tangent to the two circles Γ_1' and Γ_2'. Here Γ_1' and Γ_2' are the images of circles Γ_1 and Γ_2 under the inversion performed.

Constructing the common tangents to circles Γ_1' and Γ_2' and thereafter repeating the same inversion, we obtain the solution of the original problem, which thus has not more than four solutions.

The student should investigate this problem and perform the construction as an exercise. He should also give the constructions and carry out the investigations for the remaining limiting cases of the problem of Apollonius.

Returning to the general problem of Apollonius, we note that it can easily be reduced to the foregoing limiting case number 6. To do this we can, without altering the location of the centers of the three given circles or of the center of the required circle, shrink one of the given circles to a point, reducing its radius to zero and appropriately altering the radii of the remaining circles, without destroying the tangencies.

If two circles are tangent to each other internally, and we change the radius of one of the circles while maintaining the condition of tangency and the position of the centers, the radius of the other circle will change in the same manner, either both increasing or both decreasing. When the tangency, however is external, then as one of the radii increases the other will decrease, and vice versa.

Suppose now that the radius of one of the given circles, say Γ_1, decreases. Then the radius of the required circle will increase or decrease according as its tangency to Γ_1 is external or internal. (In case of internal tangency, the required circle will, of course, contain Γ_1 in its interior so that the tangency continues to be internal until Γ_1 shrinks to a point.)

As the radius of the required circle changes, the radii of the other two given circles Γ_2 and Γ_3 will also change. The change will be of the same kind -- both radii increasing or both decreasing -- when the tangencies of the two circles to the required circle are of the same kind (both internal or both external). Otherwise, one of the radii will increase and the other will decrease.

After the indicated transformation of the given circles -- Γ_1 into point A, Γ_2 and Γ_3 into circles Γ_2' and Γ_3' -- the general problem of Apollonius turns into the limiting case: <u>to draw a circle passing through a given point A and tangent to two</u>

given circles Γ_2 and Γ_3. We have already dealt with this problem. The plan for the solution of the problem of Apollonius is accordingly as follows.

Let the centers and radii of the given circles Γ_1, Γ_2, Γ_3 be respectively O_1, O_2, O_3 and r_1, r_2, r_3. Further, let us assume that $r_1 < r_2$ and $r_1 < r_3$. About point O_2 we describe circles Γ'_2 and Γ''_2 with radii $r_2 + r_1$ and $r_2 - r_1$; about point O_3 in like manner we describe circles Γ'_3 and Γ''_3 with radii $r_3 + r_1$ and $r_3 - r_1$. We then construct circles passing through point O_1 and tangent in the same manner to circles Γ'_2 and Γ'_3, as well as the circles passing through O_1 and tangent in the same manner to circles Γ''_2 and Γ''_3. We obtain four solutions.

We next construct circles passing through O_1 which are tangent to circles Γ'_2 and Γ''_3 and to circles Γ''_2 and Γ'_3, in opposite manners within each pair. We obtain four more solutions.

The eight circles which (in the general case) we have found will be concentric with the required circles and their radii may be easily determined. The problem has not more than eight solutions.

It is highly desirable that the student should familiarize himself with the various methods of solution of the problem of Apollonius given in the existing literature [3], [2], [1].

THE MEASUREMENT OF LENGTHS, AREAS AND VOLUMES.

Chapter IX

THE GENERAL PROBLEM OF MEASURING LENGTHS, AREAS AND VOLUMES

In Chapter IX we examine the comparison of magnitudes. The solution of the general problem of the measurement of the lengths of segments of straight lines and the areas of simple polygons is presented. Some idea is given of the system of measuring the areas of plane figures and the areas of curved surfaces. A brief discussion is given of the problem of measuring volumes.

53. COMPARISON OF STRAIGHT-LINE SEGMENTS

Even before introducing the concept of length, we may compare straight-line segments one with another, i.e. decide whether they are equal, which of them is larger, which is smaller, divide a segment into two equal parts and so on.

The notion of equality of segments is accepted as a primary notion not directly defined. The following axioms[1] of congruence of segments serve as an indirect definition of the concept of equality, or congruence, of segments.

III_1. Suppose A and B are two points on a line a and A' is a point on the line a' (which may or may not coincide with line a). Then it is always possible to find a point B', lying on the line a' on a given side of the point A', such that the segment AB is congruent, i.e. equal, to the segment A'B'.

(1) Footnote: In Part IV of this book we shall give a full list of the axioms of geometry. The numbers used here correspond to that list. --Translators.

The congruence of the segments AB and A'B' is denoted by

$$AB \equiv A'B'.$$

This axiom merely states that it is possible to lay off a segment satisfying the given conditions. That only one such segment may be laid off may be demonstrated as a theorem [15].

III_2. If the segments A'B' and A"B" are congruent to the same segment AB, then the segment A'B' is also congruent to the segment A"B", briefly, if two segments are congruent to a third, they are congruent to each other.

Using axioms III_1 and III_2, it is possible to prove as a theorem that every segment is equal to itself.

Proof. Suppose we are given a segment AB. We lay off on any ray a segment A'B' equal to the segment AB. This can be done by axiom III_1. We then have the congruence

$$AB \equiv A'B'. \qquad (\alpha)$$

This congruence may be written again,

$$AB \equiv A'B' \qquad (\beta)$$

Congruences (α) and (β) may be read as follows: Two segments AB and AB (in this case the two segments coincide) are congruent to a third segment A'B'. Consequently by axiom III_2 these segments are equal to each other: $AB \equiv AB$, q.e.d. Here we have an instance of the proof of a theorem whose "self-evidence" is not less than that of the simplest axiom.

The property of a segment just demonstrated, that of being equal to itself, is called reflexiveness. It is also easy to show that equality of segments has the properties of symmetry and transitivity.

By symmetry we mean that from the equality

$$AB \equiv A'B'$$

there follows the equality (or congruence)

$$A'B' \equiv AB.$$

Let us prove this: From the property of reflexiveness we get $A'B' \equiv A'B'$. In addition, we are given that $AB \equiv A'B'$. This means that the two segments $A'B'$ and AB are congruent to a third segment $A'B'$. By axiom III_2 we conclude that

$$A'B' \equiv AB,$$

q.e.d.

By transitivity we mean that from the equalities

$$AB \equiv A'B', \quad A'B' \equiv A''B''$$

there follows the equality

$$AB \equiv A''B''.$$

The proof is obtained by noting that from the second equality, $A'B' \equiv A''B''$, we have by symmetry $A''B'' \equiv A'B'$. We then apply axiom III_2.

The third axiom pertaining to the properties of equality of segments states:

III_3. Suppose that AB and BC are two segments of a line a having no interior points in common, and suppose further that $A'B'$ and $B'C'$ are two segments on the same line or another line a', also having no interior points in common (fig. 216); if we have

$$AB \equiv A'B' \text{ and } BC \equiv B'C'$$

then

$$AC \equiv A'C'$$

is also true.

This axiom implies the possibility of <u>adding</u> segments.

Let us now define the concept of <u>greater</u> and <u>smaller</u> as applied to segments.

We say that the segment CD is greater than the segment AB, and write

$$CD > AB,$$

if upon laying off (in accordance with axiom III_1) on the segment CD a segment A'B' equal to AB we obtain a point B' lying <u>between</u> points C and D (fig. 217).

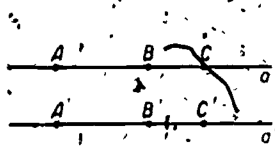


Fig. 216.

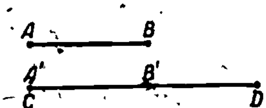


Fig. 217.

If, as the segment AB is laid off on the line CD, the point B' falls <u>outside</u> the segment CD (in which case the point D will lie <u>between</u> the points C and B') we say that the segment CD is <u>less than</u> segment AB.

For two given segments *a* and *b* one and only one of the following three statements is true:

1. $a < b$ and $b > a$;
2. $a = b$;
3. $a > b$ and $b < a$.[1]

The comparison of segments with respect to magnitude is <u>transitive</u>; that is, from each of the three propositions:

(1) Footnote: It must be emphasized that we are here simply denoting segments by single letters; it by no means follows that these letters denote numbers. --Translators.

1. $a > b$, $b > c$;

2. $a > b$, $b \equiv c$;

3. $a \equiv b$, $b > c$

it follows that

$$a > c.$$

We omit the proof of these propositions.

In analogous fashion axioms are set up which permit the comparison of angles and also the establishment of the congruence of triangles.[2]

Employing the axioms of congruence we shall prove, as an example, the following theorem:

Theorem. Every segment can be divided into two equal parts, and this division is unique.

Proof of existence: Let AB be the segment which is to be divided into two equal parts (fig. 218). We construct the congruent angles MAB and NBA so that their sides AM and BN are located in the same plane and on opposite sides of line AB. The possibility of such a construction is coverd by an axiom concerning angles, analogous to axiom III_1 for segments.

We further select on the rays AM and BN the points C and D such that $AC \equiv BD$ (axiom III_1). The point O of the intersection of lines AB and CD will be the midpoint of segment AB. Proof: From the congruence of triangles ABC and ABD ("by two sides and the included angle") it follows that $CB \equiv AD$.

(2) Footnote: A full list of axioms of congruence is given in Part IV of this book. --Translators.

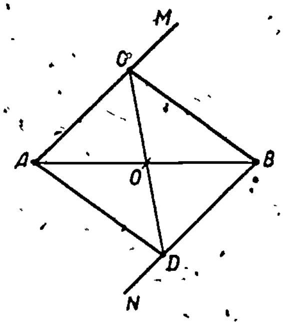


Fig. 218.

From this we derive the congruence of triangles ACD and BDC ("by three sides") and consequently:

$$\angle ACD \equiv \angle BDC.$$

The congruences so far established permit us to conclude that triangles ACO and BDO are congruent ("by one side and the adjacent angles"). Whence it follows that $AO \equiv OB$.

The proof of uniqueness, that is, that the segment has only one mid-point, will be omitted [24]. We shall make use of this theorem in solving the general problem of the measurement of segments.

Having at our command the notion of the equality of segments, we can now define the concept of a motion in the geometric sense.

The reduction of the concept of a motion to the concept of equality of segments was set forth in 29. and 34. above.

But the matter may also be approached in another way. The concept of a motion may be regarded as fundamental and undefined. We may formulate axioms pertaining to the concept of "motion" and subsequently define equality of segments and equality of angles in terms of "motion". For example, with respect to segments we have: two segments are equal, or congruent, if there exists a motion carrying the first segment over into the second. In this development, axioms III_1 and III_2 become theorems.

The axioms of motions are presented in Part IV of this book, where there is also shown the equivalence of the two sets of axioms: the axioms of motions and the axioms of congruence. By the equivalence of two systems of axioms we mean that the totality of the facts of geometry remains unaltered by the exchange of one system of axioms for the other.

Thus, in our example the first system of axioms consists of all the axioms of geometry -- including the axioms of congruence presented in the first list in Part IV, while the second system differs from the first only in that we adopt the axioms of motion in place of the axioms of congruence.

To prove the equivalence of two systems of axioms it is sufficient to do the following: on the basis of the axioms of the first system, to prove the assertions of those axioms of the second system which are not included in the first, and conversely, to prove the axioms of the first system using the axioms of the second.

54. THE GENERAL PROBLEM OF THE MEASUREMENT OF SEGMENTS

The problem of measuring straight-line segments consists in assigning to every segment AB a unique non-negative number $\rho(AB)$, called the length of segment AB, such that the following conditions are fulfilled:

1. $\rho(AB) = 0$ when and only when points A and B coincide.
2. $\rho(AB) = \rho(BA)$.
3. If $AB \equiv CD$, then $\rho(AB) = \rho(CD)$, that is, equal lengths must be assigned to equal segments.
4. If AB and BC are two segments of a line and have no internal points in common, then $\rho(AB) + \rho(BC) = \rho(AC)$, that is, the sum of two segments must have a length equal to the sum of the lengths of the added segments (the condition of additivity.)

The problem of assigning to every segment a number fulfilling the conditions 1 through 4 is known as the direct problem of the theory of measurement of segments.

From the stated conditions it follows that

1) the length of the sum of a finite number of segments must equal the sum of the lengths of the segments and hence must be not less than the length of any one of the added segments.

2) a larger segment must have a greater length.

It is suggested that the proof of these two propositions be carried out as an exercise.

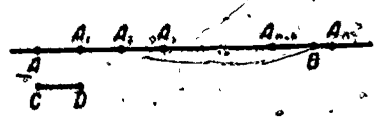


Fig. 219.

If the association with every segment AB of a number $\rho(AB)$ satisfying the stated conditions 1 through 4 has been successfully carried out, we say that a system of measurement of segments -- or a system of lengths of segments -- has been established. In this connection it is indispensable that two questions be answered: does even one system of measurement exist, and if so, how can we obtain all the systems of measurement?

The establishment of a system of measurement requires the acceptance of an additional axiom. Customarily we employ for this purpose the axiom of Archimedes.

V_1 (axiom of measurement, or axiom of Archimedes): Let AB and CD be any two segments whatsoever; then on the line AB there exists a finite number of points $A_1, A_2, A_3, \ldots, A_n$ such that the segments $AA_1, A_1A_2, A_2A_3, \ldots, A_{n-1}A_n$ are each congruent to segment CD and point B lies between A and A_n (fig. 219).

The axiom of Archimedes is frequently given the following readily grasped formulation: Any segment CD may be added to itself a sufficient number of times so that the sum exceeds any arbitrary segment AB.

We shall now prove that conditions 1 through 4 uniquely determine the length of every segment if a number equal to 1 is assigned to some given segment PQ with distinct endpoints P and Q.

We call the segment PQ the unit of measurement of length, or the unit segment.

First we shall show that if there exists a system of measurement in which a segment PQ is the unit of measurement, such a system is unique.

Thus we shall start from the assumption that to every non-null segment AB there has been assigned some positive number a, that those numbers satisfy conditions 2 through 4, and that to some given segment PQ there has been assigned the number 1, i.e. $\rho(PQ) = 1$.

Let O be the midpoint of segment PQ (see 53.). Since the lengths $\rho(PO)$ and $\rho(OQ)$ of the congruent segments PO and OQ are equal (condition 3), we have from condition 4:

$$1 = \rho(PQ) = \rho(PO) + \rho(OQ) = 2\,\rho(OQ).$$

Thus, each of the segments PO and OQ has a length equal to $\frac{1}{2}$. Analogously, we conclude that each of the halves PO_1 and O_1O of the segment PO has a length equal to $\frac{1}{2^2}$, and so on.

On the line AB, to which the segment AB belongs, we lay off from point A in the direction of point B the segments AA_1, A_1A_2, A_2A_3, ... congruent to the unit segment PQ (axiom III_1). If one of the points A_n coincides with point B, the length a of segment AB will be equal to:

$$a = \rho(AB) = \rho(AA_1) + \rho(A_1A_2) + \dots + \rho(A_{n-1}A_n) = n,$$

in accord with the first inference from the conditions for a system of measurement of segments.

If on the other hand no one of the points $A_1, A_2, A_3, \ldots$ coincides with point B, there follows from the axiom of Archimedes the existence of two points A_n and A_{n+1} such that B lies between them. Since segment AA_n is less than segment AB and segment AA_{n+1} is greater than segment AB, we have:

$$\rho(AA_n) = n, \quad \rho(AA_{n+1}) = n + 1,$$

$$n < a < n + 1$$

in accord with the second inference from the conditions for a system of measurement of segments.

Thus the length a of segment AB is determined to within one unit.

We now divide the segment A_nA_{n+1} into two equal parts by means of the point P_1. One of three possibilities will take place:

1. Point B coincides with point P_1. In this case

$$a = \rho(AB) = \rho(AP_1) = \rho(AA_n) + \rho(A_nP_1) = n + \frac{1}{2}$$

and the process of measuring segment AB is finished.

2. Point B lies between A_n and P_1. In this case segment AB is greater than segment AA_n and less than segment AP_1. Consequently:

$$n < a < n + \frac{1}{2}.$$

3. Point B lies between P_1 and A_{n+1}. In this case segment AB is greater than segment AP_1 and less than segment AA_{n+1}. Consequently:

$$n + \frac{1}{2} < a < n + 1.$$

In the last two cases the process of measuring segment AB is not finished; we say that length a of segment AB has been determined to within $\frac{1}{2}$ unit. Continuing the process of measurement, that one of the segments A_nP_1 and P_1A_{n+1} in which point B is found is divided into two equal parts by a point P_2. Here again one of three possibilities will hold, and so on.

The result of the measurement can be written in the form

$$a = n + \frac{n_1}{2} + \frac{n_2}{2^2} + \frac{n_3}{2^3} + \ldots + \frac{n_k}{2^k} + \ldots$$

Here n -- the integral part -- indicates how many units of measurement are contained in segment AB. The numbers n_1, n_2, n_3, ..., n_k, ... are equal to 0 or to 1 depending on which of the successive halves contains point B. For example, $n_1 = 0$ if point B lies in segment A_nP_1 with the exclusion of its endpoint P_1; $n_1 = 1$ if point B is a point of segment P_1A_{n+1} with the exclusion of its endpoint A_{n+1}. We determine analogously the value of any n_k.

The above expression for a will have a finite number of terms if point B coincides with one of the "midpoints P_s" which we construct in the process of measuring segment AB; or an infinite number of terms if point B does not coincide with any one of these points of division.

Remark: In practice the unit segment PQ is divided not into halves but into ten equal parts by points of division P_1, P_2, ..., P_8, P_9 and the result of the measurement is written out in the form

$$a = n + \frac{n_1}{10} + \frac{n_2}{10^2} + \frac{n_3}{10^3} + \ldots + \frac{n_k}{10^k} + \ldots$$

Here n is the integral part and numbers $n_1, n_2, n_3, \ldots, n_k, \ldots$ are each equal to one of the numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, depending on which of the ten segments in each successive subdivision into ten parts contains point B. For example, if point B is found between points A_n and A_{n+1} and after the division of segment A_nA_{n+1} into ten equal parts point B lies in the first segment A_nP_1 then $n_1 = 0$, but if it lies in segment P_7P_8 then $n_1 = 7$, and so on.

The foregoing expression can be written, of course, in the form of a decimal fraction:

$$a = n.n_1n_2n_3\ldots n_k\ldots \quad .$$

In the same way the expression

$$a = n + \frac{n_1}{2} + \frac{n_2}{2^2} + \frac{n_3}{2^3} + \ldots + \frac{n_k}{2^k} + \ldots,$$

where $n_1, n_2, \ldots, n_k, \ldots$ are equal to 0 or 1 and n is a non-negative integer, appears as the binary fraction

$$a = n.n_1n_2n_3\ldots n_k\ldots, \quad .$$

From the above process of measurement of segment AB it follows that it is possible to find any desired digit in the binary (or decimal) representation of its length a.

From all this it follows that the length of segments is uniquely determined by conditions 1 through 4; more precisely, if there exists a system of measurement of segments in which a given non-null segment PQ has a length equal to 1, then this system of measurement is unique.

It remains to prove the existence of a system of measurement of segments, that is, to show that to every segment there can be assigned a number such that conditions 1 through 4 are fulfilled.

Let us assign the number 1 to an arbitrarily given, non-null segment PQ. The process of measurement, applied to the segment PQ, will give the number 1. In general, let us assign to every segment AB the number obtained as the result of the previously described process of measurement.

By this process of measurement a number will indeed be assigned to every segment -- the number zero having been assigned beforehand to null segments.

It is necessary to prove that the numbers thus assigned to segments fulfill the conditions 1 through 4.

The fulfillment of conditions 1 and 2 is self-evident. It is also clear that the process of measurement, applied to equal segments, will yield equal numbers; hence condition 3 is also fulfilled.

Before proving the fulfillment of condition 4, we shall establish the validity of two lemmas [24].

Lemma 1. It is always possible to select a natural number n so large that by dividing the unit of measurement PQ into 2^n equal parts we obtain segments each of which is smaller than a given segment MN.

Proof: If we suppose that for any n the 2^n-th part of the unit segment PQ will be greater than or equal to segment MN, then if we add the segment MN to itself a number of times equal to an arbitrary 2^n we shall not be able to obtain a segment which exceeds the unit segment PQ, thus contradicting the axiom of Archimedes.

Lemma 2. If segment A^*B^* is smaller than segment AB, and the numbers b^* and b respectively are obtained by measurement of these segments, then $b^* < b$.

Proof: Let

$$b^* = n^* + \frac{n_1^*}{2} + \frac{n_2^*}{2^2} + \frac{n_3^*}{2^3} + \ldots + \frac{n_k^*}{2^k} + \ldots$$

and

$$b = n + \frac{n_1}{2} + \frac{n_2}{2^2} + \frac{n_3}{2^3} + \ldots + \frac{n_k}{2^k} + \ldots$$

Since $A^*B^* < AB$, in virtue of axiom III_1 and of the definition of "smaller" for segments there exists on segment AB a point B' such that $AB' = A^*B^*$.

We construct, starting from point A in the direction of point B, the segments AA_1, A_1A_2, A_2A_3, ..., equal to the unit of measurement PQ. If even one point A_i is located between B' and B, the integral part n^* of the number b^* is less than the integral part n of the number b, whence $b^* < b$. If on the other hand both points B' and B lie within any segment A_iA_{i+1} then b^* and b will have identical integral terms $n^* = n$. Dividing segment A_iA_{i+1} in half, we shall find either that points B' and B belong to different halves of this segment and thus $n_1^* < n_1$, which means that $b^* < b$, or that points B' and B belong to the same half of segment A_iA_{i+1}, whereupon we in turn divide this half of segment A_iA_{i+1} into halves, and so on.

In the end we shall establish the inequality $b^* < b$ providing that points B' and B are not found to lie always in the same half after every bisection in the process of measurement. The latter result, however, is impossible, since then the segment $B'B$ would have to be smaller than any 2^n-th part of the unit segment, which contradicts lemma 1.

We return to the completion of the proof of the existence of a system of measurement of segments. We must establish the fulfillment of condition 4. Note that to each segment AB we have assigned the result of the measurement of AB -- the number $\rho(AB)$.

Let AB be any segment, and C a point between A and B. Then

$$AB = AC + CB$$

in accord with the meaning of addition of segments.

We shall show that the numbers $\rho(AB)$, $\rho(AC)$ and $\rho(CB)$ satisfy the condition

$$\rho(AB) = \rho(AC) + \rho(CB).$$

Let n be a fixed natural number and $\frac{PQ}{2^n}$ the 2^n-th part of the unit segment PQ. From point C in the direction toward A we lay off segments CA_1, A_1A_2, A_2A_3, ..., A_sA_{s+1}, ... (figure 220) each equal to the 2^n-th part of the unit segment PQ.

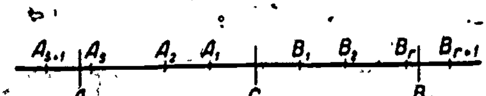


Fig. 220.

It follows from the axiom of Archimedes that there exist two points A_s and A_{s+1} such that A will either coincide with A_s or lie between A_s and A_{s+1}. In either case we have

$$CA_s \leq AC < CA_{s+1}.$$

or, passing to numbers, we have by lemma 2:

$$\frac{s}{2^n} \leq \rho(AC) < \frac{s+1}{2^n} \qquad (*)$$

We now lay off from point C in the direction toward B segments CB_1, B_1B_2, B_2B_3, ..., B_rB_{r+1}, ..., likewise each equal to the 2^n-th part of the unit segment PQ. From the axiom of Archimedes we have:

$$CB_r \leq CB < CB_{r+1},$$

or, in accordance with lemma 2, numerically:

$$\frac{r}{2^n} \leq \rho(CB) < \frac{r+1}{2^n} \qquad (**)$$

We also have

$$A_sB_r \leq AB < A_{s+1}B_{r+1},$$

or, passing to numbers, by lemma 2:

$$\frac{s+r}{2^n} \leq \rho(AB) < \frac{s+r+2}{2^n} \qquad (***)$$

Adding the corresponding sides of the inequalities (*) and (**), we obtain

$$\frac{s+r}{2^n} \leq \rho(AC) + \rho(CB) < \frac{s+r+2}{2^n}$$

From this inequality together with inequality (***) it follows that

$$|\rho(AC) + \rho(CB) - \rho(AB)| \leq \frac{1}{2^{n-1}}$$

But n is any natural number. Consequently,

$$\rho(AB) = \rho(AC) + \rho(CB).$$

The property of additivity is proved. The numbers $\rho(AB)$ assigned to each segment AB will be the lengths of these segments.

The question, whether or not there exists even one system of measurement of segments, has an affirmative answer. More than this, every choice of an arbitrary segment PQ as a unit segment determines a system of measurement of segments. There exists an infinite set of systems of measurement.

We shall now solve the converse problem of the theory of measurement of segments.

It follows from the foregoing that the set of all segments is mapped by any system of measurement into the set of real non-negative numbers: to every segment there corresponds a real number -- the length of the segment. If, under a given system of measurement, some number is the length of a given segment, then all equal segments have as their image that same number. The question remains open, whether or not the lengths of all non-null segments exhaust the entire series of real non-negative numbers. The answering of this question forms the content of the converse problem of the theory of measurement of segments; and a positive answer is possible only after the introduction of a new axiom.

It is sufficient to adopt the following axiom of Kantor on a nested sequence of closed segments.

V_2. Kantor's axiom. If on a line a there is given an infinite sequence of closed segments A_1B_1, A_2B_2, A_3B_3, ..., of which each successive $A_{n+1}B_{n+1}$ is contained in the preceding A_nB_n (fig. 221) and if no segment exists which is contained in all the segments of the given sequence, then there exists one and only one point X which belongs to all segments of the sequence.

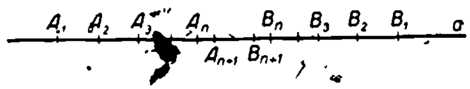


Fig. 221.

In this axiom, all segments are understood to be non-null, that is, to have distinct endpoints.

It must be noted that with this formulation the conclusion in the axiom is somewhat stronger than necessary.

It is sufficient to require the existence of at least one point X belonging to all segments of the series.

The uniqueness of point X will then follow.

Drawing upon the concept of the intersection of figures, the assertion of the axiom can also be expressed as follows: the intersection of all the segments of the indicated series A_1B_1, A_2B_2, A_3B_3, ... is the point X.

Remark: The axioms of Archimedes and Cantor are called the axioms of continuity. With their aid we may prove theorems like the following: A line a, passing through a point A interior to circle Γ, intersects the latter in two points.

The converse problem of the theory of measurement of segments is solved by the following theorem: Whatever be the real number $a > 0$, there exists under any given system of measurement a segment whose length is equal to a.

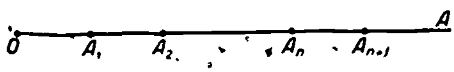


Fig. 222.

Proof: Let the number a be represented in the form

$$a = n + \frac{n_1}{2} + \frac{n_2}{2^2} + \ldots + \frac{n_k}{2^k} + \ldots,$$

where n is a natural number or zero, and the numbers $n_1, n_2, \ldots, n_k, \ldots$ are equal to 0 or 1. This is the representation of the number a in the form of a binary fraction:

$$a = n.n_1n_2n_3\ldots n_k\ldots$$

On the ray OA we lay off segments OA_1, A_1A_2, A_2A_3, ..., A_nA_{n+1}, each equal to the unit segment PQ (fig. 222). We bisect segment A_nA_{n+1} by means of point B_1. If $n_1 = 0$, we choose the segment A_nB_1; if $n_1 = 1$, we choose B_1A_{n+1}.

The chosen segment, s_1, is then bisected by means of point B_2. If $n_2 = 0$, from the two halves we select the segment nearer to point O; if $n_2 = 1$, we select the other half.

This second segment s_2 we bisect by means of point B_3, and so on.

The closed non-null segments $s_1, s_2, \ldots, s_k$ obtained in this manner form an infinite nested sequence of segments each lying within the preceding one and fulfilling the conditions of Kantor's axiom, since by lemma 1 no non-null segment can belong to all the segments of the series. From this it follows that the process of measurement of the segment OX, where X is the point common to all segments of the series, yields the length a. The theorem is proved.

Remark: The reasoning cited does not depend on whether the binary representation of the number a is finite or infinite. In the case of a finite binary fraction one of the points B obtained in the successive bisections will give a segment OB having a length precisely equal to a. In this case we can in fact dispense with Kantor's axiom.

More than this, if $a = \frac{p}{q}$, where p and q are natural numbers, that is, if a is a rational number, by dividing the unit segment PQ into q equal parts and taking the q-th part p times we obtain a segment of length a. This method of finding a segment of length $\frac{p}{q}$ can be carried out without recourse to Cantor's axiom.

In this exposition of the theory of measurement of segments we have made use of the existing theory of real numbers [37], [34]. Historically, the theory of measurement began its development first, and later the two theories developed simultaneously.

Only the solution of the direct and converse problems of the theory of measurement permits us, having selected a unit of measurement, to speak of a one-to-one correspondence between the set of real numbers and the set of points on a line.

This correspondence establishes a system of coordinates on a line. Proceeding from this to coordinates on a plane and in space involves no difficulties.

As we shall see in the next section, the solution of the converse problem of the theory of measurement of segments provides at the same time solution of the problem previously stated of finding the set of all systems of measurement of segments.

55. THE DEPENDENCE OF THE LENGTH OF A SEGMENT ON THE CHOICE OF THE UNIT OF MEASURE

The length $\rho(AB)$ of segment AB is determined only when we have selected a unit of measurement -- the segment PQ. A second choice of a unit segment also changes the length of the given segment AB.

Let us pose a problem: Given the length x of a segment AB under a system of measurement having the unit of measurement PQ, to find the length y of the same segment AB under a system of measurement with the unit P'Q'.

Evidently y is a function of x:

$$y = f(x),$$

defined on the set of all non-negative numbers.

To see this, we note that to every $x \geq 0$ there corresponds a segment AB, the length of which is x. Measuring segment AB with the unit P'Q', we obtain a unique real number y equal to the length of segment AB in the new system of measurement. Since, evidently, $f(0) = 0$, it remains to find the function $f(x)$ for $x \neq 0$.

We observe, firstly, that the function $f(x)$ is monotonically increasing, since to a larger value of the length x there corresponds a larger segment, and consequently also a larger value of y. Secondly, letting $y_1 = f(x_1)$ and $y_2 = f(x_2)$, we obtain by virtue of condition 4

$$y_1 + y_2 = f(x_1 + x_2),$$

or

$$f(x_1) + f(x_2) = f(x_1 + x_2). \qquad (\alpha)$$

Equation (α) is a very simple example of a functional equation.

We have now to find the function f on the basis of the property of this function expressed in (α). A precise formulation of the problem is as follows: to find all the monotonically increasing functions $f(x)$, defined on the set of real numbers ≥ 0, which satisfy the functional equation

$$f(x_1 + x_2) = f(x_1) + f(x_2)$$

for all $x_1 \geq 0$ and $x_2 \geq 0$.

By mathematical induction we find:

$$f(x_1 + x_2 + \ldots + x_n) = f(x_1) + f(x_2) + \ldots + f(x_n).$$

Putting

$$x_1 = x_2 = \ldots = x_n = x,$$

we have:

$$f(nx) = nf(x) \qquad (\beta)$$

for any natural number n and any $x \geq 0$.

Setting $x = 1$, we have, for any natural n,

$$f(n) = nf(1).$$

Evidently, $f(1)$ is the length of the unit segment PQ of the first system of measurement when measured by the unit P'Q' of the second system of measurement. Denoting this length by a,

$$f(1) = a,$$

we obtain:

$$f(n) = an. \qquad (1)$$

In equation (β) we now set $x = \frac{1}{n}$:

$$f(1) = nf(\frac{1}{n}),$$

from which follows:

$$f(\frac{1}{n}) = a\frac{1}{n} . \qquad (2)$$

Setting, in equation (β), $x = \frac{m}{n}$, where $m \geq 0, n > 0$ and both are integers, we find:

$$f(m) = nf(\frac{m}{n}),$$

whence by eq. (1) $am = nf(\frac{m}{n})$

or

$$f(\frac{m}{n}) = a\frac{m}{n} . \qquad (3)$$

From the foregoing we conclude that for all rational $x = \frac{m}{n} \geq 0$ the function $f(x)$ has the form

$$f(x) = ax.$$

Let z be an irrational positive number. Then for an arbitrary integer $n > 0$ there can be found an integer $m \geq 0$ such that

$$\frac{m}{n} < z < \frac{m+1}{n} .$$

But since the function $f(x)$ is monotonically increasing,

$$f(\frac{m}{n}) < f(z) < f(\frac{m+1}{n}),$$

that is,

$$\frac{m}{n} a < f(z) < \frac{m+1}{n} a .$$

After passing to the limit as $n \to \infty$ we obtain

$$f(z) = az .$$

In this manner, for any $x \geq 0$ we have:

$$f(x) = ax,$$

or

$$y = ax. \qquad (\gamma)$$

Here x is the "old" length of the given segment, y is its "new" length and $a = f(1)$ is the length of the "old unit" when measured by the "new unit".

The functional equation (α) has been solved, and at the same time the dependence (γ) of the length of the segment upon the choice of the unit of measurement has been found. Namely: <u>In passing from one system of measurement to another the lengths of all segments are multiplied by one and the same number, which is equal to the length of the old unit of measurement as measured by the new unit of measurement.</u>

Let there be given two segments AB and CD the lengths of which are x_1 and $x_2 > 0$ in one system of measurement, and y_1 and y_2 in a second.

Equation (γ) gives:

$$y_1 = ax_1, \quad y_2 = ax_2,$$

whence $\frac{y_1}{y_2} = \frac{x_1}{x_2}$.

Accordingly, the ratio of the lengths of two segments is independent of the choice of the unit of measurement.

This ratio of the lengths of segments, independent of the system of measurement, is called the <u>ratio of the segments</u>:

$$\frac{AB}{CD} = \frac{x_1}{x_2}.$$

Remark: Our problem has not been concerned with establishing a system of measurement of the lengths of curvilinear arcs. In this connection we shall merely note the following: If a definite number is assigned to any arc of a curve, it does not yet follow that this number can be called the length of the given arc. It is necessary that to each arc of a definite class of curves containing all straight-line segments there be assigned a number, and that this system of numbers satisfy the following conditions:

A. To congruent arcs there must correspond equal numbers (the property of invariance).

B. The number assigned to the union of two arcs having no interior points in common must be equal to the sum of the numbers assigned to the arcs added (the property of additivity).

Only after meeting these requirements can the numbers in question be called the lengths of the corresponding arcs.

For example, if to every circle is assigned a number equal to the common limit of the perimeters of the inscribed and the circumscribed simple polygons as the number of sides of each polygon approaches infinity and the length of each side approaches zero, this number cannot be considered and called the length of the circle until there has been established a system of measurement of length for some class of curves [34].

56. COMPARISON OF AREAS ON A PLANE

In 53. we saw that for the comparison of magnitudes of straight-line segments it was not necessary to introduce the notion of the length of a segment.

For a certain class of plane figures there can be introduced three concepts, which may bear the names equal, greater and less.

Here the concepts of congruence and equality may not be identical, as they were in the case of linear segments.

The congruence of figures could, for example, be defined as follows: two plane figures are said to be congruent if there exists a motion (of the first or the second kind) in the plane which will transform one figure into the other.

We know that the transformations of a plane into itself known as motions can be defined merely by means of the axioms of congruence. These axioms relate to the concept of congruence of segments and of angles.

The definition just given of the congruence of figures is a definition of a relationship between figures which satisfies all the conditions imposed upon the concept of "equality". These are:

1. reflexiveness: every figure is congruent with itself;
2. symmetry: if one figure is congruent with a second, then the second is congruent with the first;
3. transitivity: if one figure is congruent with a second, and the second with a third, then the first figure is congruent with the third.

If we should call only congruent figures equal, we should be deprived of the possibility of comparing the magnitudes of many important classes of figures, since the relationship of congruence frequently cannot be supplemented by the relationships <u>greater</u> and <u>less</u> in such a manner that the class of figures becomes a class of magnitudes.*

For example, let the given class of figures be the class of squares. We shall call two squares equal if and only if they are congruent; a square whose sides are larger we shall consider <u>larger</u> than a square with smaller sides, and so on. Using these conditions the class of squares can be considered a class of magnitudes with respect to the given meaning of <u>equal</u>, <u>greater</u> and <u>less</u>.

Let us consider a second example. Let the given class of figures be the set of all circles. It is easy to establish a meaning for the relationships <u>equal</u>, <u>greater</u> and <u>less</u> which converts the class of circles into a class of magnitudes with respect to these relationships.

Again, let the given class of figures be the set of all rectangles. If we consider congruent and only congruent rectangles equal, the question arises, how to

---

* By this is meant that for any two figures a and b belonging to the class, one and only one of the following statements is true: $a < b$, $a \equiv b$, or $a > b$; for any three figures a, b and c the fulfillment of any one of the conditions (1) $a > b$ and $b > c$; (2) $a > b$ and $b \equiv c$; or (3) $a \equiv b$ and $b > c$, implies that $a > c$. Furthermore, if figure a is a subset of figure b we have $a < b$.
--Translators.

formulate a meaning of the relationships <u>greater</u> and <u>less</u> which would satisfy all the conditions for converting the class of rectangles into a class of comparable magnitudes. More than this, it is necessary to decide whether it is even possible to formulate such meanings.

Analogous questions may be posed with reference to the class A of figures consisting of the set of all rectangles and of all circles, or to the class B consisting of the set of all finite plane figures.

It is not difficult to show that if figures of class A are considered equal only on condition of their being congruent, it will be impossible to find meanings for the relationships <u>greater</u> and <u>less</u> such that all these relationships will satisfy the requirements for the comparison of magnitudes.

In other words, it is impossible to assign in any manner a meaning to the concepts <u>greater</u> and <u>less</u> whereby under the already given meaning of the relationship <u>equal</u> (namely, congruent) class A would become a class of magnitudes.

For class B it is not possible in any manner whatsoever to establish notions of equal, greater and less which will fulfill the requirements for the comparison of magnitudes. We shall not adduce the proof of this assertion.

The concepts of equicomposition and equivalence play a fundamental role in establishing the relationships equal, greater and less for classes of plane two-dimensional figures, in particular for simple polygons.

Definition. A polygon is said to be simple if all its vertices are distinct, none of them forms an interior point of a side of the polygon, and no two of its sides have any interior point in common.

There exists the following important theorem:

Theorem. Every simple polygon divides those points of the plane in which it lies which do not belong to the polygon itself (considered as a system of segments) into two regions -- the interior and the exterior which have the following properties: if A is a point of the interior and B a point of the exterior, then every broken line lying in the plane of the polygon and joining points A and B will have at least one point in common with the polygon; and if on the other hand A and A' are points interior to the polygon and B and B' points exterior to it, then there always exist in the plane of the polygon broken lines joining A with A' and B with B' and having no points in common with the polygon.

We shall not stop to prove this theorem [15].

If two points A and B of a simple polygon P are joined by a broken line not self-intersecting and lying entirely within this polygon, two new polygons P_1 and P_2 are obtained (fig. 223). We say in this case that polygon P is decomposed into polygons P_1 and P_2, and also that polygons P_1 and P_2 together compose polygon P.

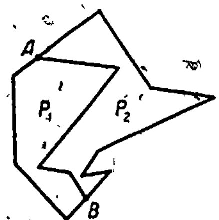


Fig. 223.

Definition: Two simple polygons are said to be equicomposed, or equivalent by decomposition, if it is possible to decompose them into a finite number of pairwise congruent polygons -- and hence into a finite number of pairwise congruent triangles.

Intuitively, we may say that two equicomposed simple polygons are polygons which may be cut into parts such that either polygon can be constructed from the parts of the other.

We shall consider a number of theorems concerning the equicomposition of some simple polygons.

Theorem. A triangle is equicomposed with a parallelogram having as its base the base of the triangle and an altitude equal to one-half the altitude of the triangle.

The proof is self-evident (fig. 224).

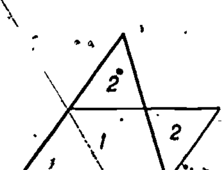


Fig. 224.

Theorem. A parallelogram is equicomposed with a rectangle the base of which consists of the longer side of the parallelogram and whose altitude is the corresponding altitude of the parallelogram (fig. 225).

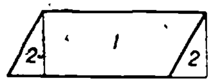


Fig. 225.

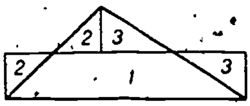


Fig. 226.

It is suggested that the student prove this independently. Why is the longer side specified?

Theorem. A triangle is equicomposed with a rectangle whose base consists of the longest side of the triangle and whose altitude is one-half the corresponding altitude of the triangle (fig. 226).

It is suggested that the student prove this independently.

Theorem. Parallelograms with a common base and equal altitudes are equicomposed.

We draw through E, the point of intersection of sides AD and BC, the line MN parallel to AB (fig. 227) and lay off MP equal to AM, and so on. It is suggested that the proof be completed independently.

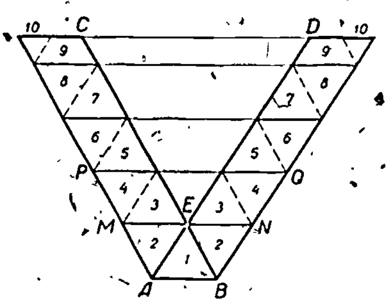


Fig. 227.

Theorem of Pythagoras. The figure consisting of the union of the two squares constructed upon the legs of a right triangle is equicomposed with the square constructed upon the hypotenuse (fig. 228).

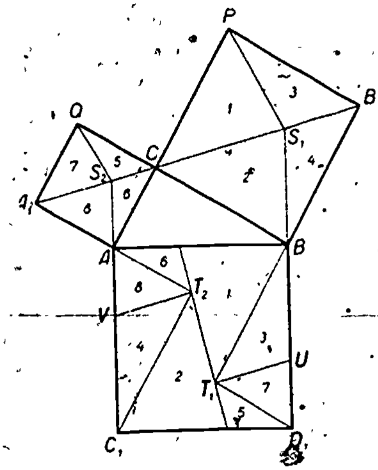


Fig. 228.

Proof. We draw the diagonals A_1C and CB_1 of the squares constructed upon the legs. These diagonals constitute one segment A_1B_1. We prolong the sides C_1A and D_1B of the square constructed upon the hypotenuse to their points of intersection S_2 and S_1 with segment A_1B_1. We join S_1 and S_2 with the vertices P and Q respectively. We obtain eight triangles into which the squares upon the legs are decomposed. We prolong the side BB_1 of the square BB_1PC to its intersection in T_1 with the line D_1T_1 parallel to BC. Analogously we construct the point of intersection T_2 of the line A_1A with the line C_1T_2 parallel to AC. We join T_1 and T_2 by a line and we draw the segments T_1U and T_2V parallel to A_1B_1.

The segments constructed divide the square upon the hypotenuse into eight triangles respectively congruent to the first eight triangles. It is suggested that the pairwise congruence of the triangles be proved independently.

If as the relation of equality we adopt equicomposition, or equivalence by decomposition, then this relation within the class of simple polygons will possess the following properties:

1) <u>reflexivity</u>: every simple polygon is equicomposed with itself;

2) <u>symmetry</u>: if a simple polygon is equicomposed with any other simple polygon, then the second polygon is equicomposed with the first;

3) <u>transitivity</u>: if a simple polygon is equicomposed with any other simple polygon, and this second polygon is equicomposed with a third, then the first polygon is equicomposed with the third.

The proof of this last property is not entirely self-evident because, for example, the second polygon may have been shown to be equicomposed with the first by means of one decomposition and equicomposed with the third by means of another decomposition not containing triangles congruent with those of the first.

The first two properties are easily proved; the second having been proved, the proof of the third property is reduced to the proof of the theorem:

<u>Two simple polygons, each equicomposed with a third, are equicomposed with each other.</u>

Proof. Let the simple polygons P_1 and P_2 each be equicomposed with the simple polygon P_3 (fig. 229).

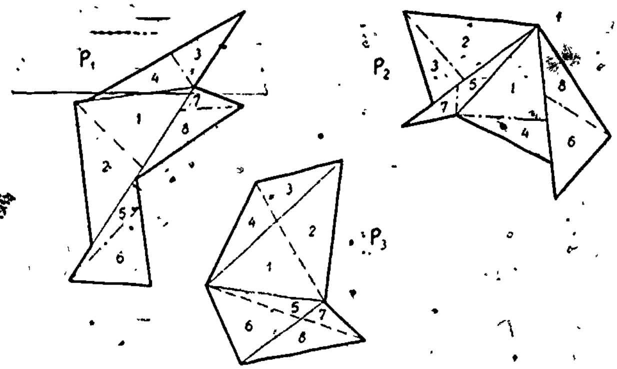


Fig. 229.

Since polygon P_1 is equicomposed with polygon P_3, there exists a decomposition of polygons P_1 and P_3 into pairwise congruent triangles. For the same reason polygons P_2 and P_3 can be decomposed into pairwise congruent triangles.

Inspecting the two decompositions of polygon P_3, we observe that the segments of the second decomposition decompose the triangles of the first decomposition into polygons. We add to the decomposition of polygon P_1 the segments of the second decomposition of P_3 in such a manner that each segment entering into the decomposition of any triangle in P_3 enters in exactly the same way into the congruent triangle of P_1. We proceed analogously with the triangles into which polygon P_2 was decomposed.

The two polygons P_1 and P_2 will evidently be decomposed into pairwise congruent polygons and are consequently equicomposed. The theorem is proved.

Thus we see that in the class of simple polygons equivalence by decomposition possesses all the properties which are required of the relationship of equality.

We can, however, define the relationship of equality within the class of simple polygons in another way.

Definition. Two simple polygons are said to be equivalent by completion, or simply equivalent, if it is possible to join to each of the given polygons one member from each of a finite number of pairs of congruent polygons to such effect that the two resultant polygons will be equicomposed (fig. 230).

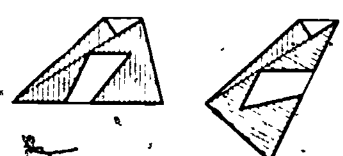


Fig. 230.

Lobachevskii writes in his "Geometrical investigations concerning the theory of parallels" [31]: "In order to judge in general as to the equality of two surfaces, the following proposition is taken as fundamental: two surfaces are equal if they are formed by the composition or the separation of equal parts."

Lobachevskii applies this definition to simple plane and spherical polygons, and also to the simple polygons of Lobachevskian geometry.

The class of simple polygons consisting of all those polygons equivalent to a given simple polygon evidently contains as a part of itself the class of all simple polygons equicomposed with the given polygon.* Whether this part is a proper part or not has to be decided by appropriate investigation.

We shall introduce some simple theorems concerning equivalence by completion.

Theorem. Two parallelograms with equal bases and equal altitudes are equivalent.

We adduce the proof[1] in the form given by Euclid [22]. Let ABCD and EBCF be two parallelograms with common base (fig. 231).

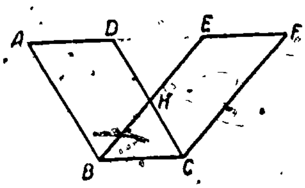


Fig. 231.

Euclid demonstrates first of all the congruence of triangles ABE and DCF, thereupon adding parallelogram ABCD to triangle DCF and parallelogram EBCF to triangle ABE to produce one and the same trapezoid ABCF. From this follows that the given parallelograms are equivalent.

(1) Footnote: Although we have already a previous theorem stating that such parallelograms are equicomposable and a fortiori equivalent, the author prefers to include this proof which does not depend upon the axiom of Archimedes. --Translators.

Remark. Euclid does not introduce the notion of equicomposition nor that of equivalence of polygons, but makes use of the concept of the equality of triangles with respect to area. The notion of area is regarded by Euclid as a primary, undefinable concept.

The proof of the following theorem is analogous: Two triangles with equal bases and equal altitudes are equivalent.

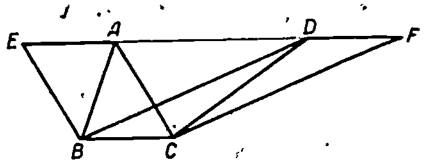


Fig. 232.

Euclid (in proposition 37) considers the equivalent parallelograms AEBC and DFCB (fig. 232) and then passes to their respective halves, triangles ABC and DBC. The proof, based only upon the concept of equivalence, is suggested as an exercise.

Proposition 43 of Euclid's "Elements" may be formulated as follows:

Theorem. If from any point K of the diagonal AC of parallelogram ABCD (fig. 233) the lines KH and KF are drawn parallel to the sides of the parallelogram, then those two of the four parallelograms into which ABCD is thus decomposed through which the diagonal AC does not pass are equivalent.

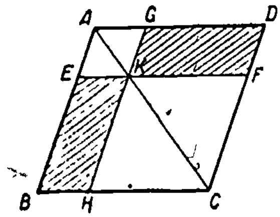


Fig. 233.

This is one of the fundamental theorems of the Euclidean theory of area. From it can be directly derived, for example, the reduction of any polygon to an equivalent square. Here this theorem is introduced merely as an example. The proof is left to the student.

The following fundamental propositions hold:

1) Reflexivity: every simple polygon is equivalent to itself.

2) Symmetry: if one simple polygon is equivalent to a second, the second is likewise equivalent to the first.

3) Transitivity: if one simple polygon is equivalent to a second, and the second to a third, then the first polygon is also equivalent to the third.

The first two properties are self-evident. We shall not stop to prove the third property [41], [15].

The notions of equicomposition and equivalence by completion were needed in order to establish the entire theory of area without using the axioms of continuity, in particular the axiom of Archimedes. Hilbert [15] showed that without the axiom of Archimedes it is impossible to prove that two equivalent simple polygons will also be equicomposed; it was for this reason that it was necessary to introduce the more general notion of equivalence by completion.

Having successively established theorems on the equivalence of two parallelograms with equal altitudes and equal bases, on the equivalence of two triangles with equal altitudes and equal bases, and on the fact that it is always possible to construct a right triangle having a given leg and equivalent to any given simple polygon, Hilbert remarks that all these theorems become significant only if we succeed in showing that not all simple polygons are equivalent to each other.

In order to elucidate this observation of Hilbert's let us suppose, with respect to some previously unknown set having some algebraic operation, that it had been established that this set constituted a group. From this fact, that the set was a group, there had then been deduced far-reaching consequences, and so on; in a word, there had been constructed a whole theory. Of course, the theory would lose all significance if it were afterwards discovered that the group studied consisted solely of one neutral element.

However, Hilbert succeeded in establishing, without using the axiom of Archimedes, that if two triangles equivalent by completion have equal bases then they also have equal altitudes. It follows that two triangles with equal bases can not be equivalent, thus establishing that not all simple polygons are equivalent.

The theorem just mentioned is none other than the 39th proposition of Euclid's "Elements". As we see, Euclid's proof, based upon the idea that areas are magnitudes, can not be accepted until after the introduction of the relationships "equal", "greater" and "less" for the class of simple polygons. Without the use of the axiom of Archimedes, Hilbert proves the theorem: If by means of straight lines a rectangle is divided into triangles and if even one of these triangles is removed, it is not possible to reconstitute this rectangle with the remaining triangles.

Hilbert introduces the notions of greater and less and defines a measure of area in the class of simple polygons without using the axioms of Archimedes and Cantor. The theory of area can thus be considered to have been established without the use of the axioms of continuity. Only thereafter can the set of all simple polygons be regarded as a class of magnitudes. It has been shown that it is likewise possible to establish the entire theory of area without using either Euclid's axiom on parallels or the corresponding axiom of Lobachevskii.

57. THE TRANSFORMATION OF POLYGONS

If we admit without proof, or regard as established, that two triangles equivalent by completion and having equal bases have also equal altitudes, then it is possible to prove the theorem: Any two equivalent simple polygons are equicomposed.[1]

In proving this theorem we shall make no special point of limiting ourselves in the use of axioms. The fundamental properties of motions will also be assumed to be known.

The proof is divided into several propositions.

Proposition one. Two polygons equicomposed with a third are equicomposed with each other.

The proof was given in the preceding section.

Proposition two. Two triangles ABC and ADC having a common vertex A, a common side AC and equal bases BC and CD lying on one line (fig. 234) are equicomposed.

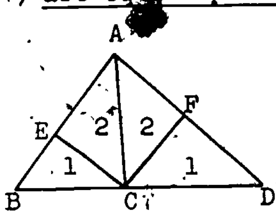


Fig. 234.

Proof. Through point C, the midpoint of segment BD, we draw lines CE and CF, respectively parallel to sides AD and AB. By this construction the given triangles are decomposed into

(1) The proof of this theorem, which was formulated somewhat differently, was given by the Hungarian mathematician F. Bolyai in the first half of the nineteenth century. The proof has been repeatedly simplified. A very simple version of one of these proofs was given by B. F. Kagan [25]. It is this latter version which, with some refinements, we shall introduce here.

pairwise congruent triangles:

$$\triangle BEC = \triangle CFD,$$

$$\triangle AEC = \triangle CFA,$$

q.e.d.

Proposition three. If two equivalent triangles have a side of one equal to a side of the other, they are equicomposed.

Proof. Let us arrange the given triangles, A_1BA_2 and A_1CA_2, so that the equal sides, which we shall regard as the bases, coincide and the triangles lie on opposite sides of the common base A_1A_2, and so that the angles of the respective triangles which adjoin each other at A_1 are both acute (fig. 235).

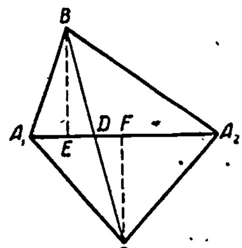


Fig. 235.

Since the triangles are equivalent, their altitudes BE and CF are equal. Therefore, the diagonal BC of the quadrilateral which has been formed is bisected at its point of intersection D with the common base A_1A_2.

The quadrilateral formed by the two triangles may turn out to be convex, or it may not.

1° Let us suppose first that the quadrilateral A_1BA_2C is convex. Diagonal BC, lying entirely within the quadrilateral, decomposes triangles A_1BA_2 and A_1CA_2 into the triangles A_1BD, BDA_2, A_1DC and DCA_2. But, by virtue of proposition two, triangles A_1BD and A_1DC are equicomposed. Triangles BDA_2 and

DCA_2 are likewise equicomposed. It is suggested that the student construct all the pairs of congruent triangles into which the triangles A_1BA_2 and A_1CA_2 can be "cut".

2° It may happen that the quadrilateral A_1BA_2C has a reëntrant angle (greater than two right angles) at vertex A_2. The diagonal BC will lie outside the quadrilateral (fig. 236).

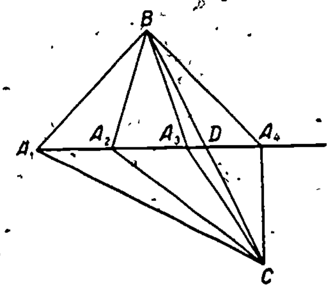


Fig. 236.

We lay off on the line A_1A_2, in the direction from A_1 to A_2, the segments A_2A_3, A_3A_4, ..., $A_{n-1}A_n$ each congruent to segment A_1A_2.

By virtue of the axiom of Archimedes there exists an n such that point D is found between A_{n-1} and A_n, the consequence being that quadrilateral $A_{n-1}BA_nC$ will be convex (in the diagram $n = 4$). By what has already been proved, the triangles A_3BA_4 and A_3CA_4 are equicomposed. But since, by virtue of proposition two, the triangles

$$A_1BA_2,\ A_2BA_3,\ \ldots,\ A_{n-1}BA_n$$

are equicomposed with each other, exactly as are the triangles

$$A_1CA_2,\ A_2CA_3,\ \ldots,\ A_{n-1}CA_n\ ,$$

from this and the transitivity of equicomposition it follows that the triangles A_1BA_2 and A_1CA_2 are equicomposed.

Proposition four. Any two equivalent triangles are equicomposed.

Proof. Let triangles ABC and $A_1B_1C_1$ be equivalent. If side AB is congruent to side A_1B_1, then our assertion follows from proposition three. Suppose that BA is greater than A_1B_1. From vertex B_1 (fig. 237) of triangle $A_1B_1C_1$ we draw line a parallel to side A_1C_1, and on a we find a point B' such that $A_1B' = AB$.

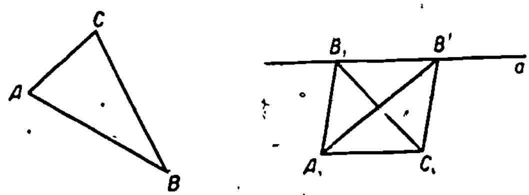


Fig. 237.

The triangles $A_1B_1C_1$ and $A_1B'C_1$ are equivalent, since they have a common base A_1C_1 and equal altitudes. Having the common side A_1C_1 they will also be, by virtue of proposition three, equicomposed.

Since triangle $A_1B'C_1$ is equivalent to triangle $A_1B_1C_1$ and the latter is equivalent to triangle ABC, it follows by the transitivity of the equivalence relation that triangles ABC and $A_1B'C_1$ are equivalent. But since these triangles have the equal sides AB and A_1B' they are also equicomposed. Whence, by the transitivity of equidecomposition it follows that the given triangles ABC and $A_1B_1C_1$ are equicomposed.

Proposition five. Every polygon is equicomposed with some triangle which is equivalent to the polygon.

Proof. We shall first show how to construct a convex quadrilateral which is equicomposed with the union of two arbitrary triangles ABC and $A_1B_1C_1$.

If in the given triangles ABC and $A_1B_1C_1$ no side of one is equal to a side of the other, then, assuming $AB > A_1B_1$, we construct triangle $A_1B'C_1$ equicomposed with $A_1B_1C_1$ and so that $A_1B' = AB$ (fig. 237). Then placing together the equal bases AB and A_1B' of triangles ABC and $A_1B'C_1$, we arrange their vertices C and C_1 on opposite sides of the common base. If we thus obtain a convex quadrilateral (as in fig. 235) the construction is finished. If the quadrilateral is not convex, we apply the construction used in the proof of proposition three (fig. 236).

By this method it is possible to replace any two triangles by a convex quadrilateral equicomposed with their union.

We proceed by Euclid's method to reduce the convex quadrilateral to an equicomposed triangle. Through the vertex A of the convex quadrilateral ABCD we draw line AE parallel to diagonal BD (fig. 238). If E is the point of intersection of AE and CD, triangle BCE will be equicomposed with quadrilateral ABCD, since triangles ABD and BDE are equicomposed (by proposition three).

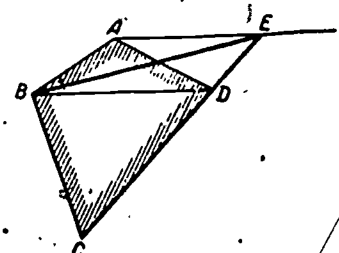


Fig. 238.

In this manner any two triangles can be replaced by one triangle which is equicomposed with their union.

Decomposing any simple polygon into triangles and constructing from two triangles of this decomposition a single triangle equicomposed with their union, we join to the latter one more triangle of the decomposition, and so on, until we reduce the whole of the given polygon into a single triangle equicomposed with the polygon.

Proposition six. Any two equivalent simple polygons are equicomposed.

Proof. Let P_1 and P_2 be two equivalent simple polygons. We construct triangles $\triangle_1$ and $\triangle_2$ respectively equicomposed with polygons P_1 and P_2. By the transitivity ~~of equivalence we find~~ that triangles $\triangle_1$ and $\triangle_2$ are equivalent. By virtue of proposition four triangles $\triangle_1$ and $\triangle_2$ are equicomposed. But since equicomposition is symmetrical and transitive we find that polygon P_1 is equicomposed with triangle $\triangle_1$ and, consequently, with triangle $\triangle_2$ and with polygon P_2, q.e.d.

From the foregoing investigation it follows that, with the use of the axiom of Archimedes, the class of simple polygons equicomposed with a given simple polygon is identical with the class of simple polygons equivalent to the given polygon.

Hilbert has shown that without the axiom of Archimedes, using the remaining axioms only, the assertion of the identity of these classes would be untrue.

Hereinafter we shall be able to omit making any distinction between simple polygons equivalent by decomposition, i.e., equicomposed, and simple polygons equivalent by completion or simply equivalent. Polygons for which it is possible to find a motion of whatever kind which will bring one into coincidence with the other will be spoken of as "congruent" (identical).

Lobachevskii formulates this distinction as follows: "Geometrical magnitudes are said to be identical when one may be taken without distinction in place of the other. The criterion of identity is this, that when one is placed upon the other the lines and vertices of the first coincide with the lines and vertices of the second. If however, two geometric magnitudes are identical only part by part then they are equal"[30].

58. THE PROBLEM OF MEASUREMENT OF THE AREA OF SIMPLE POLYGONS

As in the case of segments, the problem of the measurement of any class of simple figures consists of assigning to each figure of the given class a number in such a way that the following requirements are met:

1. One and the same number shall correspond to congruent figures (the property of invariance with respect to motions).

2. The number attributed to the sum of two figures shall be equal to the sum of the numbers attributed to each of the figures (the property of additivity).

Under "the sum of figures" is here understood the union of figures (5) having no interior points in common but which may have boundary points in common.

If we find it possible to assign to each figure of a given class a number such that conditions 1 and 2 are satisfied, we shall say that we have a system of measurement of figures of the given class, and the number corresponding to a figure will be referred to as the measure of the figure.

We already know how to establish the measure of a linear segment; this measure is called the length of the segment.

It must be noted that to establish a system of measurement for the class of all figures, even of all finite ones, is impossible. We shall limit ourselves to the class of simple polygons. The measure of a polygon will be called its area.

Remark. To degenerate triangles, i.e. those whose three vertices all lie upon one line, we shall always assign the number 0.

Here, as in the case of segments, it is necessary first of all to decide the question of the possibility of a system of measurement of simple polygons and, upon confirming such a possibility, to find all the possible systems of measurement.

Assuming a system of measurement of segments to have been chosen, we shall establish the following theorems [41], [1].

Theorem 1. In any triangle the product of a side and the corresponding altitude is independent of the choice of side.

Remark: Hereinafter, instead of "the product of the lengths of segments" we shall simply speak of "the product of segments".

Proof. We draw the altitudes AE and CD in triangle ABC (fig. 239). The right triangles AEB and CDB are similar since they have the angle B in common. Because of their similarity

$$\frac{h_a}{c} = \frac{h_c}{a},$$

or, $ah_a = ch_c$. We can likewise write

$$ah_a = bh_b = ch_c, \quad \text{q.e.d.}$$

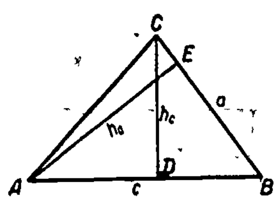


Fig. 239.

We select some fixed positive number k and assign to every triangle a number S equal to the product of any side times the corresponding altitude times k, that is,

$$S = kah_a.$$

It is evident that equal numbers will be assigned to congruent triangles.

One may go around any triangle ABC in two different directions, clockwise and counterclockwise. Assigning to a one-dimensional triangle one of these directions is called orienting the triangle, or assigning an orientation to it. One of these two possible orientations is called positive and the other negative, and as indicated in fig. 240 it is the counterclockwise orientation which is customarily called positive.

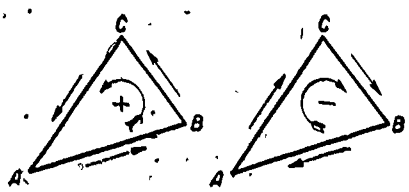


Fig. 240.

Selecting the orientation (AB or BA) of one of the sides of triangle ABC completely determines the orientation of this triangle.

All statements in the remainder of this section will refer to figures lying in one plane. The given triangles, unless otherwise specified, are given with positive orientation.

Theorem. If an arbitrary point O is joined to the vertices of a given oriented triangle ABC and the orientations of all the resulting triangles are determined by the orientations of the sides AB, BC and CA of triangle ABC, then the difference between the sum of the "numbers S" of all the resulting triangles having a vertex at O and positive orientation and the sum of the "numbers

S" of such triangles with negative orientation is equal to the "number S" of the given triangle.

Here it is necessary to distinguish five cases:

1°. Point O lies on one of the sides (fig. 241). Using the notation already introduced we have:

$$S_{ABC} = kch_c;\ S_{AOC} = k\cdot AO\cdot h_c;\ S_{OBC} = k\cdot OB\cdot h_c,$$

whence

$$S_{AOC} + S_{OBC} = k(AO + OB)h_c = kch_c = S_{ABC}.$$

2°. Point O lies on the prolongation of one of the sides (fig. 242). We have:

$$S_{OCA} = k\cdot AO\cdot h_c;\ S_{OBC} = k\cdot BO\cdot h_c,$$

whence

$$S_{OCA} - S_{OBC} = k(AO - BO)h_c = kch_c = S_{ABC}.$$

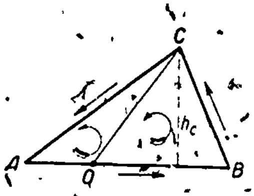


Fig. 241.

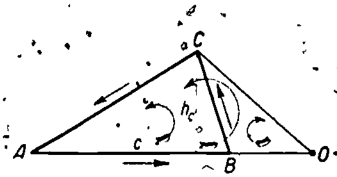


Fig. 242.

3°. Point O lies within the triangle (fig. 243). Prolonging, say, CO to the point D of its intersection with side AB, in accord with the previous cases we find:

$$S_{OCA} = S_{DCA} - S_{DOA};\ S_{OBC} = S_{DBC} - S_{DBO};$$

$$S_{OAB} = S_{DOA} + S_{DBO}.$$

From this we obtain:

$$S_{OCA} + S_{OBC} + S_{OAB} = (S_{DCA} - S_{DOA}) + (S_{DBC} - S_{DBO}) + (S_{DOA} + S_{DBO}) = S_{DCA} + S_{DBC} = S_{ABC}.$$

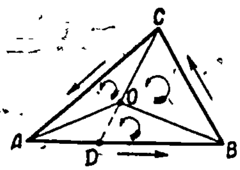


Fig. 243.

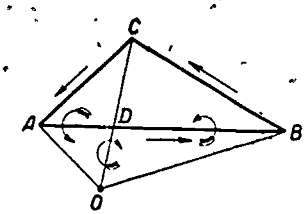


Fig. 244.

4°. Point O lies outside the given triangle but within one of its angles (fig. 244).

Let the angle in question be angle C. Denoting by D the point of intersection of side AB with the ray OC, in accord with the previous cases we find:

$$S_{OCA} = S_{DAO} + S_{DCA}; \quad S_{OBC} = S_{DOB} + S_{DBC}; \quad S_{OAB} = S_{DAO} + S_{DOB},$$

whence we obtain:

$$S_{OCA} + S_{OBC} - S_{OAB} = (S_{DAO} + S_{DCA}) + (S_{DOB} + S_{DBC}) - (S_{DAO} + S_{DOB}) = S_{DCA} + S_{DBC} = S_{ABC}.$$

5°. Point O lies within an angle vertical to one of the angles of the given triangle (fig. 245).

By the preceding we find:

$$S_{OAB} = S_{DOA} + S_{DBO}; \quad S_{OCA} = S_{DOA} - S_{DCA}; \quad S_{CBO} = S_{DBO} - S_{DBC},$$

whence we obtain:

$$S_{OAB} - S_{OCA} - S_{CBO} = (S_{DOA} + S_{DBO}) - (S_{DOA} - S_{DCA}) - (S_{DBO} - S_{DBC}) = S_{DCA} + S_{DBC} = S_{ABC}.$$

Now let there be given any simple polygon $A_0A_1A_2A_3...A_n$ with positive orientation (fig. 246). Furthermore, let this polygon be decomposed in any manner into N triangles and let O be an arbitrary point. Then the following holds:

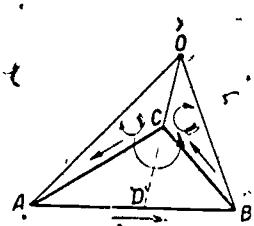


Fig. 245.

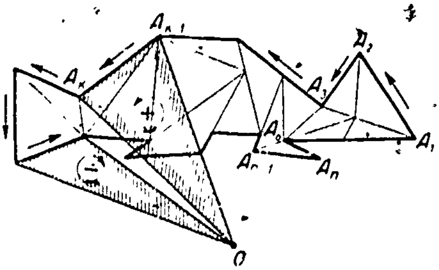


Fig. 246.

Theorem 3. The difference between the sum of the "numbers S" of the triangles with positive orientation having as base an oriented side of the polygon and vertex at the point O and the sum of the "numbers S" of such triangles with negative orientation is equal to the sum Σ of the "numbers S" of all the triangles of the decomposition.

Suppose the theorem to be true for each of two contiguous simple polygons P_1 and P_2, i.e. polygons having only a side or part of a side, but no interior points, in common (fig. 247). Further, let AB be one of the common sides or common parts of sides of polygons P_1 and P_2.

Remark. The set of common sides or common parts of sides of polygons P_1 and P_2 may consist merely of individual points or may even be empty.

The triangle with vertex at O and base AB has one orientation with reference to polygon P_1, but the opposite one with reference to polygon P_2. In figure 247, with reference to polygon P_1 the triangle OAB has positive orientation, while with reference to polygon P_2 the same triangle OAB has negative orientation.

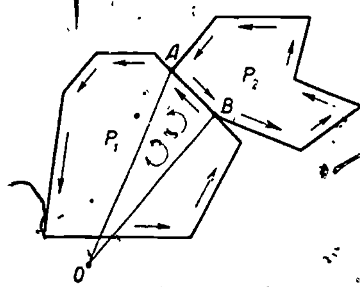


Fig. 247.

In the algebraic sum S* of the "numbers S" of the triangles with vertex at O and a side of one of the polygons as base, the terms referring to the triangles having the common side AB cancel each other out. There remain only the terms referring to the sides of a polygon P which is the sum of polygons P_1 and P_2. At the same time it is evident that the sums Σ_1 and Σ_2 of the "numbers S" of the triangles of decomposition give

$$\Sigma = \Sigma_1 + \Sigma_2 ,$$

where Σ is the sum of the "numbers S" for the triangles in the decomposition of polygon P generated by the decompositions of polygons P_1 and P_2. Since, by our assumption,

$$S_1^* = \Sigma_1 \quad \text{and} \quad S_2^* = \Sigma_2 ,$$

then

$$S_1^* + S_2^* = \Sigma_1 + \Sigma_2 = \Sigma .$$

But, as has just been shown,

$$S_1^* + S_2^* = S^*,$$

where S^* is the sum of the "numbers S" of the triangles with vertices in O and the sides of polygon P as bases; whence

$$S^* = \Sigma .$$

The proof of theorem 3 is completed as follows: theorem 3 is valid for a single ($n = 1$) triangle since in this case it is identical with theorem 2. Assuming the theorem valid for polygons decomposed into n triangles, we see that it will be valid for polygons decomposed into $n + 1$ triangles, since a polygon decomposed into $n + 1$ triangles can be regarded as the sum of one triangle and a polygon composed of n triangles.

Corollary 1. The number S^* is independent of the choice of the point O, since the number Σ is independent of this choice.

Corollary 2. The number Σ is independent of the way in which the polygon is decomposed into triangles.

We shall now assign to every simple polygon a number $\Sigma = S^*$. This attribution of a number Σ to every simple polygon actually establishes a system of measurement of areas of simple polygons. This is true because the numbers thus assigned possess, firstly, the property of invariance -- one and the same number is assigned to each of two congruent simple polygons, since congruent polygons can be decomposed into pairwise congruent triangles; and, secondly, the property of additivity -- the number assigned to a polygon consisting of the union of two polygons is equal to the sum of the numbers assigned to each part.

We are now able to speak of the number Σ, derived as above, as the <u>area</u> of a simple polygon. At the same time there has been demonstrated the existence of an infinite set of systems of measurement of the area of simple polygons.

To every number $k > 0$ there corresponds a particular system of measurement. In the expression

$$S_{\triangle} = kah_a$$

for the area of a triangle, the number $k > 0$ is a matter of choice. The choice of this number k fixes the unit of area. The unit of area is customarily taken as the square, the side of which is equal to the unit of length. Decomposing such a unit square into two triangles by means of a diagonal, we have by our condition

$$1 = 2 \cdot k \cdot 1 \cdot 1 ,$$

whence $k = \frac{1}{2}$. For the area of a triangle we have the usual expression

$$S_{\triangle} = \frac{1}{2} ah_a .$$

Under this choice of a unit of area the area of any simple polygon will be expressed <u>in square units</u>. It is, furthermore, not difficult to show that the system of measurement we have described is the <u>only</u> system of measurement satisfying the conditions of invariance and additivity and assigning to the square constructed upon the unit of length the area unity.

From the expression for the area of a triangle

$$S_{\triangle} = kah_a$$

it follows, for any positive value of k, that if two triangles with equal bases have equal area their corresponding altitudes are equal.

Remark. If as the unit of area we select the area of the equilateral triangle with sides equal to the unit of length, then the number k is defined by the relationship

$$1 = k \cdot 1 \cdot \frac{\sqrt{3}}{2} .$$

The area of a triangle measured in triangular units, for example, in triangular meters will be given by the formula

$$S_{\triangle} = \frac{2\sqrt{3}}{3} ah_a .$$

It is also possible to take as the unit of area the area of a circle with radius equal to the unit of length. The area of a triangle in such circular units, for example in circular meters, will, as may readily be verified, take the form

$$S_{\triangle} = \frac{1}{2\pi} ah_a .$$

From the system of measurement of the areas of simple polygons which we have now established it follows that simple polygons equivalent by completion have identical areas. This follows from the fact that polygons equivalent by completion are also equivalent by decomposition, i.e. equicomposed.

Since every simple polygon is equicomposed with some triangle, it follows that if two simple polygons have identical areas they are equicomposed and a fortiori equivalent by completion.

59. ON THE NOTION OF THE AREA OF A PLANE FIGURE

Analogously to our establishment of a system of measurement of the areas of simple polygons, we may pose the problem of the measurement of the areas of any other class of plane figures. As already mentioned, it is impossible to establish a system of measurement of area for all plane figures, not even for all finite ones, i.e. those, each of which lies within some square.

In order to establish a system of measurement of the areas of plane figures it is necessary first of all:

A. to set up a sufficiently broad class of plane figures, containing all simple polygons;

B. to assign to each figure of the class thus set up a positive number such that:

C. equal numbers are attributed to any two congruent figures of the given class (property of invariance);

D. to every figure of the given class consisting of the union of two figures of this class having no interior points in common there is assigned a number equal to the sum of the numbers of each of the added figures (property of additiveness).

After thus assigning to every figure of the class in question of a positive number possessing the properties of invariance and additiveness, we speak of each number as the area of the corresponding figure.

Beyond the properties mentioned, still another one is required:

E. the numbers, i.e. areas, shall be completely and uniquely determined when the area of one figure of the given class is known.

Let us note the following: From the fact that to a figure there is related a determinate number it does not follow that this number can be considered and called the area of this figure. It is required that a number shall be assigned to each plane figure of a selected class in such a manner that the correspondence possesses the properties "C", "D" and "E". For example, if to every circle there were to be assigned a number equal to the common limit of the areas of the inscribed and circumscribed simple polygons under the usual assumptions, it would not be permissible to consider or designate this number as the area of the circle until and unless there had been established a system of measurement of areas of the figures of some class which would include simply polygons, circles, sectors, segments of circles and so on. Only when we have a system of measurement of areas which fulfills conditions A, B, C, D and E can we speak of, say, the area of a circle.

The requirements enumerated for a system of measurement of the areas of plane figures naturally extend also to a system of measurement of the areas of a class of curved surfaces [34]. In this connection it must be emphasized that the intuitive conception of the area of a curved surface as the limit of the area of a polyhedron inscribed in the given surface may not give the desired result. The limit thus obtained may not yield a number which, in the given system of measurement of areas, can be regarded as the area of the given curved surface.

Let us consider an example, given by Schwartz. Let the radius of a right circular prism be R, and its altitude H. We divide the altitude H into n equal parts and through the points of division we draw n - 1 cylindrical parallels.

Further, we inscribe in the upper base of the cylinder a regular polygon of m sides, and in the first (counting from the top) of the parallels just drawn we inscribe a polygon congruent to the first one but in such a manner that the vertices of the latter lie in the meridional planes passing through the mid-points of the sides of the first polygon of m sides (fig. 248a).

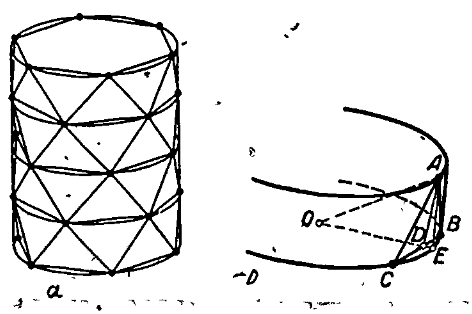


Fig. 248.

In the second parallel we inscribe a regular polygon of m sides so that its vertices lie on the meridians of the vertices of the polygon of the upper base. In the third polygon we inscribe a polygon of m sides with vertices on the meridians of the vertices of the polygon of the first parallel, and so on. Each polygon is obtained from the one directly above it by a translation of the latter along the axis of the cylinder to a distance of $\frac{H}{n}$ and a rotation through the angle $\frac{\pi}{m}$ around that axis.

We join the vertices of each polygon of m sides with the vertices of the neighboring ones and calculate the sum of the areas of the congruent isosceles triangles ABC which we have constructed (fig. 248b).

The base BC of triangle ABC will be

$$BC = 2R \sin \frac{\pi}{m},$$

and the altitude

$$AD = \sqrt{AE^2 + DE^2},$$

or

$$AD = \sqrt{\left(\frac{H}{n}\right)^2 + \left(R - R\cos\frac{\pi}{m}\right)^2} = \sqrt{\left(\frac{H}{n}\right)^2 + 4R^2 \sin^4 \frac{\pi}{2m}}.$$

In this manner we obtain for the area of the polyhedral surface inscribed in the lateral surface of the given cylinder:

$$S(m,n) = 2Rmn \sin\frac{\pi}{m}\sqrt{\frac{H^2}{n^2} + 4R^2 \sin^4 \frac{\pi}{m}}.$$

This latter expression may be rewritten in the form

$$S(m,n) = 2\pi R\left(\frac{\sin\frac{\pi}{m}}{\frac{\pi}{m}}\right)\sqrt{H^2 + \frac{\pi^4 R^2}{4}\left(\frac{n}{m^2}\right)^2\left(\frac{\sin\frac{\pi}{2m}}{\frac{\pi}{2m}}\right)^4}.$$

From this it can be seen that the limit of $S(n,m)$ as $n \longrightarrow \infty$ and $m \longrightarrow \infty$ depends upon the particular manner in which m and n tend toward infinity.

Suppose that m and n vary in such a manner that the ratio $\frac{n}{m^2}$ has a determinate limit q. In that case we have:

$$\lim_{\substack{n \to \infty \\ m \to \infty}} S(m,n) = 2\pi R\sqrt{H^2 + \frac{\pi^4 R^2}{4}q^2}$$

If, for example, n and m vary in such a way that $\lim \frac{n}{m^2} = q = 1$, then the limit of $S(m,n)$ will be equal to $2\pi R\sqrt{H^2 + \frac{\pi^4 R^2}{4}}$. We obtain this limit if we inscribe polygons of m sides in the parallels and at the same time divide the altitude of the cylinder into $n = m^2$ parts.

If n and m vary in such a manner that $\lim \frac{n}{m^2} = q = 0$, which can be brought about by taking, for instance, $n = m$, we arrive at the result

$$\lim_{\substack{n \to \infty \\ m \to \infty}} S(m,n) = 2\pi RH.$$

Only with $q = 0$ is a limit obtained which ~~is equal~~ to the well-known area of the lateral surface of the cylinder. Strictly speaking, the limit of $S(m,n)$ as $m \to \infty$, $n \to \infty$ does not exist.

The example discussed is instructive in that it vividly draws attention to the care and the caution with which one must approach the problem of establishing a system of measurement of areas.

60. A BRIEF DISCUSSION OF THE PROBLEM OF THE MEASUREMENT OF VOLUMES

The formulation of the general problem of the measurement of volumes in no way differs from the formulation of the problem of the measurement of areas.

To establish a system of measurement of the volumes of any definite class of figures it is necessary

A) to be able to assign to each figure of the given class a positive number such that the following conditions will be fulfilled:

B) that equal numbers will be attributed to congruent figures (the property of invariance);

C) that to the union of two figures having no interior points in common there will be assigned a number equal to the sum of the two numbers assigned to the given figures (the property of additiveness).

Only after fulfilling conditions A), B) and C) can the numbers in question be considered and called the volumes of figures of the given class. Over and above this the following condition must also be fulfilled:

D) the numbers representing volumes must be completely and uniquely determined when we know the volume of one figure of the given class.

In elementary geometry we establish first of all a system of measurement of the volumes of simple polyhedrons. The establishment of a system of measurement of volumes can be carried out on quite the same lines as in the case of areas:

It can be shown that in any tetrahedron (not necessarily regular) the product of the area of one of the sides and the corresponding altitude is independent of the choice of the side.

Theorems (entirely analogous to the theorems on the decomposition of simple polygons into triangles) can then be proved concerning the decomposition of simple polyhedrons into tetrahedrons, by which means there can be established a system of measurement of the volumes of simple polyhedrons [1], by assigning to each tetrahedron a number kSh, and so on. However, the measurement of volumes is not completely analogous to the measurement of areas. Namely, two tetrahedrons having equal volumes -- in the sense just established -- may be neither equicomposed nor equivalent by completion.

In other words, there exist two tetrahedrons with equivalent bases and equal altitudes which it is impossible either to decompose in any way into pairwise congruent tetrahedrons or to enlarge by the addition of pairwise congruent tetrahedrons so as to produce polyhedrons which could be decomposed into pairwise congruent tetrahedrons. Hilbert proposed as a problem the proof of this assertion [15]. The proof was first given by Dehn. The simplest proof is due to V. F. Kagan [25].

The remarks made in 59. concerning a system of measurement of the areas of plane and curved surfaces hold good also with regard to a system of measurement of the volumes of solids belonging to a class of figures containing as a proper part the class of simple polyhedrons..

PART FOUR

ELEMENTS OF THE FOUNDATIONS OF GEOMETRY AND THE GEOMETRY OF LOBACHEVSKII

Chapter X

EUCLID'S "ELEMENTS"

In Chapter X is given a short survey of Euclid's "Elements", pointing out their virtues and shortcomings. Certain axioms of order and continuity are formulated, and a number of important theorems are proved without recourse to Euclid's fifth postulate on parallels.

61. A SURVEY OF EUCLID'S "ELEMENTS"

By about the fourth century before our era, geometry had so far developed that treatises giving a systematic exposition of the geometrical science of the time could make their appearance.

The first treatises of this kind of which we have historic evidence have not themselves come down to us. These works were neglected after the appearance of the celebrated treatise called "The Elements" by the Alexandrine geometer Euclid (ca. 300 B.C.).

In this extensive scholarly treatise on elementary geometry Euclid endeavored to give a rigorously logical development of geometry starting from a few principles. The Euclidean "Elements" consist of fifteen "Books", that is to say, Chapters. The last two Books are not now considered to have been written by Euclid.

Let us briefly examine the contents of Euclid's "Elements" [22], [23], [11].

Book I. The first book begins with definitions.

1. A point is that which has no part.

2. A line is length without breadth.

3. The extremities of a line are points.

4. A straight line is one which lies evenly with the points on itself.

5. A surface is that which has length and breadth only.

6. The extremities of a surface are lines.

7. A plane surface is one which lies evenly with the straight lines on itself.

There follow definitions of the angle between lines, of a straight angle, of right, obtuse and acute angles, of a disk, its center and diameter, etc.

The following definitions are worth noting:

13. A boundary is that which is the extremity of anything.

14. A figure is that which is contained by any boundary or boundaries.

15. A circle is a plane figure contained by one line [called a circumference], such that all the straight lines falling upon it from one point among those lying within the figure are equal to one another.

20. Of trilateral figures, an equilateral triangle is that which has its three sides equal; an isosceles triangle -- that which has two of its sides alone equal; a scalene triangle -- that which has its three sides unequal.

22. Of quadrilateral figures a square is that which is both equilateral and right-angled; an oblong -- that which is right-angled but not equilateral; a rhombus -- that which is equilateral but not right-angled; a rhomboid [parallelogram] -- that which has

its opposite sides and angles equal to one another, but is neither equilateral nor right-angled. And let quadrilaterals other than these be called trapezia.

Thus, Euclid uses the word trapezium (literally "table") in the sense of a quadrilateral of general form.

The final definition in the first book reads:

23. Parallel straight lines are straight lines which, being in the same plane and being produced indefinitely, do not meet one another in either direction.

The word parallel signifies drawn alongside each other.

The next section contains a list of postulates.

Postulates. Let the following be postulated:

1. [It is possible] to draw a straight line from any point to any point.

2. [It is possible] to produce a finite straight line continually in a straight line.

3. [It is possible] to describe a circle with any center and distance.

4. (Axiom 10) That all right angles are equal to one another.

5. (Axiom 11) That if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines if produced indefinitely meet on that side on which the angles are less than two right angles.

The last two postulates are designated in other lists as Axioms 10 and 11.

The words in parentheses are not in the Greek original.

We next find:

Common Notions (axioms):

1. Things equal to the same thing are equal to each other.

2. If equals be added to equals, the wholes are equal.

3. If equals be subtracted from equals, the remainders are equal.

[4. And if equals are added to unequals, the sums will be unequal.]

5. And things equal to twice the same thing are equal to each other.

6. And halves of the same thing are equal to each other.

7. Things which coincide with one another are equal to one another.

8. The whole is greater than the part.

[9. Two straight lines do not enclose (or contain) a space.]

The first six axioms are the prototypes of the axioms of arithmetic, and also the prototypes of the axioms of congruence[22].

The sentences in square brackets are those of which Euclid's authorship is doubtful.

Next come propositions 1 through 48.

In order to give an idea of the style of the "Elements", we quote in full Proposition 1 and its proof.

Proposition 1. On a given finite straight line to construct an equilateral triangle.

Let AB be the given finite straight line (fig. 249). It is required to construct on line AB an equilateral triangle. With center A and radius AB we describe the circle BCD (postulate 3), and then with center B and radius BA describe the circle ACE (postulate 3); and from point C, at which the circles intersect each other, we draw lines CA, CB joining point C with points A and B (postulate 1).

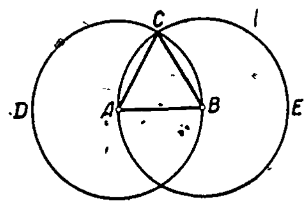


Fig. 249.

Now, since point A is the center of circle CDB, AC is equal to AB (definition 15); further, since point B is the center of circle CAE, BC is equal to BA (definition 15). But it has already been shown that CA is equal to AB; that is, CA and CB are each equal to AB. But things equal to the same thing are equal to each other (axiom 1); this means that CA is also equal to CB.

Therefore, the three lines CA, AB and BC are equal to each other. Hence, triangle ABC is equilateral (definition 20) and constructed on the given finite straight line AB, which is that which was required to do.

In order to have an idea of the nature of the first 28 propositions of the "Elements" it should be kept in mind that these 28 propositions are proved by Euclid without recourse to the fifth postulate on parallels.

Euclid further shows how to lay off, add and subtract segments (propositions 2 and 3); then follow theorems on the equality of triangles resulting from the equality of two sides and the included angle (proposition 4), on the equality of the base angles of an isosceles triangle (proposition 5) and the converse theorem (proposition 6). Proposition 8 proves the equality of triangles resulting from the equality of a side and the two adjacent angles. After the bisection of the angle and of a segment, the drawing of perpendiculars, and theorems on adjacent and vertical angles (propositions 9 - 15) there follows a theorem on the external angle of a triangle.

Proposition 16. In any triangle, if one of the sides be produced, the exterior angle is greater than either of the interior and opposite angles.

Euclid's proof is repeated in every textbook of elementary geometry.

We emphasize again that this theorem on the exterior angle of a triangle was proved by Euclid without the aid of the fifth postulate.

Then follows

Proposition 17. In any triangle two angles taken together in any manner are less than two right angles.

After the proofs of theorems asserting that in a triangle the larger side lie opposite the larger angle, and that if two sides of one triangle are equal to two sides of another while the included angles are unequal, the side lying opposite the greater angle is the longer, there follows

Proposition 27. If a straight line falling on two straight lines make the alternate angles equal to one another, the straight lines will be parallel to one another. The proof, not using the fifth postulate, is already familiar (2, fig. 2).

Proposition 28. If a straight line falling on two straight lines make the exterior angle equal to the interior and opposite angle on the same side, or the interior angles on the same side equal to two right angles, the straight lines will be parallel to one another. This proposition is a direct corollary of proposition 27 and hence the fifth postulate was not used here either.

Beginning with proposition 29 Euclid makes use of his fifth postulate on parallels:

Proposition 29. A straight line falling on parallel straight lines makes the alternate angles equal to one another ....

This is the converse of proposition 27. The remaining part of proposition 29 refers to interior angles on the same side etc. and is the converse of proposition 28.

Proposition 47 is the Pythagorean theorem. The final proposition 48 of the first book is the converse of the Pythagorean theorem.

Book II. The second book consists of the geometrical exposition of what we might call algebraic identities. For example, proposition 4 of Book II reads: If a straight line be cut at random, the square on the whole (line) is equal to the squares on the segments and twice the rectangle contained by the segments (fig. 250).

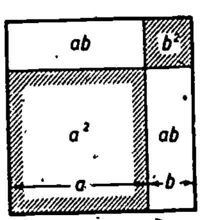


Fig. 250.

This is nothing other than the identity

$$(a + b)^2 = a^2 + b^2 + 2ab .$$

In figure 250 the modern notation is shown.

In all, Book II contains two definitions and 14 propositions. Proposition 11, for example, deals with the division of a segment into mean and extreme proportionals (the golden section).

Proposition 11. To cut a given straight line so that the rectangle contained by the whole and one of the segments is equal to the square on the remaining segment.

Proposition 14, the last of this book, reads: To construct a square equal to a given rectilineal figure.

Here he performs the quadrature of any rectilinear figure, that is, in a certain sense, of any polygon.

Book III. In the third book are set forth the basic properties of the circle, of chords, of inscribed and circumscribed angles, of tangents etc. It contains 11 definitions and 37 propositions.

Book IV. The fourth book consists of propositions relating to equilateral triangles, squares and regular pentagons, hexagons and fifteen-sided figures inscribed in and circumscribed about circles. It contains 7 definitions and 16 propositions.

Book V. In book five the arithmetic of the ancients is expounded in geometric dress. It must be remembered that fractions, not to mention irrationals, were not recognized as numbers by Euclid.

A number is a multitude composed of units -- so reads the second definition in Book VII.

Thus, for Euclid there is no one-to-one correspondence between geometric magnitudes and numbers; he did not reduce the ratio of two segments, areas or volumes to a ratio of numbers. Euclid, therefore, constructs two theories where we see one: a theory of magnitudes in Book V and a theory of numbers in Book VII.

Of particular interest in Book V is

Definition 5. Magnitudes are said to be in the same ratio, the first to the second and the third to the fourth, when, if any equimultiples whatever be taken of the first and third, and any equimultiples whatever of the second and fourth, the former equimultiples alike exceed, are alike equal to, or alike fall short of the latter equimultiples respectively taken in corresponding order.

In language more familiar to us this means:

$$\frac{a}{b} = \frac{c}{d}$$

if for all natural numbers m and n,

$$ma > nb \text{ when } mc > nd ,$$

$$ma = nb \text{ when } mc = nd ,$$

and

$$ma < nb \text{ when } mc < nd .$$

Euclid's definition is equivalent to the definition of real positive numbers. In modern form it completely corresponds to Dedekind's method of cuts [18].

The fifth book contains 18 definitions and 25 propositions.

Book VI. With the sixth book the treatment of plane geometry is concluded. In it are set forth the theory of the ratios of areas, of the similarity of figures and of proportions. For example, proposition 12 consists of the problem: To three given straight lines to find a fourth proportional.

Proposition 13. To two given straight lines to find a mean proportional.

In modern textbooks these and similar propositions serve as the basis of the application of algebra to geometry, and yield the construction of the segments $x = \frac{ab}{c}$, $x = \sqrt{ab}$, etc.

It is with just these constructions of Euclid that Descartes begins his Geometry (1635) [19], giving to them a numerical meaning.

The sixth book contains five definitions and 33 propositions.

Books VII, VIII and IX are devoted to the theory of whole numbers given in partly geometric form. Euclid represents numbers by segments.

In Book VII, containing 23 definitions and 39 propositions, we find the definition of a prime number.

Definition 12. A prime number is that which is measured by a unit alone.

In the same book is set forth the familiar method of finding the greatest common divisor of two whole numbers -- the algorithm of Euclid.

The relevant theorems are these:

Proposition 1. Two unequal numbers being set out, and the less being continually subtracted from the greater, if the number which is left never measures the one before it until a unit is left, the original numbers will be prime to one another.

Proposition 2. Given two numbers not prime to one another, to find their greatest common measure.

Today the algorithm of Euclid is presented by the system of equations

$$a = bq + r_1$$

$$b = r_1q_1 + r_2$$

$$r_1 = r_2q_2 + r_3$$

$$\cdots\cdots\cdots$$

$$r_{n-3} = r_{n-2}q_{n-2} + r_{n-1}$$

$$r_{n-2} = r_{n-1}q_{n-1} + r_n$$

Proposition 1 states that if $r_n = 1$, then a and b are numbers prime to each other. If $r_n = 0$, then by proposition 2 the greatest common denominator of numbers a and b is r_{n-1}.

Also expounded in the seventh book is a theory of divisibility, leading to a fundamental theorem on the uniqueness of the decomposition of a number into prime factors. The following are some of the relevant theorems:

Proposition 23. If two numbers be prime to one another, the number which measures the one of them will be prime to the remaining number.

In modern notation: if $(a,b) = 1$ and $a = mc$, then $(c,b) = 1$ [12].

Proposition 24. If two numbers be prime to any number, their product also will be prime to the same number. Or, if $(a,c) = 1$ and $(b,c) = 1$, then $(ab,c) = 1$.

After several more propositions which we may express as follows:

If $(a,b) = 1$, then $(a^n,b^n) = 1$;

If $(a,b) = 1$, then $(a + b, b) = 1$ and $(a + b, a) = 1$;

If p is a prime number and not a divisor of a, then $(p, a) = 1$, there follows the fundamental proposition in the theory of divisibility,

Proposition 30. If two numbers by multiplying one another make some number, and any prime number measures the product, it will also measure one of the original numbers.

This means: if $ab = np$, where p is a prime number, then either $a = m_1p$ or $b = m_2p$, that is, if a prime number is a divisor of the product of two numbers it is a divisor of at least one of the factors.

Then follow the theorems:

Proposition 31. Any composite number is measured by some prime number.

Proposition 32. Any number either is prime or is measured by some prime number.

From these propositions follow the uniqueness of the decomposition of a number into prime factors.

Book VIII consists of 27 propositions and no definitions. It contains, in particular, a theorem asserting in effect that there exists no fraction the square of which equals a whole number, and an analogous theorem regarding the cube. Euclid expresses these theorems in his characteristic geometrical form.

Book IX gives a proof of a theorem asserting that the set of prime numbers is infinite:

Proposition 20. Prime numbers there are more than any assigned multitude of prime numbers.

In the ninth book there are 36 propositions and no definitions.

Book X presents, in four definitions and 115 propositions, a theory of irrationals and a classification of some irrationals. This book is the most difficult in Euclid's work for the student. In the first eighteen propositions of Book X Euclid deals with the commensurability and incommensurability of magnitudes. The remainder of the book contains a theory of what we would now call the irrationals involved in the solution of quadratic and biquadratic equations.

We shall quote as examples the text of the first definition and part of the third, as well as several propositions from Book X.

Definition 1. Those magnitudes are said to be commensurable which are measured by the same measure, and those incommensurable which cannot have any common measure.

Definition 3. ... Let then the assigned straight line be called rational, and those straight lines which are commensurable with it, whether in length or in square or in square only, rational but those which are incommensurable with it, irrational.

Proposition 1. Two unequal magnitudes being set out, if from the greater there be subtracted a magnitude greater than its half and from that which is left, a magnitude greater than its half, and if this process be repeated continually, there will be left some magnitude which will be less than the lesser magnitude set out.

Proposition 2. If when the less of two unequal magnitudes is continually subtracted in turn from the greater, that which is left never measures the one before it, the magnitudes will be incommensurable.

Proposition 5. Commensurate magnitudes have to one another a ratio which a number has to a number.

It must be remembered that for Euclid irrational numbers did not exist.

Proposition 6. If two magnitudes have to one another the ratio which a number has to a number, then these quantities will be commensurable.

Proposition 7. Incommensurable magnitudes have not to one another the ratio which a number has to a number.

This last fact is indeed the reason for the appearance of the arithmetical books (as we should now call them) of the "Elements".

Proposition 8. If two magnitudes have not to one another the ratio which a number has to a number, the magnitudes will be incommensurable.

Books XI, XII and XIII deal with solid geometry.

Book XI contains 31 definitions and 40 propositions. Here we find the definition of a solid sphere, a cone, a cylinder and, in particular, of the five regular polyhedrons. The book contains propositions dealing with the relative positions of straight lines, of planes, and of lines and planes in space. It concludes with an investigation of the properties of equivalent parallelepipeds.

Book XII consists of 18 propositions and deals with equivalent prisms and pyramids, concluding with a proposition on the ratio of the volumes of spheres. In particular, we find the following propositions in this book:

Proposition 5. Pyramids having triangular bases and identical altitudes have the same ratio as their bases.

Proposition 7. Every prism which has a triangular base is divided into three pyramids equal to one another which have triangular bases.

Book XIII, in 18 propositions, gives the elements of the theory of regular polyhedrons.

We shall cite some examples of the propositions in this book(1):

Proposition 13. To inscribe a tetrahedron in a given sphere.

Proposition 14. To inscribe an octahedron in a given sphere.

Proposition 15. To inscribe a cube in a given sphere.

Proposition 16. To inscribe an icosahedron in a given sphere.

Proposition 17. To inscribe a dodecahedron in a given sphere.

Proposition 18. To find the sides of the five bodies mentioned in the foregoing propositions.

These propositions give methods of constructing the spatial figures and proofs of their correctness.

Euclid's remark on proposition 1 of Book XIII is worth noting:

"Remark. A proposition is proved analytically if, we take that which is sought as known, and by means of consequences derived from

(1) M. E. Vashchenko-Zakharchenko, "Euclid's Elements, with an explanatory Introduction and interpretations", Kiev 1880. 747 pages. In 82 pages of the Introduction are set forth the foundations of Lobachevskian geometry and an account of its development up to that time. This translation of Euclid had a great influence on the Russian mathematicians occupied with the foundations of geometry. No less an influence was exerted by the introduction to the book. In Vashchenko-Zakharchenko's translation Books VII, VIII and IX do not appear. These had been translated by F. Petrushevskii in 1835. He had in fact, translated the first six books and XI and XII in 1819. Translations of Euclid into Russian began to be published in the first half of the eighteenth century. All the books of Euclid have now appeared in a new translation by D. D. Mordukhai-Boltovskii (1948-1949-1950) [11], [22], [23].

it arrive at known truths. On the other hand, a proposition is proved synthetically if with the aid of known truths we arrive at that which is sought."

Then follows an example of an analytic and a synthetic proof of the same proposition.

Book XIV contains seven propositions in which are set forth the properties of regular polyhedrons. Like Book XV, Book XIV is considered by some historians not to be the work of Euclid; they ascribe it to Hypsicles, the Alexandrian.

Book XV contains seven propositions dealing with the problem of inscribing one regular polyhedron in another. For example:

Proposition 5. To inscribe a dodecahedron in a given icosahedron.

With this we conclude our short survey of this celebrated book.

It must be remembered that the "Elements" do not exhaust all that was known to the science of geometry in Euclid's day. The "Elements" were not the only work even by Euclid himself. Although history has not preserved them to us, he also wrote four books "On Conic Sections" two books "On Surfaces", "Porisms" and "On False Conclusions".

The theory of conic sections was developed by Apollonius of Perga (ca. 200 B.C.). Seven books of his "On Conic Sections" have been preserved, but many of his works are lost.

The great mathematician of antiquity Archimedes (killed 212 B.C.) wrote the treatises "On the Sphere and the Cylinder", "The Measurement of the Circumference of the Circle", "On Spirals", "On Conoids and Spheroids", "On the Equilibrium of Plane Figures", "The Quadrature of the Parabola" and others (1).

(1) All the surviving works of Archimedes are available in Russian. Gostekhizdat is preparing for the press a new Complete Collected Works of Archimedes.

62. THE HISTORIC SIGNIFICANCE AND THE MERITS AND SHORTCOMINGS OF THE "ELEMENTS"

From the foregoing survey of Euclid's "Elements" it can be seen what an enormous influence this work had upon the development of the mathematical sciences.

When we consider that secondary school geometry is a restatement in somewhat modernized form of the "Elements", we recognize the significance of this book in general culture as well.

It must also be emphasized that the geometry of Euclid, including its more advanced branches which received their development at the end of the eighteenth and in the nineteenth century, is a part of the foundation of astronomy, of mechanics, of many branches of physics and consequently also of technology.

It is sufficient to glance through any textbook of the statics of structures, the dynamics of machines, electrotechnics, radio technology, optics, gas and hydrodynamics -- particularly airfoil and airscrew theory -- and many other theoretical and applied disciplines in order to form an idea to what extent modern life is permeated by the facts set forth in the "Elements".

The "Elements" formed part of the education of Copernicus, of Galileo, Descartes, Newton, Lomonosov, Lagrange, Lobachevskii, Ostrogradskii, Chebyshev, Liapunov, Zhukovskii [Joukowsky] -- and it may be said, all the great mathematical scholars of past centuries or of our day.

A different view of geometric science and the development of a new geometry, different from the Euclidean, came from the great Russian mathematician Nikolai Ivanovich Lobachevskii. The two eras of the history of geometry are from Euclid to Lobachevskii and from Lobachevskii onward.

In order to arrive at an understanding of geometry in its development after the work of Lobachevskii, as well as the work of Lobachevskii himself, let us briefly consider the merits and shortcomings of the "Elements".

Without question we must cite as one of the fundamental merits of the "Elements" that it formulated the question of the logical structure of geometric science.

For centuries the "Elements" were considered the model of logical perfection. While this claim is not true for every line of Euclid's text, much of his work is indeed irreproachable. With respect to logic the significance of the "Elements" lay in formulating a program for the rigorously logical development of the science of geometry. This development can not be said to have been completed until the beginning of the twentieth century, and in it the new geometry founded by Lobachevskii played a vital part. The immense practical importance of the "Elements", tested by two millenia, is their decisive merit.

Passing to the shortcomings, we note first of all that the "definitions" of basic concepts in Euclid are not outstanding for their clarity, and indeed often define nothing. Many "definitions" in the first book, e.g. those of a point, a curve, a straight line, and others, are not logically operative in the remainder of the work; that is, Euclid does not make use of these "definitions" in the proofs of theorems. He does not make such references to them as, say, "but since a point is that which has no parts..." or "but since a line is length without breadth, it follows from this..." These definitions can be removed from the text, and the proofs lose neither their meaning nor their truth, insofar as they are actually logical.

Again, how are we to understand definition 4: "A straight line is one which lies evenly with the points on itself"? The sense of this "definition" is obscure. Would not the circumference of a circle satisfy it? It is sometimes claimed that in saying "lies evenly with the points on itself", Euclid had in mind all the properties of a line including direction [22]. There is no general agreement on this matter.

On the subject of such "definitions" Lobachevskii writes:

"Thus, regardless of their great antiquity and regardless of all our subsequent brilliant successes in mathematics, the Euclidean "Elements" preserve down to our day their primeval shortcomings.

"Indeed, all must agree that no mathematical science should start from such obscure notions as those which we, repeating Euclid, begin geometry [32]".

In the second half of the nineteenth century it was noticed that Euclid had no axioms of position. Such concepts as "lying between", "lying inside", "outside" were not given any logical treatment by Euclid, but were used by him on the basis of intuitive not precisely determined notions. There is for example, in Euclid, no axiom such as the one formulated by Pasch.

Pasch's postulate. Let A, B and C be three points which are not collinear, and let a be a line in the plane ABC, not passing through any of these three points; then if the line a passes through a point of the segment AB, it must pass through a point of segment AC or else through a point of segment BC.

This axiom turns out to be a basic logical tool for operations with the ideas of between, within, outside, etc.

Let us consider some theorems indispensable to us for what will follow. In their proofs we shall not make use of the axiom on parallels.

Theorem. If in the quadrilateral ABCD having two right angles A and B (fig. 251) the sides AD and BC are such that $AD > BC$, $AD = BC$ or $AD < BC$, then the angles $\gamma = \angle C$ and $\delta = \angle D$ have in the respective cases the relationships $\gamma > \delta$, $\gamma = \delta$ and $\gamma < \delta$.

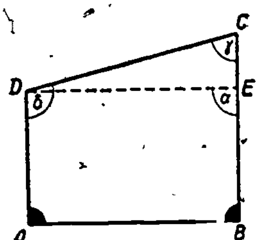


Fig. 251.

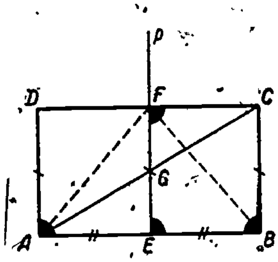


Fig. 252.

Proof. To begin with let $AD = BC$ (fig. 252). We shall show that angles D and C are equal. Let point E bisect AB and let a line p perpendicular to AB be erected at E. This perpendicular intersects line CD. This fact is demonstrated as follows: If we join points A and C and apply Pasch's postulate to the triangle ABC and line p we find that:

1) the line p does not pass through any of the points A, B or C,

and 2) it does not intersect side BC, since two perpendiculars to the same line do not meet; consequently p intersects AC in a point G, lying between A and C.

We then apply the same postulate to the triangle ACD and line p. We find: firstly, p and AD do not intersect; secondly, p does not pass through any of the points A, C or D. Consequently, line p intersects segment CD in a point F lying between C and D.

We shall now show that point F bisects segment CD.

From the equality of triangles AEF and BEF we find that

$$AF = BF \text{ and } \angle FAE = \angle FBE ,$$

whence it follows that

$$\angle FAD = \angle FBC,$$

and consequently, that the triangles ADF and BCF are congruent. From this we find that

$$\angle D = \angle C \text{ and } DF = FC.$$

Furthermore, from the equality of the triangles in question it follows that $\angle$ EFC is equal to its adjacent angle EFD and consequently that line p intersects the upper base CD of quadrilateral ABCD at right angles.

We have shown that angles C and D are equal. But that these angles are right angles can be proved only with the aid of Euclid's postulate on parallels.

Now let AD < BC (fig. 251)

We lay off on BC the segment BE = AD and join D and E by the segment DE. (The student should prove that segment DE lies within the angle AD .)

By the preceding we find:

$$\alpha = \angle BED = \angle ADE < \delta .$$

But α is an exterior angle of triangle DEC; consequently,

$$\delta > \alpha > \gamma .$$

In the same way we can show that for $AD > BC$ we shall have $\gamma > \delta$.

Since the three relationships

$$AD < BC, \quad AD = BC, \quad AD > BC$$

are mutually exclusive, there follows the validity of the

Converse theorem. If in the quadrilateral ABCD having two right angles A and B (fig. 251) angles δ and γ are such that $\delta > \gamma$, $\delta = \gamma$ or $\delta < \gamma$, then the sides BC and AD have in the respective cases the relationships $BC > AD$, $BC = AD$ and $BC < AD$.

Remark. The quadrilateral ABCD with right angles at A and B and the equal sides AD and BC is known as Saccheri's quadrilateral. We have demonstrated that in Saccheri's quadrilateral the angles at the upper base are equal and that a perpendicular erected on the lower base at its midpoint bisects the upper base and is perpendicular to it.

A substantial shortcoming of the "Elements" lies in the absence of any axioms on continuity. Such axioms were introduced into geometry only in the second half of the nineteenth century. The logical treatment of concepts and propositions dealing with continuity was absent not only from the "Elements" but from all geometric treatises until the introduction of Dedekind's axiom on continuity (1872), or of its equivalent the axiom of Archimedes and together with the axiom of Kantor. The latter two we already know. We shall now cite the text of Dedekind's axiom [18].

Axiom on continuity (of Dedekind). If all the points of a line are divided into two classes in such a manner that each class contains points and that all the points of one class lie on the same side of every point of the second class, then there exists one point dividing the line into two rays upon one of which lie all points of the first class and upon the other all points of the second class, while the point dividing the line into two rays belongs either to the first class or to the second. This point is said to define a Dedekind section of the line.

Using the axiom of Dedekind it is possible to prove the axioms of Archimedes and Kantor as theorems, but for this the whole list of the axioms of geometry is necessary. Conversely, from the axioms of Archimedes and Kantor there follows the validity of Dedekind's axiom as a theorem. The equivalence previously mentioned consists in the truth of these direct and converse assertions, which we shall demonstrate later.

The first proposition in the "Elements" already contains a most serious gap. The existence of the point of intersection C of the two circles (with centers at A and B and having radius AB, in fig. 249) is assumed without justification. Until an axiom on continuity is introduced, it is impossible to assert that the circumferences in question intersect. It might turn out that at the location of point C there occurred in one (or both) of the circumferences a gap, and that the two circles had no common point (fig. 253).

Fig. 253.

It has also been noted long since that some of Euclid's axioms could be proved. However, to proceed with such proofs, we should have a complete list of basic axioms. Otherwise, it remains unclear upon what the proof is based.

Thus, for example, the fourth postulate -- on the equality of all right angles -- began to be the object of proofs already in the eighteenth century, but in this it was intuitively assumed that angles constituted a system of continuous magnitudes. At the turn of the twentieth century Hilbert, in his "Foundations of Geometry", gave a proof independent of the axioms of continuity.

In like manner, efforts were also made to prove the fifth Euclidean postulate on parallels. Attempts to prove this postulate indeed go well back into antiquity. Euclid himself was at pains to postpone the use of the fifth postulate as faraas possible. As we have seen, the proofs of the first 28 propositions of the "Elements" were not based on this postulate.

It would be possible to cite a considerable list of additional propositions from plane and solid geometry which are or could be proved without the postulate on parallels.

Concerning the attempts of more than two thousand years to prove the postulate on parallels, Lobachevskii wrote in 1823:

"To the present day they have not succeeded in finding a rigorous proof of this truth. Such as have been given can only be called explanations, but do not deserve to be esteemed Mathematical proofs in the full sense"[30].

In Lobachevskii's work "On the Elements of Geometry", published in 1829 consisting of extracts from a "lecture read at the session of the Department of Physico-Mathematical Sciences 11th February (Old Style) 1826" [32], in which were set forth the principles of the new geometry created by Lobachevskii, we read:

"...it is impossible to tolerate anywhere in Mathematics such a lack of rigor as we have been constrained to admit in the theory of parallel lines. It is true that we are guarded against false conclusions due to the vagueness of primary and general concepts in Geometry by the conception of the objects themselves in our imaginations; and we have convinced ourselves without proof of the correctness of the received truths by their simplicity and by experience, for example, by astronomical observations. But all this cannot satisfy the intellect trained to rigorous judgment. Indeed that intellect has not the right to neglect finding the answer to a question as long as it remains unknown and [illegible] long as we do not know that it might not lead us to further results.

"It is my intention here to explain in what manner I propose to fill up these gaps in Geometry. To set forth all my researches in proper order and connection would require too much space and the presentation of the whole science of geometry in an entirely new form."

Chapter XI

THE GEOMETRY OF LOBACHEVSKII

In Chapter XI, after a theorem on the sum of the angles of a triangle, various forms of the axiom on parallels in Euclidean geometry are examined.

After stating the axiom of Lobachevskii, the chapter presents a considerable number of the simplest facts of Lobachevskian geometry, and the behavior of a line in the Lobachevskian plane is studied in detail. At the end of the chapter some idea is given of the measurement of area in Lobachevskian geometry.

63. THEOREM ON THE SUM OF THE ANGLES OF A TRIANGLE

Following his "On the Principles of Geometry" (1829), Lobachevskii published in 1835 the article "Imaginary Geometry". This work begins with the words:

"Proposition XII [XI] of Euclid's "Elements" is accepted in Geometry as a palpable truth which mathematicians have for two thousand years vainly labored to prove rigorously. Legendre in particular occupied himself with this problem and in the Memoires of the French Academy he collected everything which seemed to him to be fairly satisfactory." (Here Lobachevskii cites the relevant work by Legendre, published in 1833.) "All who thought to have found an answer to this difficult question without exception fell into error, because they were convinced beforehand of the correctness of that which cannot yet follow directly from our conceptions of bodies without the aid of observation, as I believe I have shown unquestionably in my treatise "On the Principles of Geometry". After expounding a new theory of parallels, I asserted that it is permissible, independently of actual measurement, to assume that the sum of the angles of a rectilinear triangle is less than half a circle, and to found upon this assumption a new geometry, which I called imaginary and which, even if it does not exist in nature, should at least be accepted in Analysis.'

With the help of geometric constructions alone, equations were derived which express the functional relationships of the sides and angles of a rectilinear triangle; finally, expressions were given for the elements of a line, a surface and the volume of a solid, so that imaginary geometry as a new branch of Mathematical science was embraced to its full extent, so that no further doubt might remain as to the correctness and adequacy of it's principles."

In 1832 the Hungarian mathematician Janos Bolyai (1802 - 1860) published a remarkable work in Latin as an appendix to a book by his father Farkas Bolyai. The full title of the son's work reads: "An Appendix, containing the absolutely true science of space, independent of the truth or falsity of Euclid's XI axiom (which a priori can never be decided), to which is added, for the case of its falsity, a geometric squaring of the circle" [10].

In this work Janos Bolyai independently of Lobachevskii, but some years later, set forth in extremely terse form the fundamentals of that non-euclidean geometry, the ideas of which N. I. Lobachevskii submitted on February 23, 1826 to the session of the Physico-Mathematical Faculty of Kazan University. In 1839 appeared "On the Principles of Geometry, in which the great Russian savant first published his discovery.

In order to convey the fundamentals of this new geometry, now called the non-euclidean geometry of Lobachevskii, we shall, partially following Lobachevskii himself [31], set forth some important propositions on the sum of the angles of a triangle. These will be proved without recourse to the fifth postulate on parallels.

Proposition 1. In a rectilinear triangle the sum of the three angles cannot exceed two right angles.

Proof. Let us assume that in triangle ABC (fig. 254) -

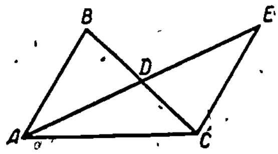


Fig. 254.

the sum of the three angles is equal to $\pi + \alpha$ (where π denotes two right angles); if its sides are not equal, let BC be the shortest. We bisect BC at D, draw the line AD and beyond D take a point E on this line such that DE = AD. We join E and C. In the equal triangles ADB and CDE, angle ABD = angle DCE and angle BAD = angle DEC. It follows that in triangle ACE the sum of the angles must also be equal to $\pi + \alpha$.

The smallest angle BAC of triangle ABC has gone over into the new triangle AEC, divided into two parts:

$$\angle BAC = \angle EAC + \angle AEC .$$

One of these parts is not greater than one-half of angle BAC. Continuing in this way, we shall be able to construct a triangle, the sum of whose angles must be $\pi + \alpha$, in which will be found two such angles, each less than $\frac{1}{2}\alpha$; since, however, the third angle cannot be greater than π, it follows that α cannot be greater than zero.

<u>Proposition 2.</u> <u>If in any rectilinear triangle the sum of the three angles is equal to two right angles, then this must also be the case in any other triangle.</u>

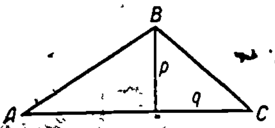


Fig. 255.

<u>Proof.</u> Let us assume that in triangle ABC (fig. 255) the sum of the angles is equal to π. At least two angles of this triangle are acute. From the vertex B of the third angle we let fall a perpendicular p to the opposite side. Triangle ABC is decomposed into two right triangles,(1) in each of which the sum of the angles must likewise be equal to π, since if in one of these right triangles the sum of the angles is less than π, in the other the sum of the angles will be greater than π, which is impossible.

Thus we obtain a right triangle with legs p and q and an angle sum equal to π. From this triangle we obtain a quadrilateral having its opposite sides equal and right angles (fig. 256). By repeated opposition we can obtain from this quadrilateral a quadrilateral whose opposite sides are equal and each of whose angles are right angles.

(1) Since the aim of the author is not to write a systematic treatise, but to acquaint the students with important geometric ideas, the author does not undertake to justify every step in the proofs by reference to a basic set of axioms. Neither does he attempt in this chapter to show that every step in the proof may be justified without the use of Euclid's fifth postulate. Thus, for instance the author merely states but does not prove the fact that the acuteness of angles A and B in fig. 255 implies that triangle ABC is divided by p into two parts. —Translators.

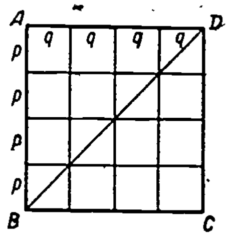


Fig. 256.

The sides of this quadrilateral can be made equal to np and mq, i.e.

$$AB = np \text{ and } AD = mq ,$$

m and n being arbitrary natural numbers (fig. 256). This quadrilateral is divided by the diagonal BD into two equal right triangles BAD and BCD, in each of which the sum of the angles is equal to π.

Thus, if in any one triangle the sum of the angles is equal to π, it is always possible to construct such a right triangle, with legs as large as desired and with the sum of its angles equal to π.

It is now easy to show that in every right triangle PQR the sum of the angles is equal to π if in any one triangle ABC the sum of the angles is equal to π.

Let us construct the right triangle P_1RQ_1, with angle sum equal to π and legs RP_1 and RQ_1 respectively longer than the corresponding legs RP and RQ of triangle PQR (fig. 257).

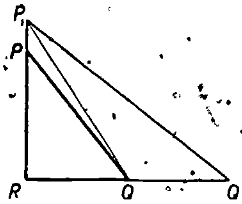


Fig. 257.

Drawing segment P_1Q, we note (using the same reasoning as for fig. 255) that in right triangle P_1RQ the sum of the angles is likewise equal to π. On the same grounds the sum of the angles of right triangle PRQ is also equal to π.

Since any triangle $A_1B_1C_1$ can be decomposed into two right triangles by an altitude drawn from the vertex of its largest angle, it follows that <u>the sum of the angles of any triangle will be equal to π if it is equal to π in some one triangle</u>. The theorem is proved.

From the two theorems just proved there follows

<u>Proposition 3.</u> <u>There are only two possibilities: either the sum of the angles is equal to π in every rectilinear triangle, or it is less than π in every triangle.</u>

<u>Proposition 4.</u> <u>If the sum of the angles is identical in all rectilinear triangles, then that sum is equal to π.</u>

<u>Proof</u>. Let us denote the sum of the angles of every triangle by x.

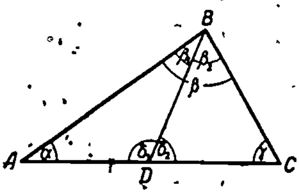


Fig. 258.

In an arbitrary triangle ABC we draw the transversal BD (fig. 258). Using for the angles the notation shown in the diagram, we have:

$$x = \alpha + \beta + \gamma,$$

$$x = \alpha + \beta_1 + \delta_1,$$

$$x = \gamma + \beta_2 + \delta_2.$$

Adding the last two equations, we obtain

$$2x = \alpha + (\beta_1 + \beta_2) + \gamma + (\delta_1 + \delta_2),$$

or,

$$2x = \alpha + \beta + \gamma + \pi,$$

whence

$$x = \pi, \qquad \text{q.e.d.}$$

From this theorem follows

Proposition 5. If for every triangle the sum of the angles is less than π, then this sum cannot be identical for all triangles. In other words, this sum differs from triangle to triangle.

We shall now prove an important lemma.

Lemma. From a given point it is always possible to draw a straight line in such a manner that it forms with a given straight line an arbitrarily small angle.

From the given point A we drop upon the given line a the perpendicular AB (fig. 259). We take an arbitrary point B_1 on a and join it with A. We denote angle AB_1B of right triangle ABB_1 by α. In the direction BB_1 we lay off from point B_1 the segment B_1B_2 equal to segment AB_1 and draw AB_2.

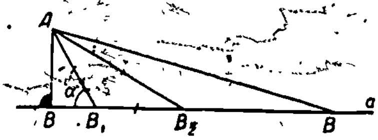


Fig. 259.

In the isosceles triangle AB_1B_2, angle AB_2B_1 must be either equal to or less than $\frac{1}{2}\alpha$, depending upon what is the possible sum of the angles of a triangle. Continuing in this manner we finally arrive at an angle AB_nB which is less than any given angle.

Proposition 6. If through a given point it is possible to draw only one parallel to a given line, then the sum of the angles of any triangle is equal to π. The proof is familiar.

Proposition 7. If the sum of the angles of a triangle is equal to π, then through a given point only one parallel can be drawn to a given line.

Proof. Let A be the given point not lying on the given line a. We draw AB perpendicular to a and Ab perpendicular to AB (fig. 260).

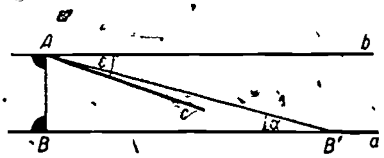


Fig. 260.

Lines a and b cannot meet without forming a triangle the sum of whose angles is greater than π, and are therefore parallel. Let us assume that line c is also parallel to line a. Let the acute angle be angle BAc, equal to $\frac{\pi}{2} - \epsilon$. We take point B' on line a such that angle $AB'B = \alpha$ is less than angle ϵ (by the lemma) and so that point B' lies on the same side of line AB as the acute angle cAB.

The sum of the angles of right triangle ABB' being equal to π by assumption, it follows that

$$\angle B'AB = \frac{\pi}{2} - \alpha > \frac{\pi}{2} - \epsilon = \angle cAB$$

since $\alpha < \epsilon$. Whence it follows that line Ac, entering triangle ABB' through vertex A, intersects the base BB', i.e. line a. The contradiction thus obtained proves the theorem. Line b is the only line passing through point A and parallel to line a.

Proposition 8. If the sum of the angles of a triangle is less than π, then it is possible to draw through a given point more than one line which does not intersect a given line.

Proof. Let the sum of the angles of a triangle be less than π. The assumption that through a given point A only one line can be drawn parallel to a given line a leads to the conclusion that the sum of the angles of a triangle is equal to π (proposition 6), that is, to a contradiction.

Thus we have the following possibilities or assumptions:

A. It is possible to draw only one parallel to a given line through a given point.

B. It is possible to draw more than one parallel to a given line through a given point.

Under the first assumption the sum of the angles of a triangle equals π; under the second it is less than π.

Conversely, statements A and B follow respectively from the corresponding assumptions about the sum of the angles of a triangle.

"The first assumption", wrote Lobachevskii [31], "serves as the basis of ordinary geometry and plane trigonometry.

"The second assumption may equally well be adopted without leading to any contradictory results; it underlies the new geometrical theory to which I have given the name 'imaginary geometry' and of which I here intend to carry the exposition as far as the derivation of equations for the relationships between the sides and angles of rectilinear and spherical triangles."

64. DIVERSE FORMS OF THE AXIOM OF PARALLELISM IN EUCLIDEAN GEOMETRY

The foregoing investigation shows that the axiom on parallels may be replaced by a proposition on the sum of the angles of a triangle. Namely, if we accept without proof that the sum of the angles of any one triangle is equal to π, the axiom on parallels will follow as a theorem.

There exist any number of such propositions capable of replacing the axiom on parallels. We shall cite a number of them which are of importance for the further exposition of Lobachevskian geometry.

First of all, from the axiom on the uniqueness of parallels there follows the fifth postulate of Euclid, and conversely.

In proof: Let there pass through point A a line b, the unique parallel to the line a (fig. 261).

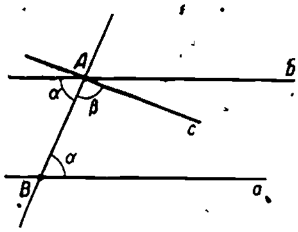


Fig. 261.

Since a transversal forms equal alternate interior angles with two parallels, and since b is the only parallel to a passing through A, line AB forms with lines b and a equal alternate interior angles. Consequently, any line c forming with the transversal AB an angle β such that $\alpha + \beta < \pi$ intersects line a. Moreover, since the sum of the angles of a triangle cannot be greater than π, it intersects it on that side of AB on which the sum of the interior angles on the same side is less than π. The converse proposition can also readily be proved.

The proposition asserting that a perpendicular and an inclined line to the same line always intersect can also take the place of the axiom on parallels.

The student should prove independently that it is sufficient to require the uniqueness of the parallel through some one given point A to some one given line a. The uniqueness will then follow for any line and any point.

We know from secondary school geometry that if the fifth postulate of Euclid is correct, then similar triangles exist.

The converse proposition allows us to replace the fifth postulate by the assertion of the existence of two unequal similar triangles.

Theorem (of Wallis). If there exist two similar but unequal triangles, then Euclid's postulate on parallels follows.

Proof. In triangles ABC and A'B'C' let $\angle A = \angle A'$, $\angle B = \angle B'$, $\angle C = \angle C'$ and $AB > A'B'$. We lay off on sides AB and AC of triangle ABC the segments $AD = A'B'$ and $AE = A'C'$ (fig. 262).

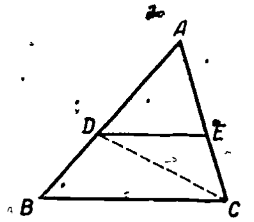


Fig. 262.

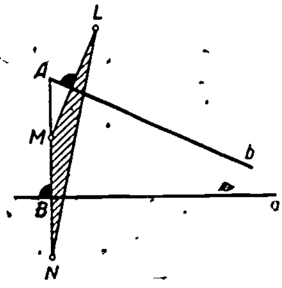


Fig. 263.

Since $AD = A'B' < AB$, point D falls between points A and B. If point E were to coincide with point C, angle ACD, equal to angle AED ($= \angle A'C'B'$) would be less than angle C of triangle ABC. In the same way if point E were to lie outside segment AC, then here also angle AED would be less than angle C. But angle AED is equal to angle C; consequently point E lies between points A and C.

But then the sum of the angles in quadrilateral BCED is equal to 2π. Consequently, in each of the triangles BDC and DCE the sum of the angles is equal to π, since in neither triangle can the sum of the angles be greater than π.

But from the equality of the sum of the angles of even one triangle to π there follows the fifth postulate. The theorem is proved.

It is clear from the above how closely the theory of similarity is bound up with the axiom on parallels.

Theorem. If the fifth postulate of Euclid is true, then a circle may be circumscribed about any triangle.

The proof is familiar from secondary school geometry.

Converse theorem (of F. Bolyai). If it is assumed that a circle can be circumscribed about any triangle, the fifth postulate of Euclid follows.

Proof. Let it be supposed possible to circumscribe a circle about every triangle. We shall show that a perpendicular aB and an inclined line bA to a line AB invariably intersect (fig. 263), which is equivalent to the fifth postulate. We take on segment AB an arbitrary point M and find the points L and N symmetrical with point M about lines b and a respectively. Since line ML is perpendicular and line MA is inclined (1) to line b, the lines MA and ML are distinct, and consequently the three points L, M and N are not collinear. Since line a is the locus of points equidistant from vertices N and M of triangle LMN, the center O of the circle circumscribed around this triangle must lie on line a. Since line B consists of points equidistant from vertices M and L, this center O must lie on line b also. Hence lines a and b intersect, q.e.d.

Thus the proposition that a circle may be circumscribed about any triangle can take the place of the fifth postulate on parallels.

Theorem. If Euclid's fifth postulate is true, then three points equidistant from a given line and lying on the same side of this line are collinear.

The proof is familiar from secondary school geometry.

Converse theorem. If it is assumed that three points lying on the same side of a given line and equidistant from it are collinear, Euclid's fifth postulate follows.

Proof. Let A, B, C be three points lying on the same side of a line a and such that

$$AA_1 = BB_1 = CC_1$$

where A_1, B_1 and C_1 are the feet of the perpendiculars from points A, B and C to line a (fig. 264).

(1) For, if lines MA and ML were both perpendicular to b they would be parallel to each other. (This theorem is independent of the fifth postulate.)

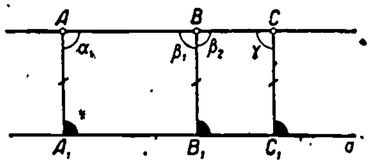


Fig. 264.

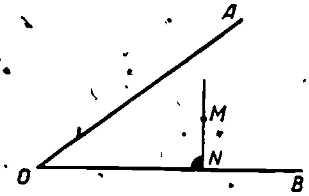


Fig. 265.

The quadrilaterals A_1B_1BA, A_1C_1CA and B_1C_1CB are Saccheri quadrilaterals, and consequently angles α_1, β_1; α_1, γ; β_2, γ at their upper bases will be related as follows:

$$\alpha_1 = \beta_1 ; \alpha_1 = \gamma; \beta_2 = \gamma .$$

From these equalities it follows that

$$\beta_1 = \beta_2 .$$

But equal adjacent angles are right angles; consequently, in quadrilateral A_1B_1BA all the angles are right angles. From this we find that in each of the triangles A_1B_1B and A_1BA the sum of the angles is equal to π, whence the fifth postulate follows.

This theorem shows that the assumption that two parallel lines are at a uniform distance from each other is simply another form of the axiom on parallels.(1)

Now let there be given an acute angle AOB (fig. 265) and a point M lying within it. Dropping a perpendicular MN to the side OB, we see that the perpendicular NM and the line OA inclined to the line OB always intersect if we accept the fifth

(1) For, let line b be parallel to a and let A,B,C be three points on line b. Then these points are equidistant from a as well as collinear. The student will now observe that the hypothesis in the theorem need only be satisfied for one specific case for the conclusion to hold. --Translators.

postulate. This result may be formulated in another way:

Theorem. If the fifth postulate of Euclid is true, then through any point lying within an arbitrary angle it is always possible to draw a line which will intersect both sides of the angle.

Here the condition that the angle be acute may be discarded. If the angle is obtuse, we need only draw its bisector, a perpendicular to which will then intersect both sides of the given angle.

Converse theorem. If it is assumed that through any point M lying within any given angle AOB it is always possible to draw a line which will intersect both sides of the angle, then Euclid's postulate on parallels will follow.

The proof is by contradiction. Let us assume that through any point M lying within the angle a line can be drawn intersecting both sides of the angle, but that the fifth postulate is false.

Since the fifth postulate is false by assumption, the sum of the angles of any triangle ABC is less than π. We shall denote by $\delta(ABC)$ the difference between π and this angle-sum:

$$\delta(ABC) = \pi - (A+B+C),$$

where $\delta > 0$.

This difference $\delta(ABC)$ is called the defect of this triangle. Let A be an angle of triangle ABC not smaller than either of the other two.

We construct the point A', symmetrical to point A with respect to line BC (fig. 266).

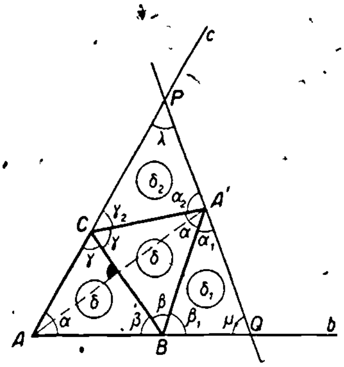


Fig. 266.

By our assumption with regard to angle A, point A' will lie within angle cAb. We draw through point A' the line PQ <u>intersecting</u> both sides ACc and ABb of angle cAb in the points P and Q.

Joining point A' with points B and C and using the notation shown in fig. 266 for the angles, we find

$$\delta(ABC) = \delta = \pi - (\alpha + \beta + \gamma),$$
$$\delta(BCA') = \delta = \pi - (\alpha + \beta + \gamma),$$
$$\delta(A'BQ) = \delta_1 = \pi - (\alpha_1 + \beta_1 + \mu),$$
$$\delta(A'CP) = \delta_2 = \pi - (\alpha_2 + \gamma_2 + \lambda).$$

Adding these equations, we obtain

$$2\delta + \delta_1 + \delta_2 = 4\pi - (\gamma + \gamma + \gamma_2) - (\beta + \beta + \beta_1) -$$
$$- (\alpha + \alpha_1 + \alpha_2) - (\alpha + \lambda + \mu)$$

or,

$$2\delta + \delta_1 + \delta_2 = 4\pi - \pi - \pi - \pi - (\alpha + \lambda + \mu)$$

whence $\quad \delta(APQ) = \pi - (\alpha + \lambda + \mu) = 2\delta + \delta_1 + \delta_2 .$

From this it follows that

$$\delta(APQ) > 2 \cdot \delta(ABC),$$

since

$$\delta_1 > 0 \text{ and } \delta_2 > 0.$$

The result we have obtained can be expressed in words as follows:

From the simultaneous assumptions that through any point within an angle a line can be drawn intersecting both sides of the angle, and that the fifth postulate is false, it follows that for any triangle ABC whatsoever it is always possible to find a triangle APQ whose defect $\delta(APQ)$ is greater than twice the defect $\delta(ABC)$ of triangle ABC.

From this it follows that triangles exist with arbitrarily large defects. But this is impossible, since the difference between π and a positive number less than π cannot be greater than π. This contradiction proves the theorem.

We shall cite one more equivalent of Euclid's axiom on parallels. For this purpose let us consider two theorems.

Theorem. If Euclid's fifth postulate is true, then the side of a regular hexagon inscribed in a circle is equal to the radius of that circle. The proof is familiar from secondary school geometry.

Converse theorem. If it is assumed that the side of a regular hexagon inscribed in a circle is equal to the radius of that circle, then Euclid's fifth postulate follows.

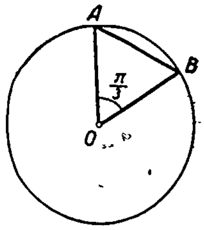


Fig. 267.

Proof. In fig. 267, let side AB of the inscribed regular hexagon be equal to the radius. That is to say,

$$AB = OA = OB.$$

But in an equilateral triangle the angles are equal, and since

$$\angle AOB = \frac{\pi}{3},$$

the sum of the angles of triangle AOB is equal to π. From this the fifth postulate follows. The theorem is proved.

With this we conclude the enumeration of the different forms of the axiom on parallels in Euclidean geometry.

65. THE AXIOM OF LOBACHEVSKII

We know that through a point A not on a line a it is possible to draw, in the plane determined by point A and line a, a line b not intersecting line a.

This fact follows directly from the theorem on the exterior angle of a triangle (see 2.).

Euclidean geometry postulates the uniqueness of this line.

Lobachevskii constructed a new non self-contradicting geometry, starting from the assumption that there is more than one such line passing through point A and not intersecting line a.

Axiom of Lobachevskii. Let a be an arbitrary line, and A a point not lying thereon; then in the plane determined by line a and point A there exist not less than two straight lines passing through point A and not intersecting line a.

In the geometry of Lobachevskii all the axioms of Euclidean geometry retain their force excepting only one - Euclid's fifth postulate - which is discarded and replaced by Lobachevskii's axiom.

It follows from the foregoing that Lobachevskii's axiom can be replaced by the assertion:

There exists a triangle the sum of whose angles is less than π. Here π denotes two right angles.

In his last work, "Pangeometry", published in 1855, a year before his death, Lobachevskii wrote:

"I published a complete theory of parallels under the title 'Geometrische Untersuchungen zur Theorie der Parallellinien Berlin 1840. In der Finke'schen Buchhandlung'.

"In this work I set forth the proofs of all the propositions in which it was not necessary to resort to the aid of parallel lines. Among these propositions the one giving the ratio of the surface of a spherical triangle to that of the whole sphere deserves particular attention (Geometr. Unters. 27.). If A, B, C denote the angles of a spherical triangle, the ratio of the surface of this spherical triangle to the surface of the whole sphere to which it belongs will be equal to the ratio of

$$\frac{1}{2}(A + B + C - \pi)$$

to four right angles. Here π denotes two right angles."

In this passage Lobachevskii is speaking of spherical geometry and points out that the proposition in question is independent of the axiom on parallels, that is, this proposition in spherical geometry will be equally true whether the fifth postulate or Lobachevskii's axiom be adopted.

"I then show," continues Lobachevskii, "that the sum of the three angles in a rectilinear triangle can not be more than two right angles (Geometr. Unters., 19), and that if this sum is equal to two right angles in a single rectilinear triangle it must be so in all rectilinear triangles (Geometr. Unters., 20). Thus, only two assumptions are possible: either the sum of the three angles in every rectilinear triangle equals two right angles -- this assumption characterizes ordinary geometry -- or else in every rectilinear triangle this sum is less than two right angles, and this latter assumption is the basis of the special geometry to which I have given the name imaginary geometry, but which may more properly be called pangeometry, since this name denotes geometry in a generalized form, of which ordinary geometry is a particular case" [32]. At present this new geometry created by Lobachevskii is familiar under the name of <u>the non-Euclidean geometry of Lobachevskii</u>.

66. THE ELEMENTS OF THE GEOMETRY OF LOBACHEVSKII

Upon adopting the axiom of Lobachevskii in place of Euclid's fifth postulate, the following propositions of Lobachevskian geometry follow:

Theorem. The sum of the three angles of a rectilinear triangle is less than two right angles and is a variable quantity, in general changing from triangle to triangle.

Theorem. A perpendicular and an inclined line to a given straight line do not always intersect. For, if they always intersected, there would follow Euclid's fifth postulate, which cannot be, since we have adopted the axiom of Lobachevskii.

Theorem. Similar but unequal triangles do not exist. If three angles of one triangle are equal to the three corresponding angles of another, these two triangles are congruent.

For, if there existed two unequal but similar triangles, the fifth postulate would hold good, which is impossible in Lobachevskian geometry.

Thus in Lobachevskian geometry there are no similar (and unequal) figures.

Theorem. There exist triangles about which it is impossible to circumscribe a circle.

For, the possibility of circumscribing a circle about every triangle is a proposition equivalent to the fifth postulate.

From the above theorem it follows that there exist triangles, perpendiculars to whose sides at their midpoints are parallel, i.e., do not intersect.

Theorem. The equal angles at the upper base of Saccheri's quadrilateral are acute.

For, if they were right angles, then there would exist a quadrilateral in which all four angles were right angles, which is impossible.

Theorem. The locus of the points in a plane lying on one side of a given straight line and whose distance from this line is equal to a given segment is a curve, no three points of which are collinear.

For, if three points of this locus lay in a straight line, then Euclid's postulate would be true (64.), which is not the case.

This curve is called a curve of equal distances or an equidistant curve. Ordinarily, by the equidistant curve is meant the locus of points at a fixed distance h from a straight line a, not necessarily on only one side of it. There are two branches of the equidistant curve, the upper and the lower. The line a is called the base of the equidistant curve.

Theorem. The side of a regular hexagon inscribed in a circle is greater than the radius of this circle.

Proof: The sum of the three angles of triangle AOB (fig. 267) is less than two right angles and consequently angle AOB, equal to $\frac{\pi}{3}$, is the largest angle in triangle AOB. But the largest side of a triangle lies opposite the largest angle, that is,

$$AB > OA .$$

Theorem. For any acute angle whatsoever, there always exists a perpendicular to one side of the angle which does not intersect the other side.

Proof. Let $\angle aOb$ be any acute angle (fig. 268). Let us assume that every perpendicular to side b intersects side a. From an arbitrary point A of side a we drop a perpendicular AB to side b, and we find the point B_1 on side b, which is symmetrical to point O about line AB. At the point B_1 we erect a perpendicular to side b. By our assumption, this perpendicular intersects side a in some point A_1. It can easily be calculated that the defect $\delta(OA_1B_1)$ of triangle OA_1B_1 is more than twice the defect $\delta(OAB)$ of triangle OAB:

$$\delta(OA_1B_1) > 2\delta(OAB).$$

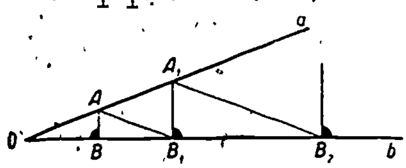


Fig. 268.

Taking then the point B_2 on line b, symmetrical to point O about line A_1B_1, we erect at B_2 a perpendicular to line b. By our assumption, this perpendicular intersects side a in some point A_2. Exactly as before, we find

$$\delta(OA_2B_2) > 2\,\delta(OA_1B_1).$$

Continuing thus, we obtain

$$\delta(OA_nB_n) > 2\,\delta(OA_{n-1}B_{n-1}) > 2^2\,\delta(OA_{n-2}B_{n-2}) > \ldots$$
$$\ldots > 2^{n-1}\,\delta(OA_1B_1) > 2^n\,\delta(OAB).$$

But the defect of a triangle cannot be greater than π; consequently, since $\delta(OAB) > 0$, not all perpendiculars to one side of an acute angle intersect the other side. The theorem is proved.

We shall hereafter refer to this important theorem as the theorem of the non-intersecting perpendicular.

67. THE RELATIVE POSITION OF LINES IN THE LOBACHEVSKIAN PLANE

In Lobachevskian geometry the plane is designated specifically as the Lobachevskii plane. In this plane the axiom of Lobachevskii is in force. The plane of ordinary geometry is called the Euclidean plane. In it Euclid's fifth postulate holds good.

We shall now investigate the relative position of straight lines in the Lobachevskii plane.

It is clear that through a point A not on a given line a there passes an infinite set of straight lines which do not intersect the given line a.

In proof, let b and c be two straight lines passing through A and not intersecting a (fig. 269); the existence of at least 2 such lines is asserted by the axiom of Lobachevskii.

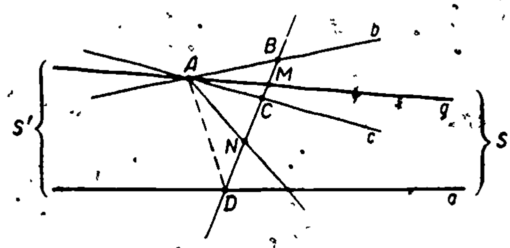


Fig. 269.

Through a point B on line b, so chosen as to lie on that side of line c on which there are no points belonging to line a, and through an arbitrary point D of line a we draw the line BD. The segment BD intersects line c in the point C. If M is an arbitrary point of segment BC, the line AM does not intersect line a. For if lines g (=AM) and a intersected in a point S on line g in the direction from A to M, then applying Pasch's postulate to triangle MDS and line c we should have to conclude that c intersected a; but this is excluded. If, on the other hand, it were assumed that g and a intersected in a point S' lying in the direction from M to A, then from the application of Pasch's postulate to triangle MDS' and line b we should have to conclude that b intersected a; but this is likewise excluded. Thus, none of the lines g lying within the angle BAC intersects a.

The lines of the pencil with center at A fall into two classes with respect to the line a. To the first class we assign those lines of the pencil which intersect line a ; to the second class, those which do not intersect a.

Each class contains an infinite set of lines and each line of the pencil belongs to a determinate class.

Turning our attention to the segment BD (fig. 269), if we assign to each line lying within angle BAD its point of intersection with segment BD,(1) then the points of this segment will fall into two classes which determine a Dedekind cut. To the first class belong those points N of segment BD such that line AN intersects a; to the second belong all the rest. The point D, for example, belongs to the first class, while the points B, M, C belong to the second. We complete the first class with the points of line BD lying outside segment BD in the direction from B to D, and the second class with the exterior points lying in the direction from D to B. It is not difficult to show that every point of the first class lies on the same side of any point in the second class.

(1) That each line lying within angle BAD intersects segment BD, may be proved [illegible]eans of Pasch's axiom. --Translators.

By the axiom of Dedekind there exists a point R on the segment BD which effects the division of line BD into two portions: the first portion DR containing all the points of the first class, and the second containing the points of the second class. The line AR (fig. 270) is seen to be the boundary of those lines lying within angle BAD which intersect line a. The line AR itself does not intersect a. For if AR intersected a in a point S, then taking a point T on line a outside segment DS in the direction from D to S, we should obtain a line AT belonging to the first class, which is impossible.

Lobachevskii, like Euclid, has no axiom of continuity, nor any axioms of position. Relying on intuition, Lobachevskii writes:

"All straight lines issuing from one point and lying in a given plane can be divided into two classes with respect to a given line of the same plane, namely into those intersecting the latter and those not intersecting it. The boundary line between these two classes is said to be parallel to the given line."(1)

We see that by a line parallel to a given line Lobachevskii meant not any line which does not intersect the given line a, but only the boundary line.

If from point A (fig. 270) a perpendicular is dropped to line a, then line AR', symmetrical with line AR about this perpendicular, will likewise be parallel to a in Lobachevskii's sense. Lobachevskii further distinguishes the side or direction of parallelism. Line AR is parallel to a in one direction and

---

(1) These words are, in essence, equivalent to the rigorous formulation of the axiom of discontinuity advanced by Dedekind half a century later.

line AR' in the other.

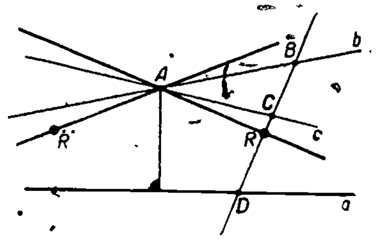


Fig. 270.

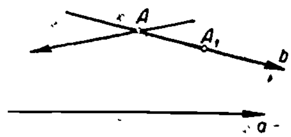


Fig. 271.

We shall conventionally represent the direction of parallelism by means of arrows (fig. 271).

Through every point A external to a given line a there pass two and only two lines parallel to the given line a in Lobachevskii's sense.

The lines g, not intersecting a (fig. 269) and which are not boundary or parallel lines with respect to a, are said to be superparallel to a.

Lobachevskii uses a different terminology [33]:

"Lines issuing from one point either intersect a given line in the same plane, or they never meet it, no matter how far prolonged. It is therefore necessary to distinguish between such lines with respect to a given line: meeting or converging lines and non-meeting or non-converging lines, to which belong the parallels, which constitute the transition from the one group to the other — the diverging lines."

Diverging in Lobachevskii's sense corresponds to what we call superparallel. The sense of the term used by Lobachevskii will become clear later.

If we limit ourselves to a specified direction, or more properly, orientation, of line a, then through point A there passes in <u>the given direction</u> a unique parallel in the Lobachevskian sense.

"In saying that a line is parallel to another, we shall hereinafter be referring" - writes Lobachevskii - "to that case only in which both extend to one side of some third line which joins them.

"Consequently, there can be only one parallel drawn from a given point to a given line; its distinctive property is that the slightest change of direction to one side will make it a converging line, and to the other side - a diverging line.

"With this approach" - continues Lobachevskii - "we are considering parallelism in its complete generality. Euclid, not being in a position to give a satisfactory proof, admitted in ordinary geometry only that special case in which two parallels must both be perpendicular to one line.

"Euclid's followers only made the matter more difficult with supplementary propositions, either arbitrary or very obscure, in the attempt to convince themselves of the correctness of an accepted truth which in the very nature of geometry it was impossible to prove."

The lines parallel and superparallel to a are determinate for a given point A. The question arises, whether or not line b, parallel to line a at point A, is also parallel to a at <u>every one</u> of its own points A'. Line b does not intersect a. But may not line b be parallel to a at one of its points, A, and superparallel to a at any other of its points A_1? Lobachevskii proved the following.

<u>Theorem</u>. <u>A straight line preserves the criterion of parallelism at all its points</u>.

<u>Proof</u>. Let line b be parallel to a at point A in the indicated direction (fig. 272), and let A_1 be an arbitrary point of ray Ab. We take any point B on line a and join it with points A and A_1. Since b is parallel to a at point A, any ray within the angle bAB intersects line a. At point A_1, line b is a line not intersecting line a. We shall show that any ray A_1c lying within angle bA_1B intersects line a; this will also signify the parallelism of line b to line a at point A_1.

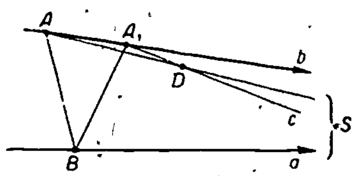


Fig. 272.

We select on ray A_1c an arbitrary point D, lying within angle bA_1B, and draw line AD. Since AD lies within angle bAB, it intersects line a in some point S. Considering triangle ABS and line c, we find that c does not pass through any vertex of triangle ABS and that it intersects side AS in point D.

It may be easily shown that line c does not intersect the segment AB; consequently, c intersects BS, that is, it intersects a. Thus line b is parallel to line a at point A_1.

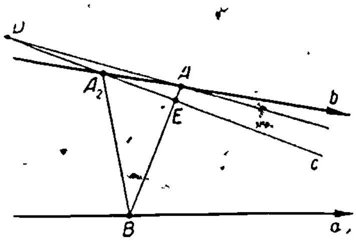


Fig. 273.

If point A_2 of line b lies on the other side of A, as in fig. 273, then drawing line A_2c within angle bA_2B we select a point D on this line such that A_2 lies between E and D, where E is the intersection of line c with segment AB.

The line DA, entering within the angle bAB, intersects a in a point S. Applying Pasch's postulate to the triangle ABS and line c we find that c intersects a. The theorem is proved.

After proving this theorem it is possible to say that line b is parallel -- in Lobachevskii's sense -- to line a, without having to state at what point.

The parallelism of lines in a specified direction possesses the properties of <u>symmetry</u> and <u>transitivity</u>.

<u>Theorem</u>. <u>Parallelism is always reciprocal, i.e., if line b is parallel to line a, then a is also parallel to b.</u>

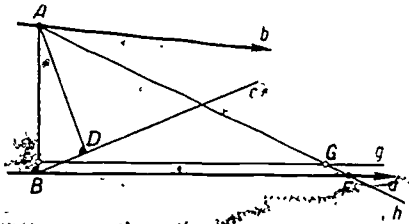


Fig. 274.

Proof. Let b (fig. 274) be parallel to a. We shall show that a is parallel to b in the same direction. From an arbitrary point A on line b we drop the perpendicular AB to line a (fig. 274). Since line a does not intersect b, it will be sufficient to show that a is parallel to b for point B. Let us draw an arbitrary line Bc within the angle aBA; we shall show that c intersects b. We let fall from A the perpendicular AD to line a. Since the hypotenuse AB of the right triangle ADB is greater than the side AD, there is a point E on segment AB such that AE = AD. We then erect at point E the perpendicular gE to segment AB.

After a rotation about point A through the angle DAB the segment AD will occupy the position AE; line Dc will lie along Eg; and line b will take the position Ah and intersect line a in point F, since b is parallel to a.

Applying Pasch's postulate to triangle ABF and line g, we find that line g intersects AF in a point G lying between A and F; lines a and g, being two perpendiculars to line AB, do not intersect.

A rotation about point A opposite to the first rotation transforms AE into AD; line g into c; line h into b; and, consequently, G, the point of intersection of lines g and h is transformed by the rotation into G*, the point of intersection of lines c and b, that is, c intersects b in the point G*. The theorem is proved.

Theorem. Two lines, parallel in the same direction to a third line, are parallel to each other in that same direction.

Proof. Let lines b and c be parallel to a and lie on opposite sides of the latter (fig. 275). This being the case, b and c will not intersect. We join the arbitrary points B and C on lines b and c by the segment BC and denote by A the intersection of BC with line a. Since b is parallel to a, an arbitrary line g within the angle bBA intersects line a in

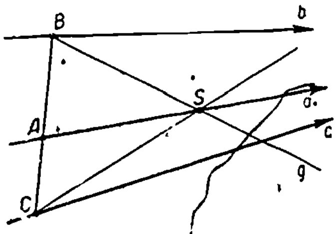


Fig. 275.

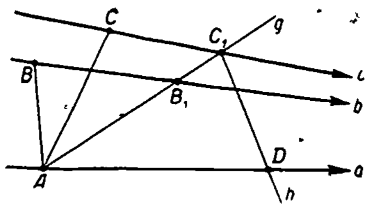


Fig. 276.

a point S and enters angle aSC through its vertex. But since a is parallel to c by symmetry, line g intersects c, that is, in this case b is parallel to c.

Now let lines b and c lie on the same side of line a (fig. 276). Lines b and c cannot intersect, since if they did intersect in some point there would be two lines passing through this point and parallel to a in the same direction, which is impossible. On line a we take an arbitrary point A and join it with any points B and C of lines b and c respectively by segments AB and AC.

Since the angles aAB and aAC have a common side Aa, it is possible to draw through point A a line g lying within both angles. By virtue of the fact that a is parallel to b and c, line g intersects the latter lines in points B_1 and C_1. For definiteness, let B_1 lie between A and C_1. An arbitrary line h drawn through C_1 and lying within the angle cC_1A intersects line a in point D, since c is parallel to a. Applying Pasch's postulate to triangle AC_1D and line b we find that h intersects b. This means that c is parallel to b. The theorem is proved.

From this and the preceding theorem there follows the transitivity of parallelism in the same direction. It then follows that the set of all oriented lines in the plane falls into classes of mutually parallel lines. Every oriented line determines a class of lines parallel to it -- a pencil of parallel, oriented lines. The same line, oriented in the opposite direction, determines another pencil.

Theorem. If two lines form with a third equal alternate interior angles, the two lines are superparallel.

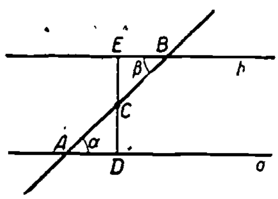


Fig. 277.

Proof. Let lines a and b form equal alternate interior angles α and β with the third line AB (fig. 277). From the midpoint C of segment AB we let fall the perpendiculars CD and CE to lines a and b. The right triangles ADC and BEC have equal hypotenuses and acute angles and are therefore equal. Consequently, ∠ ACD = ∠ BCE and points E, C, D lie in a straight line. Thence it follows that lines a and b are perpendiculars to the same line ED and consequently are superparallel.[1]

<u>Theorem</u>. <u>For any acute angle whatsoever, there always exists a line perpendicular to one side of the angle and parallel to the other</u>.

<u>Proof</u>. Let aOb be an acute angle (fig. 278). From the theorem on the non-intersecting perpendicular (66.) it follows that the points of the ray Oa fall into two classes. The first class consists of those points M_1 at which the perpendiculars to a intersect side b; the second, of those points M_2 at which the perpendiculars to a do not meet side b. All the points M_1 of the first class lie on the same side of every point M_2 of the second.

By Dedekind's axiom there exists a point R of the ray Oa such that the segment OR contains the points of the first class, and the ray Ra those of the second.

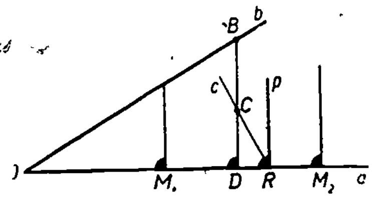


Fig. 278.

The perpendicular p to line a at the point R <u>does not itself intersect side b</u>. In proof: if p intersected b at a point K, then, taking the point L on side b beyond the point

(1) Line b cannot be <u>parallel</u> to line a. For if b were parallel to a, then by symmetry it would be parallel in both directions. Thus there would be only one line through E parallel to line a.
--Translators.

K and letting fall from point L a perpendicular to side b, we would obtain a perpendicular of the second class intersecting side b, which is impossible.

We shall now show that the first perpendicular p to one side Oa of the acute angle bOa which does not intersect the other side b is parallel to the second side b.

We draw line Rc within angle ORp. We shall show that Rc intersects Ob. Letting fall the perpendicular CD to a from an arbitrary point C on Rc, we find that the foot D of the perpendicular CD belongs to the segment OR and consequently CD intersects Ob in a point B. Applying Pasch's postulate to triangle ODB and line c we find that line c intersects b. The theorem is proved.

It follows from this that in the geometry of Lobachevskii every acute angle aOb uniquely determines a segment OR such that the perpendicular to line OR at R is parallel to the other side. Conversely, to every segment OR there corresponds an unique angle ROb, such that the line Ob forming this angle with OR will be parallel to the perpendicular to OR at R. To find this angle it is only necessary to draw from the endpoint O of segment OR the parallel b to p, the perpendicular to segment OR erected at the endpoint R.

Denoting the measure of segment OR by x and that of angle bOR by α, we see that α is a function of x.

Lobachevskii writes this function in the form:

$$\alpha = \Pi (x).$$

Angle α is called the angle of parallelism for point A and line a (fig. 279). The function $\Pi(x)$ is called the Lobachevskii function.

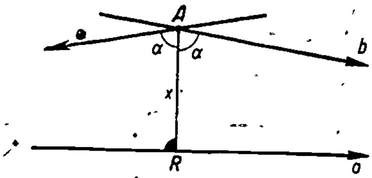


Fig. 279.

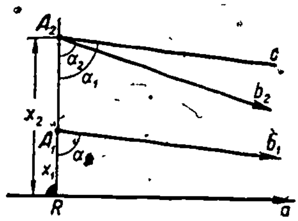


Fig. 280.

With increasing distance x of point A from line a the angle of parallelism decreases without limit:

$$\Pi(x) \to 0 \text{ as } x \to \infty .$$

In proof: Let $RA_1 = x_1$ and $RA_2 = x_2$ (fig. 280). Let us assume that $x_1 < x_2$. We have: $\alpha_1 = \Pi(x_1)$ and $\alpha_2 = \Pi(x_2)$. If we draw through point A_2 the line c forming the angle α_1 with A_2R, line c will, by a previous proposition, be superparallel to line b_1 and consequently line b_2, parallel to b_1, forms with line A_2R an angle α_2 less than angle α_1.

Thus, $\Pi(x_2) < \Pi(x_1)$ if $x_1 < x_2$. The Lobachevskii function is a decreasing function.

Since for an arbitrarily small angle ε there can be found an x such that $\Pi(x) = \varepsilon$, it follows that $\Pi(x)$ tends toward zero as $x \to \infty$.

With reference to the absolute dependence between segments and angles in non-Euclidean geometry, Lobachevskii wrote in "Principles of Geometry" (1829):

"The theory of parallels which we have set forth assumes lines and angles to be subject to a dependence, the existence of which in nature no one is in a position to prove or disprove. At any rate, astronomical observations convince us that all the lines which we are capable of measuring, even the distances between heavenly bodies, are so small in comparison to the line taken in theory as the unit that the equations hitherto used in rectilinear plane Trigonometry are correct to the extent of having no perceptible errors." Later, after setting forth the results of measurement of the angles of an astronomical triangle the vertices of which were stars, Lobachevskii continues:[3]

"After this it is impossible any longer to maintain that the supposition that the measure of lines depends upon angles -- a supposition which many Geometers have wished to accept as strict truth without demanding proof -- might be proved to be perceptibly false before the day when we shall have gone beyond the limits of the universe visible to us."

The symmetry of the relation of superparallelism can be established without difficulty.

Theorem. If one line is superparallel to a second, the second is superparallel to the first.

Proof. Let line a be superparallel to line b. Line b does not intersect a. If we assumed that b was parallel to a we would have to conclude that a was parallel to b. This, however, is not true. The theorem is proved.

Superparallelism does not, however, possess the property of transitivity.

Theorem. Two superparallel lines always have one and only one common perpendicular.

Proof. Let lines a and b be superparallel (fig. 281). Taking a point A on line a, we draw from A the lines p_1 and p_2 parallel to line b in opposite directions [29].

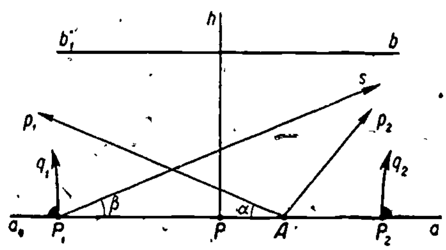


Fig. 281.

One of the angles aAp_2, $a_1\hat{A}p_1$ is necessarily acute. Let this be angle $a_1Ap_1 = \alpha$. Further, let segment AP_1 be such that $\Pi(AP_1) = \alpha$. Then the perpendicular q_1 to a at point P_1 will be parallel to p_1 and consequently parallel also to line bb_1 in the indicated direction. From the point P_1 we draw a line s parallel to b in the direction b_1b. Angle $sP_1a = \beta$ is acute and therefore a segment P_1P_2 can be found such that $\beta = \Pi(P_1P_2)$.

The perpendicular q_2 to line a at P_2 is parallel to s and hence also to b in the direction b_1b.

Thus, on one of the superparallel lines, a, we have found two points P_1 and P_2 such that the perpendiculars at these points to line a are parallel to the other superparallel line, b: one in the direction bb_1, the other in the direction b_1b. We bisect segment P_1P_2 by means of point P and erect at P the perpendicular h to line a. By virtue of the symmetry of the figure consisting of line a, perpendiculars q_1 and q_2 and line b with respect to line h, line h intersects line b and does so at right angles.

Two lines can not have two common perpendiculars because in that case there would exist a quadrilateral having four right angles, which is impossible.

<u>Theorem.</u> <u>In the direction of their parallelism two parallel lines converge without limit, while in the opposite direction they diverge without limit.</u>

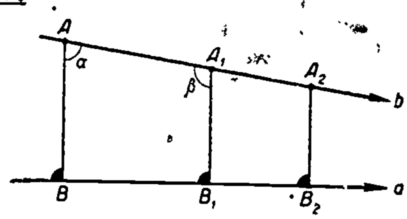


Fig. 282.

<u>Proof</u>. Let a and b be two parallel lines (fig. 282). Letting fall from points A and A_1 of line b to line a the perpendiculars AB and A_1B_1, we find that in the quadrilateral BB_1A_1A, which has the two right angles B and B_1, the angle $A = \alpha$ is acute while the angle $B = \beta$ is obtuse, whence

$$A_1B_1 < AB.$$

In like manner $A_2B_2 < A_1B_1$, that is, the distance from points of one parallel to the other parallel decreases in the direction of parallelism; in the opposite direction it increases.

We shall show that parallels converge <u>without limit</u>, or, in the usual terminology, <u>asymptotically</u>.

Let there be given an arbitrarily small segment ε. From point A on a line b (fig. 283) we let fall the perpendicular AB to the line a parallel to b. If AB is equal to or less than ε, the theorem is proved. Suppose then, $AB > \varepsilon$. We lay off on BA from point B the segment $BE = \varepsilon$ and draw through E the lines c_1 and c_2 parallel to a in opposite directions. Line c_1 makes with segment EB the acute angle α, while line c_2 makes with the ray EA the obtuse angle β.

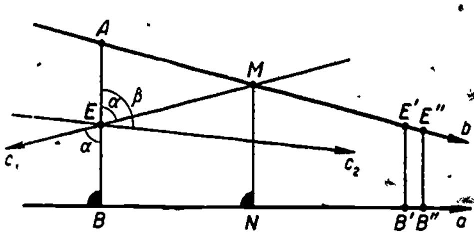


Fig. 283.

Consequently, line c_1 lies within angle $c_2EA = \beta$.

Since c_2 is parallel to b, line c_1 intersects line b in a point M. Letting fall from M the perpendicular MN to a, we find that lines b and c_1, as being the parallels to a from point M, are symmetrical about NM. Segment E'B' symmetrical to segment EB about NM will be the distance from point E' of line b to line a, parallel to b, and will be equal to ε. The distance E"B" (fig. 283) will be less than ε. q.e.d.

The divergence without limit of the parallel lines in the opposite direction is demonstrated analogously.

Theorem. Two superparallel lines diverge without limit in both directions from their common perpendicular.

Proof. Any two superparallel lines a and b (fig. 284) have a common perpendicular PQ. From an arbitrary point A on line b we let fall the perpendicular AB to line a. Since in quadrilateral PBAQ the angles P, B and Q are right angles, angle A is acute. Thence it follows that $AB > PQ$. In like manner, letting fall from point A_1 belonging to ray QA but exterior to segment QA the perpendicular A_1B_1, we find that $A_1B_1 > AB$.

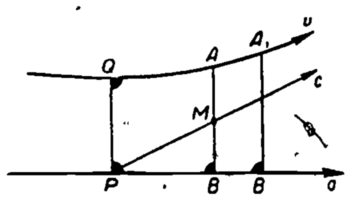


Fig. 284.

From this it follows that the distances from points of one superparallel line to the other increase in both directions from their common perpendicular.

The segment PQ of the common perpendicular of two superparallel lines is the shortest distance between them.

From point P we draw line c parallel to line b. Line c intersects the segments AB of all perpendiculars from points on ray Qb to line a, as can be shown with the aid of Pasch's postulate. Consequently, $AB > MB$. But MB, being the perpendicular from point M of side Pc of the acute angle cPa, increases without limit as point M moves along side c of the angle in the direction away from its vertex P. This signifies that AB also increases without limit with increasing distance of point A from point Q. The theorem is proved.

The relative positions of parallel and superparallel lines are reflected in Lobachevskii's terms <u>converging</u> and <u>diverging</u>.

68. THE CONCEPT OF MEASUREMENT OF AREA

The theories of the measurement of segments in the geometries of Euclid and Lobachevskii coincide.

The system of measurement of area in the geometry of Lobachevskii is constructed in a manner analogous to that in Euclidean geometry. The concepts of the equicomposition and the equivalence by completion of simple polygons is introduced in Lobachevskii's geometry entirely in the same manner as in Euclidean geometry. All the general theorems likewise preserve their formulation with a single exception: to the triangle is assigned, not a number proportional to the product of base times altitude -- in the geometry of Lobachevskii this product turns out not to be independent of the choice of a side of the triangle -- but a number $\triangle$, proportional to the defect of the triangle:

$$\triangle = k^2 (\pi - A - B - C).$$

Here k^2 is a coefficient of proportionality, A, B and C are the angles of triangle ABC and π denotes two right angles. The number $\triangle$ is then the area of the triangle.

The expression for the area of a triangle in the geometry of Lobachevskii calls to mind the expression for the area of a spherical triangle (65.); and this is not fortuitous. Let us briefly sketch the development of the system of measurement of the area of simple polygons.

1. To each triangle we assign a positive number $\triangle = k^2 (\pi - A - B - C)$, which we shall call the area of the triangle.

2. Examining oriented triangles (58.), we shall prove the theorem:

If in the plane of an oriented triangle an arbitrary point 0 is selected and joined to the three vertices of the given triangle, the difference between the sum of the areas of the triangles having the same orientation as the given one and the sum of the areas of the triangles having the opposite orientation is equal to the area of the original triangle.

Omitting a detailed analysis of all the possible cases, we shall consider only the case in which point 0 lies outside the given triangle but within one of its angles (fig. 285).

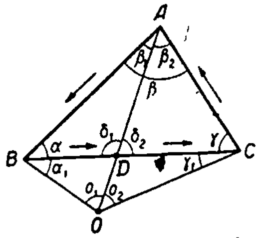


Fig. 285.

Nor shall we reduce this case to a simpler one, but immediately carry out the entire computation. Designating the angles by the symbols shown in the diagram, we find for the positively oriented triangles;

$$\triangle(OAB) = k^2\ (\pi - \alpha - \alpha_1 - \beta_1 - o_1);$$

$$\triangle(OCA) = k^2\ (\pi - \gamma - \gamma_1 - \beta_2 - o_2).$$

And for the negatively oriented triangles:

$$\triangle(OBD) = k^2\ (\pi - \alpha_1 - \delta_2 - o_1);$$

$$\triangle(ODC) = k^2\ (\pi - \gamma_1 - \delta_1 - o_2).$$

For the required difference we find:

$$\triangle(OAB) + \triangle(OCA) - \triangle(OBD) - \triangle(ODC) =$$
$$= k^2 \{\delta_1 + \delta_2 - \alpha - \beta_1 - \beta_2 - \gamma\} = k^2 (\pi - \alpha - \beta - \gamma) = \triangle(ABC).$$

3. Beyond this the treatment of areas of simple polygons proceeds entirely as in Euclidean geometry (58).

The student should carry out precisely and in detail all the reasoning relevant to the foregoing, as an exercise.

Chapter XII

THE AXIOMATIC STRUCTURE OF GEOMETRY

Chapter XII presents a complete set of geometric axioms. Examples are given to elucidate the fundamental idea of the axiomatic method. The equivalence of various axioms of continuity is established, and the axioms of motion are introduced.

69. THE FUNDAMENTAL OBJECTS OF GEOMETRY; THE BASIC RELATIONSHIPS BETWEEN THE OBJECTS

In the axiomatic development of geometry we choose certain fundamental concepts and fundamental relationships between these concepts. The former as well as the latter are neither defined nor explained. An indirect definition of the basic concepts and the basic relationships between them is, however, provided by the axioms. The abstract geometry thus developed may then be applied to sets of objects of diverse kinds.

In abstracting from the particular and concrete to fundamental concepts and their basic interrelationships, geometry takes those general qualities which underlie the laws and regularities to which spatial forms are subject. More than that, on the succeeding level of abstraction, geometry is abstracted as well from the particular form of the object of its study, and emerging beyond the bounds of this form, acquires thereby the possibility of reflecting nature more deeply and completely.

"The abstractions matter, natural law, value and so on, in a word all scientific (correct, serious, not absurd) abstractions reflect nature more deeply, more truly, more fully. From the concrete observation of nature to abstract thought and from this to practice -- such is the dialectical course of the apprehension of truth, the cognition of objective reality" (Lenin, Philosophical Notebooks, edn. 1947, pp. 146-147).

In order better to elucidate the nature of abstract geometry, we may note the following comparison of geometry and grammar:

"The distinguishing characteristic of grammar is that it gives the rules of modification of words -- referring not to particular words, but to words in general; it gives the rules for the formation of sentences, not particular concrete sentences -- with, let us say, a concrete subject, a concrete predicate, and so on -- but all sentences in general, irrespective of the concrete form of any sentence in particular. Hence, abstracting itself, as regards both words and sentences, from the particular and the concrete, grammar takes those general qualities which lie at the basis of the modification of words and their combination into sentences, and builds it into grammatical rules, grammatical laws. Grammar is the outcome of a prolonged work of abstraction of human thought; it is an indicator of the tremendous achievement of thought.

"In this respect grammar resembles geometry, which creates its own laws by a process of abstraction from particular concrete objects, regarding objects as bodies without any particularity, and defining the relations between them not as the particular relations of particular objects, but as the relations of bodies in general, without any concreteness." (Stalin: "On Marxism in Linguistics.")

In order correctly to understand the structure of the science of geometry one must remember that geometry "is the result of a prolonged work of abstraction of human thought" (Stalin); that "man in his practical activity has before him the objective world, depends upon it, determines his activity by it"(Lenin).

The axioms of geometry themselves, like the axioms of logic, are the result of man's practical activity.

"...The practical activity of man had to lead human consciousness to the million-fold repetition of various logical forms before they could acquire the significance of axioms ..."; "by his practice man proves the objective validity of his ideas, his concepts, his knowledge, his science". (Lenin, Philosophical Notebooks, 1947, pp. 161, 164).

In the study of the axiomatic method and of the application of abstract geometry to objects of various kinds it must be borne in mind that the only requirement is that all the axioms be satisfied. The conclusions flowing from the axioms -- the established theorems -- will then automatically hold. In the practical utilization of an axiomatically developed geometry it is necessary to decide in each case the question of whether the axioms are fulfilled and the degree of exactness of this fulfillment.

In Hilbert's axiomatic system, which (with some modifications) we shall here adopt, three categories of things are taken as fundamental objects [15].

The things of the first category are called points and are represented by the symbols A, B, C, ...; those of the second category are called straight lines and the symbols a, b, c, ... are used for them; those of the third category are called planes and are designated by $\alpha, \beta, \gamma, \ldots$.

The basic objects are associated in definite relationships designated by the expressions lie on, lie between, be congruent or equal.

The exact, and for mathematical purposes complete, description of these relationships is achieved by the axioms of geometry.

Thus, points, straight lines and planes are in Hilbert's axiomatics self sufficient and are not described in terms of other objects.

For example, a plane is not necessarily the set of points belonging to it. From this point of view, the term figure may designate a set of points, straight lines and planes.

The basic objects and their basic relations are not defined, not explained, not described. However, whatever be the nature of these objects, to whatever branch of knowledge they relate, they and their properties must be so far known that it will be possible to verify that the axioms relating to them are satisfied.

One and the same geometric system may be developed by means of different axiomatic systems dealing with different fundamental objects and different basic relationships.

In the axiomatics of Hilbert there are twenty axioms, falling into five groups.

The geometry of Lobachevskii exerted a decisive influence on the development of axiomatic method in mathematics, particularly in geometry.

70. THE FIRST GROUP OF AXIOMS: AXIOMS OF CONNECTION (BELONGING)

The fundamental relationship is: lie on, or belong to.

I_1. For any two points A, B there exists a straight line a belonging to both of these two points A, B.

I_2. For any two points A, B there exists not more than one line belonging to both of the points A, B.

Instead of the term belonging to, other expressions are also used. For example, instead of: line a belongs to both of the points A and B, we may say: line A passes through points A and B, or, line a joins point A with point B; instead of A belongs to a, we may say A lies on a, or, A is a point of a, and so on.

I_3. On a line there exist at least two points. There exist at least three points not lying on the same line.

I_4. For any three points A, B, C not lying on the same line there exists a plane α belonging to all three points A, B, C. For any plane there always exists a point belonging to it.

I_5. For any three points A, B, C not lying on the same line there exists not more than one plane belonging to all of these points.

I_6. If two points A, B of line a lie in plane α, then every point of line a lies in plane α.

I_7. If two planes α and β have a common point A, then they have at least one additional common point B.

I_8. There exist at least four points not lying in the same plane.

In order briefly to elucidate the fundamental idea of the abstract development of geometry, let us consider the following example. We construct a model as follows. We shall take four ordinary lines lying in the same plane, and intersecting each other in pairs, and shall consider them to be objects of the first category, call them "points" and designate them A, B, C, D (fig. 286).

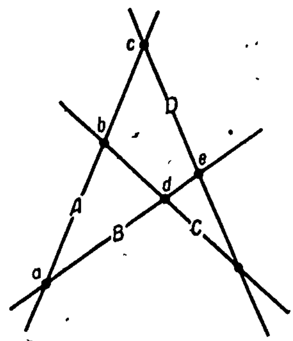


Fig. 286.

The six ordinary points of intersection of these four lines we shall consider to be objects of the second category; we shall call them "lines" and designate them a, b, c, d, e, f .

The fundamental objects having been indicated (we shall consider only the planar axioms I_{1-3}), we introduce the relationship of belonging in the following manner. To each pair of objects of the first category we assign objects of the second category as shown below:

$$(A,B)\to f;\quad (B,C)\to c;\quad (C,D)\to a;$$
$$(A,D)\to d;\quad (B,D)\to b;\quad (A,C)\to e.$$

To each two ordinary lines of our model is assigned that ordinary point which does not lie on them.

We shall now consider that a "line" "belongs" to those "points" to which it has been assigned.

Axioms I_{1-3} are seen to be true for our model. But we have neither the usual representation of the fundamental objects nor the usual meaning of the fundamental relation of belonging to.

We shall say that "line" a "passes through" "points" C and D; "points" C and D "lie on" "line" a; "lines" a and b "intersect" in "point" D, and so on.

Hereinafter, in describing the construction of a model we shall omit the quotation marks.

Let us take another example. Let the set of ordered pairs (x,y) of real numbers x,y be the set of objects of the first category. The elements of this set are points. Let the set of all ratios (u:v:w) of three real numbers u,v,w of which the first two can not both vanish, be the set of objects of the second category. The elements of this set are lines.

We shall consider that a point (x,y) belongs to a line (u:v:w) if the equality

$$ux + vy + w = 0$$

is true. For example, point (1,3) lies on line (2:-1:1).

In this model each of the axioms I_{1-3} is satisfied.

The first model (fig. 286) was introduced as an illustrative example. The second -- the numerical model -- has a greater scientific and practical significance; in it the plane axioms of groups I-V are satisfied; for this, of course, it is necessary to define appropriately the meaning of the terms lie between and equal.

71. THE SECOND GROUP OF AXIOMS: AXIOMS OF ORDER

The axioms of this group pertain to the concept "between".

II_1. If a point B lies between a point A and a point C, then A, B, C are three distinct points of a straight line, and B also lies between C and A (fig. 287).

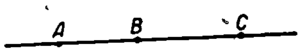


Fig. 287.

Fig. 288.

II_2. For any two points A and C on a straight line AC there exists at least one point B such that point C lies between A and B (fig. 288).

II_3. Of any three collinear points not more than one lies between the other two.

Definition. A system of two points A and B on a line a is called a segment. The points lying between A and B are called points of the segment AB, or interior points of the segment AB, the points A and B - the endpoints of the segment. All the remaining points on the line a are called points exterior to the segment AB.

Axioms II_{1-3} are the axioms of order on a line.

II_4. (Pasch's postulate.) Let A, B, C be three non-collinear points, and a line in plane ABC, not passing through any of the points A, B, C; if, then, line a passes through one of the points of segment AB, it must pass either through one of the points of segment AC or through one of the points of segment BC.

Following are examples of the rigorous proof of theorems with the aid of axioms [15].

<u>Theorem</u>. <u>For any two points A and C on line AC there exists at least one point D lying between A and C</u>.

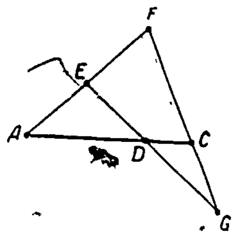


Fig. 289.

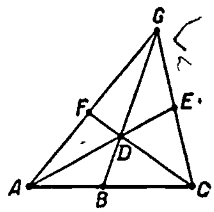


Fig. 290.

<u>Proof</u>. By axiom I_3 there exists outside line AC (fig. 289) some point E, and by axiom II_2 there exists on line AE a point F such that E lies between A and F. Again by axiom II_2, there exists on line FC a point G not lying within segment FC. It is easy to show that line EG does not pass through any of the points A, F or C. It follows, therefore, by axiom II_4 that line EG intersects segment AC in point D.

<u>Theorem</u>. <u>Of three points A, B, C on the same straight line, one always lies between the other two</u>.

<u>Proof</u>. Suppose that A does not lie between B and C and C does not lie between A and B. We draw a line through point D, not lying on line AC, and point B, and we take -- in accordance with axiom II_2 -- a point G on this line such that D lies between B and G (fig. 290). Applying II_4 to the triangle BCG and line AD, we find that AD and CG intersect in some point E lying between C and G. We find also that lines CD and AG intersect in point F, lying between A and G. Applying II_4 to the triangle AEG and line CF, we find that D lies between A

and E. Applying the same axiom to triangle AEC and line BG, we see that point B lies between A and C.

All the axioms of groups I_{1-8} and II_{1-4} will be operative in, for example, the following model. Let the set of objects to be called points be the set of points interior to some sphere (fig. 291). The points of the sphere itself and points outside it will not be considered objects of the first category. Let the set of lines be the set of all chords of the sphere. Let the planes be those open disks within the sphere whose circumferences are on the sphere itself. Let the notions belong to and between have their customary meanings.

It can readily be verified that all the axioms of groups I and II hold good in this model. The student should prove this to himself by carefully verifying the validity of this model in relation to each of these axioms.

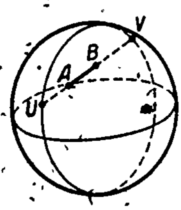


Fig. 291.

72. THE THIRD GROUP OF AXIOMS: AXIOMS OF CONGRUENCE

These axioms pertain to the concept of "congruence" and thus serve to define the concept of motion (29, 34).

III_1. If A and B are two points on a line a and A' is a point on the same line or on another line a', it is always possible to find a point B' lying on line a' on a given side[1] from point A', and such that segment AB is congruent, or in other words equal, to segment A'B'. The congruence of segment AB to segment A'B' is symbolized by

$$AB \equiv A'B'.$$

III_2. If segments A'B' and A"B" are congruent to the same segment AB, then segment A'B' is also congruent to segment A"B"; that is, if two segments are congruent to a third, they are congruent to each other.

III_3. Let AB and BC be two segments of a line a having no interior points in common, and let A'B' and B'C' be two segments of the same line or of a second line a', likewise having no interior points in common (fig. 216) if in such a case

(1) Let A, B, C, D denote any four points on the same line $\mathcal{L}$. Then the pair of points (B,C) is said to lie on the same side of A, if A is not between B and C. Using Pasch's postulate we may prove the following: 1) if each of the pairs (B,C) and (C,D) lies on the same side of A then the pair (B,D) also lies on the same side of A; 2) if the pair (B,C) does not lie on the same side of A then either the pair (C,D) or the pair (B,D) lies on the same side of A. Using 1) and 2) as well as axiom II_2, we may show that the points of L, other than A, are divided into two mutually exclusive classes such that any two points are in the same class if and only if they are on the same side of A. Each of these classes, augmented by the point A is called a ray issuing from A. All points on the ray, other than A, are said to lie on the same side from A. --Translators.

$$AB \equiv A'B' \text{ and } BC \equiv B'C',$$

then

$$AC \equiv A'C'.$$

We have already made use of axioms III_{1-3} (53).

Definition. Let α be an arbitrary plane, and h and k any two distinct rays in this plane issuing from the same point O and belonging to different lines. We shall call such a system of rays h and k an angle and write it as $\angle(h,k)$ or $\angle(k,h)$. The rays h, k are called the sides of the angle and point O its vertex.

Let the ray h belong to the line $\overline{h}$ and the ray k to the line $\overline{k}$. The rays h and k together with the point O divide the remaining points of plane α into two regions: one region is composed of the points lying on the same side of $\overline{k}$ (1) as h and on the same side of $\overline{h}$ as k, and these points are said to lie within the angle denoted by $\angle(h,k)$; the remaining points form the second region and are said to lie outside this angle.

III_4. Let there be given an angle, $\angle(h,k)$, in the plane α and a line a' in the plane α', as well as a specifically indicated side of α' relative to line a'. Let h' designate a

(1) Let ℓ be any given line and let A, B, C be points not on ℓ. Then the pair (A,B) is said to lie on the same side of ℓ if the segment AB does not intersect ℓ. Using Pasch's postulate, we may show: 1) if the pairs (A,B) and (B,C) each lies on the same side of ℓ, then the pair (A,C) also lies on the same side of ℓ; 2) if the pair (A,B) does not lie on the same side of ℓ, then either the pair (B,C) or the pair (A,C) lies on the same side of ℓ. From this and axiom II_2 it follows that all the points in the plane which do not belong to ℓ are divided into two classes. Any two points A, B will be in the same class of the pair (A,B) lies on the same side of ℓ. Each class is said to form a half-plane, lying on one side of line ℓ. --Translators.

ray of line a' issuing from a point O'. With these things given, there exists in plane α' one and only one ray k' having the following property: the angle $\angle(h,k)$ is congruent, or in other words equal, to the angle $\angle(h',k')$, while all the interior points of $\angle(h',k')$ are located in plane α' on the given side of line a'.

The congruence of angles is denoted thus:

$$\angle(h,k) \equiv \angle(h',k').$$

Every angle is congruent to itself, i.e. in every case

$$\angle(h,k) \equiv \angle(h,k).$$

In short, every angle can be laid off in one and only one way in a given plane, from a given ray, and on a given side of this ray.

P. K. Rashevskii [15] gives the following explanation: "According to Hilbert, by 'laying off' an angle we are to understand, not the construction of the angle by some instruments, such as compasses and straight-edge, but the fact of the existence of a ray defining an angle congruent to the given one. Accordingly, by the 'unique manner of construction' we are to understand the existence of only one such angle."

III_5. If in two triangles ABC and A'B'C' we have the congruences:

$$AB \equiv A'B', \quad AC \equiv A'C', \quad \angle BAC \equiv \angle B'A'C',$$

then the following congruence also exists:

$$\angle ABC \equiv \angle A'B'C'.$$

Appropriately changing the notation used for the vertices of the given triangles, we find, that if the conditions of axiom III_5 are fulfilled, there are always two congruences:

$$\angle ABC \equiv \angle A'B'C' \text{ and } \angle ACB \equiv \angle A'C'B'.$$

Thus, Hilbert takes as an axiom a part of the assertion familiar under the name of the first criterion for the equality of triangles.

Axioms I_{1-8}, II_{1-4}, III_{1-5} are valid both for the customary secondary school model of Euclidean geometry and also for points, lines and planes and the relationships lying on, between, being congruent segments, being congruent angles in the geometry of Lobachevskii.

As another example, we shall introduce the concept of congruence in the model constructed of the interior points of a sphere (71). To every pair of points A, B of line UV (fig. 292) we shall assign the number

$$\rho(AB) = k \cdot \log\left(\frac{UB}{BV} : \frac{UA}{AV}\right)$$

and shall consider segments congruent or equal if the numbers so assigned to them are equal in absolute value. We shall call these numbers the lengths of the segments. In the equation k denotes a positive coefficient. It can easily be verified that with the concept of congruence of segments thus introduced, all the relevant axioms of group III will be valid.

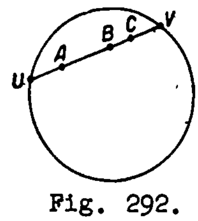


Fig. 292.

In order to verify the truth of axiom III_3 it is sufficient to establish the additiveness of lengths:

$$\rho(AC) = \rho(AB) + \rho(BC),$$

where B lies between A and C.

We find (fig. 292):

$$\rho(AB) = k\cdot\log\left(\frac{UB}{BV} : \frac{UA}{AV}\right)$$

$$\rho(BC) = k\cdot\log\left(\frac{UC}{CV} : \frac{UA}{AV}\right)$$

Adding, we obtain:

$$\rho(AB) + \rho(BC) = k\cdot\log\left(\frac{UB}{BV} : \frac{UA}{AV}\right) + k\cdot\log\left(\frac{UC}{CV} : \frac{UB}{BV}\right) =$$

$$= k\cdot\log\left\{\left(\frac{UB}{BV} : \frac{UA}{AV}\right) \cdot \left(\frac{UC}{CV} : \frac{UA}{AV}\right)\right\} = k\cdot\log\left(\frac{UC}{CV} : \frac{UA}{AV}\right) = \rho(AC).$$

After this it is not difficult to verify axiom III_3 and all the remaining axioms relevant to the congruence of segments.

It should be noted that if the endpoints of a segment coincide, its length $\rho(AA)$ is equal to zero. If point B approaches point V while point A remains fixed, the length $\rho(AB)$ approaches infinity.

Into this model there can also be introduced the congruence of angles and the validity of all axioms of the first three groups can be fully established. The idea of constructing models plays a great role not only in purely scientific but also in technological problems. The study of non-Euclidean geometry and of the foundations of geometry will provide the best insight into this idea.

73. THE FOURTH GROUP OF AXIOMS: THE AXIOM ON PARALLELS

We have already examined in detail the question of the axiom on parallels. As an axiom of parallelism we can introduce either the Euclidean one (or an equivalent), in which case we have Euclidean geometry, or the axiom of Lobachevskii, in which case we have the geometry of Lobachevskii. The only difference between the sets of axioms of these two geometries lies in the axiom on parallels.

For completeness we introduce once more the text of each.

IV. (The Euclidean axiom). Let a be an arbitrary line and A a point not lying thereon; then in the plane determined by line a and point A there exists not more than one line passing through point A and not intersecting line a.

If we add to the four groups of axioms so far given the axioms of group V on continuity, we obtain the complete set of axioms of Euclidean geometry.

In the geometry of Lobachevskii the axioms I_{1-8}, II_{1-4}, III_{1-5} and V_{1-2} are retained unchanged, but the Euclidean axiom IV is replaced by Lobachevskii's axiom IV*.

IV* (the axiom of Lobachevskii). Let a be an arbitrary line and A a point not lying thereon; then in the plane determined by line a and point A there exist not less than two lines passing through point A and not intersecting line a.

We can construct more than one model in which all the axioms I_{1-8}, II_{1-4}, III_{1-5}, IV* and V_{1-2} are valid.

In the model constructed from the interior points of a sphere the axiom of Lobachevskii holds good along with the axioms of groups I, II, III and V.

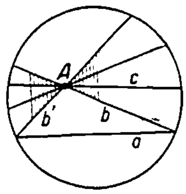


Fig. 293.

In the drawing (fig. 293) there is given a plane determined by line a and point A. Line b is parallel to a in one direction and line b' in the other; line c is superparallel to line a.

In this plane model it is not difficult to see intuitively the relative positions of lines in the plane of Lobachevskii. This model was first worked out by the Italian mathematician Beltrami as a map of the Lobachevskii plane.

It is also not difficult to obtain from a spatial model of Lobachevskii geometry (fig. 291) a full insight into the relative positions of lines, of planes, and of lines and planes. It is only necessary to keep in mind that insofar as this model of Lobachevskian geometry is constructed of Euclidean objects, angles and segments in the sense of Lobachevskii will also be angles and segments in the Euclidean sense. However, the measure of segments and angles in the Lobachevskii geometry will, in general, not coincide with their measure in the Euclidean sense. Two angles or two segments which are equal to each other in the Euclidean sense will not, in general, be equal in the Lobachevskii sense.

74. THE FIFTH GROUP OF AXIOMS: AXIOMS OF CONTINUITY

The axioms of continuity are, as we have seen, the following:

V_1 (axiom of measurement, or axiom of Archimedes). Let AB and CD be any two segments; then on line AB there exists a finite number of points $A_1, A_2, A_3, \ldots A_n$ such that the segments $AA_1, A_1A_2, A_2A_3, \ldots, A_{n-1}A_n$ are each congruent to segment CD and point B lies between A_{n-1} and A_n. (Fig. 219.)

V_2 (axiom of Kantor). If on a line a there is given an infinite sequence of closed segments $A_1B_1, A_2B_2, \ldots,$ such that each successive segment $A_{n+1}B_{n+1}$ is contained in the preceding segment A_nB_n (fig. 221) and if there exists no segment contained in all the segments of the given sequence, then there exists one and only one point X belonging to every segment of the sequence.

The role of this axiom in the structure of geometry has been made clear in previous chapters (54, 67). In Hilbert's system there appears instead of Kantor's axiom under the number V_2 the axiom of completeness.

In "The Foundations of Geometry" Hilbert set himself the task, in particular, of investigating as completely as possible all the conclusions which could be reached without using the axioms of continuity -- and, consequently, without the axiom of Archimedes. Hilbert shows in one of his works that the geometry of Lobachevskii can be built upon a foundation consisting exclusively of the axioms relating to the plane, and without the use of the axioms of continuity. For example, Hilbert gave a proof of the theorem on the existence of a perpendicular to one side of an acute angle not intersecting the other side which does not make use of the axioms of group V. On the other hand, the proofs given in 66 rely on the axioms of continuity.

The problem of "The Foundations of Geometry" consists in formulating the axioms of geometry and investigating their interrelationships.

75. THE CONCEPT OF THE EQUIVALENCE OF AXIOMS

In 64. we saw that Euclid's axiom on parallels can be replaced by other assertions, for example, by the assertion that the sum of the angles of any rectilinear triangle is equal to two right angles. The axiom of Lobachevskii can be replaced by the equivalent assumption stating that the sum of the angles of a rectilinear triangle is less than two right angles; and so on.

In the foregoing statements, however, it remains unclear just what are the conditions under which the asserted equivalence holds good. We shall present some of the results of the researches of Hilbert and his students.

"If we adopt the axiom of Archimedes, then the axiom on parallels can be replaced by the assumption that the sum of the angles of a triangle equals two right angles."

This result was also obtained in 64., but without explicit reference to the axiom of Archimedes.

On the basis of axioms I-III, that is, not using the axioms of continuity, one can prove the theorem: "If in any one triangle the sum of the angles is greater than, equal to or less than two right angles, then the same is true of every triangle."

"If we discard the axiom of Archimedes, then from the assumption that through a single point an infinite number of lines can be drawn not intersecting a given line it by no means follows that the sum of the angles of a triangle is less than two right angles."

We point out again the equivalence of the axioms of Archimedes and Kantor to the single axiom of Dedekind (62.) This means that using the axioms of groups I-III and V the axiom of Dedekind can be proved as a theorem, and inversely that using axioms I-III and that of Dedekind the axioms of Archimedes and Kantor can be proved as theorems.

Let us derive, for example, the axiom of Archimedes from axioms I-III and the axiom of Dedekind [24].

The proof is by contradiction. Let us suppose that for some segment AB the axiom of Archimedes is not true. This means that there exists an infinite sequence of congruent segments $AA_1 \equiv A_1A_2 \equiv \ldots \equiv A_nA_{n+1} \equiv \ldots$, lying within segment AB.

Selecting the direction from A to B along the line, we divide the points of line AB into two c sses. To the first we assign every point which precedes (counting in the above-mentioned direction) some point A_n (and thus precedes points $A_{n+1}, A_{n+2}, \ldots$ also); to the second class, all the remaining points of line AB.

In this division into classes the requirements of the axiom of Dedekind are fulfilled. Firstly, every point of line AB belongs to one of the classes, and each class contains points; the first class contains $A_1, A_2, \ldots, A_n, \ldots$ while the second class contains point B. Secondly, every point of the first class precedes all points of the second class.

By Dedekind's axiom there exists a point C which effects the division of the points into the two classes. There is no last point in the first class; consequently point C is the first point of the second class. By axiom III_1 it is possible to lay off in the direction from B to A a segment CD congruent to each of the segments $AA_1, A_1A_2, \ldots$. Point D will belong to the first class. This means that there can be found a point A_n such that point D precedes it. Thus the segment A_nA_{n+1} is a proper part of segment DC.

Therefore

$$A_nA_{n+1} < DC.$$

But

$$A_nA_{n+1} = DC.$$

The contradiction obtained proves the proposition of Archimedes.

Remark. The terms precede and follow can be precisely defined on the basis of the relationship between. Conversely, it is possible to reformulate all the axioms of order on the basis of the relationship precede and then to define the term between.

76. THE AXIOMS OF MOTION

In 72. were formulated the axioms of congruence on which we based our definition of a (geometric) motion as a certain kind of one-to-one mapping of space (or of the plane) onto itself (29,34). We can, however, proceed in a different fashion: instead of regarding congruence as a fundamental relation and formulating congruence axioms, we can regard motion as a fundamental relation and formulate axioms on motion, thereafter establishing the equivalence of the resulting system of axioms.

Let us turn to the formulation of axioms of motion in the plane [5].

III^*_1. In the group of all one-to-one mappings of the set of all points of a plane onto itself there exists a subgroup of mappings called motions.

III^*_2. Under every motion the image of a segment is a segment.

From this it follows that the image of a line under any motion is a line. The proof of this is recommended as an exercise.

III^*_3. Let there be given two arbitrary points O and O', two arbitrary rays OA and O'A', and associated respectively with each, the half-planes α and α' selected from each pair of half-planes into which the lines OA and O'A' respectively divide the plane. There exists then one and only one motion simultaneously taking point O into point O', the half-line OA into the half-line O'A' and the half-plane α into the half-plane α'.

"Motion" as defined by axioms III_1^*, III_2^* and III_3^* is nothing other than the "motion in the extended sense" (36) for the plane.

With the aid of these axioms it can easily be proved that under every motion an angle goes over into an angle, a triangle into a triangle, etc.

Definition. Two segments, two angles, two triangles and in general two figures are said to be congruent if there exists a motion taking one of these figures into the other.

With the aid of these axioms we can readily prove the propositions III_{1-5} on congruence.

For example, from the group property of motions it can easily be shown that two segments (or angles) congruent to a third are congruent to each other.

Proof. Let $A'B' \equiv AB$, $A''B'' \equiv AB$. This means that, firstly, there exist motions f_1 and f_2 such that

$$A'B' = f_1(AB) \text{ and } A''B'' = f_2(AB).$$

From the group property it follows, secondly, that $AB = f_1^{-1}(A'B')$, where f_1^{-1} is the motion inverse to f_1. Hence

$$A''B'' = f_2[f_1^{-1}(A'B')].$$

But the product $f_2 f_1^{-1}$ of two motions is a motion, and consequently $A''B'' \equiv A'B'$, q.e.d.(1).

(1) The axioms on motions and their consequences, containing propositions III_{1-5} may be studied in greater detail in V. I. Kostin's book [28].

CHAPTER XIII

THE IDEA OF AN INTERPRETATION OF A GEOMETRIC SYSTEM

In Chapter XIII some interpretations of Euclidean geometry are studied; in particular, Feodorov's interpretation of the solid geometry of Euclid is examined. Three interpretations of the geometry of Lobachevskii are given. One of these is studied in considerable detail, and its isomorphism with the others is established. In conclusion the requirements imposed upon a system of axioms are set forth.

77. DIFFERING INTERPRETATIONS OF THE EUCLIDEAN GEOMETRY OF THE PLANE

To give an interpretation of a geometrical system means to set up some model of this geometry, that is, to assign a concrete meaning to its fundamental objects -- points, lines, planes -- and to its fundamental relationships -- belonging to, lying between, being congruent -- and to do this in such a way that all the axioms of the given geometrical system are satisfied.

One of the models of the geometry of Euclid is the plane geometry to which we are accustomed, with its familiar basic objects and relationships.

A. Let us construct another model of the Euclidean geometry of the plane. We shall consider a sheaf consisting of parallel lines and planes. Each line of the sheaf will be considered a point, and each plane a line. We thus have the fundamental objects of plane geometry.

We shall employ the customary meanings of belonging to and lying between. By the term motions we shall understand the subgroup of all motions in space which take the sheaf of lines into itself in such a manner that every plane perpendicular to a line of the sheaf slides along itself.

The verification of the planar axioms of groups I, II, III*, IV, V offers no difficulty (figs. 294, 295).

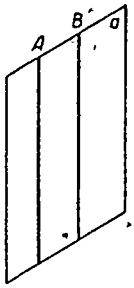


Fig. 294

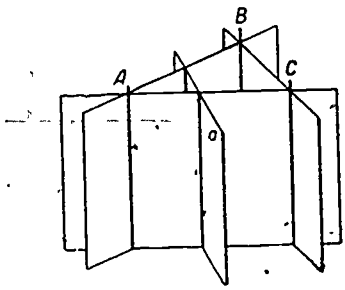


Fig. 295

In fig. 295, for example, is represented the application of Pasch's postulate to triangle ABC and line a.

B. Let us consider still another model of Euclidean plane geometry.

We take the customary Euclidean plane π and "puncture" it in a given point O, that is, we remove the point O from the plane. In order not to disturb thereby the topological structure of the plane, we add to it one improper point, which we shall agree to consider the inverse image of the pole S under the stereographic projection of plane π onto a sphere tangent to the plane at point O (figs. 100, 118, 124). On the sphere the image of the improper point is placed at S, the hitherto existing perforation.

We shall consider as <u>points</u>, all the points of the plane with the point at O omitted, plus the improper point.

As <u>lines</u> we shall take those lines and circles in plane π which pass through the perforation O (fig. 296).

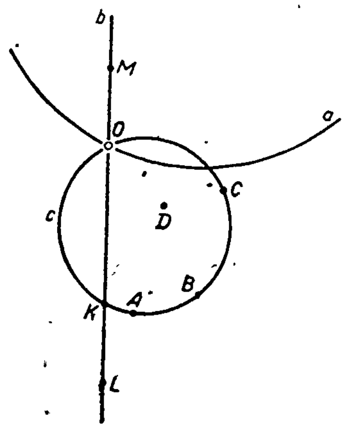


Fig. 296

Belong to and lie between will be understood in the usual sense. Angles whose vertices are ordinary points are said to be equal if they are equal in the ordinary Euclidean sense. If the vertex of an angle α is the improper point, then its sides are two ordinary rays issuing from O. Let α' be the ordinary angle with vertex O, formed by the two rays. Then the angle α is compared with other angles by means of the angle α'. The interpretation of equality for segments will be omitted for the present.

Our model is ready. 1)

It must be emphasized that the choice of a model is not arbitrary: we are obliged to insure that the axioms are satisfied.

A geometrical system can be applied to objects and relationships of any kind exhibiting regularities which may be expressed by means of the axioms of the geometry.

In the above mentioned model, all the axioms will hold. Let us verify some of them.

Axiom I_1. For every two points A and B there exists a line a, belonging to both of these points A and B. If points A and B are not collinear with O they determine a circle c passing through O (fig. 296); but in our model, c is a line.

1) As the reader can see, the points and lines of our model are the images of the points and lines of the Euclidean plane under an inversion with respect to a circle with center at O, the improper point serving as the image of O under the inversion. Two angles - or two "segments" - are regarded as equal if their inverse images under the inversion are equal. -- the Translators.

If points K and L lie on the same ordinary line b with point O , then they determine this line b which is also a line of our model. If one point K is a proper point, but the other, L , is the improper point, then the ordinary line b determined by O and K will be the line of our model determined by O, K and L (fig. 296).

In exactly the same way we verify, for example, axiom II_1. Point B lies on line c between points A and C . (fig. 296). Point L lies between points K and M . In this latter case line b is "cut" at O and "pasted together" by means of the improper point. This can be readily represented by means of the stereographic projection mentioned above.

The Euclidean form of axiom IV is operative here (fig. 297).

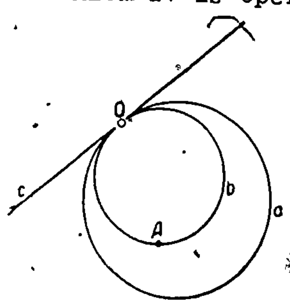


Fig. 297

The drawing shows a line a and the unique line b passing through the proper point A and parallel to a . Since the point O has been removed, lines a and b do not intersect. On the other hand, line c is the unique line passing through the improper point and parallel to a .

We emphasize again that all ordinary lines passing through O are lines in our model passing through the improper point and intersecting in it at the same angle as in the puncture O .

As an exercise the student should, using this model, carry through the proofs of some elementary theorems.

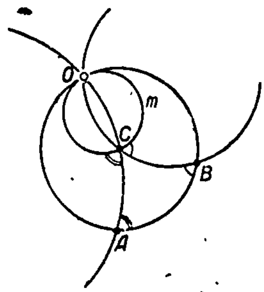


Fig. 298

By way of example we shall prove that the sum of the angles of the rectilinear triangle ABC (fig 298) is equal to two right angles. We shall follow the usual proof. We draw through point C the line m, parallel to AB. The alternate angles marked in the diagram are equal, etc. The equality of these angles can also be verified directly in the usual way.

It is also possible, without drawing line m, to find at O angles equal to the interior angles of triangle ABC and to obtain their sum directly.

In order to obtain a better understanding of the constructed model and to make clear what must be understood by <u>motion</u> or <u>congruence</u>, we shall proceed as follows. We shall pass from one model to another by means of stereographic projection.

Let us suppose that we have the ordinary plane π and on it the usual points, lines, motions and so on, in short the customary interpretation.

We project stereographically the whole plane π onto the sphere perforated at the pole S (fig. 117). To every point A of the plane there will correspond a point B on the sphere; to each ordinary line a , a circle a' of the sphere passing through the puncture S . To each ordinary motion in plane π there will correspond some transformation into itself of the sphere punctured at S .

Thus, with the aid of stereographic projection, we have constructed a new model of Euclidean plane geometry. Its points are the points of the sphere perforated at S . Its lines are the circles of the sphere which pass through the puncture S . Belonging to and lying between have their usual meaning.

We shall regard as congruent segments, those circular arcs whose inverse images are equal. Similarly, angles are considered to be equal when their inverse images are equal. Since stereographic projection preserves angles, the congruence of angles will have its usual meaning.

It is also easy to see what must be understood by motions in this model. In this connection we must note that under all motions other than rotations of the sphere about the diameter SP (which correspond to rotations of plane π about point P) the point P leaves its place and goes to a different location on the perforated sphere.

Clearly, all the planar axioms are valid on this model.

By means of a second stereographic projection we return to the model with which we began. We do this by passing a plane π' tangent to the given sphere at the puncture S , and from P (fig. 117) as the pole, projecting the sphere onto this plane π' . For the sake of one-to-one correspondence we consider plane π' to be perforated at the point of tangency S (this puncture is point

O of the original model) and to be completed by an improper point (the image of point P). All circles on the sphere passing through S go over in the plane π' into circles or into lines passing through the puncture S (or O). The point P has as its image the improper point of plane π' .

The original model has been reconstructed. It now becomes easy to see the meaning of the congruence of angles. The meanings of congruence of segments and motion are also readily perceived.[1]

C. We have already encountered the analytical interpretation of Euclidean plane geometry. Let us construct a model.

As points we take all the ordered pairs (x , y) of real numbers x , y ; as lines, all the ratios (u:v:w) of three real numbers u , v , w , where u and v are not simultaneously zero. Thus, (2:3:-1) and (4:6:-2) are not distinct lines.

The point (x , y) belongs to the line (u:v:w) if

$ux + vy + w = 0$.

Three points (x_1 , y_1) , (x_2 , y_2) , (x_3 , y_3) are collinear if

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0 .$$

---

1) Quite often when removing a point O from the Euclidean plane its replacement by an improper point is omitted, thus committing an error. See, for example, V. F. Kagan: Lobachevskii, published by USSR Academy of Science, 1944, p. 301-302. In this book the Euclidean plane punctured at point O is even given a special name, "the plane of Poincaré," which is a misconception.

See also N. M. Beskin: A Methodology of Geometry, State Pedagogical Press, 1947, p. 19 and elsewhere; and V. Molodshii: "Is the Geometry of Lobachevsky True?" in Mathematics and Physics in the School, No. 1 of 1936, where this gross error is repeated.

The point (x, y) of line $(u:v:w)$ lies between the points (x_1, y_1) and (x_2, y_2) collinear with it if

$$x_1 > x > x_2$$

or

$$x_1 < x < x_2$$

(When $x_1 = x_2$, if $y_1 < y < y_2$ or $y_1 > y > y_2$.)

A purely analytical definition of the congruence of segments and angles and of motions can easily be formulated.

This interpretation is customarily studied under the name of the analytical geometry of the plane.

78. FEODOROV'S INTERPRETATION OF EUCLIDEAN SOLID GEOMETRY

Academician E. S. Feodorov (1853-1919), eminent Russian crystallographer and mathematician, in his work on the structure of crystals devised the following method of representing space in the plane [46] [47] [48].

In a plane π we assign to each point A in space a circle Γ with center at the foot A' of the perpendicular from point A to plane π, with a radius equal to the distance AA' of point A from plane π, and so oriented that, looking from point A towards plane π, the orientation arrow points in a counter-clockwise direction (fig. 299).

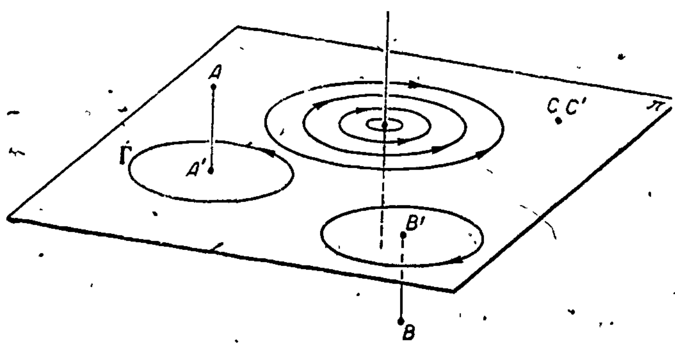


Fig. 299

The same circle but with opposite orientation corresponds to the point symmetrical to A with respect to plane π. To any point C on plane π there corresponds a circle of zero radius, i.e., the point C itself. A ray perpendicular to plane π is represented by a set of concentric circles with the same orientation. A line a intersecting plane π not at a right angle is represented by a set M' of homothetic oriented circles with center of similarity at the point of intersection of line a with plane π (fig. 300).

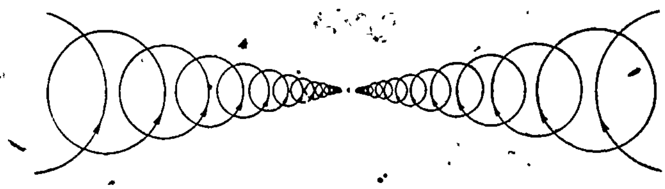


Fig. 300

To a plane α intersecting plane π there corresponds a set M" of oriented circles whose radii are proportional to the distances of their centers from the line of intersection of planes α and π. After establishing this one-to-one mapping of space onto the set of all oriented circles in the plane, we shall construct Feodorov's model.

The points are the oriented circles of plane π.

The lines are the above described sets M' of oriented circles.

The planes are the above described sets M" of oriented circles.

To belong to is understood in the set-theoretic sense, that is, it means to be an element of a set.

A circle is said to lie between two other circles when its inverse image is a point lying between the two points corresponding to the other circles. It is possible of course to give the meaning of this term directly, without referring to the inverse image.

Sets of circles are considered congruent if their inverse images form congruent figures. Feodorov gives the meaning of these terms directly without going outside the model. However, we shall not go into this point.

It is readily seen that all the axioms, and consequently all the theorems, will be true in this model of solid geometry.

For example, the axiom that a plane is completely determined by three points which are not collinear here takes in ordinary terminology the form: given three oriented circles not having a common center of similarity, then the three centers of similarity, corresponding to each pair of circles, lie on one line (fig. 301). The student should independently establish for each orientation of the circles what centers are to be considered corresponding, and should investigate all the possible cases.

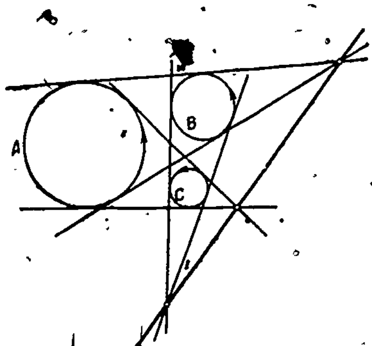


Fig. 301

We see that a simple axiom when applied to a model yields, when regarded from the customary point of view, a quite complicated proposition. As to this Feodorov says: "Give us a new theorem, and we shall obtain from it an endless number of other theorems."

Note that the foregoing mapping of Euclidean solid geometry onto a plane of oriented circles is, by reason of its one-to-one nature, an interpretation. On the other hand, for example, descriptive geometry, representing a point of space by a pair of points, does not constitute an interpretation because of the existence of exceptions which destroy the one-to-one property of the mapping.

For this very reason the Feodorov model possesses, for many problems, greater advantages than the method of Monge in descriptive geometry [46] [47] [39].

79. INTERPRETATIONS OF THE GEOMETRY OF LOBACHEVSKII IN THE EUCLIDEAN PLANE

A. Poincaré's first interpretation of Lobachevskii's geometry.

Using the forms of ordinary Euclidean plane geometry we shall construct the following model. We select an arbitrary circle Γ and consider the set of points interior to the circle, the set of all circular arcs interior to and orthogonal to circle Γ, and the set of all diameters of circle Γ. We establish the categories of basic objects and basic relations as follows: Points are those points interior to circle Γ. Lines are the circular arcs interior and orthogonal to circle Γ together with the diameters of that circle. The term belonging to has its customary meaning, as does also the term between. The term congruent segments will be given a concrete meaning later. Congruent angles will be those which are congruent in the Euclidean sense. The meaning of the term motion will be given later.

The applicability of the planar axioms of the first two groups is immediately verifiable, as soon as these axioms are formulated with the customary names for the elements of the model substituted for presently assigned conventional names.

I_1. For any two points A and B interior to circle Γ there exists an arc orthogonal to Γ (or a diameter of Γ) passing through both of these two points A and B.

I_2. For two points A and B interior to circle Γ there exists not more than one arc orthogonal to Γ (or diamater of Γ) passing through both of the points A and B.

I_3. On an arc orthogonal to Γ (or a diameter of Γ) there exist at least two points interior to Γ. There exist at least three points interior to Γ not lying all on the same arc orthogonal to Γ 1), or on the same diameter.

The student should verify for himself the axioms of order II_{1-4}.

We shall for the moment skip the axioms of congruence, or of motions, and turn to the axiom on parallelism. Here the axiom of Lobachevskii is valid. For the purpose of verifying this axiom we shall return to the original meanings of the basic forms and shall not hold ourselves to the exact text of the Lobachevskian axiom.

IV.* Through a point A interior to circle Γ there passes an infinite set of arcs orthogonal to Γ (or of diameters of Γ) which do not intersect a given arc orthogonal to Γ (or diameter of Γ) in any point interior to Γ. In fig. 302 the center O of circle Γ plays the role of the point A.

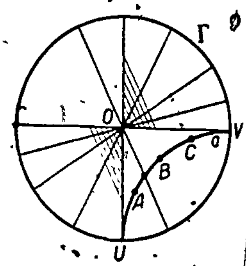


Fig. 302

In fig. 209 -- using the language of our present model -- two lines are drawn through point A parallel 2) to line g.

---

1) Circular arcs are to be understood in all cases.

2) It is easy to see that two "lines" are parallel in the sense of Lobachevskii if and only if they intersect on the circumference of Γ. -- the Translators.

In fig. 207 is drawn the unique line AB through the points A and B . In fig. 204 are shown a pencil of lines with center at point A ; a pencil of lines parallel to line a_2 ; and a pencil of superparallel lines. In fig. 208 a perpendicular is dropped from point A to line g. Fig. 210 illustrates the theorem on the perpendicular p to the side g_1 of acute angle g_1Ag_2 parallel to the other side g_2 .

In fig. 211 the common perpendicular to the superparallel lines g_1 and g_2 is constructed. In like manner all the theorems relating to the relative positions of lines in the geometry of Lobachevskii can be verified. Each of these theorems is at the same time a proposition or a good construction problem in Euclidean geometry.

Fig. 303 illustrates the theorem on the sum of the angles of a triangle. Here the sum of the angles between the chord BC and the tangents to arc BC at its endpoints is equal to the defect of the triangle ABC .

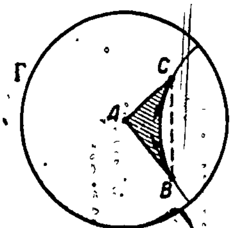


Fig. 303

Let us elucidate the meaning of a motion in this model of the geometry of Lobachevskii.

In studying the group of motions in the plane in Euclidean geometry (29), having at our command the concept of congruence of segments we first of all defined a motion of the first kind.

A motion of the first kind was completely determined by two congruent segments AB and A'B' located in the given plane (fig. 147). Point A under this motion goes over into point A'; point B into point B'; and any point M into a point M' such that triangles AMB and A'M'B' are equal and of like orientation. The whole of this construction is also part of the geometry of Lobachevskii. In Euclidean geometry we can represent every motion as the sum of a translation and a rotation (29). In the geometry of Lobachevskii there exist rotations about a point, but there are no parallel translations.

The plane may be translated along a straight line a, but then the points not lying on line a slide along equidistant curves having a as their base line, and not along straight lines as in Euclidean geometry.

However, every motion of the non-Euclidean plane can be represented as the sum of a sliding along a straight line and a rotation (in that order). Namely, we first translate the plane along the line AA' (fig. 147) so that point A is taken into point A' and then rotate the plane about point A' until point B coincides with point B'.

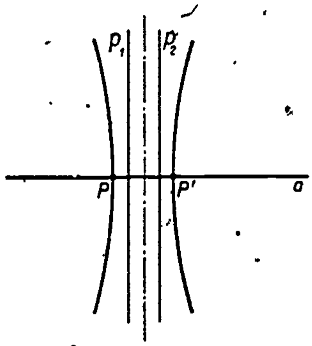


Fig. 304

A rotation about a point can be represented as the sum of two reflections, as in the Euclidean case (fig. 177). A sliding along line PP' (fig. 304) such that point P goes over into point P' can also be represented as the sum of two reflections about the lines p_1 and p_2 which are perpendicular to segment PP' at points distant $\frac{1}{4}$ PP' from its ends.

This shows that the group of motions -- of the first and the second kinds -- of the plane of Lobacheyskii consists of all reflections and all their possible sums.

From this we conclude that the set of all motions in the Lobachevskian plane in Poincaré's model consists of all inversions with respect to those circles, arcs of which are lines of the model. A reflection about a diameter of circle Γ is regarded as the limiting case of inversion.

In all these inversions circle Γ is transformed into itself, as are the points interior to it, by reason of the orthogonality of circle Γ to the circles of inversion a (fig. 302).

We have already learned how to obtain images under such transformations (49, 51; figs. 202, 204).

After the introduction of the group of motions it is easy to verify the applicability of the axioms on motions, and thereafter, in the usual manner, to define the congruence of figures. Since under inversions angles remain unchanged, it follows that in the Poincaré model, angles congruent in the sense of Lobachevskii will also be congruent in the ordinary sense.

Poincaré's model may be regarded as a conformal map of the plane of Lobachevskii.

There will now be no difficulty in constructing a system of measurement of segments within this model of Lobachevskian geometry.

We know (49) that the double ratio

$$(ABCD) = \frac{AC}{AD} : \frac{BC}{BD}$$

of four points is invariant under inversions, and consequently also under motions in our model.

Having in mind the aim of defining the length of a segment AB, which must be an invariant under motions, let us consider the end-points U and V of the orthogonal arc constituting the line AB in the model (fig. 302). The double ratio (ABUV), which is invariant under motions, can not, however, be the length of the segment, since the property of additivity will not exist. As the length $\rho(AB)$ of segment AB we must take the number

$$\rho(AB) = k \cdot \log (ABUV),$$

or

$$\rho(AB) = k \cdot \log \left(\frac{AU}{AV} : \frac{BU}{BV}\right).$$

With this definition additivity will exist.

In proof of this (referring to fig. 302):

$$\rho(BC) = k \cdot \log\left(\frac{BU}{BV} : \frac{CU}{CV}\right)$$

$$\rho(AB) + \rho(BC) = k \cdot \log\left\{\left(\frac{AU}{AV} : \frac{BU}{BV} \cdot \frac{BU}{BV} : \frac{CU}{CV}\right)\right\} =$$

$$= k \log\left(\frac{AU}{AV} : \frac{CU}{CV}\right) = \rho(AC)$$

It will now be evident that the axioms of motion and the axioms of congruence are operative in the model under consideration.

The applicability in our model, of the axioms on continuity is obvious.

The construction of a model of the geometry of Lobachevskii using the forms of Euclidean geometry shows that we shall not be able to discover any contradictions in the geometry of Lobachevskii since otherwise we should obtain, by way of the model, a contradiction in Euclidean geometry.

The geometry of Lobachevskii is free from contraditions to the same degree as is the geometry of Euclid. But since Euclidean geometry is non-contradictory to the same degree as arithmetic -- this follows from the existence of the analytical model --, the geometry of Lobachevskii can no more lead to self-contradictions than can arithmetic.

It is to be noted that an analytical (numerical) model for Lobachevskian geometry can also be constructed directly.

From the foregoing there also follows this important fact: The Euclidean postulate on parallels can not be proved on the basis of the first three and the fifth groups of axioms since if the postulate could be proved then its negation -- the axiom of

Lobachevskii -- would not be compatible with the axioms I_{1-8}, II_{1-4}, III_{1-5} and V_{1-2}. The existence of the model, however, proves this compatibility.

This first model can readily be extended to solid geometry also. For our categories of fundamental objects we take: as points -- the ordinary points interior to a given sphere Γ; as lines -- the arcs interior to the sphere belonging to circles orthogonal to the sphere, and the diameters of the sphere (without their endpoints); as planes -- the portions interior to Γ of all spheres orthogonal to Γ, and the open disks passing through the center of Γ. Belonging to and between are understood in the customary sense. As motions we take the group of all inversions with respect to those spheres which play the role of planes in the model, as well as all reflections about customary planes passing through the center of sphere Γ.

B. Poincaré's second model of Lobachevskii's geometry

This model is obtained from the first when the circle Γ degenerates into a line Γ^* and the interior of the circle into a half-plane (fig. 305); or when the fundamental circle Γ of the first model is inverted with respect to a center of inversion lying on the circumference Γ.

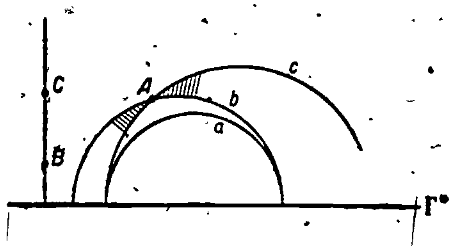


Fig. 305

The lines are all semicircles in the half-plane whose centers lie on Γ^* and all rays in the half-plane perpendicular to Γ^*. The points are the ordinary points of the half-plane. Belonging to and between have their ordinary meanings.

In fig. 305 are constructed the two parallels b and c to line a . In fig. 212 is constructed the common perpendicular p' to two superparallel lines g'_1 and g'_2 .

Motions here are defined in exactly the same way as in the first model.

In this model all the axioms of Lobachevskii's plane hold good.

The introduction of the theorems of Lobachevskii's geometry into this second Poincaré model provides, just as in the case of the first, a fine source of geometrical problems which are interesting from the Euclidean viewpoint.

This model also may readily be extended to solid geometry. The categories of fundamental objects will be: as points -- the ordinary points above a given plane π ; as lines -- the rays perpendicular to plane π and the semicircles with centers on plane π and lying in planes perpendicular to π ; as planes -- the hemispheres with centers on plane π and the half-planes perpendicular to plane π [29].

C. The Beltrami-Klein interpretation of Lobachevskian geometry

Let us construct the model. The categories of basic objects will be as follows. The points will be the ordinary points interior to a circle Γ . The lines will be the chords of circle Γ .

Belonging to and between will have their usual meanings. (fig. 293)

What is to be understood by congruence and motion will be made clear later.

The axiom of Lobachevskii holds good in this model.

If there are two interpretations of the same set of axioms such that a one-to-one correspondence can be established between the objects of the respective categories in the one and the other model with the preservation of the corresponding relationships, these two interpretations are said to be isomorphic.

The customary model of Euclidean geometry taught in school and its numerical model, analytical geometry, are thus isomorphic. Also isomorphic are the customary interpretation of plane geometry and the interpretation (77) in a plane punctured at a point 0 and completed with an improper point, in which the lines are defined as the circles and lines passing through 0 .

We shall establish the isomorphism of each of Poincare's models with the Beltrami-Klein model. For this purpose we construct a sphere tangent to the plane π of the Beltrami circle at the center of this circle, and having a radius equal to that of the circle.

We first project orthogonally the entire Beltrami map onto the lower hemisphere (fig. 306). The map of Lobachevskian geometry thus obtained on the hemisphere is next stereographically projected from the upper pole S onto the original plane π . This latter projection will be seen to yield the first Poincaré interpretation of the geometry of Lobachevskii [16] [29].

It is readily seen that the correspondence thus established between the two models is isomorphic. In particular, every motion in the Poincaré model induces a corresponding motion in the Beltrami-Klein model.

If we construct a hemisphere the equator of which coincides with the circumference of the Beltrami map and orthogonally project this map upon the hemisphere, and if we then stereographically project the hemisphere from a pole S which lies on the equator onto a plane α perpendicular to the diameter passing through the pole of projection S, we obtain on the plane α Poincaré's second model of the geometry of Lobachevskii [29].

The isomorphism of all three models has been established.

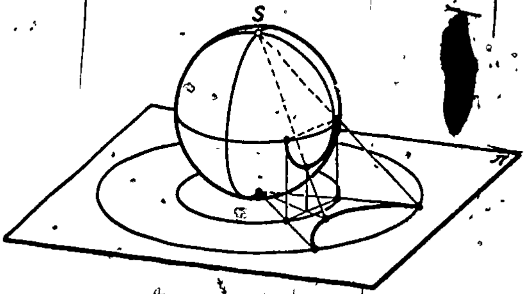
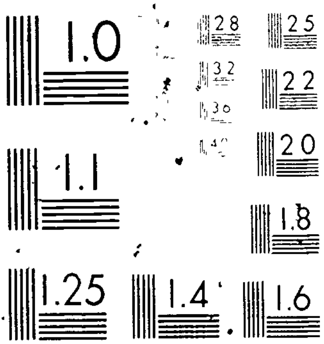


Fig. 306


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6

CONCLUSION

80. THE REQUIREMENTS IMPOSED UPON A SYSTEM OF AXIOMS

We shall briefly consider what is demanded of a system of axioms. First of all, a system of axioms is required to be consistent. In a consistent system it can not happen that a theorem and its negation should be simultaneously demonstrable. In order to prove non-contradictoriness, a model or interpretation of the system of axioms is constructed out of objects and relations between them which are considered as evidently existing, correct, true, non-contradictory.

Thus, for example, to prove the absence of contradiction in Euclidean geometry we construct the analytical or numerical model. In so doing we accept the objects of arithmetic -- numbers and their relationships -- as true.

To prove non-contradiction in the system of axioms of Lobachevskii geometry we likewise construct a model out of objects and relations between them held to be real, correct, true, non-contradictory.

Thus, for the proof of consistency, or non-contradiction, in any system of axioms we have need of concrete, well-tested things and relations between them (see 2, 69).

The second fundamental requirement imposed upon a system of axioms is that as far as possible each axiom be independent of the others. The independence of each axiom means that one axiom cannot be derived from the others as their logical consequence, and that if one axiom is withdrawn from the system it is impossible to use the other axioms in any way which would restore to the system the effect of the axiom which was removed.

The axiomatics of geometry are, however, not such that we can speak of the independence of all axioms. The fact is, for example, that in the formulation of the axioms of the second group it is assumed that there is already established the concept of "belonging to," the properties of which are in fact described in the axioms of the first group. In the systems of axioms which we have been studying it is only possible to raise the question of the independence of some axioms.

The independence of some axiom from all the rest, that is, the impossibility of proving it by means of all the others, is proved by replacing the axiom by its negation (in one or another form) and proving the non-contradictoriness of the resulting set of axioms.

Thus, proof of the independence of Euclid's axiom on parallels is found in the fact that it is replaced by the axiom of Lobachevskii and then the consistency of Lobachevskii's geometry is proved. Hilbert in his "Foundations of Geometry" examines many axioms and indeed entire groups of axioms as to their independence. [15].

Another problem of geometric axiomatics is that of the completeness of a system of axioms. A system of axioms is said to be complete if any two of its interpretations are isomorphic. It is not required that a system of axioms necessarily be complete. Thus, the axiomatic system of the theory of groups is not complete: there exist non-isomorphic groups.

BIBLIOGRAPHY

(Note: All entries refer to Russian-language editions.)

1. J. Hadamard: Elementary Geometry, Parts I and II, Uchpedgiz 1936-1948.
2. A. Adler: Theory of Geometrical Constructions, Uchpedgiz 1940.
3. I. I. Alexandrov: Collection of Geometrical Problems In Construction, Uchpedgiz 1950.
4. P. S. Alexandrov: Introduction To The Theory of Groups, Uchpedgiz 1939.
5. P. S. Alexandrov: What Is Non-Euclidean Geometry? "N. I. Lobachevskii" Series. (1793-1943). GTI, 1943.
6. P. S. Alexandrov and V. A. Efremovich: On the Simplest Notions of Contemporary Topology, ONTI, 1935.
7. P. S. Alexandrov and V. A. Efremovich: Outline Of The Basic Concepts Of Topology, ONTI, 1936.
8. S. A. Bogomolov: Geometry (Systematic Course), Uchpedgiz 1949.
9. S. A. Bogomolov: Foundations of Geometry, GIZ, 1923.
10. János Bolyai: Appendix. GITTL, 1950.
11. M. E. Vashchenko-Zakharchenko: Euclid's Elements, with an expository introduction and interpretations. Kiev, 1880.
12. I. M. Vinogradov: Fundamentals of the Theory of Numbers, GTI, 1949.
13. V. M. Voronets: Geometry With a Pair of Compasses, ONTI, 1934.
14. A. B. Gedymin: Cartography, Uchpedgiz 1946.
15. D. Hilbert: Foundations of Geometry, GTI 1948.
16. D. Hilbert and S. Kohn-Vossen: Intuitive Geometry, ONTI, 1936.
17. N. A. Glagolev: Elementary Geometry, Parts I and II.
18. R. Dedekind: Continuity and Irrational Numbers, Odessa, 1923.
19. R. Descartes: Geometry, ONTI 1938.
20. B. Delaunay and O. Zhitomirskii: Workbook in Geometry, Gostekhizdat 1949.
21. B. N. Delaunay and D. A. Raikov: Analytical Geometry, vols. I and II, Gostekhizdat 1948-1949.
22. Euclid: Elements, Books I-VI; VII-X; XI-XV, GTI, 1948-1949-1950.
23. The Eighth Book of Euclid's Elements. Translated by F. Petrushevskii. St. Petersburg, 1819.
24. N. B. Efimov: Advanced Geometry, GTI 1949.
25. V. F. Kagan: On the Transformation of Polygons. ONTI, 1933.
26. V. F. Kagan: Lobachevskii. Acad. Sci. USSR, 1944.

BIBLIOGRAPHY

(Note: All entries refer to Russian-language editions.)

1. J. Hadamard: Elementary Geometry, Parts I and II, Uchpedgiz 1936-1948.
2. A. Adler: Theory of Geometrical Constructions, Uchpedgiz 1940.
3. I. I. Alexandrov: Collection of Geometrical Problems In Construction, Uchpedgiz 1950.
4. P. S. Alexandrov: Introduction To The Theory of Groups, Uchpedgiz 1939.
5. P. S. Alexandrov: What Is Non-Euclidean Geometry? "N. I. Lobachevskii" Series. (1793-1943).GTI, 1943.
6. P. S. Alexandrov and V. A. Efremovich: On the Simplest Notions of Contemporary Topology, ONTI, 1935.
7. P. S. Alexandrov and V. A. Efremovich: Outline Of The Basic Concepts Of Topology, ONTI, 1936.
8. S. A. Bogomolov: Geometry (Systematic Course), Uchpedgiz 1949.
9. S. A. Bogomolov: Foundations of Geometry, GIZ, 1923.
10. János Bolyai: Appendix. GITTL, 1950.
11. M. E. Vashchenko-Zakharchenko: Euclid's Elements, with an expository introduction and interpretations. Kiev, 1880.
12. I. M. Vinogradov: Fundamentals of the Theory of Numbers, GTI, 1949.
13. V. M. Voronets: Geometry With a Pair of Compasses, ONTI, 1934.
14. A. B. Gedymin: Cartography, Uchpedgiz 1946.
15. D. Hilbert: Foundations of Geometry, GTI 1948.
16. D. Hilbert and S. Kohn-Vossen: Intuitive Geometry, ONTI, 1936.
17. N. A. Glagolev: Elementary Geometry, Parts I and II.
18. R. Dedekind: Continuity and Irrational Numbers, Odessa, 1923.
19. R. Descartes: Geometry, ONTI 1938.
20. B. Delaunay and O. Zhitomirskii: Workbook in Geometry, Gostekhizdat 1949.
21. B. N. Delaunay and D. A. Raikov: Analytical Geometry, vols. I and II, Gostekhizdat 1948-1949.
22. Euclid: Elements, Books I-VI; VII-X; XI-XV, GTI, 1948-1949-1950.
23. The Eighth Book of Euclid's Elements. Translated by F. Petrushevskii. St. Petersburg, 1819.
24. N. B. Efimov: Advanced Geometry, GTI 1949.
25. V. F. Kagan: On the Transformation of Polygons. ONTI, 1933.
26. V. F. Kagan: Lobachevskii. Acad. Sci. USSR, 1944.

27. F. Klein: Elementary Mathematics From the Higher Point of View vols. I and II, GTI 1933-1938.

28. V. I. Kostin: Foundations of Geometry. Uchpedgiz 1948.

29. B. V. Kutuzov: The Geometry of Lobachevskii and the Elementary Foundations of Geometry. Uchpedgiz 1950.

30. N. I. Lobachevskii: Geometry. Kazan, 1909.

31. N. I. Lobachevskii: Geometrical Researches on the Theory of Parallel Lines. Academy of Sciences of USSR, 1945.

32. N. I. Lobachevskii: Complete Collected Works On Geometry. Kazan, 1883.

33. N. I. Lobachevskii: Complete Collected Works, vols. I, II, IV. GTI, 1946-1949.

34. A. Lebegue: On the Measurement of Magnitudes. Uchpedgiz 1938.

35. Lemaire: Methodological Guide to the Solution of Geometrical Problems. Problems in Construction. 1907.

36. N. N. Luzin: Introduction to the Theory of the Functions of Real Variables. Uchpedgiz 1948.

37. A. I. Markushevich: Real Numbers and the Fundamental Principles of the Theory of Limits. Acad. of Pedagogical Sciences, 1948.

38. P. S. Modenov: Collection of Examination Problems in Mathematics with Analysis of Mistakes Made. State Publishers of Soviet Science, 1950.

39. Gaspard Monge: Descriptive Geometry. Academy of Sciences of USSR, 1947.

40. S. I. Novosiolov: Algebra and Elementary Functions. Uchpedgiz 1950.

41. D. I. Perepiolkin: Course in Elementary Geometry, Parts I and II, GTI 1948-1949.

42. D. I. Perepiolkin: Geometrical Construction in the Middle School: Acad. of Pedagog. Sciences, 1947.

43. Rademacher and Teplitz: Numbers and Figures. ONTI 1936.

44. F. Rudio: On the Squaring of the Circle. ONTI 1936.

45. E. S. Feodorov: Symmetry and the Structure of Crystals. Academy of Sciences of USSR 1949.

46. E. S. Feodorov: The Exact Representation of the Points of Space on a Plane. Memoirs of the Mining Institute, vol. I, St. Petersburg 1907.

47. E. S. Feodorov: Representing the Structure of Crystals by Means of Vectorial Circles. Memoirs of the Mining Institute, vol. I, 1908.

48. E. S. Feodorov: A New Geometry as the Basis of Drawing. St. Petersburg 1907.

49. E. S. Feodorov: The Simple and Exact Representation of the Points of Four-Dimensional Space on the Plane by Means of Vectors. Memoirs of the Mining Institute, vol. II, 1909.

50. G. G. Zeiten: History of Mathematics in Antiquity and the Middle Ages. ONTI, 1932.

51. N. F. Chetverukhin: Methods of Geometrical Construction. Uchpedgiz 1938.

52. P. A. Shirokov: Brief Outline of the Fundamentals of the Geometry of Lobachevskii. Series "Nikolai Ivanovich Lobachevskii (1793-1856)". Acad. Sci. USSR, 1943.

53. Ia. Shteiner: Geometrical Constructions Done With the Straight Line and the Fixed Circle. Uchpedgiz 1939.