M. Postnikov

Lectures in Geometry

SEMESTER I

Analytic Geometry

Mir
Publishers
Moscow

This textbook comprises lectures read by the author to the first-year students of mathematics at Moscow State University.
The book is divided into two parts containing the texts of lectures read in the first and second semesters, respectively.
Part One contains 29 lectures and read in the first semester.
The subject matter is presented on the basis of vector axiomatics of geometry with special emphasis on logical sequence in introduction of the basic geometrical concepts.
Systematic exposition and application of bivectors and trivectors enables the author to successfully combine the above course of lectures with the lectures of the following semesters.
The book is intended for university undergraduates majoring in mathematics.

M. POSTNIKOV

LECTURES IN GEOMETRY

SEMESTER I

ANALYTIC GEOMETRY

ИЗДАТЕЛЬСТВО МИР

М. М. ПОСТНИКОВ

ЛЕКЦИИ ПО ГЕОМЕТРИИ

СЕМЕСТР I

АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ

МОСКВА «НАУКА»
Главная редакция физико-математической литературы

M. POSTNIKOV

LECTURES IN GEOMETRY

SEMESTER I

ANALYTIC GEOMETRY

Translated from the Russian by Vladimir Shokurov

MIR PUBLISHERS
Moscow

First published 1982
Revised from the 1979 Russian edition

На английском языке

CONTENTS

PREFACE TO THE RUSSIAN EDITION

This book is a faithful record of lectures which the author delivered in the first semester of a two-year course in geometry at the Mathematics-Mechanics Faculty of Moscow State University to students studying mathematics. The contents of these lectures were determined by the curriculum, by the established traditions of the Faculty's Department of Higher Geometry and Topology, by the needs of the second semester's course and by the author's personal aims. The sequence of presentation was governed by the necessity of agreement with the concurrently delivered courses in algebra and analysis, by the requirements of the assistants conducting seminars and by other similar considerations of no fundamental consequence but of paramount practical importance. For example, the decision to consider some question or other at one of the last lectures was dictated by the impracticability of consolidating the material of those lectures with the aid of exercises. The contents of the very last lecture were determined by the fact that owing to the postponement of some lectures because of the intervening holidays it would often fall on the examination period and is sometimes not delivered at all for lack of time, etc. etc.

Only two features of the book seem to deserve special mention. The first is that from the outset the exposition is based on axioms and geometric visualization is made use of solely for propaedeutic purposes. For obvious reasons, from the many possible systems of axioms the "vector-point" one developed by H. Weyl has been chosen. This accounts for the unusually early introduction of the general concept of vector space. Experience has shown that as a rule students learn this material without difficulty.

The other, more controversial, feature of the book is a systematic development and use of bivectors and trivectors. This makes it possible to separate distinctly the affine part of the theory from its metric part and provides a background for a general theory of multivectors in the second semester.

Each "lecture" in the book is really a two hours' discourse, as a rule. This explains why a previous topic often gives way to a new one in the middle of a lecture. One exception is the last, 28th, lecture which is a combination of two different versions of the concluding lecture. Because of the specific character of oral and written forms of presentation, the "isochronous" lectures have turned out to be of different lengths in the book. Their number is accounted for by the fact that although the curriculum assigns 36 lectures to a course in analytic geometry, in practice it has to be ended as early as the 28th lecture or earlier.

I also wish to express my gratitude to T.P. Aldatova for her prompt and excellent typing of the original manuscript.

M.M. Postnikov

PREFACE
TO THE ENGLISH EDITION

After the Russian edition of this book appeared some of my fellow lecturers asserted that many of the lectures in the book are far too long to be physically delivered during the allowed two teaching periods. By the right of friendship I had to remind them that a lecturer must prepare for his lectures—even if he has been lecturing for over a dozen years—and make in advance an elaborate, practically minute-by-minute plan of every lecture. It is necessary to consider beforehand the rhythm of the lecture to be delivered—what portions of it are to be read slowly, almost at dictation speed, and what may be said quicker—and its pattern of intonation—where to raise the voice and where to lower it. One also needs a joke somewhere about the middle of the lecture to rouse the tired students and it should be prepared yet at home, and in every detail, up to a play of facial muscles. It goes without saying that one must plan in advance what to write on the blackboard and in what order and where, and when to delete anything, and coordinate all this with everything else. It is surprising how all this extends the limits of lecture time and how much it is then possible to say in an outwardly unhurried and thorough manner, with numerous repetitions and explanations.

Some reviewers have reproached me for a systematic use of bivectors and trivectors saying that one may well do without them. Some well-known physicist, Max Planck, I think, once said that new ideas (he meant scientific ideas but this can be fully applied to methodical ideas as well), could win only when their opponents have retired from the stage as a result of a natural change of generations. An excellent example illustrating this thesis is the introduction of vectors into the courses of analytic geometry half a century ago. Now only a few people remember the fierce discussions concerning this matter and the present generation does not know how many a lance was broken and how much ink split in attempts to prove that vectors were a harmful thing because replacing three equations in coordinates by one vector equation they saved paper but proportionally hampered comprehension. The last of the authoritative opponents of vectors in the USSR died soon after the war but some ten years more passed before diffidently excusatory reservations disappeared altogether from vectorial presentations of geometry (as well as from mechanics and physics where, however, this happened a little earlier). Now bivectors and trivectors are awaiting their turn.

I have taken the opportunity to introduce some minor improvements in the text. The most serious one is perhaps a simpler construction of the complexification of an affine space in Lecture 19. It is true that it contains a certain element of arbitrariness (which was what restrained me at first) but experience has shown that this arbitrariness is perfectly harmless. Besides, the last lecture has been divided into two since two versions of the concluding lecture were combined in it for purely technical, internal editorial reasons. So the book contains 29 lectures now.

As far as I can judge with my poor knowledge of English the translation is well done and conveys all the nuances of my thought.

M.M. Postnikov

May 1, 1980

Lecture 1

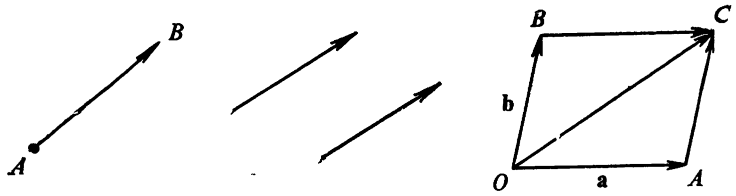
The subject-matter of analytic geometry • Vectors • Vector addition • Multiplication of a vector by a number • Vector spaces • Examples • Vector spaces over an arbitrary field

Analytic geometry, the subject to which these lectures are devoted, is not a definite branch of mathematics but a curriculum (course of studies), with varying contents, which is in the main centred round the concept of coordinates. By tradition this course in any case includes the theory of first- and second-degree curves (straight lines and conic sections) in a plane and, to a smaller extent as a rule, the theory of first- and second-degree surfaces in space. In other respects its contents vary rather greatly.

Before proceeding to analytic geometry proper, i.e. to coordinates, we shall endeavour to provide a reliable axiomatic basis for geometry. It is usual to base an axiomatic construction of elementary geometry, essentially following Euclid in this respect, on the concepts of a point, a straight line, and a plane. Experience shows that this results in a rather complicated axiomatics containing more than a score of axioms which, what is still worse, are not used either entirely or partially anywhere else in mathematics.

It turns out that a much more convenient and simpler system of axioms can be obtained if we base it on the concept of a *vector*. Here simplicity is attained owing to the fact that a "vector" system of axioms uses the theory of real numbers, an extensive portion of which has willy-nilly to be reproduced in "Euclidean"-type axiomatics. Besides, separate fragments of this system of axioms play an exceptionally important role in modern mathematics and must be of necessity studied sooner or later. Thus an axiomatic construction of geometry based on the concept of vector kills two birds with one stone.

A vector is none other than a directed line segment. The direction of a vector is fixed by assuming one of its end points to be the initial point and the other the terminal point.



Vector $\overrightarrow{AB}$ *Equal vectors* *Addition of vectors*

A vector with initial point A and terminal point B is denoted by the symbol $\overrightarrow{AB}$. In a drawing a vector is represented by an arrow.

In physics vectors are forces, velocities, and accelerations.

Two vectors are considered to be equal if they have the same length and the same direction, i.e. are on parallel lines and point in the same direction. We emphasize that here and in what follows *we consider coinciding straight lines to be parallel.*

It is known from physics (mechanics) that the action of two forces upon a particle is equivalent to that of a single force determined from the well-known parallelogram law. Accordingly the sum $\mathbf{a}+\mathbf{b}$ of the two vectors $\mathbf{a}=\overrightarrow{OA}$ and $\mathbf{b}=\overrightarrow{OB}$ is a vector that forms the diagonal $\overrightarrow{OC}$ of the parallelogram whose sides are the vectors **a** and **b**.

Since $\mathbf{b}=\overrightarrow{AC}$, the definition of a sum can be written as a simple formula:

$$\overrightarrow{OA}+\overrightarrow{AC}=\overrightarrow{OC}.$$

The rule of vector addition expressed by this formula is often called the "triangle law".

The operation of vector addition is associative, i.e.

(1) $$\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$$

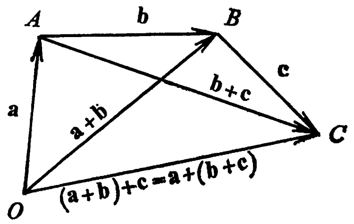
for any vectors **a**, **b**, and **c**. Indeed, if $\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{AB}$ and $\mathbf{c} = \overrightarrow{BC}$, then

$$\mathbf{a} + (\mathbf{b} + \mathbf{c}) = \overrightarrow{OA} + \overrightarrow{AC} = \overrightarrow{OC}$$

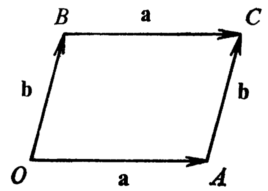
and

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \overrightarrow{OB} + \overrightarrow{BC} + \overrightarrow{OC}.$$

Therefore in the sum of three (or more) vectors brackets



Associativity of vector addition



Commutativity of vector addition

may be omitted: the symbol $\mathbf{a} + \mathbf{b} + \mathbf{c}$ has only one meaning.

The operation of vector addition is also commutative, i.e.

(2) $$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$$

for any vectors **a** and **b**. Indeed, if $\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{OB}$ and $\mathbf{a} + \mathbf{b} = \overrightarrow{OC}$, then $\mathbf{a} = \overrightarrow{BC}$, and therefore

$$\mathbf{b} + \mathbf{a} = \overrightarrow{OB} + \overrightarrow{BC} = \overrightarrow{OC}.$$

The vector $\overrightarrow{AB}$ is also considered when $A = B$. Such a vector $\overrightarrow{AB}$ is called a *zero vector*. It is independent of A and is denoted by the symbol **0**. The length of a zero vector is by definition considered to be zero and it has no direction (it may also be assumed that it has an arbitrary direction).

The formula

$$\overrightarrow{AA} + \overrightarrow{AB} = \overrightarrow{AB}$$

shows that the vector **0** is the zero of addition, i.e.

(3) $$\mathbf{0} + \mathbf{a} = \mathbf{a}$$

for any vector **a**.

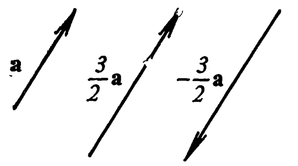
On rearranging the end points of the vector $\mathbf{a} = \overrightarrow{AB}$ we obtain a vector $\overrightarrow{BA}$ which is denoted by the symbol $-\mathbf{a}$. The formula

$$\overrightarrow{AB} + \overrightarrow{BA} = \overrightarrow{AA}$$

shows that the *vector* $-\mathbf{a}$ *is the negative of the vector* **a** *with respect to addition*, i.e.

(4) $$\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$$

for any vector **a**.



Multiplication of a vector by a number

The *product* of a vector **a** by a (real) number k is a vector $k\mathbf{a}$ whose length is equal to the length of the vector **a** multiplied by the absolute value of the number k and whose direction coincides with the direction of the vector **a** if $k > 0$ and is opposite to it if $k < 0$. The cases $\mathbf{a} = \mathbf{0}$ or $k = 0$ are not excluded from this definition. In either of them $k\mathbf{a} = \mathbf{0}$.

The definition of the multiplication of a vector by a number is compatible with that of addition in that, as is easy to see directly from the figure,

$$n\mathbf{a} = \underbrace{\mathbf{a} + \ldots + \mathbf{a}}_{n \text{ times}}$$

for any natural number n. In addition, it is clear that

$$(-1)\,\mathbf{a} = -\mathbf{a}.$$

One can easily see that

(5) $$(k + l)\,\mathbf{a} = k\mathbf{a} + l\mathbf{a}$$

and that

(6) $$(kl)\,\mathbf{a} = k\,(l\mathbf{a})$$

for any numbers k, l and any vector $\mathbf{a}$ (the proof reduces to the enumeration of all possible cases of distributing the signs of the numbers k and l, the statement being obvious in each of the cases).

Also, from the fact that under homothetic transformation (and central symmetry) a parallelogram transforms into a parallelogram it immediately follows that

$$k(\mathbf{a}+\mathbf{b}) = k\mathbf{a} + k\mathbf{b} \tag{7}$$

for any number k and any vectors $\mathbf{a}$ and $\mathbf{b}$.

Finally, it is obvious that

$$1\mathbf{a} = \mathbf{a} \tag{8}$$

for any vector $\mathbf{a}$.

Having established these visual-geometric, intuitive facts, we can now reverse the point of view and accept them as axioms. In these axioms (see relations 1° to 8° below), for reasons which will presently become clear, the symbol $\mathbb{K}$ denotes the set $\mathbb{R}$ of all real numbers.

Definition 1. Let $\mathscr{V}$ be some set whose members we shall call *vectors*, although they may be arbitrary in nature. Suppose that with any two vectors $\mathbf{a} \in \mathscr{V}$ and $\mathbf{b} \in \mathscr{V}$ is in some way associated a third vector denoted by the symbol $\mathbf{a}+\mathbf{b}$ and called the *sum* of the vectors $\mathbf{a}$ and $\mathbf{b}$. In addition suppose that with any number $k \in \mathbb{K}$ and any vector $\mathbf{a} \in \mathscr{V}$ is somehow associated a new vector denoted by the symbol $k\mathbf{a}$ and called the *product* of the vector $\mathbf{a}$ by the number k. If the above properties (1) to (8) hold, i.e. if

1° $\mathbf{a}+(\mathbf{b}+\mathbf{c}) = (\mathbf{a}+\mathbf{b})+\mathbf{c}$ for any vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in \mathscr{V}$;

2° $\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}$ for any vectors $\mathbf{a}$, $\mathbf{b} \in \mathscr{V}$;

3° there exists a vector $\mathbf{0} \in \mathscr{V}$ such that $\mathbf{0}+\mathbf{a} = \mathbf{a}$ for any vector $\mathbf{a} \in \mathscr{V}$;

4° for any vector $\mathbf{a} \in \mathscr{V}$ there exists a vector $-\mathbf{a} \in \mathscr{V}$ such that $\mathbf{a}+(-\mathbf{a}) = \mathbf{0}$;

5° $(k+l)\mathbf{a} = k\mathbf{a} + l\mathbf{a}$ for any numbers k, $l \in \mathbb{K}$ and any vector $\mathbf{a} \in \mathscr{V}$;

6° $(kl)\mathbf{a} = k(l\mathbf{a})$ for any numbers k, $l \in \mathbb{K}$ and any vector $\mathbf{a} \in \mathscr{V}$;

7° $k(\mathbf{a}+\mathbf{b}) = k\mathbf{a} + k\mathbf{b}$ for any number $k \in \mathbb{K}$ and any vectors $\mathbf{a}$, $\mathbf{b} \in \mathscr{V}$;

$8°$ $1\mathbf{a} = \mathbf{a}$ for any vector $\mathbf{a} \in \mathcal{V}$, then the set $\mathcal{V}$ is said to be *vector* (or *linear*) *space.*

When $k = 1/\lambda$ the vector $k\mathbf{a}$ is denoted by the symbol $\mathbf{a}/\lambda$.

We stress that Definition 1 imposes no restrictions on the nature of the members of the set $\mathcal{V}$ (vectors) or on particular realizations of the operations $(\mathbf{a}, \mathbf{b}) \mapsto \mathbf{a} + \mathbf{b}$ (addition) and $(k, \mathbf{a}) \mapsto k\mathbf{a}$ (multiplication by a number). Therefore there may (and do) exist many different vector spaces.

Examples.

1. Even "geometric" vectors (directed segments), with which we began, allow several different vector spaces $\mathcal{V}$ to be constructed. That is, it is possible to consider either various vectors in space, or only vectors in a plane, or only vectors on a straight line. This gives us three different vector spaces.

2. The simplest vector space is $\{\mathbf{0}\}$ consisting of only one zero vector $\mathbf{0}$. To simplify the notation we shall use the same symbol $\mathbf{0}$ for this vector space and for the only vector it contains. We do not think that any confusion will arise.

3. Let X be an arbitrary set and let $\mathcal{F}(X)$ be the set of all (real-valued) functions defined on X. On defining the sum $f + g$ of two functions f and g and the product kf of the function f by the number k in the usual way ("by values"), i.e. by the formulas

$$(f + g)(x) = f(x) + g(x), \quad x \in X,$$
$$(kf)(x) = k(f(x)), \quad x \in X, \tag{9}$$

we verify without difficulty that all the axioms 1° to 8° hold. This means that the set $\mathcal{F}(X)$ is a vector space with respect to operations (9).

Thus it is functions that are "vectors" in this example.

4. When $X = \mathbb{R}$ (or, more generally, when X is an arbitrary subset of the number axis $\mathbb{R}$), it makes sense to speak of the set $\mathcal{S}(X)$ of all continuous functions defined on X. It is known from the course in analysis that the sum $f + g$ of continuous functions and the product kf of a continuous function by a number are continuous functions. Therefore the set $\mathcal{S}(X)$ of all continuous functions is (with respect

to operations (9)) a vector space (axioms 1° to 8° need not be verified, since $\mathscr{S}(X) \subset \mathscr{F}(X)$ and hence these axioms hold automatically).

Similarly, vector spaces are sets of differentiable (a given number of times) functions, sets of functions satisfying a Lipschitz condition and many other classes of functions which mathematical analysis deals with.

These examples of the so-called functional vector spaces explain why in modern function theory the concept of a vector space plays perhaps even a greater part than in geometry. For obvious reasons we shall not concern ourselves here with these vector spaces. The branch of mathematics dealing with them is called "functional analysis". Its foundations are included in the Analysis III course.

5. Various polynomials (involving one variable) must also constitute a vector space.

A vector space will of course be the totality of all polynomials whose degree does not exceed a given number n. (Thus, depending on n we obtain infinitely many different vector spaces.)

We see that vector spaces are of paramount importance in algebra as well.

6. Let n be an arbitrary natural number. Consider the set $\mathbb{R}^n$ of all n-term sequences $(a_1, \ldots, a_n)$ of real numbers. On defining the vector operations "componentwise", i.e. by the formulas

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n),$$
$$k(a_1, \ldots, a_n) = (ka_1, \ldots, ka_n),$$

we obviously turn $\mathbb{R}^n$ into a vector space.

This vector space will play a very important part in what follows and it is therefore necessary to give particular attention to it.

Remark 1 (which should be read when fields and rings have been introduced in the parallel course in algebra). The fact that $\mathbb{K}$ is the set (field) of real numbers is practically not used in Definition 1 at all. Literally the same statement will make sense if by $\mathbb{K}$ we mean an arbitrary field (even of with a finite characteristic). The result is a definition of the vector (linear) space *over the field* $\mathbb{K}$. When $\mathbb{K} = \mathbb{R}$ we return to the former vector spaces.

For definiteness (and with geometric applications in mind) we shall consider in what follows that $\mathbb{K} = \mathbb{R}$. However, all the theory to be developed later is in fact (unless the opposite is explicitly stated) valid for an arbitrary field $\mathbb{K}$.

As a matter of fact, Definition 1 makes sense also when $\mathbb{K}$ is only a ring (with unity), since in axioms 1° to 8° nothing is said about division. In this case the object introduced by Definition 1 is called a *module* (over a ring $\mathbb{K}$). The theory of modules is substantially more complicated than the theory of vector spaces (because of the absence of division) and we shall not be concerned with it.

Remark 2. Axioms 1° to 4° mean that a vector space $\mathcal{V}$ is an Abelian (i.e. commutative) *group* with respect to addition. As to axioms 5° to 8°, they express the fact that a field (ring) $\mathbb{K}$ is a field (ring) of the *operators* of this group. Vector spaces (modules) over $\mathbb{K}$ may thus be said to be (additively written) Abelian groups with a field (ring) of operators $\mathbb{K}$.

Lecture 2

The simplest consequences of the vector space axioms • Independence of the sum of any number of vectors on brackets arrangement • The concept of a family

Geometric intuition is invaluable in the theory of vector spaces and it must be extensively used in both geometrical interpretation of established results and the formulation of new theorems. However, it should be resolutely excluded from proofs, basing these exclusively on axioms 1° to 8°. That is why, in particular, before passing on to really interesting and important concepts and constructions we shall have to derive from axioms 1° to 8° a number of geometrically "obvious" consequences.

Let $\mathscr{V}$ be an arbitrary vector space.

1. Axiom 3°, while stating the existence in $\mathscr{V}$ of a zero vector, says nothing about its uniqueness. Nevertheless it turns out that *there exists only one zero vector*, i.e., if $\mathbf{0}_1$ and $\mathbf{0}_2$ are vectors of $\mathscr{V}$ such that

$$\mathbf{0}_1 + \mathbf{a} = \mathbf{a} \quad \text{and} \quad \mathbf{0}_2 + \mathbf{a} = \mathbf{a}$$

for any vector $\mathbf{a} \in \mathscr{V}$, then $\mathbf{0}_1 = \mathbf{0}_2$. Indeed, setting $\mathbf{a} = \mathbf{0}_1$ in the relation $\mathbf{0}_2 + \mathbf{a} = \mathbf{a}$, we obtain $\mathbf{0}_2 + \mathbf{0}_1 = \mathbf{0}_1$ and setting $\mathbf{a} = \mathbf{0}_2$ in the relation $\mathbf{0}_1 + \mathbf{a} = \mathbf{a}$, we obtain $\mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_2$. Hence $\mathbf{0}_1 = \mathbf{0}_2$. □

2. Similarly, although axiom 4° says nothing about the uniqueness of the negative of $\mathbf{a}$, denoted $-\mathbf{a}$, *this vector is unique*, i.e. if

$$\mathbf{a} + \mathbf{b} = \mathbf{0} \quad \text{and} \quad \mathbf{a} + \mathbf{c} = \mathbf{0},$$

then $\mathbf{b} = \mathbf{c}$. Indeed,

$$\mathbf{b} = \mathbf{0} + \mathbf{b} = (\mathbf{a} + \mathbf{c}) + \mathbf{b} = (\mathbf{a} + \mathbf{b}) + \mathbf{c} = \\ = \mathbf{0} + \mathbf{c} = \mathbf{c}. \ \square$$

3. For any two vectors $\mathbf{a}, \mathbf{b} \in \mathscr{V}$ *the equation*

$$\mathbf{a} + \mathbf{x} = \mathbf{b}$$

has a unique solution $\mathbf{x} = \mathbf{b} + (-\mathbf{a})$. Indeed, if $\mathbf{a} + \mathbf{x} = \mathbf{b}$, then

$$\mathbf{x} = (\mathbf{a} + \mathbf{x}) + (-\mathbf{a}) = \mathbf{b} + (-\mathbf{a})$$

and, on the other hand,

$$\mathbf{a} + (\mathbf{b} + (-\mathbf{a})) = (\mathbf{a} + (-\mathbf{a})) + \mathbf{b} = \mathbf{0} + \mathbf{b} = \mathbf{b}. \ \square$$

We shall write $\mathbf{b} - \mathbf{a}$ instead of $\mathbf{b} + (-\mathbf{a})$ in what follows.

4. Multiplying the directed segment $\mathbf{a}$ by the number 0 obviously yields a zero vector:

$$0\mathbf{a} = \mathbf{0}.$$

It is easy to see that this holds for vectors $\mathbf{a}$ of an arbitrary vector space as well. Indeed,

$$0\mathbf{a} = (0 + 0)\,\mathbf{a} = 0\mathbf{a} + 0\mathbf{a}$$

and thus

$$0\mathbf{a} = 0\mathbf{a} - 0\mathbf{a} = \mathbf{0}. \ \square$$

5. Similarly,

$$k\mathbf{0} = \mathbf{0}$$

for any $k \in \mathbb{K}$. Indeed,

$$k\mathbf{0} = k\,(\mathbf{0} + \mathbf{0}) = k\mathbf{0} + k\mathbf{0}$$

and thus

$$k\mathbf{0} = k\mathbf{0} - k\mathbf{0} = \mathbf{0}. \ \square$$

6. The formula

$$(-1)\,\mathbf{a} = -\mathbf{a}$$

is also valid in any vector space. Indeed,

$$\mathbf{a} + (-1)\,\mathbf{a} = 1\mathbf{a} + (-1)\,\mathbf{a} = (1 - 1)\,\mathbf{a} = 0\mathbf{a} = \mathbf{0}$$

and therefore, by virtue of the uniqueness of the negative of $\mathbf{a}$, $(-1)\,\mathbf{a} = -\mathbf{a}$. $\square$

It is interesting that axiom 2° on the commutativity of addition follows from the remaining axioms and the uniqueness of the negative. Indeed, by virtue of Statement 6, whose proof uses only the uniqueness of the negative,

$$(\mathbf{a}+\mathbf{b})-(\mathbf{b}+\mathbf{a})=\mathbf{a}+\mathbf{b}+(-1)\,(\mathbf{b}+\mathbf{a})=$$
$$=\mathbf{a}+(\mathbf{b}-\mathbf{b})-\mathbf{a}=\mathbf{a}-\mathbf{a}=0,$$

and therefore

$$\mathbf{b}+\mathbf{a}=\mathbf{a}+\mathbf{b}.$$

7. Axiom 1° states that the sum of three vectors does not depend on bracketing, i.e. on the order in which it is computed. It turns out that *a similar statement is also valid for the sum of any number of terms.* In contrast to the previous statements, the proof of this statement is not quite trivial and calls for the introduction of a number of auxiliary concepts.

Adding n terms, we perform $n-1$ additions with any bracketing, one and only one addition being the last to be performed. This means that in the addition of n terms $\mathbf{a}_1, \ldots, \mathbf{a}_n$ for any bracket arrangement there is a uniquely determined index k, $2 \leqslant k \leqslant n$, such that the sum of the terms $\mathbf{a}_1, \ldots, \mathbf{a}_n$ corresponding to that arrangement is of the form

$$\mathfrak{a}+\mathfrak{b}$$

where $\mathfrak{a}$ is the sum of the terms $\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}$ (corresponding to a certain bracket arrangement) and $\mathfrak{b}$ is the sum of the terms $\mathbf{a}_k, \ldots, \mathbf{a}_n$ (also corresponding to a certain bracket arrangement). When $k=2$ the sum $\mathfrak{a}$ reduces to a single term $\mathbf{a}_1$ and when $k=3$ it is the sum $\mathbf{a}_1+\mathbf{a}_2$ without brackets; similarly, when $k=n$ the sum $\mathfrak{b}$ reduces to the term $\mathbf{a}_n$ and when $k=n-1$ it is the sum $\mathbf{a}_{n-1}+\mathbf{a}_n$ without brackets.

We shall call the index k the *rank* of the bracket arrangement under consideration (or the rank of the corresponding sum).

We now define by induction the sum of $n \geqslant 3$ terms *with normal bracketing* (or, briefly, the *normal sum*). For $n=3$ we shall consider normal the sum $(\mathbf{a}_1+\mathbf{a}_2)+\mathbf{a}_3$.

Suppose that the normal sum of $n-1$ terms has already been defined. Then the normal sum of n terms $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is the sum of the form $\mathfrak{a} + \mathbf{a}_n$, where $\mathfrak{a}$ is the normal sum of $n-1$ terms $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$.

Thus the normal sum of n terms is of the form

$$\underbrace{((\ldots}_{n\text{-}2 \text{ brackets}}(\mathbf{a}_1+\mathbf{a}_2)+\mathbf{a}_3)\ldots+\mathbf{a}_n).$$

It is obviously enough for us to prove that *the sum of $n \geqslant 3$ terms with an arbitrary bracket arrangement is equal to the sum of the same terms with the normal bracket arrangement.*

To this end we shall proceed by induction on the number n. For $n = 3$ the statement reduces to axiom 1°.

Suppose it has been proved that each sum of at most $n-1$ terms is equal to their normal sum. Consider an arbitrary sum of n terms $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Let k be its rank. If $k < n$, then our sum is of the form $\mathfrak{a} + \mathfrak{b}$, where $\mathfrak{b}$ is the sum of at least two (and at most $n-1$) terms $\mathbf{a}_k, \ldots, \mathbf{a}_n$. It can be assumed by induction that the sum $\mathfrak{b}$ does not depend on bracket arrangement, so that we can arrange the brackets in it any way we like, without changing the final result (when $k = n-1$ the induction assumption is inapplicable, but in that case the sum $\mathfrak{b}$ has no brackets at all and so there is nothing to speak of).

In particular, we may assume that

$$\mathfrak{b} = \mathbf{a}_k + \mathfrak{c},$$

where $\mathfrak{c}$ is a certain sum (no matter with what bracket arrangement) of the terms $\mathbf{a}_{k+1}, \ldots, \mathbf{a}_n$ (for $k = n-1$, there is a single term, $\mathbf{a}_n$). But then, by axiom 1°,

$$\mathfrak{a} + \mathfrak{b} = \mathfrak{a} + (\mathbf{a}_k + \mathfrak{c}) = (\mathfrak{a} + \mathbf{a}_k) + \mathfrak{c}$$

on the right of which there has resulted a sum of rank $k+1$.

Thus, raising step-by-step the rank of a given sum we obtain a sum of rank n, i.e. a sum of the form $\mathfrak{a}' + \mathbf{a}_n$, where $\mathfrak{a}'$ is a certain sum of $n-1$ terms. By the induction assumption, the last sum is equal to the normal sum $\mathfrak{a}''$. But then the given sum is also equal to the normal sum $\mathfrak{a}'' + \mathbf{a}_n$. $\square$

By virtue of the assertion that has been proved the sum of any number of terms may be written without brackets.

At the next lecture we shall pass on to the more interesting concepts and theorems having a nontrivial geometrical meaning. To do this we shall need the general concept of a family of elements. It is worth recalling it now, since it is often confused with the concept of a subset.

Let X be an arbitrary set. A *family* (or *sequence*) of n elements (or members) of the set X is an arbitrary mapping

$$[1, \ldots, n] \to X \tag{1}$$

of the set $[1, \ldots, n]$ of the first n natural numbers into the set X.

The image of the number i, $1 \leqslant i \leqslant n$, under this mapping, is generally denoted by the symbol x_i and called the *i-th member* of the family and the entire family is denoted by symbol $(x_1, x_2, \ldots, x_n)$ or simply $x_1, x_2, \ldots, x_n$.

Family (1), which is an injection, i.e. such that $x_i \neq x_j$, with $i \neq j$, is called a *nonrecurrent family*. It determines an n-member subset in X consisting of the members $x_1, x_2, \ldots, x_n$. We say that this subset corresponds to a family (it being nothing else but the image of mapping (1)) and that the family has been obtained by a certain *numbering* of the subset.

It is obvious that for any n-member subset there exist $n!$ nonrecurrent families that subset corresponds to. We say that these families are obtained from one another by *renumbering*.

Finally, recall that a *subfamily of the family* $(x_1, \ldots, x_n)$ is an arbitrary family of the form $(x_{i_1}, \ldots, x_{i_m})$, where $1 \leqslant i_1 < \ldots < i_m \leqslant n$.

Lecture 3

Linear dependence and linear independence • Linearly independent sets • The simplest properties of linear dependence • Linear-dependence theorem

Let $\mathscr{V}$ be an arbitrary vector space and let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be a certain family of its elements.

Definition 1. *A linear combination of vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_m$ *with the coefficients* $k_1, \ldots, k_m$ is a vector

$$k_1\mathbf{a}_1 + \ldots + k_m\mathbf{a}_m. \tag{1}$$

This vector is also said to be *linearly expressible* in terms of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

It is clear that the zero vector $\mathbf{0}$ can be linearly expressed in terms of any vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ (it is sufficient to set $k_1 = 0$, $k_2 = 0$, $\ldots$, $k_m = 0$).

Definition 2. The linear combination (1) is said to be *nontrivial* if at least one of the coefficients $k_1, \ldots, k_m$ is nonzero. Otherwise the linear combination is said to be *trivial*.

The trivial linear combination of an arbitrary family of vectors is clearly zero (a zero vector).

Definition 3. A family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is said to be *linearly dependent*, if there is a zero nontrivial linear combination of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$, i.e. if there are such numbers $k_1, \ldots, k_m$, **not all zero**, that

$$k_1\mathbf{a}_1 + \ldots + k_m\mathbf{a}_m = \mathbf{0}.$$

Otherwise the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is said to be *linearly independent*.

To avoid a frequent and annoying discussion of individual cases it is convenient to add to the class of linearly independent families of vectors an *empty family* of vectors corresponding to the case $m = 0$. Thus by definition the *empty family is linearly independent*. This, perhaps somewhat paradoxical, definition agrees quite well with all the statements concerning nonempty families.

It is clear that *any recurrent family is linearly dependent*. Indeed, if, for example, $\mathbf{a}_1 = \mathbf{a}_2$, then

$$1\mathbf{a}_1 + (-1)\,\mathbf{a}_2 + 0\mathbf{a}_3 + \ldots + 0\mathbf{a}_m = \mathbf{0}$$

and two coefficients of this linear combination are nonzero.

Now let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be an arbitrary nonrecurrent family and let $\mathbf{a}'_1, \ldots, \mathbf{a}'_m$ be a family which has been obtained by renumbering it in some way (so that to both families there corresponds the same set of vectors). It is obvious that the family $\mathbf{a}'_1, \ldots, \mathbf{a}'_m$ is linearly dependent if and only if so is the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$. This ensures the correctness of the following definition:

Definition 4. A finite subset of a vector space $\mathscr{V}$ is said to be *linearly independent* (*dependent*), if at least one (and hence every) numbering of its members yields a linearly independent (dependent) family.

According to what has been stated above an *empty set is linearly independent*. As to a *one-member subset* (consisting of a single vector $\mathbf{a}$), *it is linearly independent if and only if* $\mathbf{a} \neq \mathbf{0}$. Indeed, if $k\mathbf{a} = \mathbf{0}$, where $k \neq 0$, then $\mathbf{a} = k^{-1}(k\mathbf{a}) = k^{-1}\mathbf{0} = \mathbf{0}$, and if $\mathbf{a} = \mathbf{0}$, then, for example, $1\mathbf{a} = \mathbf{0}$. □

We now establish some simple but useful properties of the concepts we have introduced.

1. *If a vector* $\mathbf{a}$ *is linearly expressible in terms of vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_m$ *and if every vector* $\mathbf{a}_1, \ldots, \mathbf{a}_m$ *is linearly expressible in terms of vectors* $\mathbf{b}_1, \ldots, \mathbf{b}_n$, *then the vector* $\mathbf{a}$ *can be linearly expressed in terms of the vectors* $\mathbf{b}_1, \ldots, \mathbf{b}_n$.

The proof is obvious. □

This property is called the **transitivity** of linear expressibility.

2. *A family (set) of vectors possessing a linearly dependent subfamily (subset) is linearly dependent.*

It is sufficient to add to the zero nontrivial linear combination of vectors of the subfamily all the other vectors of the family, having furnished them with zero coefficients. □

This property justifies the following definition which is sometimes useful:

Definition 5. An infinite set of vectors is said to be *linearly dependent*, if it has a finite linearly dependent subset, and it is called *linearly independent* if any of its finite subsets is linearly independent.

3. *A family (set) of vectors containing a zero vector is linearly dependent.*

Indeed, in order to obtain a zero nontrivial linear combination, it is sufficient to provide a zero vector with the coefficient 1 and furnish all the other vectors with the coefficients 0. One may also refer to statement 2, since the zero vector constitutes a linearly dependent family. □

4. *A family (set) of vectors is linearly dependent if and only if at least one of its vectors is linearly expressible in terms of the other vectors.*

Indeed, if

$$k_1\mathbf{a}_1 + k_2\mathbf{a}_2 + \ldots + k_m\mathbf{a}_m = \mathbf{0},$$

where, for example, $k_1 \neq 0$, then

$$\mathbf{a}_1 = \left(-\frac{k_2}{k_1}\right)\mathbf{a}_2 + \ldots + \left(-\frac{k_m}{k_1}\right)\mathbf{a}_m.$$

Conversely, if

$$\mathbf{a}_1 = l_2\mathbf{a}_2 + \ldots + l_m\mathbf{a}_m$$

then

$$(-1)\,\mathbf{a}_1 + l_2\mathbf{a}_2 + \ldots + l_m\mathbf{a}_m = \mathbf{0},$$

where $-1 \neq 0$. □

For families (but of course not for sets) of vectors the more exact result also holds.

5. *A family of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is linearly dependent if and only if a certain vector $\mathbf{a}_i$, $1 \leqslant i \leqslant m$, of the family is linearly expressible in terms of the previous vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}$.*

Proof. If a vector $\mathbf{a}_i$, $1 \leqslant i \leqslant m$, is linearly expressible in terms of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}$ (when $i = 1$, in accordance with the general convention accepted above this means that $\mathbf{a}_1 = \mathbf{0}$), then the family $\mathbf{a}_1, \ldots, \mathbf{a}_i$ is linearly dependent (statement 4) and therefore so is the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ (statement 2).

Conversely, let a family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be linearly dependent. Then there exists the smallest i, $1 \leqslant i \leqslant m$, such that the family $\mathbf{a}_1, \ldots, \mathbf{a}_i$ is linearly dependent ($i = 1$ if and only if $\mathbf{a}_1 = \mathbf{0}$). Let

$$k_1\mathbf{a}_1 + \ldots + k_i\mathbf{a}_i = \mathbf{0}$$

be a zero nontrivial linear combination of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_i$. It is clear that $k_i \neq 0$, for when $k_i = 0$ the family $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}$ is linearly dependent. But that contradicts the hypothesis. Consequently it is possible to divide by k_i and we derive that

$$\mathbf{a}_i = \left(-\frac{k_1}{k_i}\right)\mathbf{a}_1 + \ldots + \left(-\frac{k_{i-1}}{k_i}\right)\mathbf{a}_{i-1}. \ \square$$

6. *A family of vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_m$ *is linearly independent if and only if any vector linearly expressible in terms of these vectors can be expressed in terms of them in a unique way.*

Indeed, let a vector $\mathbf{a}$ be expressible in two different ways in terms of given vectors:

$$\mathbf{a} = k_1\mathbf{a}_1 + \ldots + k_m\mathbf{a}_m,$$
$$\mathbf{a} = l_1\mathbf{a}_1 + \ldots + l_m\mathbf{a}_m,$$

where $k_i \neq l_i$ for at least one $i = 1, \ldots, m$. Then, subtracting one equation from the other, we obtain a zero nontrivial linear combination

$$(k_1 - l_1)\,\mathbf{a}_1 + \ldots + (k_m - l_m)\,\mathbf{a}_m = \mathbf{0}$$

of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

Conversely, if there exists a zero nontrivial linear combination

$$\lambda_1\mathbf{a}_1 + \ldots + \lambda_m\mathbf{a}_m = \mathbf{0}$$

then any vector

$$\mathbf{a} = k_1\mathbf{a}_1 + \ldots + k_m\mathbf{a}_m$$

that can be linearly expressed in terms of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ may be expressed in terms of these vectors in the other way as well:

$$\mathbf{a} = (k_1 + \lambda_1)\,\mathbf{a}_1 + \ldots + (k_m + \lambda_m)\,\mathbf{a}_m. \ \square$$

All these properties of linear dependence are more or less trivial. The following property, on the contrary, is by no means trivial. To emphasize this fact we shall formulate it as a theorem.

Theorem 1 (linear dependence). *Let each vector of the family*

$$(2) \qquad \mathbf{a}_1, \ldots, \mathbf{a}_m$$

be linearly expressible in terms of vectors

$$(3) \qquad \mathbf{b}_1, \ldots, \mathbf{b}_n$$

Then, if $m > n$, the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is linearly dependent.

To prove the theorem we shall need the following lemma:

Lemma. *Supposing as before that each vector of family* (2) *is linearly expressible in terms of the vectors* (3), *assume that family* (2) *is linearly independent. Then for any $s = 0, \ldots, n$ there exists a family of vectors:*

$$(4) \qquad \mathbf{c}_1^{(s)}, \ldots, \mathbf{c}_n^{(s)}$$

with the following properties.

(a) *each vector of family* (2) *is linearly expressible in terms of vectors* (4);

(b) *the first s vectors of family* (4) *coincide with the first s vectors of family* (2):

$$\mathbf{c}_1^{(s)} = \mathbf{a}_1, \ \ldots, \ \mathbf{c}_s^{(s)} = \mathbf{a}_s.$$

We shall prove the lemma by induction on the number s. For $s = 0$ it is obvious (family (3) may be taken as family (4)). Suppose it has been proved for s. We prove it for $s + 1$. Consider the family

$$(5) \qquad \mathbf{a}_{s+1}, \mathbf{c}_1^{(s)}, \ldots, \mathbf{c}_n^{(s)}.$$

This family is linearly dependent, for the vector $\mathbf{a}_{s+1}$ can be linearly expressed in terms of the vectors $\mathbf{c}_1^{(s)}, \ldots, \mathbf{c}_n^{(s)}$. Therefore, according to property 5 of linear dependence, some vector of family (5) is linearly expressible in terms of the previous vectors. It cannot be the first vector $\mathbf{a}_{s+1}$, for it is nonzero (under the hypothesis family (2) is linearly independent). If $i \leqslant s$, then the vector $\mathbf{a}_i = \mathbf{c}_i^{(s)}$ is expressible in terms of the vectors $\mathbf{a}_{s+1}, \mathbf{a}_1, \ldots, \mathbf{a}_{i-1}$ and family (2) turns out to be linearly dependent, which is contrary to the assumption. Therefore $i > s$. Consequently, there exists a number i such that the vector $\mathbf{c}_i^{(s)}$ is linearly expressible in terms of the vectors $\mathbf{a}_{s+1}, \mathbf{c}_1^{(s)}, \ldots, \mathbf{c}_{i-1}^{(s)}$.

We now remove from family (5) the vector $\mathbf{c}_i^{(s)}$. The family obtained consists of n vectors and obviously has the property that each vector of family (4) and hence (property 1) each vector of family (2) can be linearly expressed in terms of it. Thus it satisfies condition (a).

This property is obviously preserved if we move the vector $\mathbf{a}_{s+1}$ to the $(s+1)$th position. But then the family will also satisfy condition (b) with respect to the number $s+1$. It completes proof of lemma by induction. □

Proof of Theorem 1. Suppose that the theorem is incorrect, i.e. that family (2) is linearly independent under its hypothesis. Then the lemma proved above would apply to this family. But for $s = n$ it follows from the lemma that each vector of family (2), and in particular the vector $\mathbf{a}_m$, can be linearly expressed in terms of the vectors $\mathbf{c}_1^{(n)} = \mathbf{a}_1, \ldots, \mathbf{c}_n^{(n)} = \mathbf{a}_n$. Since $m > n$, this means that family (1) is linearly dependent. The contradiction obtained proves the theorem. □

Another Proof of Theorem 1. It is known from the algebra course that a system of homogeneous linear equations, with a number of unknowns *greater* than the number of equations, has necessarily a nonzero solution. Theorem 1 follows from this statement almost automatically. Under the hypothesis there are indeed numbers

$$\begin{array}{l} k_{11}, \ \ldots, \ k_{1n}, \\ k_{21}, \ \ldots, \ k_{2n}, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ k_{m1}, \ \ldots, \ k_{mn} \end{array}$$

such that

$$\begin{aligned} \mathbf{a}_1 &= k_{11}\mathbf{b}_1 + \ldots + k_{1n}\mathbf{b}_n, \\ \mathbf{a}_2 &= k_{21}\mathbf{b}_1 + \ldots + k_{2n}\mathbf{b}_n, \\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \mathbf{a}_m &= k_{m1}\mathbf{b}_1 + \ldots + k_{mn}\mathbf{b}_n. \end{aligned}$$

Consider the system of equations

$$\begin{aligned} k_{11}x_1 + k_{21}x_2 + \ldots + k_{m1}x_m &= 0, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ &\cdot \\ k_{1n}x_1 + k_{2n}x_2 + \ldots + k_{mn}x_m &= 0. \end{aligned}$$

Since under the hypothesis $m > n$, by the indicated algebraic theorem there exist numbers $x_1^{(0)}, x_2^{(0)}, \ldots, x_m^{(0)}$, not all zero, satisfying these equations. But then

$$\begin{aligned}
&x_1^{(0)}\mathbf{a}_1 + x_2^{(0)}\mathbf{a}_2 + \ldots + x_m^{(0)}\mathbf{a}_m = \\
&= x_1^{(0)}(k_{11}\mathbf{b}_1 + \ldots + k_{1n}\mathbf{b}_n) + x_2^{(0)}(k_{21}\mathbf{b}_1 + \ldots + k_{2n}\mathbf{b}_n) + \ldots \\
&\qquad \ldots + x_m^{(0)}(k_{m1}\mathbf{b}_1 + \ldots + k_{mn}\mathbf{b}_n) = \\
&\qquad = (k_{11}x_1^{(0)} + k_{21}x_2^{(0)} + \ldots + k_{m1}x_m^{(0)})\mathbf{b}_1 + \ldots \\
&\qquad \ldots + (k_{1n}x_1^{(0)} + k_{2n}x_2^{(0)} + \ldots + k_{mn}x_m^{(0)})\mathbf{b}_n = \\
&\qquad\qquad = 0\mathbf{b}_1 + \ldots + 0\mathbf{b}_n = \mathbf{0},
\end{aligned}$$

and, consequently, the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is linearly dependent. □

Of course, the burden of the proof here falls on algebra.

Lecture 4

Collinear vectors • Coplanar vectors • The geometrical meaning of collinearity and coplanarity • Complete families of vectors, bases, dimensionality • Dimensionality axiom • Basis criterion • Coordinates of a vector • Coordinates of the sum of vectors and those of the product of a vector by a number

In the previous lecture we have established that a one-member vector set is linearly independent if and only if it consists of a nonzero vector. We shall now consider a similar question for two- and three-member sets.

Definition 1. Two vectors are said to be *collinear*, if they constitute a linearly dependent set.

Let $\mathbf{a}$ and $\mathbf{b}$ be collinear vectors. Under the hypothesis there are numbers k and l, at least one of which is nonzero, such that

$$k\mathbf{a} + l\mathbf{b} = \mathbf{0}.$$

If $k \neq 0$, then $\mathbf{a} = h\mathbf{b}$, where $h = -\frac{l}{k}$, and if $l \neq 0$, then $\mathbf{b} = h\mathbf{a}$, where $h = -\frac{k}{l}$. It is clear that, conversely, if $\mathbf{a} = h\mathbf{b}$ or $\mathbf{b} = h\mathbf{a}$, then the vectors $\mathbf{a}$ and $\mathbf{b}$ are collinear.

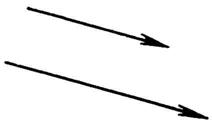
Definition 2. A vector $\mathbf{b}$ is said to be *proportional* to a vector $\mathbf{a}$, if there is a number h such that $\mathbf{b} = h\mathbf{a}$.

We have thus proved that *two vectors are collinear if and only if at least one of them is proportional to the other.* □

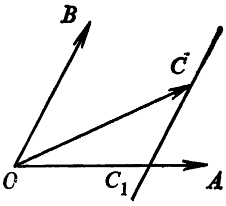
It is useful to keep in mind that if vectors $\mathbf{a}$ and $\mathbf{b}$ are nonzero and if one of them is proportional to the other then the second is proportional to the first. The words "at least one" are hence necessary in order not to exclude the case where one (and only one) of the given vectors is zero.

Definition 3. Three vectors **a**, **b**, **c** are said to be *coplanar*, if they constitute a linearly dependent set.

According to the general theory (see property 4 in the previous lecture) vectors **a**, **b**, **c** are coplanar if and only if one of them is linearly expressed in terms of the others. If none two of the vectors **a**, **b**, **c** are collinear, then each of



Collinear vectors



Vectors lying in the same plane

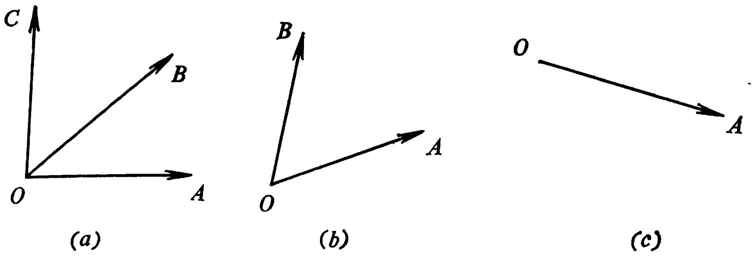
these vectors can be expressed in terms of the others. But if, for example, the vectors **a** and **b** are collinear, then the three vectors **a**, **b**, **c** are coplanar with the vector **c** being arbitrary.

In terms of descriptive geometry collinearity obviously means that both vectors are parallel to the same straight line or (what is the same) are located on the same straight line. What is the geometrical meaning of coplanarity?

Let vectors $\mathbf{a} = \overrightarrow{OA}$ and $\mathbf{b} = \overrightarrow{OB}$ be noncollinear. Then the three points O, A, B determine a unique plane. This plane contains the vectors **a**, **b** and therefore any vector **c** of the form $k\mathbf{a} + l\mathbf{b}$. Thus if vectors **a**, **b**, and **c** are coplanar, then they all lie in some plane (or, what is the same, are parallel to it). It is clear that this conclusion remains valid also when the vectors **a** and **b** are collinear.

Conversely, let three vectors $\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{OB}$, $\mathbf{c} = \overrightarrow{OC}$ be located in the same plane. We prove that they are coplanar. If the vectors **a** and **b** are collinear there is nothing to prove. We may therefore assume that the vectors $\mathbf{a} = \overrightarrow{OA}$ and $\mathbf{b} = \overrightarrow{OB}$ are noncollinear and that therefore the plane

OAB is determined. Under the hypothesis the point C lies in this plane. The straight line passing through the point C



Linearly independent vectors: (a) in space; (b) in the plane; (c) on a straight line

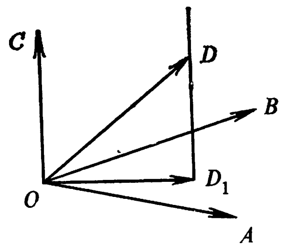
and parallel to the straight line OB intersects the straight line OA in some (unique) point C_1. By the definition of vector addition

$$\overrightarrow{OC} = \overrightarrow{OC_1} + \overrightarrow{C_1C}.$$

But the vector $\overrightarrow{OC_1}$ is parallel to the vector $\overrightarrow{OA} = \mathbf{a}$ and, therefore, proportional to it, so that $\overrightarrow{OC_1} = k\mathbf{a}$. Similarly $\overrightarrow{C_1C} = l\mathbf{b}$. This proves that

$$\mathbf{c} = k\mathbf{a} + l\mathbf{b},$$

i.e. that the vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are coplanar. □



Four vectors in space

If we now take four points O, A, B, and C not in the same plane, then, by what has been proved, the vectors

$$\mathbf{a} = \overrightarrow{OA}, \quad \mathbf{b} = \overrightarrow{OB}, \quad \mathbf{c} = \overrightarrow{OC}$$

will be linearly independent. Thus there are linearly independent triples of vectors in space.

Similarly there are linearly independent pairs of vectors in the plane (it is sufficient to take three points O, A, B not on the same straight line) and linearly independent one-member vector sets on the straight line (it is sufficient to take any two distinct points O and A).

Any two vectors on a straight line, as well as any three vectors in a plane, will be linearly dependent. So will any four vectors in space.

It is indeed sufficient to consider the case where of the four vectors

$$\mathbf{a}=\overrightarrow{OA},\quad \mathbf{b}=\overrightarrow{OB},\quad \mathbf{c}=\overrightarrow{OC},\quad \mathbf{d}=\overrightarrow{OD}$$

the first three vectors **a**, **b**, **c** are noncoplanar, i.e. the points O, A, B, C are not in the same plane (and therefore the points O, A, B are not on the same straight line). Then the straight line passing through the point D and parallel to the straight line OC will meet the plane OAB at some (uniquely determined) point D_1. Since the vectors **a**, **b**, and $\overrightarrow{OD_1}$ are coplanar and the vectors **a** and **b** are noncollinear, the vector $\overrightarrow{OD_1}$ is linearly expressed in terms of the vectors **a** and **b**. Consequently, since

$$\overrightarrow{OD}=\overrightarrow{OD_1}+\overrightarrow{D_1D}$$

and since the vector $\overrightarrow{D_1D}$ is proportional, by construction, to the vector $\mathbf{c}=\overrightarrow{OC}$, the vector $\mathbf{d}=\overrightarrow{OD}$ can be linearly expressed in terms of the vectors **a**, **b**, and **c**.

In order to formulate uniformly the results obtained we set $n = 3$ if we deal with geometry in space (stereometry) and $n = 2$ if we confine ourselves to plane geometry (planimetry). Then, by what has been proved, the following statements hold.

1. *Any family of vectors consisting of more than n vectors is linearly dependent.*

2. *There are linearly independent families of vectors consisting of n vectors.*

For vectors on a straight line these statements hold when $n = 1$.

Statements 1 and 2 do not follow from axioms 1° to 8° of vector space. Therefore in an axiomatic construction of geometry it is necessary either to take them as axioms or to introduce some other axiom with the aid of which it is possible to prove them.

To formulate this axiom it is convenient to introduce the following definition:

Definition 4. The family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ of vectors of a vector space $\mathscr{V}$ is said to be *complete* if any vector of $\mathscr{V}$ is linearly expressible in terms of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

It is easy to see that *if a complete family of vectors is linearly dependent, then it is possible to remove from it one vector in such a way that the remaining family is also complete.*

Indeed, if the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is linearly dependent, then at least one of its vectors can be linearly expressed in terms of the rest. Let for definiteness $\mathbf{a}_m$ be this vector. Then each vector of the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ can be linearly expressed in terms of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{m-1}$ and therefore, by the property of transitivity of linear dependence (see property 1 in the previous lecture), any vector linearly expressible in terms of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ can also be linearly expressed in terms of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{m-1}$. Consequently, if the family of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is complete, so is the family $\mathbf{a}_1, \ldots, \mathbf{a}_{m-1}$. $\square$

Definition 5. A vector space $\mathscr{V}$ in which there are (finite) complete families of vectors is called *finite-dimensional.*

On the strength of the statement we have just proved, *there are complete linearly independent families of vectors in any finite-dimensional vector space $\mathscr{V}$.* Indeed, to obtain such a family it is sufficient to remove from an arbitrary complete family a required number of vectors seeing to it that completeness is preserved.

Definition 6. Every complete linearly independent family of vectors is called a *basis* of a vector space $\mathscr{V}$.

It should be emphasized that by definition it is a family (and not a set) of vectors that is a basis. At the same time, completeness and linear independence are preserved in any renumbering (any interchange) of vectors in a basis. Therefore on interchanging the vectors of a basis we again obtain a basis, but it is a different basis.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be some basis and let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be an arbitrary family of vectors in a vector space $\mathscr{V}$.

Proposition 1. *If the family* $\mathbf{a}_1, \ldots, \mathbf{a}_m$ *is* (a) *linearly independent, then* $m \leqslant n$; *if it is* (b) *complete, then* $m \geqslant n$.

Proof. Since the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a complete family of vectors, any vector of the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is linearly expressible in terms of it. Consequently, by Theorem 1 of the previous lecture, if the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is linearly independent, then $m \leqslant n$.

If the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is complete, then any vector of the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is linearly expressible in terms of it. Since the basis is a linearly independent family of vectors, then $m \geqslant n$ by the same Theorem 1. □

Corollary. *All bases of a finite-dimensional vector space $\mathscr{V}$ contain the same number of vectors.*

Definition 7. This number is called the *dimension* of a vector space $\mathscr{V}$ and denoted by $\dim \mathscr{V}$.

If $\dim \mathscr{V} = n$, then the vector space $\mathscr{V}$ is said to be *n-dimensional* and is usually denoted by $\mathscr{V}^n$.

We remark that $\dim \mathscr{V} = 0$ if and only if $\mathscr{V} = 0$.

We can now formulate an additional axiom 9°. There appears some natural number n in it which we assume to equal either 3 (the "stereometric" version) or 2 (the "planimetric" version). However, formally the case $n = 1$ (where one obtains a trivial "geometry of a straight line") is also possible, as is even the completely degenerate case $n = 0$ ("the geometry of a point").

Axiom 9° (dimensionality axiom). *A vector space $\mathscr{V}$ is finite-dimensional and $\dim \mathscr{V} = n$.*

We shall remark that in such a vector space statements 1 and 2 are valid in an obvious way. Indeed, statement 1 is equivalent to item (a) of Proposition 1 and statement 2 is nothing else but the statement about the existence of bases.

In an n-dimensional vector space each basis
(a) consists of n vectors,
(b) is a complete family,
(c) is a linearly independent family.

It is remarkable that if property (a) holds for some family of vectors, either of properties (b) and (c) follows from the other.

Proposition 2. *A family of vectors of an n-dimensional vector space $\mathscr{V}$ consisting of n vectors is a basis if and only if it is either complete or linearly independent.*

Proof. It is necessary to show that if the family $\mathbf{e}_1, \ldots, \mathbf{e}_n$

is complete, it is also linearly independent, and if it is linearly independent it is also complete.

But if the family $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is complete and linearly dependent, then, by removing from it a suitable vector, we can obtain a complete family consisting of $n-1$ vectors, which by virtue of Proposition 1, item (b), is impossible. Consequently, the complete family $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is linearly independent.

Now let the family $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be linearly independent. It is necessary for us to show that it is complete, i.e. that any vector $\mathbf{a} \in \mathscr{V}$ can be expressed in terms of it. Consider to this end the family $\mathbf{e}_1, \ldots, \mathbf{e}_n, \mathbf{a}$. It contains $n+1$ vectors and is therefore (Proposition 1, item (a)) linearly dependent. Consequently, one of the vectors of this family is linearly expressible in terms of the previous ones. Since the family $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is linearly independent under the hypothesis, that vector can be only the vector $\mathbf{a}$. $\square$

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be an arbitrary basis of a vector space $\mathscr{V}$. Then for any vector $\mathbf{a} \in \mathscr{V}$ there exist uniquely determined numbers $a^1, \ldots, a^n$ (where the superscripts are numbers and not exponents) such that

$$\mathbf{a} = a^1\mathbf{e}_1 + \ldots + a^n\mathbf{e}_n. \tag{1}$$

The existence of the numbers $a^1, \ldots, a^n$ is ensured by the completeness of the basis and their uniqueness by its linear independence.

Definition 8. The numbers $a^1, \ldots, a^n$ are called the *coordinates* of a vector $\mathbf{a}$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Formula (1) is called the *expansion of a vector* $\mathbf{a}$ *with respect to the basis* $\mathbf{e}_1, \ldots, \mathbf{e}_n$. We shall write it (and similar formulas) in the form

$$\mathbf{a} = a^i\mathbf{e}_i$$

assuming that summation is taken over the two repeating indices, the superscript and the subscript, from 1 to n. This abbreviated notation was suggested by Einstein.

Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors and let $a^1, \ldots, a^n$ and $b^1, \ldots, b^n$ be their coordinates (in the same basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$). Then

$$\mathbf{a} = a^1\mathbf{e}_1 + \ldots + a^n\mathbf{e}_n, \quad \mathbf{b} = b'\mathbf{e}_1 + \ldots + b^n\mathbf{e}_n$$

and therefore

$$\begin{aligned}\mathbf{a}+\mathbf{b}&=a^1\mathbf{e}_1+\ldots+a^n\mathbf{e}_n+b^1\mathbf{e}_1+\ldots+b^n\mathbf{e}_n=\\&=a^1\mathbf{e}_1+b^1\mathbf{e}_1+\ldots+a^n\mathbf{e}_n+b^n\mathbf{e}_n=\\&=(a^1+b^1)\,\mathbf{e}_1+\ldots+(a^n+b^n)\,\mathbf{e}_n.\end{aligned}$$

(In Einstein's notation this calculation is written quite briefly: $\mathbf{a}+\mathbf{b}=a^i\mathbf{e}_i+b^i\mathbf{e}_i=(a^i+b^i)\,\mathbf{e}_i$.) This proves that the coordinates of the sum of vectors are the sums of the corresponding coordinates of the summands or, briefly, that in adding vectors their corresponding coordinates are added.

Similarly for any number k we have

$$k\mathbf{a}=k\,(a^1\mathbf{e}_1+\ldots+a^n\mathbf{e}_n)=(ka^1)\,\mathbf{e}_1+\ldots+(ka^n)\,\mathbf{e}_n$$

(i.e. $k\mathbf{a}=k\,(a^i\mathbf{e}_i)=(ka^i)\,\mathbf{e}_i$). Thus when a vector is multiplied by a number its coordinates are multiplied by the same number.

Lecture 5

Isomorphisms of vector spaces • Coordinate isomorphisms • The isomorphism of vector spaces of the same dimension • The method of coordinates • Affine spaces • The isomorphism of affine spaces of the same dimension • Affine coordinates • Straight lines in affine spaces • Segments

The properties of coordinates established in the previous lecture can be stated in more invariant terms.

Definition 1. Let $\mathscr{V}$ and $\mathscr{V}'$ be two vector spaces. The bijection

$$\varphi : \mathscr{V} \to \mathscr{V}'$$

of the space $\mathscr{V}$ onto the space $\mathscr{V}'$ is said to be an *isomorphism* if it transforms a sum into a sum and a product by a number into a product by the same number, i.e. if

$$\varphi (\mathbf{a} + \mathbf{b}) = \varphi (\mathbf{a}) + \varphi (\mathbf{b})$$

and

$$\varphi (k\mathbf{a}) = k\varphi (\mathbf{a})$$

for any vectors $\mathbf{a}, \mathbf{b} \in \mathscr{V}$ and any number k.

The vector spaces $\mathscr{V}$ and $\mathscr{V}'$ are said to be *isomorphic* if there is at least one isomorphism $\mathscr{V} \to \mathscr{V}'$. In this case we write $\mathscr{V} \approx \mathscr{V}'$.

It is clear that the identity mapping $\mathscr{V} \to \mathscr{V}$ is an isomorphism, that the inverse of an isomorphism is an isomorphism, and that a composition (product) of isomorphisms is an isomorphism. It follows that the isomorphism relation is an equivalence relation, i.e. it is reflexive ($\mathscr{V} \approx \mathscr{V}$), symmetrical (if $\mathscr{V} \approx \mathscr{V}'$, then $\mathscr{V}' \approx \mathscr{V}$) and transitive (if $\mathscr{V} \approx \mathscr{V}'$ and $\mathscr{V}' \approx \mathscr{V}''$, then $\mathscr{V} \approx \mathscr{V}''$). Therefore the collection of all vector spaces (over a given field $\mathbb{K}$) falls into nonoverlapping classes of isomorphic spaces.

In the axiomatic theory of linear spaces we are (and can be) only interested in those of their properties which can be expressed in terms of the operation of addition and that of multiplication by a number. It is clear that two isomorphic spaces have identical properties of this kind. Therefore the axiomatic theory treats isomorphic spaces as identical. This makes it possible to extend statements proved for one space to all spaces isomorphic to it. It is virtually on this fact that the **method of coordinates** analytic geometry rests on is based.

Indeed, the coordinates $a^1, \ldots, a^n$ of each vector $\mathbf{a} \in \mathscr{V}$ (in a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$) constitute a sequence $(a^1, \ldots, a^n)$ that is an element of the space $\mathbb{R}^n$ (or an element of a similarly constructed space $\mathbb{K}^n$ in the case of an arbitrary ground field $\mathbb{K}$). Therefore the formula

$$\varphi(\mathbf{a}) = (a^1, \ldots, a^n) \tag{1}$$

determines some, obviously bijective (one-to-one) mapping

$$\varphi\colon \mathscr{V} \to \mathbb{R}^n.$$

The fact that the coordinates of a sum are the sums of the coordinates is obviously expressed by the formula

$$\varphi(\mathbf{a} + \mathbf{b}) = \varphi(\mathbf{a}) + \varphi(\mathbf{b}),$$

and the fact that the coordinates of a vector $k\mathbf{a}$ are the products of the coordinates of the vector $\mathbf{a}$ by the number k is expressed by the formula

$$\varphi(k\mathbf{a}) = k\varphi(\mathbf{a}).$$

Thus we see that the mapping $\varphi\colon \mathscr{V} \to \mathbb{R}^n$ is an isomorphism.

Definition 2. An isomorphism given by formula (1) is called the *coordinate isomorphism* determined by the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Of particular importance in the space $\mathbb{R}^n$ are the n vectors

$$\begin{array}{c}(1, 0, 0, \ldots, 0),\\ (0, 1, 0, \ldots, 0),\\ (0, 0, 1, \ldots, 0),\\ \cdots\cdots\cdots\\ (0, 0, 0, \ldots, 1).\end{array}$$

It is clear that they constitute a basis (called the *standard basis* for the space $\mathbb{R}^n$), the coordinates of an arbitrary vector $(a^1, \ldots, a^n) \in \mathbb{R}^n$ in this basis being its components $a^1, \ldots, a^n$, i.e.

$$(a^1, \ldots, a^n) =$$

$$= a^1 (1, 0, \ldots, 0) + a^2 (0, 1, \ldots, 0) + \ldots + a^n (0, 0, \ldots, 1).$$

The coordinate isomorphism (1) is obviously uniquely characterized as an isomorphism $\mathscr{V} \to \mathbb{R}^n$ transforming a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ for a space $\mathscr{V}$ into the standard basis for a space $\mathbb{R}^n$. It follows that *any isomorphism $\mathscr{V} \to \mathbb{R}^n$ is a coordinate isomorphism corresponding to some basis* (namely, to the basis consisting of vectors transforming under a given isomorphism into the vectors of the standard basis for a space $\mathbb{R}^n$).

The fact of the existence of isomorphisms $\mathscr{V} \to \mathbb{R}^n$ means that the following theorem holds:

Theorem 1. *Every n-dimensional vector space $\mathscr{V}^n$ is isomorphic to a space $\mathbb{R}^n$.* □

We see, in particular, that *any two vector spaces of the same dimension are isomorphic* (and, of course, any two of different dimensions are not).

Thus, although there are an immense number of different vector spaces, there exist only a countable number of classes of isomorphic finite-dimensional spaces. Moreover, for any nonnegative whole number $n \geqslant 0$ there exists one and only one such class; it contains all n-dimensional vector spaces.

Each isomorphism $\varphi\colon \mathscr{V} \to \mathscr{V}'$ of two vector spaces is determined by two bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$ for these spaces and transforms a vector $\mathbf{x} = x^i\mathbf{e}_i$ into the vector $\mathbf{x}' = x^i\mathbf{e}'_i$ having in the basis $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$ the same coordinates as the vector $\mathbf{x}$ has in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. On these grounds it is sometimes said that isomorphism φ is established *from the equality of coordinates.*

The coordinate method of analytic geometry (as applied to vector spaces) consists exactly in using the coordinate isomorphism (1) to replace an arbitrary vector space $\mathscr{V}$ by the quite definite space $\mathbb{R}^n$. The benefit derived from this is that in proving theorems in the space $\mathbb{R}^n$ we can employ all the analytic techniques for handling numbers, which of course significantly simplifies the proofs and often makes it possible to prove the theorems by almost automatic calculation (whereas their derivation from axioms, i.e. what is said to be their "synthetic" proof, nearly always involves a certain degree of ingenuity).

Care must be taken here, however, that the final conclusion has a "geometrical meaning", i.e. that it is stated only in terms of the basic operations and therefore can be transferred into the original space $\mathscr{V}$ by means of the inverse isomorphism φ^{-1}: $\mathbb{R}^n \to \mathscr{V}$. If this condition is not fulfilled, then in general a statement proved in $\mathbb{R}^n$ makes no sense in $\mathscr{V}$ (regardless of the choice of a basis).

The thing is that *isomorphism* (1) *depends on the choice of a basis* and that therefore working in $\mathbb{R}^n$ we automatically include the basis into our investigation. Only those statements have a "geometrical meaning" which do not depend on any arbitrariness in the choice of a basis. Such is, for example, the statement that some vector is zero, but not the statement that its first coordinate is zero.

The tempting idea of completely algebraizing geometry by identifying once and for all, by means of isomorphism (1), every vector space $\mathscr{V}^n$ with the space $\mathbb{R}^n$ does not work exactly because the identification is accomplished with great arbitrariness the restriction of which in any way does not appear possible.

Vectors alone do not suffice of course to construct geometry; one more thing is needed at the minimum, points. To axiomatize the construction of a vector from two points, we introduce the following definition.

Definition 3. *An affine space* is a set $\mathscr{A}$ of members of an arbitrary nature, called *points*, for which there are given

(a) some vector space $\mathscr{V}$;

(b) a mapping associating with any two points $A, B \in \mathscr{A}$ some vector of $\mathscr{A}$ denoted by $\overrightarrow{AB}$ and called a vector with its *initial point* at A and the *terminal point* at B. The following two axioms must be satisfied:

10°. For any point $A \in \mathscr{A}$ and any vector $\mathbf{a} \in \mathscr{V}$ there is a unique point $B \in \mathscr{A}$ for which

$$\overrightarrow{AB} = \alpha.$$

11°. For any three points $A, B, C \in \mathscr{A}$ the following equation holds

$$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}.$$

Setting in axiom 11° $A = B = C$ we find, as was to be expected, that $\overrightarrow{AA} = \mathbf{0}$. Setting now $C = A$ we find that $\overrightarrow{BA} = -\overrightarrow{AB}$.

A vector space $\mathscr{V}$ is said to be *associated* with an affine space $\mathscr{A}$. Its dimension $\dim \mathscr{V}$ is called the *dimension* of the affine space $\mathscr{A}$ and denoted by the symbol $\dim \mathscr{A}$.

For the time being we shall naturally be interested in affine spaces of dimension $n \leqslant 3$. A space of dimension 1 is called a *straight line*, a space of dimension 2 is called a *plane* and a space of dimension 3 is called, alas, a *space*. The confusion is aggravated by the fact that, as we shall see later, the terms straight lines and planes are also used to refer to some subsets of a space. Of course, all this sounds very unpleasant, but such is the established usage.

The part of mathematics concerned with affine spaces is called *affine geometry*. When constructed in the way we do it, this geometry contains two primary undefinable concepts (a point and a vector) and three undefinable relations (the relation between three vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ meaning that the vector $\mathbf{c}$ is the sum of the vectors $\mathbf{a}$ and $\mathbf{b}$; the relation between two vectors $\mathbf{a}$, $\mathbf{b}$ and a number k meaning that $\mathbf{b} = k\mathbf{a}$; the relation between two points A, B and a vector $\mathbf{a}$ meaning that $\overrightarrow{AB} = \mathbf{a}$). These relations must satisfy the eight "vector" axioms 1° to 8°, the dimensionality axiom 9° (for a given n) and the two "affine" axioms 10° and 11°.

Examples of affine spaces.

1. Let $\mathscr{V}$ be an arbitrary vector space. We define an affine space $\mathscr{A}$ by setting $\mathscr{A} = \mathscr{V}$ and $\overrightarrow{\mathbf{ab}} = \mathbf{b} - \mathbf{a}$. Axioms 10° and 11° are obviously satisfied (for any "point" $\mathbf{a} \in \mathscr{A}$ and any vector $\mathbf{c} \in \mathscr{V}$ the "point" $\mathbf{b} = \mathbf{a} + \mathbf{c}$ is obviously the

only point for which $\overrightarrow{\mathbf{ab}} = \mathbf{c}$; for any three points $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ the following equation holds: $\mathbf{c} - \mathbf{a} = (\mathbf{b} - \mathbf{a}) + (\mathbf{c} - \mathbf{b})$). Thus *any vector space $\mathscr{V}$ can be regarded as an affine space.* In this capacity it will sometimes be denoted by $\mathscr{V}_{aff}$.

2. In particular we find that *the set $\mathbb{R}^n$ is naturally an affine space* with which is associated the vector space $\mathbb{R}^n$. (Thus we use the symbol $\mathbb{R}^n$ to denote two different objects: an affine space and a vector space. When these are to be distinguished, one can write, for example, $\mathbb{R}^n_{aff}$ and $\mathbb{R}^n_{vec}$.) If $A = (a^1, \ldots, a^n)$ and $B = (b^1, \ldots, b^n)$ are points of $\mathbb{R}^n_{aff}$, then the vector $\overrightarrow{AB}$ of $\mathbb{R}^n_{vec}$ is defined by the formula

$$\overrightarrow{AB} = (b^1 - a^1, \ldots, b^n - a^n).$$

Let $\mathscr{A}$ and $\mathscr{A}'$ be two affine spaces and let $\mathscr{V}$ and $\mathscr{V}'$ be associated vector spaces.

Definition 4. An *isomorphism* of a space $\mathscr{A}$ onto a space $\mathscr{A}'$ is a bijection

$$\psi: \mathscr{A} \to \mathscr{A}'$$

considered together with some isomorphism $\varphi: \mathscr{V} \to \mathscr{V}'$ of associated vector spaces, such that for any two points $A, B \in \mathscr{A}$ there holds the equation

$$\overrightarrow{\psi(A)\psi(B)} = \varphi(\overrightarrow{AB}).$$

The spaces $\mathscr{A}$ and $\mathscr{A}'$ are said to be *isomorphic* if there exists at least one isomorphism of the space $\mathscr{A}$ onto the space $\mathscr{A}'$. It is clear that the isomorphism relation of affine spaces is an equivalence relation.

It is easy to see that any n-dimensional vector space $\mathscr{V}^n$ regarded as an affine space (see example 1) is isomorphic to the affine space $\mathbb{R}^n$. The corresponding isomorphism ψ: $\mathscr{V}^n_{aff} \to \mathbb{R}^n_{aff}$ is an arbitrary coordinate isomorphism $\mathscr{V}^n \to \mathbb{R}^n$ (with the latter as isomorphism $\varphi: \mathscr{V}^n_{vec} \to \mathbb{R}^n_{vec}$). Thus the isomorphism of $\mathscr{V}^n_{aff}$ with $\mathbb{R}^n_{aff}$ is given by the choice of a basis in $\mathscr{V}$.

On the other hand, *any affine space $\mathscr{A}$ is isomorphic to an associated vector space $\mathscr{V}$ regarded as an affine space.* In order to define such an isomorphism, it is necessary to choose in $\mathscr{A}$

an arbitrary point O and set

(2) $$\psi(A)=\overrightarrow{OA}.$$

It is obvious that the isomorphism $\psi: \mathscr{A} \to \mathscr{V}$ thus defined is an isomorphism of the affine spaces $\mathscr{A}$ and $\mathscr{V}_{aff}$ (in this case φ is the identity mapping $\mathscr{V} \to \mathscr{V}$).

Vector (2) is usually called the *radius vector* of the point A. It should be emphasized that it depends on the choice of the point O.

Combining the two proved statements we obtain the following theorem.

Theorem 2. *Every n-dimensional affine space $\mathscr{A}^n$ is isomorphic to the affine space $\mathbb{R}^n$.*

In other words, *all affine spaces of the same dimension are isomorphic.*

This constitutes the **completeness of axioms in affine geometry:** to within an isomorphism they uniquely determine the corresponding space.

The isomorphism of $\mathscr{A}^n$ onto $\mathbb{R}^n$ is given by an arbitrary point $O \in \mathscr{A}$ (its choice being determined by the isomorphism of $\mathscr{A}^n$ onto $\mathscr{V}^n_{aff}$) and an arbitrary basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ for the space $\mathscr{V}^n$ (its choice being determined by the isomorphism of $\mathscr{V}^n_{aff}$ onto $\mathbb{R}^n_{aff}$).

Definition 5. A collection consisting of a point O and a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is called an *affine coordinate system* in $\mathscr{A}$. It is denoted by $O\mathbf{e}_1 \ldots \mathbf{e}_n$.

For example, when $n = 3$ (in space) the affine coordinate system has the form $O\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$, when $n = 2$ (in the plane) it has the form $O\mathbf{e}_1\mathbf{e}_2$ and when $n = 1$ (on the straight line) it has the form $O\mathbf{e}_1$.

According to what has been said every affine coordinate system $O\mathbf{e}_1 \ldots \mathbf{e}_n$ determines some isomorphism $\psi: \mathscr{A} \to \mathbb{R}^n$. It is called a *coordinate isomorphism.*

Let $A \in \mathscr{A}$ and

$$\psi(A) = (a^1, \ldots, a^n).$$

Definition 6. The numbers $a^1, \ldots, a^n$ are called the *affine coordinates* (or simply *coordinates*) of a point A in the affine coordinate system $O\mathbf{e}_1 \ldots \mathbf{e}_n$.

These coordinates are nothing other than the coordinates of the radius vector $\overrightarrow{OA}$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

$$\overrightarrow{OA} = a^1\mathbf{e}_1 + \ldots + a^n\mathbf{e}_n.$$

When $n = 3$

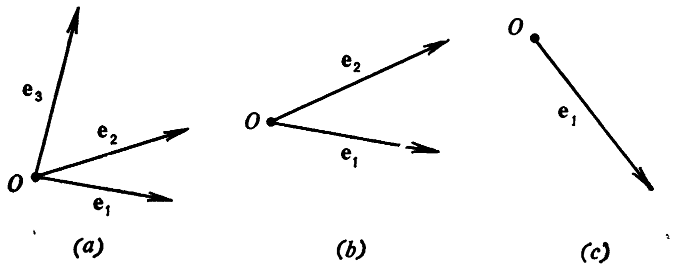
$$\overrightarrow{OA} = a^1\mathbf{e}_1 + a^2\mathbf{e}_2 = a^3\mathbf{e}_3,$$

and when $n = 2$

$$\overrightarrow{OA} = a^1\mathbf{e}_1 + a^2\mathbf{e}_2.$$

All that was said above about the role and significance of the coordinates of vectors can be said without reservation about affine coordinates.

The affine geometry based on axioms 1° to 11° is that part of elementary geometry which makes no use of the



Affine coordinate system: (a) in space; (b) in the plane; (c) on a straight line

mensuration of the lengths of segments and the values of angles. We shall see that this part is not very small. In particular the concept of a straight line makes sense in it.

How should straight lines be introduced axiomatically? To do this it is as ever necessary to consider them on the intuitive-geometric level.

It is clear on this level that any straight line (in the plane or in space) is completely determined by its arbitrary point M_0 and an arbitrary nonzero vector $\mathbf{a}$ parallel to the straight line. Then the condition that some point M should be on the straight line implies that the vector $\overrightarrow{M_0M}$ is collinear with the vector $\mathbf{a}$, i.e. that there exists a number t such that

(3) $$\overrightarrow{M_0M} = t\mathbf{a}.$$

In the axiomatic construction of geometry this statement must be reversed and taken as a definition.

Let $\mathscr{A}$ be an arbitrary affine space with an associated vector space $\mathscr{V}$.

Definition 7. *A straight line* in the space $\mathscr{A}$, given by the point $M_0 \in \mathscr{A}$ and a nonzero vector $\mathbf{a} \in \mathscr{V}$, is a set of all points $M \in \mathscr{A}$ for which the vector $\overrightarrow{M_0M}$ is collinear with the vector $\mathbf{a}$, i.e. for which equation (3) holds for some t.

The vector $\mathbf{a}$ is called the *direction vector of a straight line.* Any vector collinear with the vector $\mathbf{a}$ is said to be *parallel* to the given straight line. On the strength of this definition *vectors are collinear if and only if they are parallel to some straight line* (this fact was earlier established within the framework of intuitive theory; now we have obtained it within the axiomatic theory).

It is easy to see that *each straight line in a space $\mathscr{A}$ is in a natural way provided with the structure of an affine space of dimension* 1 (for this reason one-dimensional affine spaces, even when they are given as abstract spaces, are also called "straight lines"; see above).

It is indeed clear that

(a) all vectors parallel to a straight line form a vector space of dimension 1;

(b) for any points A, B of a straight line the corresponding vector $\overrightarrow{AB} = \overrightarrow{M_0B} - \overrightarrow{M_0A}$ is in that vector space;

(c) the resulting correspondence $(A, B) \mapsto \overrightarrow{AB}$ satisfies axioms 10° and 11°. □

Since at $t = 0$ the point M_0 is obtained we see that the point M_0 is on the straight line considered. Therefore we also say that this straight line *passes through the point M_0 and is parallel to the vector* $\mathbf{a}$.

Clearly, the point M_0 and the vector $\mathbf{a}$ constitute on the straight line (regarded as an affine space of dimension 1) an affine coordinate system. The coordinate of the point M in this system is the number t appearing in relation (3).

Proposition 1. *Let N_0 be an arbitrary point of the straight line passing through a point M_0 and parallel to a vector* **a**, *and let* **b** *be an arbitrary nonzero vector parallel to the straight line. Then the point N_0 and the vector* **b** *give the same straight line.*

Proof. Under the hypothesis there are numbers t_0 and $h_0 \neq 0$ such that

$$\overrightarrow{M_0N_0} = t_0\mathbf{a} \text{ and } \mathbf{b} = h_0\mathbf{a}.$$

Therefore if

$$\overrightarrow{M_0M} = t\mathbf{a}$$

then

$$\overrightarrow{N_0M} = \overrightarrow{M_0M} - \overrightarrow{M_0N_0} = t\mathbf{a} - t_0\mathbf{a} = \tau\mathbf{b},$$

where $\tau = \frac{t-t_0}{h_0}$.

Conversely, if

$$\overrightarrow{N_0M} = \tau\mathbf{b}$$

then

$$\overrightarrow{M_0M} = \overrightarrow{N_0M} + \overrightarrow{M_0N_0} = \tau\mathbf{b} + t_0\mathbf{a} = (\tau h_0 + t_0)\,\mathbf{a} = t\mathbf{a},$$

where $t = \tau h_0 + t_0$. □

Thus *a straight line can be given by any of its points and any nonzero vector parallel to the straight line.*

Moreover, it is easy to see that *for any two points M_0 and M_1 on a straight line the corresponding vector $\overrightarrow{M_0M_1}$ is parallel to that straight line.* Indeed, according to (3) there is t_1 such that

$$\overrightarrow{M_0M_1} = t_1\mathbf{a}.$$

Consequently, *a straight line is uniquely determined by any two of its distinct points M_0 and M_1*, since it is uniquely determined by the point M_0 and the nonzero vector $\overrightarrow{M_0M_1}$. In other words, *not more than one straight line passes through two distinct points M_0 and M_1.*

If there is such a straight line, then its points M are determined from the condition

$$\overrightarrow{M_0M} = t\overrightarrow{M_0M_1},$$

where t is an arbitrary parameter. Conversely, by the definition this relation gives some straight line for any two distinct points M_0 and M_1. Since $M = M_0$ at $t = 0$ and $M = M_1$ at $t = 1$ that straight line passes through the points M_0 and M_1.

Thus we have proved the following proposition.

Proposition 2. *One and only one straight line passes through any two distinct points M_0 and M_1 of an affine space.*

That straight line is denoted by M_0M_1.

The following definition can be introduced in the case where the ground field is the field $\mathbb{R}$ of real numbers:

Definition 8. A point M of a straight line M_0M_1 is said to *lie between* the points M_0 and M_1 if the value of the parameter t corresponding to that point satisfies the inequalities $0 < t < 1$.

The set of all points of a straight line $\overline{M_0M_1}$, which lie between the points M_0 and M_1, together with these points themselves is called a *segment* with end points M_0 and M_1. Thus for the points of the segment $0 \leqslant t \leqslant 1$. The segment is denoted by $\overline{M_0M_1}$.

Lecture 6

Parametric equations of a straight line • The equation of a straight line in a plane • The canonical equation of a straight line in a plane • The general equation of a straight line in a plane • Parallel lines • Relative position of two straight lines in a plane • Uniqueness theorem • Position of a straight line relative to coordinate axes • The half-planes into which a straight line divides a plane

Let an arbitrary point O be chosen in an affine space $\mathscr{A}$. Then relation (3) of the preceding lecture, which defines points of the straight line passing through a point M_0 and parallel to a vector $\mathbf{a}$, can be written in the form

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{a}, \tag{1}$$

where

$$\mathbf{r}_0 = \overrightarrow{OM_0}, \quad \mathbf{r} = \overrightarrow{OM}$$

(recall that $\overrightarrow{M_0M} = \mathbf{r} - \mathbf{r}_0$). As the parameter t changes from $-\infty$ to $+\infty$ the point M with radius vector $\mathbf{r}$ given by formula (1) runs over the entire straight line under consideration. On these grounds equation (1) is called the *parametric vector equation* of a straight line.

Let $n = 2$ (the case of a straight line in a plane). Having chosen an arbitrary affine coordinate system $O\mathbf{e}_1\mathbf{e}_2$, denote the coordinates of a point M (i.e. the coordinates of a vector $\overrightarrow{OM}$ in the basis $\mathbf{e}_1$, $\mathbf{e}_2$) by x, y, the coordinates of a point M_0 by x_0, y_0 and the coordinates of a vector $\mathbf{a}$ (in the basis $\mathbf{e}_1$, $\mathbf{e}_2$) by l, m. Then equation (1) will be equivalent to two numerical equations.

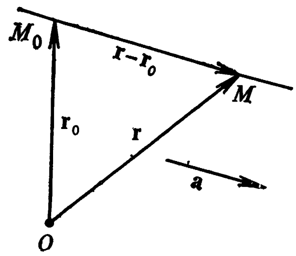
$$x = x_0 + tl,$$
$$y = y_0 + tm. \tag{2}$$

These are called (coordinate) *parametric equations* of a straight line in a plane.

When $n = 3$ (in space) another equation is added to equations (2) and the coordinate parametric equations of a straight line in space assume the form

$$x = x_0 + tl,$$
$$y = y_0 + tm, \tag{3}$$
$$z = z_0 + tn$$

(by tradition the third coordinate of the vector **a** is denoted by the letter n; we have the right to do so, since by fixing dimension 3 we have freed the letter n from denoting dimension).



A straight line given by a point and a vector

When giving a straight line by two points M_0 and M_1 with radius vectors $\mathbf{r}_0$ and $\mathbf{r}_1$ (and by the coordinates x_0, y_0, z_0 and x_1, y_1, z_1, when $n = 3$) we may assume that $\mathbf{a} = \overrightarrow{M_0M_1} = \mathbf{r}_1 - \mathbf{r}_0$ (and, accordingly, that $l = x_1 - x_0$, $m = y_1 - y_0$ and $n = z_1 - z_0$). It follows that the parametric vector equation of the straight line M_0M_1 passing through the points M_0 and M_1 is of the form

$$\mathbf{r} = \mathbf{r}_0 + t(\mathbf{r}_1 - \mathbf{r}_0),$$

i.e. of the form

$$\mathbf{r} = (1 - t)\mathbf{r}_0 + t\mathbf{r}_1, \tag{4}$$

and that its coordinate parametric equations are of the form

$$x = (1-t)x_0 + tx_1,$$
$$y = (1-t)y_0 + ty_1,$$
$$z = (1-t)z_0 + tz_1$$

(in space; in a plane the last equation is lacking).

We shall now consider more closely straight lines in a plane (i.e. in two-dimensional affine space). If an affine coordinate system $O\mathbf{e}_1\mathbf{e}_2$ is fixed in a plane, then each straight line will have parametric equations of the form (2). On eliminating from them the parameter t we obtain the relation

$$(x - x_0)\, m - (y - y_0)\, l = 0. \tag{5}$$

Thus, if a point M is on a straight line, then its coordinates x, y satisfy equation (5). Conversely, if the numbers x, y satisfy equation (5) and if, for example, $l \neq 0$, then relations (2) are satisfied at $t = \frac{x - x_0}{l}$ and if $m \neq 0$, then they are satisfied at $t = \frac{y - y_0}{m}$, i.e. the point M with the coordinates x, y is on a straight line under consideration. On these grounds relation (5) is called the *equation of a straight line* given by a point M_0 and a vector $\mathbf{a}$.

If $m \neq 0$ and $l \neq 0$, then equation (5) may be written in the form

$$\frac{x - x_0}{l} = \frac{y - y_0}{m}. \tag{6}$$

Let us agree that this equation makes sense even when $l = 0$ (but $m \neq 0$) or when $m = 0$ (but $l \neq 0$). That is, assume that when $l = 0$ the equation is equivalent to the relation $x - x_0 = 0$ and when $m = 0$ to the relation $y - y_0 = 0$. By virtue of the agreement equation (6), for any l, m (not vanishing together), is equivalent to equation (5) and hence it is also an equation of a straight line in question.

An equation of the form (6) is called a *canonical equation* of a straight line.

In what follows the symbol $M_0\,(x_0, y_0)$ always denotes that a point M_0 has coordinates x_0, y_0. Similarly, the symbol $\mathbf{a}\,(l, m)$ denotes that a vector $\mathbf{a}$ has coordinates l, m.

For the straight line M_0M_1 passing through points $M_0\,(x_0, y_0)$ and $M_1\,(x_1, y_1)$ the coefficients l and m are expressed by the formulas $l = x_1 - x_0$, $m = y_1 - y_0$.

Thus the canonical equation of a straight line passing through points $M_0\ (x_0, y_0)$ and $M_1\ (x_1, y_1)$ is of the form

$$\frac{x-x_0}{x_1-x_0}=\frac{y-y_0}{y_1-y_0}.$$

This equation may be also written in the form

$$\begin{vmatrix} x-x_0 & y-y_0 \\ x_1-x_0 & y_1-y_0 \end{vmatrix}=0.$$

Setting $A = -m$, $B = l$, we may write equation (5) in the following form:

$$A\ (x - x_0) + B\ (y - y_0) = 0.$$

This is *the general equation* of a straight line passing through a point $M_0\ (x_0, y_0)$.

Setting $C = -Ax_0 - By_0$, we can write the equation in the form

$$Ax + By + C = 0. \tag{7}$$

We remark that *for the straight line* (7) *the direction vector has the coordinates* B, $-A$.

It can easily be seen that *any equation of the form* (7), *where either* $A \neq 0$ *or* $B \neq 0$, *determines some straight line.* Indeed, find numbers x_0, y_0 such that $Ax_0 + By_0 + C = 0$ (if $A \neq 0$ one may take, for example, $x_0 = -C/A$, $y_0 = 0$ and if $B \neq 0$, for example, $x_0 = 0$, $y_0 = -C/B$) and construct a straight line passing through a point $M_0\ (x_0, y_0)$ and parallel to a vector $\mathbf{a}\ (B, -A)$. Equation (5) of this straight line differs only in sign from given equation (7). □

We now consider the question of the relative positions of two straight lines in the plane.

Definition 1. Two straight lines (no matter whether in a plane or in space) are said to be *parallel* if their direction vectors are collinear (and therefore may be chosen equal).

If parallel lines have at least one point in common, then they coincide (for a straight line is uniquely determined by a point and a direction vector). Thus the distinct parallel lines have no points in common (do not intersect).

In the plane the straight lines under consideration have equations of the form

$$Ax + By + C = 0 \tag{8}$$

and

$$A_1x + B_1y + C_1 = 0. \tag{9}$$

Since the direction vectors of these straight lines have, respectively, the coordinates B, $-A$ and B_1, $-A_1$ and vectors are collinear (i.e. proportional) if and only if their coordinates are proportional, the straight lines (8) and (9) are parallel if and only if

$$\frac{A}{A_1} = \frac{B}{B_1}.$$

Here and in what follows formulas of this kind must be perceived as "proportions" of numbers rather than their equality, i.e. as the statement about the existence of such a number $\rho \neq 0$, called the proportionality factor, that $A = \rho A_1$ and $B = \rho B_1$. Therefore it is not excluded that the "denominator" A_1 is zero, which occurs if and only if so is the "numerator" A. This is the same agreement that we employed above in relation to the canonical equations of a straight line.

If the straight lines (8) and (9) have a common point $M_0(x_0, y_0)$, i.e. the equations

$$\begin{aligned} Ax_0 + By_0 + C &= 0, \\ A_1x_0 + B_1y_0 + C_1 &= 0 \end{aligned} \tag{10}$$

hold and if these straight lines are parallel, i.e. if the equations $A = \rho A_1$ and $B = \rho B_1$ hold for some $\rho \neq 0$, then by multiplying the second of the equations (10) by ρ and subtracting from the first, we find that $C - \rho C_1 = 0$. Consequently, if

$$\frac{A}{A_1} = \frac{B}{B_1} \neq \frac{C}{C_1},$$

then equations (10) are impossible, there are no common points and the straight lines (8) and (9) do not intersect, If, on the other hand,

$$\frac{A}{A_1}=\frac{B}{B_1}=\frac{C}{C_1},$$

then either of the equations (8) and (9) is a consequence of the other, so that in this case the straight lines (8) and (9) coincide.

If, however, the straight lines (8) and (9) are nonparallel, i.e. if

$$\frac{A}{A_1}\neq\frac{B}{B_1},$$

then equations (8) and (9) have a unique solution

$$x_0=\frac{BC_1-CB_1}{AB_1-BA_1},\quad y_0=\frac{CA_1-AC_1}{AB_1-BA_1} \tag{11}$$

(these are the so-called Cramer's formulas written for a special case of two equations in two unknowns). This means that nonparallel lines intersect in a single point (with coordinates (11)).

Since the conditions obtained exhaust all the possibilities and do not overlap, each of them is necessary and sufficient. This proves the following theorem.

Theorem 1 (relative positions of two straight lines in a plane). *Two straight lines in a plane*

(a) *either have no point in common, or*

(b) *have one and only one point in common, or*

(c) *coincide.*

Case (a) *is characterized by the fact that*

$$\frac{A}{A_1}=\frac{B}{B_1}\neq\frac{C}{C_1},$$

case (b), *by the fact that*

$$\frac{A}{A_1}\neq\frac{B}{B_1},$$

case (c), *by the fact that*

$$\frac{A}{A_1}=\frac{B}{B_1}=\frac{C}{C_1}.$$

In cases (a) *and* (c) *the straight lines are parallel, and in case* (b) *they are nonparallel.* □

In particular, we see that the *two equations* (8) *and* (9) *determine the same straight line if and only if these equations are proportional.* This statement is known as the **uniqueness theorem** (for straight lines in a plane).

The affine coordinate system $O\mathbf{e}_1\mathbf{e}_2$ determines two remarkable straight lines having, respectively, equations $x = 0$ and $y = 0$. The straight line $x = 0$ is called the *axis of ordinates* (of the coordinate system under consideration) and the straight line $y = 0$ is called the *axis of abscissas.* The axis of ordinates is uniquely characterized as a straight line given by the point $O\,(0, 0)$ and the vector $\mathbf{e}_2$, and the axis of abscissas as a straight line given by the point $O\,(0, 0)$ and the vector $\mathbf{e}_1$.

It follows immediately from the theorem on the relative positions of two straight lines that the *straight line*

$$Ax + By + C = 0 \tag{12}$$

(a) *is parallel to the axis of ordinates if and only if* $B = 0$;
(b) *is parallel to the axis of abscissas if and only if* $A = 0$.

It can be added for completeness that the *straight line* (12) *passes through the origin* O *if and only if* $C = 0$.

If $B \neq 0$, i.e. if the straight line is not parallel to the axis of ordinates, then, setting $k = -A/B$ and $b = -C/B$, we can write its equation in the form familiar from school:

$$y = kx + b.$$

It is important in the definition below that the field $\mathbb{R}$ of real numbers is the ground field $\mathbb{K}$.

Definition 2. Given an arbitrary straight line (12) we say that two points M_1, M_2 of a plane not on the straight line (12) are *nonseparated* (by the straight line (12)) if those points either coincide or (if $M_1 \neq M_2$) the segment $\overrightarrow{M_1M_2}$ and the straight line (12) have no points in common.

To contract notation we set

$$F\,(x, y) = Ax + By + C.$$

Proposition 1. *Points* $M_1\,(x_1, y_1)$ *and* $M_2\,(x_2, y_2)$ *not on the straight line* (12) *are nonseparated if and only if (nonzero) numbers* $F\,(x_1, y_1)$ *and* $F\,(x_2, y_2)$ *have the same signs.*

Proof. We may assume without loss of generality that $M_1 \neq M_2$. Then a straight line M_1M_2 is determined whose coordinate parametric equations are of the form

$$x = (1-t)\,x_1 + tx_2,$$
$$y = (1-t)\,y_1 + ty_2$$

(cf. (4)). To find the points that the straight lines M_1M_2 and (12) have in common (if there are any), it is necessary to replace the expressions for x and y in equation (12) and solve the resulting equation for t. The result of the substitution has obviously the form

$$(1-t)\,F(x_1, y_1) + tF(x_2, y_2) = 0,$$

whence it follows that

$$t = \frac{F(x_1, y_1)}{F(x_1, y_1) - F(x_2, y_2)} \tag{13}$$

(if $F(x_1, y_1) = F(x_2, y_2)$, then there is no solution, i.e. the straight line M_1M_2 is parallel to the straight line (12)).

On the other hand, by definition the points M_1 and M_2 are separated by the straight line (12) if and only if the number (13) exists and satisfies the inequalities $0 < t < 1$.

Thus we see that the points M_1 and M_2 are separated by the straight line (12) if and only if $F(x_1, y_1) \neq F(x_2, y_2)$ and

$$0 < \frac{F(x_1, y_1)}{F(x_1, y_1) - F(x_2, y_2)} < 1.$$

If $F(x_1, y_1) > F(x_2, y_2)$, then this is possible if and only if $F(x_1, y_1) > 0$ and $F(x_2, y_2) < 0$, and if $F(x_1, y_1) < F(x_2, y_2)$, then it is possible if and only if $F(x_1, y_1) < 0$ and $F(x_2, y_2) > 0$. In both cases the numbers $F(x_1, y_1)$ and $F(x_2, y_2)$ have different signs. Therefore the points M_1 and M_2 are nonseparated if and only if those numbers have the same signs. □

It follows immediately from Proposition 1 that the relation of nonseparatedness is an equivalence relation and that there are exactly two respective equivalence classes.

Definition 3. These equivalence classes are called *half-planes* determined by the straight line (12).

Thus two points $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ not on the straight line (12) are in the same half-plane if and only if the segment $\overline{M_1M_2}$ does not intersect that straight line, i.e. if and only if the numbers $Ax_1 + By_1 + C$ and $Ax_2 + By_2 + C$ have the same signs.

Lecture 7

An intuitive notion of a bivector • A formal definition of the bivector • The coincidence of the two definitions • A zero bivector • Conditions for the equality of bivectors • Parallelism of the vector and the bivector • The role of the three-dimensionality condition • Addition of bivectors

Before considering straight lines in space it is appropriate to investigate the basic properties of planes in space. It turns out that the theory of planes in space is completely analogous to the theory of straight lines in the plane. For this analogy to be complete, however, the "planar" counterpart of the concept of a vector is required.

Just as geometrically a vector is a directed line segment (i.e. a part of a straight line), so a "planar vector" "floating" freely in space must be a "directed" part of the plane (an "area element"), in general of arbitrary shape, also floating freely in space. The "directedness" of such an area element means that it has the clockwise or counterclockwise direction of rotation given on it. Two area elements are, in close analogy with vectors, considered to be equal if

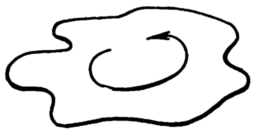
(a) they have the same area;

(b) are parallel to the same plane;

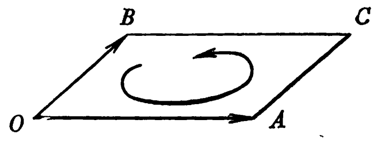
(c) have coincident directions of rotation on them. (These conditions describe what "area elements 'float freely' in space" signifies.)

In view of condition (a) we have no need to consider area elements of arbitrary shape; without loss of generality we may restrict ourselves for example to parallelograms. But the parallelogram $OACB$ is uniquely given by vectors $\mathbf{a} = \overrightarrow{OA}$ and $\mathbf{b} = \overrightarrow{OB}$. And what is more, taken in a definite order, these vectors give in the parallelogram $OABC$ the direction of rotation as well (for example, for definiteness we may agree that the vectors $\mathbf{a} = \overrightarrow{OA}$ and $\mathbf{b} = \overrightarrow{OB}$— taken exactly in that order!—determine a rotation about the point O such that the point A moves along the shortest arc to the point B).

All this means that instead of area elements we may consider ordered pairs $(\mathbf{a}, \mathbf{b})$ of vectors. Two such pairs



"*Planar vector*"



Bivector

are considered to be equal (or, rather, equivalent) if parallelograms constructed on them satisfy conditions (a), (b), and (c). The class of equivalent pairs (this is just the "planar" counterpart of a vector) is called a *bivector*.

From the intuitive point of view conditions (a), (b) and (c) are quite clear, but unfortunately we cannot include them in our axiomatic system, since it has as yet no such concepts as "area" and "plane". And what is more, it is on the basis of the concept of bivector that we are going to introduce these concepts later on. So, to avoid the vicious circle, we must introduce the equivalence relation between pairs of vectors in a different way having meaning in any vector space.

Definition 1. Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{a}_1$ and $\mathbf{b}_1$ be vectors of an arbitrary vector space $\mathscr{V}$. One says that the pair $(\mathbf{a}_1, \mathbf{b}_1)$ is obtained from the pair $(\mathbf{a}, \mathbf{b})$ by an *elementary transformation* and writes $(\mathbf{a}, \mathbf{b}) \Rightarrow (\mathbf{a}_1, \mathbf{b}_1)$ if either

(1) $\mathbf{a}_1 = \mathbf{a}$, $\mathbf{b}_1 = \mathbf{b} + k\mathbf{a}$ or $\mathbf{a}_1 = \mathbf{a} + k\mathbf{b}$, $\mathbf{b}_1 = \mathbf{b}$,

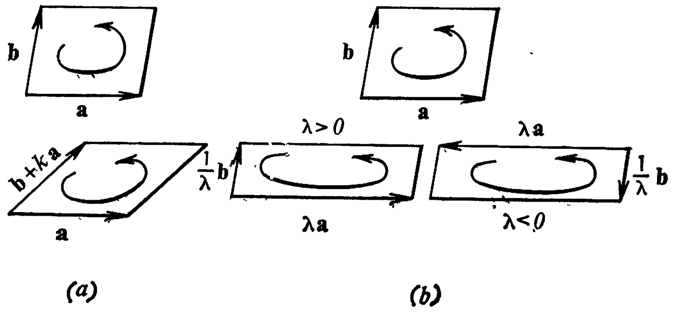
where k is an arbitrary number (an element of the ground field $\mathbb{K}$) or

(2) $$\mathbf{a}_1 = \lambda\mathbf{a}, \quad \mathbf{b}_1 = \frac{1}{\lambda}\mathbf{b},$$

where $\lambda \neq 0$ is an arbitrary nonzero number.

We draw attention to the fact that if the vectors $\mathbf{a}$, $\mathbf{b}$ are noncollinear, then so are the vectors $\mathbf{a}_1$ and $\mathbf{b}_1$. This makes meaningful the following definition.

Definition 2. Two pairs of vectors are said to be *equivalent* if either each of them consists of collinear vectors or one may be obtained from the other by means of some sequence of elementary transformations.



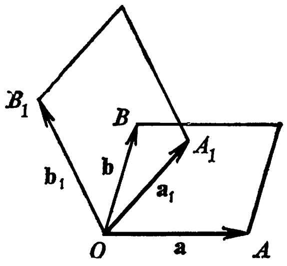
(a) Elementary transformation (1), (b) elementary transformation (2)

Clearly, $(\mathbf{a}, \mathbf{b}) \Rightarrow (\mathbf{a}, \mathbf{b})$ (it is sufficient to apply (1) with $k = 0$). Moreover, if $(\mathbf{a}, \mathbf{b}) \Rightarrow (\mathbf{a}_1, \mathbf{b}_1)$, then $(\mathbf{a}_1, \mathbf{b}_1) \Rightarrow$ $\Rightarrow (\mathbf{a}, \mathbf{b})$ in case (1) $\mathbf{a} = \mathbf{a}_1$ and $\mathbf{b} = \mathbf{b}_1 + k_1\mathbf{a}_1$ or $\mathbf{a} = \mathbf{a}_1 +$ $+ k_1\mathbf{b}_1$ and $\mathbf{b} = \mathbf{b}_1$, where $k_1 = -k$, and in case (2) $a =$ $= \lambda_1\mathbf{a}_1$ and $\mathbf{b} = \frac{1}{\lambda_1}\mathbf{b}_1$, where $\lambda_1 = \frac{1}{\lambda}$. It follows that the relation "to be obtained by means of a sequence of elementary transformations" is indeed an equivalence relation.

Definition 3. The respective equivalence classes are called ***bivectors*** of a vector space $\mathscr{V}$. A bivector determined by a pair $(\mathbf{a}, \mathbf{b})$ will be designated by the symbol $\mathbf{a} \wedge \mathbf{b}$ and the set of all bivectors by the symbol $\mathscr{V} \wedge \mathscr{V}$.

For this definition to be justified, we must return to the intuitive point of view and show that the resulting bivectors coincide (for noncollinear pairs) with those introduced above, i.e. with bivectors as area elements. In other words, we must show that the formal equivalence relation introduced in Definition 2 coincides with the "geometrical" definition based on conditions (a), (b), and (c).

The parallelograms constructed on the vectors **a**, **b** and those constructed on **a**, $\mathbf{b} + k\mathbf{a}$ obviously have the same base and the same altitude. Therefore their areas are equal (condition (a)). They are both located in the same plane (condition (b)) and, as can be seen directly from the drawing, the directions of rotations on them coincide (condition (c)). The case of the elementary transformation (2) is considered in a similar way. Thus, if pairs of (noncollinear) vectors are equivalent in the sense of Definition 2, then they are equivalent as area elements as well.



Equivalent area elements

Conversely, let the area elements (**a**, **b**) and $(\mathbf{a}_1, \mathbf{b}_1)$ be equivalent. Then all the four vectors **a**, **b**, $\mathbf{a}_1$, $\mathbf{b}_1$ lie in the same plane (condition (b)), and since the vectors **a** and **b** are noncollinear under the hypothesis, it is possible to expand the vectors $\mathbf{a}_1$ and $\mathbf{b}_1$ with respect to them. Hence we have equations of the form

$$\text{(3)} \qquad \begin{aligned} \mathbf{a}_1 &= k\mathbf{a} + l\mathbf{b}, \\ \mathbf{b}_1 &= k_1\mathbf{a} + l_1\mathbf{b}, \end{aligned}$$

with $\delta = kl_1 - lk_1 \neq 0$ (for otherwise the vectors $\mathbf{a}_1$ and $\mathbf{b}_1$ would be collinear).

In algebra the number δ is designated by the symbol $\begin{vmatrix} k & l \\ k_1 & l_1 \end{vmatrix}$ and referred to as the determinant.

Lemma 1. *If vectors* **a**, **b**, $\mathbf{a}_1$, $\mathbf{b}_1$ *of an arbitrary vector space* $\mathscr{V}$ *are connected by relations* (3), *with* $\delta \neq 0$, *then the pair* (**a**, **b**) *can be converted into the pair* $\left(\mathbf{a}_1, \frac{1}{\delta}\mathbf{b}_1\right)$ *by elementary transformations.*

Proof. If $k \neq 0$, then

$$(\mathbf{a}, \mathbf{b}) \Rightarrow \left(\mathbf{a} + \frac{l}{k}\mathbf{b}, \mathbf{b}\right) \Rightarrow \left(k\mathbf{a} + l\mathbf{b}, \frac{1}{k}\mathbf{b}\right) \Rightarrow$$

$$\Rightarrow \left(k\mathbf{a}+l\mathbf{b}, \frac{1}{k}\mathbf{b}+\frac{k_1}{k\delta}(k\mathbf{a}+l\mathbf{b})\right)=$$
$$=\left(k\mathbf{a}+l\mathbf{b}, \frac{1}{\delta}(k_1\mathbf{a}+l_1\mathbf{b})\right)=\left(\mathbf{a}_1, \frac{1}{\delta}\mathbf{b}_1\right).$$

Similarly, if $k = 0$ (and therefore $l \neq 0$ and $k_1 \neq 0$), then, since

(4) $$(\mathbf{a}, \mathbf{b}) \Rightarrow (\mathbf{a}, \mathbf{b}-\mathbf{a}) \Rightarrow (\mathbf{a}+(\mathbf{b}-\mathbf{a}), \mathbf{b}-\mathbf{a}) = = (\mathbf{b}, \mathbf{b}-\mathbf{a}) \Rightarrow (\mathbf{b}, -\mathbf{a})$$

(this intermediate result will yet prove useful to us) and

$$(\mathbf{b}, -\mathbf{a}) \Rightarrow \left(l\mathbf{b}, -\frac{1}{l}\mathbf{a}\right) \Rightarrow \left(l\mathbf{b}, -\frac{1}{l}\mathbf{a}-\frac{l_1}{l^2k_1}l\mathbf{b}\right)=$$
$$=\left(l\mathbf{b}, \frac{1}{\delta}(k_1\mathbf{a}+l_1\mathbf{b})\right)=\left(\mathbf{a}_1, \frac{1}{\delta}\mathbf{b}_1\right)$$

we again find that the pair $(\mathbf{a}_1, \frac{1}{\delta}\mathbf{b}_1)$ is obtained from the pair $(\mathbf{a}, \mathbf{b})$ by a sequence of elementary transformations. □

It follows from this lemma, by virtue of what has already been established, that the area element $(\mathbf{a}_1, \frac{1}{\delta}\mathbf{b}_1)$ is equivalent to the area element $(\mathbf{a}, \mathbf{b})$ and hence to the area element $(\mathbf{a}_1, \mathbf{b}_1)$. But clearly the area of the area element $(\mathbf{a}_1, \frac{1}{\delta}\mathbf{b}_1)$ is equal to that of the area element $(\mathbf{a}_1, \mathbf{b}_1)$ multiplied by $\frac{1}{\delta}$ and the directions of rotation on them coincide if and only if $\frac{1}{\delta} > 0$. Therefore (conditions (a) and (c)) $\delta = 1$. Hence the pair $(\mathbf{a}_1, \mathbf{b}_1)$ is obtained from the pair $(\mathbf{a}, \mathbf{b})$ by a sequence of elementary transformations. □

Thus we are fully justified in making Definitions 1 to 3.

We remark that by definition all collinear pairs give the same bivector. It is called a zero bivector and designated by the symbol 0. Thus, by definition, the *vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *are collinear if and only if* $\mathbf{a} \wedge \mathbf{b} = 0$.

Let two bivectors $\mathbf{a} \wedge \mathbf{b}$ and $\mathbf{a}_1 \wedge \mathbf{b}_1$ be given. How is one to find out whether these bivectors are equal or not? If at least one of them is zero, the answer is trivial: the bivectors are equal if the vectors $\mathbf{a}$, $\mathbf{b}$, as well as the vectors $\mathbf{a}_1$, $\mathbf{b}_1$, are collinear. Therefore without loss of generality we can seek the answer only for nonzero bivectors.

Proposition 1. *Two nonzero bivectors* $\mathbf{a} \wedge \mathbf{b}$ *and* $\mathbf{a}_1 \wedge \mathbf{b}_1$ *are equal if and only if equations of the form*

$$\mathbf{a}_1 = k\mathbf{a} + l\mathbf{b}, \qquad \mathbf{b}_1 = k_1\mathbf{a} + l_1\mathbf{b} \tag{5}$$

hold and the number $\delta = kl_1 - lk_1 = \begin{vmatrix} k & l \\ k_1 & l_1 \end{vmatrix}$ *is equal to unity.*

Proof. According to Lemma 1, if equations (5) hold, then $\mathbf{a} \wedge \mathbf{b} = \mathbf{a}_1 \wedge \frac{1}{\delta}\mathbf{b}_1$. Therefore, with $\delta = 1$, the bivector $\mathbf{a}_1 \wedge \mathbf{b}_1$ is equal to the bivector $\mathbf{a} \wedge \mathbf{b}$.

To prove the converse it is evidently sufficient to establish that

(a) if $(\mathbf{a}, \mathbf{b}) \Rightarrow (\mathbf{a}_1, \mathbf{b}_1)$, then relations (5) hold, with $\delta = 1$;

(b) if the pair $(\mathbf{a}_1, \mathbf{b}_1)$ is connected with the pair $(\mathbf{a}, \mathbf{b})$ by relations of the form (5), with $\delta = 1$, and if a pair $(\mathbf{a}_2, \mathbf{b}_2)$ is similarly connected with the pair $(\mathbf{a}_1, \mathbf{b}_1)$, then the pair $(\mathbf{a}_2, \mathbf{b}_2)$ is also connected with the pair $(\mathbf{a}, \mathbf{b})$ by relations of the form (5), with $\delta = 1$.

But statement (a) is obvious, and statement (b) is verified directly by calculation: if

$$\mathbf{a}_1 = k\mathbf{a} + l\mathbf{b}, \quad \mathbf{a}_2 = k'\mathbf{a}_1 + l'\mathbf{b}_1$$

and

$$\mathbf{b}_1 = k_1\mathbf{a} + l_1\mathbf{b} \quad \mathbf{b}_2 = k_1'\mathbf{a}_1 + l_1'\mathbf{b}_1,$$

then

$$\mathbf{a}_2 = (kk' + k_1l')\,\mathbf{a} + (lk' + l_1l')\,\mathbf{b},$$
$$\mathbf{b}_2 = (kk_1' + k_1l_1')\,\mathbf{a} + (lk_1' + l_1l_1')\,\mathbf{b},$$

with

$$(kk' + k_1l')(lk_1' + l_1l_1') - (lk' + l_1l')(kk_1' + k_1l_1') = \\ = (kl_1 - lk_1)(k'l_1' - l'k_1').$$

(An informed reader will immediately discover here the formula for matrix multiplication and the theorem on the determinant of the product and thus get rid of any calculations.) □

Definition 4. We shall say that a vector $\mathbf{e}$ is *parallel* to a nonzero bivector $\mathfrak{a} = \mathbf{a} \wedge \mathbf{b}$ and write $\mathbf{e} \parallel \mathfrak{a}$ if it is linearly expressed in terms of the vectors $\mathbf{a}$ and $\mathbf{b}$. If $\mathfrak{a} = 0$, then we shall assume by definition that $\mathbf{e} \parallel \mathfrak{a}$ for any vector $\mathbf{e}$.

It follows from Proposition 1 that the definition is correct (that it is independent of the arbitrariness in the choice of the vectors $\mathbf{a}$ and $\mathbf{b}$).

Note that according to Definition 4 the zero vector is parallel to any bivector.

Clearly, if $\mathfrak{a} = \mathbf{e} \wedge \mathbf{a}$, then $\mathbf{e} \parallel \mathfrak{a}$. The converse is true in the following form.

Proposition 2. *If* $\mathbf{e} \parallel \mathfrak{a}$ *and* $\mathbf{e} \neq \mathbf{0}$, *then there exists a vector* $\mathbf{a}$ *such that*

$$\mathfrak{a} = \mathbf{e} \wedge \mathbf{a}.$$

Proof. If $\mathfrak{a} = 0$, we may set $\mathbf{a} = \mathbf{e}$ (or $\mathbf{a} = \mathbf{0}$). Let $\mathfrak{a} = \mathbf{a}_1 \wedge \mathbf{b}_1 \neq 0$. Under the hypothesis there are numbers k and l such that

$$\mathbf{e} = k\mathbf{a}_1 + l\mathbf{b}_1.$$

Select numbers k_1 and l_1 such that $kl_1 - lk_1 = 1$ (for example, if $k \neq 0$, we may set $k_1 = 0$ and $l_1 = k^{-1}$, and if $k = 0$, we may set $k_1 = -l^{-1}$ and $l_1 = 0$; the case $k = 0$ and $l = 0$ is impossible by virtue of the condition $\mathbf{e} \neq \mathbf{0}$) and set

$$\mathbf{a} = k_1\mathbf{a}_1 + l_1\mathbf{b}_1.$$

Then, by Proposition 1, the equation $\mathfrak{a} = \mathbf{e} \wedge \mathbf{a}$ will hold. $\square$

Thus far in our discussion of bivectors we have in no way used the dimensionality axiom 9°. We now suppose that $n = \dim \mathscr{V} \leqslant 3$ (in fact only the case $n = 3$ is interesting).

Proposition 3. *For any two bivectors* $\mathfrak{a}$ *and* $\mathfrak{b}$ *there exists a nonzero vector* $\mathbf{e}$ *such that* $\mathbf{e} \parallel \mathfrak{a}$ *and* $\mathbf{e} \parallel \mathfrak{b}$.

Proof. If at least one of the bivectors, $\mathfrak{a}$ or $\mathfrak{b}$, is zero, then the existence of a vector $\mathbf{e}$ is obvious. Therefore without loss of generality we may assume that $\mathfrak{a} \neq 0$ and $\mathfrak{b} \neq 0$.

Let $\mathfrak{a} = \mathbf{a} \wedge \mathbf{a}_1$ and $\mathfrak{b} = \mathbf{b} \wedge \mathbf{b}_1$. The four vectors $\mathbf{a}$, $\mathbf{a}_1$,

$\mathbf{b}$, $\mathbf{b}_1$ must necessarily be linearly dependent in a vector space of dimension $\leqslant 3$, i.e. there exist numbers k, k_1, l and l_1, not all zero, such that

$$k\mathbf{a} + k_1\mathbf{a}_1 + l\mathbf{b} + l_1\mathbf{b}_1 = 0.$$

We set

$$\mathbf{e} = k\mathbf{a} + k_1\mathbf{a}_1 = -(l\mathbf{b} + l_1\mathbf{b}_1).$$

It is clear that $\mathbf{e} \parallel \mathfrak{a}$ and $\mathbf{e} \parallel \mathfrak{b}$. In addition $\mathbf{e} \neq \mathbf{0}$, for the vectors $\mathbf{a}$ and $\mathbf{a}_1$ are linearly independent. □

Definition 5. The *sum* of two bivectors of the form $\mathbf{e} \wedge \mathbf{a}$ and $\mathbf{e} \wedge \mathbf{b}$ is a bivector $\mathbf{e} \wedge (\mathbf{a} + \mathbf{b})$. Thus, by definition,

$$\mathbf{e} \wedge \mathbf{a} + \mathbf{e} \wedge \mathbf{b} = \mathbf{e} \wedge (\mathbf{a} + \mathbf{b}). \tag{6}$$

It follows from Propositions 3 and 2 that this construction allows determination of the sum of any two bivectors $\mathfrak{a}$ and $\mathfrak{b}$: using Proposition 3 we find a vector $\mathbf{e} \neq \mathbf{0}$ such that $\mathbf{e} \parallel \mathfrak{a}$ and $\mathbf{e} \parallel \mathfrak{b}$; then, using Proposition 2, we find vectors $\mathbf{a}$ and $\mathbf{b}$ such that $\mathfrak{a} = \mathbf{e} \wedge \mathbf{a}$ and $\mathfrak{b} = \mathbf{e} \wedge \mathbf{b}$; we finally set

$$\mathfrak{a} + \mathfrak{b} = \mathbf{e} \wedge (\mathbf{a} + \mathbf{b}).$$

This construction of the bivector $\mathfrak{a} + \mathfrak{b}$ is seen to contain a considerable amount of arbitrariness. It is therefore necessary to prove the correctness of this definition, i.e. the independence of the sum $\mathfrak{a} + \mathfrak{b}$ on the arbitrariness in the choice of the vectors $\mathbf{e}$, $\mathbf{a}$, and $\mathbf{b}$. This will be proved in the next lecture.

Lecture 8

The correctness of the definition of a bivector sum • The product of a bivector by a number • Algebraic properties of external product • The vector space of bivectors • Bivectors in a plane and the theory of areas • Bivectors in space

In proving the correctness of the definition of the sum $\mathfrak{a} + \mathfrak{b}$ of the bivectors $\mathfrak{a} = \mathbf{e} \wedge \mathbf{a}$ and $\mathfrak{b} = \mathbf{e} \wedge \mathbf{b}$ several cases have to be considered.

Let $\mathbf{e} \wedge \mathbf{a} = \mathbf{e}' \wedge \mathbf{a}'$ and $\mathbf{e} \wedge \mathbf{b} = \mathbf{e}' \wedge \mathbf{b}'$, with $\mathbf{e} \wedge \mathbf{a} \neq 0$ and $\mathbf{e} \wedge \mathbf{b} \neq 0$. Then, according to Proposition 1 of the preceding lecture, the equations

$$\mathbf{e}' = k\mathbf{e} + l\mathbf{a}, \quad \mathbf{e}' = k'\mathbf{e} + l'\mathbf{b},$$

$$\mathbf{a}' = k_1\mathbf{e} + l_1\mathbf{a}, \quad \mathbf{b}' = k_1'\mathbf{e} + l_1'\mathbf{b}$$

hold, with $kl_1 - lk_1 = 1$ and $k'l_1' - l'k_1' = 1$.

Case $l = 0$, i.e. $\mathbf{e}' = k\mathbf{e}$, where $k \neq 0$. Then $(k' - k)\,\mathbf{e} + l'\mathbf{b} = \mathbf{0}$ and therefore $k' = k$ and $l' = 0$ (for under the hypothesis the vectors $\mathbf{e}$ and $\mathbf{b}$ are linearly independent). In addition, $l_1 = l_1' = k^{-1}$. Hence in this case

$$(\mathbf{e}', \mathbf{a}' + \mathbf{b}') = (k\mathbf{e}, (k_1 + k_1')\,\mathbf{e} + k^{-1}(\mathbf{a} + \mathbf{b})) \Rightarrow$$

$$\Rightarrow (\mathbf{e}, k(k_1 + k_1')\,\mathbf{e} + (\mathbf{a} + \mathbf{b})) \Rightarrow (\mathbf{e}, \mathbf{a} + \mathbf{b}),$$

i.e.

$$\mathbf{e}' \wedge (\mathbf{a}' + \mathbf{b}') = \mathbf{e} \wedge (\mathbf{a} + \mathbf{b}).$$

This proves the correctness of addition in the case $l = 0$.

Case $l' = 0$. This case is completely symmetrical to the previous one and is investigated in exactly the same way (but the coefficients must be primed wherever they have no primes and unprimed where they have). Thus in this case, too, the sum of the bivectors $\mathbf{e}\wedge\mathbf{a}$ and $\mathbf{e}\wedge\mathbf{b}$ is defined correctly.

Case $l\neq 0$ and $l'\neq 0$. Here

$$(\mathbf{a}'+\mathbf{b}')-\left(\frac{l_1}{l}+\frac{l_1'}{l'}\right)\mathbf{e}'=$$
$$=k_1\mathbf{e}+l_1\mathbf{a}+k_1'\mathbf{e}+l_1'\mathbf{b}-\frac{l_1}{l}(k\mathbf{e}+l\mathbf{a})-\frac{l_1'}{l'}(k'\mathbf{e}+l'\mathbf{b})=$$
$$=\left(\frac{lk_1-l_1k}{l}+\frac{l'k_1'-l_1'k'}{l'}\right)\mathbf{e}=-\left(\frac{1}{l}+\frac{1}{l'}\right)\mathbf{e}$$

and

$$\left(\frac{1}{l}+\frac{1}{l'}\right)\mathbf{e}'=\frac{1}{l}(k\mathbf{e}+l\mathbf{a})+\frac{1}{l'}(k'\mathbf{e}+l'\mathbf{b})=$$
$$=(\mathbf{a}+\mathbf{b})+\left(\frac{k}{l}+\frac{k'}{l'}\right)\mathbf{e}.$$

Therefore

$$(\mathbf{e}',\mathbf{a}'+\mathbf{b}')\Rightarrow\left(\mathbf{e}',(\mathbf{a}'+\mathbf{b}')-\left(\frac{l_1}{l}+\frac{l_1'}{l'}\right)\mathbf{e}'\right)=$$
$$=\left(\mathbf{e}',-\left(\frac{1}{l}+\frac{1}{l'}\right)\mathbf{e}\right)\Rightarrow\left(\left(\frac{1}{l}+\frac{1}{l'}\right)\mathbf{e}',-\mathbf{e}\right)=$$
$$=\left((\mathbf{a}+\mathbf{b})+\left(\frac{k}{l}+\frac{k'}{l'}\right)\mathbf{e},-\mathbf{e}\right)\Rightarrow(\mathbf{a}+\mathbf{b},-\mathbf{e})\Rightarrow$$
$$\Rightarrow(\mathbf{a}+\mathbf{b}+\mathbf{e},-\mathbf{e})\Rightarrow(\mathbf{a}+\mathbf{b}+\mathbf{e},\mathbf{a}+\mathbf{b})\Rightarrow(\mathbf{e},\mathbf{a}+\mathbf{b}),$$

i.e. again

$$\mathbf{e}'\wedge(\mathbf{a}'+\mathbf{b}')=\mathbf{e}\wedge(\mathbf{a}+\mathbf{b}).$$

It remains to consider the case where at least one of the given bivectors, say the bivector $\mathbf{e}\wedge\mathbf{a}$, is zero. But if $\mathbf{e}\wedge\mathbf{a}=0$, then necessarily $\mathbf{a}=k\mathbf{e}$, where k is some number (recall that under the hypothesis $\mathbf{e}\neq 0$). Therefore

$$(\mathbf{e},\mathbf{a}+\mathbf{b})=(\mathbf{e},k\mathbf{e}+\mathbf{b})\Rightarrow(\mathbf{e},\mathbf{b}),$$

i.e.

$$0+\mathfrak{b}=\mathfrak{b},$$

where $\mathfrak{b}=\mathbf{e}\wedge\mathbf{b}$. Thus, when $\mathfrak{a}=0$ the sum $\mathfrak{a}+\mathfrak{b}$ is equal to $\mathfrak{b}$ regardless of the arbitrariness of the construction, and is, consequently, correctly defined. It is shown in a similar way that $\mathfrak{a}+0=\mathfrak{a}$ and that hence the sum $\mathfrak{a}+0$ is correctly defined.

Thus the correctness of the definition of bivector addition is proved in all the cases. □

We have also proved in the course of reasoning that, as was to be expected, the zero bivector is the zero of bivector addition, i.e. $\alpha + 0 = 0 + \alpha = \alpha$ for any bivector α.

This shows in particular that formula (6) of the preceding lecture is true also for $\mathbf{e} = \mathbf{0}$, i.e. for any vectors $\mathbf{e}$, $\mathbf{a}$, $\mathbf{b}$.

Note that for any vectors $\mathbf{e}$ and $\mathbf{a}$ and any number k we have the relation

$$(\mathbf{e}, k\mathbf{a}) \Rightarrow (k\mathbf{e}, \mathbf{a})$$

showing that

$$\mathbf{e} \wedge k\mathbf{a} = k\mathbf{e} \wedge \mathbf{a}.$$

Definition 1. *The product* $k\alpha$ of the bivector $\alpha = \mathbf{e} \wedge \mathbf{a}$ by the number k is the bivector $k\mathbf{e} \wedge \mathbf{a} = \mathbf{e} \wedge k\mathbf{a}$.

Thus, by definition,

(1) $$k(\mathbf{e} \wedge \mathbf{a}) = k\mathbf{e} \wedge \mathbf{a} = \mathbf{e} \wedge k\mathbf{a}.$$

Of course, the correctness of this definition also needs verification.

If $\mathbf{e} \wedge \mathbf{a} = 0$, then also $k\mathbf{e} \wedge \mathbf{a} = 0$, i.e. $k\alpha = 0$, if $\alpha = 0$. Thus when $\alpha = 0$ the bivector $k\alpha$ is defined correctly.

Let $\mathbf{e} \wedge \mathbf{a} \neq 0$ and let $\mathbf{e} \wedge \mathbf{a} = \mathbf{e}' \wedge \mathbf{a}'$, i.e.

$$\mathbf{e}' = k_1\mathbf{e} + l_1\mathbf{a}, \quad \mathbf{a}' = k_2\mathbf{e} + l_2\mathbf{a},$$

where $k_1l_2 - l_1k_2 = 1$. Then for $k \neq 0$

$$k\mathbf{e}' = k_1(k\mathbf{e}) + (l_1k)\mathbf{a}, \quad \mathbf{a}' = \frac{k_2}{k}(k\mathbf{e}) + l_2\mathbf{a}$$

with $k_1l_2 - l_1k\frac{k_2}{k} = k_1l_2 - l_1k_2 = 1$. Hence

$$k\mathbf{e}' \wedge \mathbf{a}' = k\mathbf{e} \wedge \mathbf{a}.$$

Since this equation is obviously satisfied for $k = 0$, the correctness of Definition 1 is completely proved. □

The operation of constructing a bivector $\mathbf{a} \wedge \mathbf{b}$, given vectors $\mathbf{a}$ and $\mathbf{b}$, may be considered as a sort of multiplication. Since in performing it we overstep the limits of vectors, it is called *external multiplication*.

Referring to the operation $\mathbf{a}, \mathbf{b} \mapsto \mathbf{a} \wedge \mathbf{b}$ as multiplication is justified by formula (6) of the preceding lecture, which is nothing other than the statement about the **distributivity** of external multiplication with respect to addition.

Similarly, formula (1) states the **homogeneity** of external multiplication with respect to multiplication by numbers.

Relation (4) derived in the preceding lecture means that

$$\mathbf{a} \wedge \mathbf{b} = \mathbf{b} \wedge (-\mathbf{a}),$$

i.e., by virtue of homogeneity, that

$$\mathbf{a} \wedge \mathbf{b} = -(\mathbf{b} \wedge \mathbf{a}).$$

This property is called **anticommutativity.**

We know that $\mathbf{a} \wedge \mathbf{b} = 0$ if $\mathbf{a}$ and $\mathbf{b}$ are collinear. Note that this follows directly from the properties of anticommutativity and homogeneity of external multiplication (provided the characteristic of the ground field is other than two). Indeed, according to the property of anticommutativity $\mathbf{a} \wedge \mathbf{a} = -\mathbf{a} \wedge \mathbf{a}$ and therefore $\mathbf{a} \wedge \mathbf{a} = 0$ for any $\mathbf{a}$. Hence $\mathbf{a} \wedge \mathbf{b} = 0$ if $\mathbf{a}$ and $\mathbf{b}$ are collinear.

However, the fact that for no other vectors $\mathbf{a}$ and $\mathbf{b}$ their external product $\mathbf{a} \wedge \mathbf{b}$ is zero does not follow from anticommutativity and is an independent property of external multiplication. It may be said that external multiplication is **free.**

Summing up all that has been said we obtain the following theorem:

Theorem 1 (*algebraic properties of external multiplication*). *External multiplication is*

(a) *distributive*

$$\mathbf{e} \wedge (\mathbf{a} + \mathbf{b}) = \mathbf{e} \wedge \mathbf{a} + \mathbf{e} \wedge \mathbf{b}$$

for any vector $\mathbf{e}$, $\mathbf{a}$, *and* $\mathbf{b}$;

(b) *homogeneous*

$$k(\mathbf{a} \wedge \mathbf{b}) = k\mathbf{a} \wedge \mathbf{b} = \mathbf{a} \wedge k\mathbf{b}$$

for any $\mathbf{a}$, $\mathbf{b}$ *and any* k;

(c) *anticommutative*

$$\mathbf{a} \wedge \mathbf{b} = -(\mathbf{b} \wedge \mathbf{a})$$

for any $\mathbf{a}$ *and* $\mathbf{b}$;

(d) *free*

$$\mathbf{a} \wedge \mathbf{b} = 0$$

if and only if **a** *and* **b** *are collinear.* □

In various calculations involving bivectors it is often helpful to bear in mind that if

$$\mathbf{a}' = k\mathbf{a} + l\mathbf{b}, \quad \mathbf{b}' = k_1\mathbf{a} + l_1\mathbf{b},$$

then

$$\mathbf{a}' \wedge \mathbf{b}' = \begin{vmatrix} k & l \\ k_1 & l_1 \end{vmatrix} (\mathbf{a} \wedge \mathbf{b}). \tag{2}$$

This formula actually constitutes the content of Lemma 1 of the preceding lecture, but its proof now reduces to a trivial calculation making use of the distributivity, homogeneity, and anticommutativity of external multiplication.

In particular, we see that *if* $\mathbf{a} \parallel \mathfrak{a}$ *and* $\mathbf{b} \parallel \mathfrak{a}$, *then the bivector* $\mathbf{a} \wedge \mathbf{b}$ *is proportional to the bivector* $\mathfrak{a}$. □

Having on the set $\mathscr{V} \wedge \mathscr{V}$ of all bivectors (only for $\dim \mathscr{V} \leqslant 3$, mind you) the operations of addition and multiplication by a number, we may raise the problem: is the set $\mathscr{V} \wedge \mathscr{V}$ a vector space with respect to these operations?

Theorem 2. *If* $\dim \mathscr{V} \leqslant 3$ *the set* $\mathscr{V} \wedge \mathscr{V}$ *is a vector space.*

Proof. Verification of all the axioms 1° to 8°, except axiom 1° of the associativity of addition, is absolutely trivial. For example, axiom 7° is verified using the following calculation (where $\mathbf{e} \wedge \mathbf{a} = \mathfrak{a}$ and $\mathbf{e} \wedge \mathbf{b} = \mathfrak{b}$):

$$k(\mathfrak{a}+\mathfrak{b}) = k(\mathbf{e} \wedge (\mathbf{a}+\mathbf{b})) = \mathbf{e} \wedge k(\mathbf{a}+\mathbf{b}) =$$
$$= \mathbf{e} \wedge k\mathbf{a} + \mathbf{e} \wedge k\mathbf{b} = k\mathfrak{a} + k\mathfrak{b}.$$

The only difficulty lies in verifying axiom 1°, i.e. in proving that for any three bivectors $\mathfrak{a}$, $\mathfrak{b}$, and $\mathfrak{c}$ we have

$$(\mathfrak{a} + \mathfrak{b}) + \mathfrak{c} = \mathfrak{a} + (\mathfrak{b} + \mathfrak{c}). \tag{3}$$

If at least one of the bivectors, $\mathfrak{a}$, $\mathfrak{b}$, or $\mathfrak{c}$, is zero, then relation (3) is obvious. Therefore without loss of generality we may assume that $\mathfrak{a} \neq 0$, $\mathfrak{b} \neq 0$, and $\mathfrak{c} \neq 0$.

By using Propositions 3 and 2 of the preceding lecture we can represent the bivectors $\mathfrak{a}$ and $\mathfrak{c}$ as

$$\mathfrak{a} = \mathbf{e} \wedge \mathbf{a}, \quad \mathfrak{c} = \mathbf{e} \wedge \mathbf{c},$$

where $\mathbf{e} \neq 0$. If $\mathbf{e} \parallel \mathfrak{b}$, i.e. (see Proposition 2 of the preceding lecture) if there is a vector $\mathbf{b}$ such that $\mathfrak{b} = \mathbf{e} \wedge \mathbf{b}$, then everything reduces to the associativity of vector addition.

$$(\mathfrak{a}+\mathfrak{b})+\mathfrak{c} = \mathbf{e} \wedge (\mathbf{a}+\mathbf{b}) + \mathbf{e} \wedge \mathbf{c} = \mathbf{e} \wedge ((\mathbf{a}+\mathbf{b})+\mathbf{c}) =$$
$$= \mathbf{e} \wedge (\mathbf{a}+(\mathbf{b}+\mathbf{c})) = \mathbf{e} \wedge \mathbf{a} + \mathbf{e} \wedge (\mathbf{b}+\mathbf{c}) =$$
$$= \mathbf{e} \wedge \mathbf{a} + (\mathbf{e} \wedge \mathbf{b} + \mathbf{e} \wedge \mathbf{c}) = \mathfrak{a} + (\mathfrak{b}+\mathfrak{c}).$$

Thus the only nontrivial case arises when $\mathbf{e} \nparallel \mathfrak{b}$.

Let $\mathfrak{b} = \mathbf{b}_1 \wedge \mathbf{b}_2$. Since $\mathfrak{b} \neq 0$, then for $\mathbf{e} \nparallel \mathfrak{b}$ the three vectors $\mathbf{e}$, $\mathbf{b}_1$, $\mathbf{b}_2$ are linearly independent. Therefore the case $\mathbf{e} \nparallel \mathfrak{b}$ is possible only for $\dim \mathscr{V} = 3$, and then the vectors $\mathbf{e}$, $\mathbf{b}_1$, $\mathbf{b}_2$ constitute a basis. Expand the vector $\mathbf{a}$ with respect to this basis

$$\mathbf{a} = k\mathbf{e} + k_1\mathbf{b}_1 + k_2\mathbf{b}_2.$$

then $\mathbf{e} \wedge \mathbf{a} = \mathbf{e} \wedge \mathbf{a}'$, where $\mathbf{a}' = k_1\mathbf{b}_1 + k_2\mathbf{b}_2$. This means that without loss of generality we may assume that $\mathbf{a} \parallel \mathfrak{b}$.

Similarly, we may assume that $\mathbf{c} \parallel \mathfrak{b}$.

Let, thus,

$$\mathbf{a} = k_1\mathbf{b}_1 + k_2\mathbf{b}_2, \quad \mathbf{c} = l_1\mathbf{b}_1 + l_2\mathbf{b}_2.$$

Then (see formula (2))

$$\mathbf{a} \wedge \mathbf{c} = \delta(\mathbf{b}_1 \wedge \mathbf{b}_2) = \delta\mathfrak{b},$$

where $\delta = k_1l_2 - k_2l_1 = \begin{vmatrix} k_1 & k_2 \\ l_1 & l_2 \end{vmatrix}$.

Therefore, if $\delta \neq 0$, then

$$\mathfrak{a}+\mathfrak{b} = \mathbf{e} \wedge \mathbf{a} + \frac{1}{\delta}(\mathbf{a} \wedge \mathbf{c}) = \left(\mathbf{e} - \frac{1}{\delta}\mathbf{c}\right) \wedge \mathbf{a}.$$

Since, on the other hand,

$$\mathfrak{c} = \mathbf{e} \wedge \mathbf{c} = \left(\mathbf{e} - \frac{1}{\delta}\mathbf{c}\right) \wedge \mathbf{c},$$

we have

$$(\mathfrak{a}+\mathfrak{b})+\mathfrak{c} = \left(\mathbf{e} - \frac{1}{\delta}\mathbf{c}\right) \wedge (\mathbf{a}+\mathbf{c}). \tag{4}$$

Similarly,

$$\mathfrak{b}+\mathfrak{c}=\frac{1}{\delta}(\mathbf{a}\wedge\mathbf{c})+\mathbf{e}\wedge\mathbf{c}=\left(\frac{1}{\delta}\mathbf{a}+\mathbf{e}\right)\wedge\mathbf{c},$$

$$\mathfrak{a}=\mathbf{e}\wedge\mathbf{a}=\left(\frac{1}{\delta}\mathbf{a}+\mathbf{e}\right)\wedge\mathbf{a}$$

and

$$\mathfrak{a}+(\mathfrak{b}+\mathfrak{c})=\left(\frac{1}{\delta}\mathbf{a}+\mathbf{e}\right)\wedge(\mathbf{a}+\mathbf{c}). \tag{5}$$

To complete the proof (for $\delta\neq 0$) it remains to note that the difference between the right-hand sides of formulas (4) and (5) is zero.

$$\left(\mathbf{e}-\frac{1}{\delta}\mathbf{c}\right)\wedge(\mathbf{a}+\mathbf{c})-\left(\frac{1}{\delta}\mathbf{a}+\mathbf{e}\right)\wedge(\mathbf{a}+\mathbf{c})=$$
$$=\left(\mathbf{e}-\frac{1}{\delta}\mathbf{c}-\frac{1}{\delta}\mathbf{a}-\mathbf{e}\right)\wedge(\mathbf{a}+\mathbf{c})=$$
$$=-\frac{1}{\delta}(\mathbf{a}+\mathbf{c})\wedge(\mathbf{a}+\mathbf{c})=0.$$

Finally, if $\delta=0$, then there is a number h such that $\mathbf{c}=h\mathbf{a}$. In addition, since $\mathbf{a}\parallel\mathfrak{b}$, there is a vector $\mathbf{b}$ such that $\mathfrak{b}=\mathbf{a}\wedge\mathbf{b}$. Therefore

$$(\mathfrak{a}+\mathfrak{b})+\mathfrak{c}=(\mathbf{e}\wedge\mathbf{a}-\mathbf{b}\wedge\mathbf{a})+\mathbf{e}\wedge h\mathbf{a}=$$
$$=(\mathbf{e}-\mathbf{b})\wedge\mathbf{a}+h\mathbf{e}\wedge\mathbf{a}=((h+1)\,\mathbf{e}-\mathbf{b})\wedge\mathbf{a}$$

and similarly

$$\mathfrak{a}+(\mathfrak{b}+\mathfrak{c})=\mathbf{e}\wedge\mathbf{a}+(-\mathbf{b}\wedge\mathbf{a}+h\mathbf{e}\wedge\mathbf{a})=$$
$$=((h+1)\,\mathbf{e}-\mathbf{b})\wedge\mathbf{a}.$$

This completes the proof of formula (3) and hence that of Theorem 2. □

When $n=1$ there is obviously only a zero bivector. Thus

$$\mathscr{V}^1\wedge\mathscr{V}^1=0,$$

and hence $\dim(\mathscr{V}^1\wedge\mathscr{V}^1)=0$.

When $n=2$ consider an arbitrary basis $\mathbf{e}_1$, $\mathbf{e}_2$ for the space $\mathscr{V}^2$. According to formula (2), if

$$\mathbf{a}=a^1\mathbf{e}_1+a^2\mathbf{e}_2,\quad \mathbf{b}=b^1\mathbf{e}_1+b^2\mathbf{e}_2,$$

then

$$\mathbf{a} \wedge \mathbf{b} = \begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix} (\mathbf{e}_1 \wedge \mathbf{e}_2). \tag{6}$$

Since $\mathbf{e}_1 \wedge \mathbf{e}_2 \neq 0$, this proves that

$$\dim (\mathscr{V}^2 \wedge \mathscr{V}^2) = 1.$$

Thus any two bivectors in $\mathscr{V}^2$ ("in a plane") are proportional and any basis $\mathbf{e}_1$, $\mathbf{e}_2$ for the space $\mathscr{V}^2$ determines the basis $\mathbf{e}_1 \wedge \mathbf{e}_2$ for the space $\mathscr{V}^2 \wedge \mathscr{V}^2$. In that basis the coordinate of the bivector $\mathbf{a} \wedge \mathbf{b}$ is equal to $\begin{vmatrix} a^1 & b^2 \\ a^1 & b^2 \end{vmatrix}$.

We see that, as was to be expected, the space $\mathscr{V}^2 \wedge \mathscr{V}^2$ is algebraically of little interest. Nevertheless it can be of help in obtaining meaningful geometric results.

Adopting again the intuitive point of view and interpreting bivectors as parallelogram area elements we immediately see that if two bivectors α_0 and α are connected by the relation $\alpha = k\alpha_0$, then the ratio of the area of α to the area of α_0 is $|k|$. In particular, if the area of α_0 is unity, i.e. α_0 is said to be the *area standard*, then the area of α is $|k|$. On the basis of formula (6) this proves the following proposition.

Proposition 1. *If a basis* $\mathbf{e}_1$, $\mathbf{e}_2$ *possesses the property that the area of a parallelogram constructed on the vectors* $\mathbf{e}_1$ *and* $\mathbf{e}_2$ *is unity, then the area of a parallelogram constructed on arbitrary vectors* $\mathbf{a}\ (a^1, a^2)$ *and* $\mathbf{b}\ (b^1, b^2)$ *is equal to the absolute value of the determinant*

$$\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix},$$

i.e. to

$$|a^1b^2 - a^2b^1|. \ \square$$

The area of a triangle OAB constructed on vectors $\mathbf{a} = \overrightarrow{OA}$ and $\mathbf{b} = \overrightarrow{OB}$ is half the area of the parallelogram constructed on them. Hence *the area of a triangle constructed on vectors* $\mathbf{a}\ (a^1, a^2)$ *and* $\mathbf{b}\ (b^1, b^2)$ *is equal to the absolute value of the number*

$$\frac{1}{2}\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}. \ \square$$

If the vertices O, A, B of the triangle have, respectively, the coordinates (x_0, y_0), (x_1, y_1), and (x_2, y_2), then $a^1 = x_1 - x_0$, $a^2 = y_1 - y_0$, $b^1 = x_2 - x_0$, $b^2 = y_2 - y_0$. Hence *the area of a triangle with vertices at the points* (x_0, y_0), (x_1, y_1), *and* (x_2, y_2) *is equal to the absolute value of the number*

$$\frac{1}{2}\begin{vmatrix} x_1-x_0 & y_1-y_0 \\ x_2-x_0 & y_2-y_0 \end{vmatrix} = \frac{1}{2}\begin{vmatrix} x_0 & y_0 & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix}. \ \square$$

In the axiomatic theory of areas the initial steps are, as ever, reversed: after some bivector $\alpha_0 \neq 0$ has been chosen as the standard, the area of an arbitrary bivector $\alpha = k\alpha_0$ is **defined** to be a number $|k|$. The resulting computational formulas are of course the same. A detailed reasoning required here is left to the reader.

In the same circle of ideas one can also define the so-called oriented area. It is k.

Now let $n = 3$ and let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be an arbitrary basis for a space $\mathscr{V}^3$. Then, on calculating for the vectors

$$\mathbf{a} = a^1\mathbf{e}_1 + a^2\mathbf{e}_2 + a^3\mathbf{e}_3, \quad \mathbf{b} = b^1\mathbf{e}_1 + b^2\mathbf{e}_2 + b^3\mathbf{e}_3$$

their external product $\mathbf{a} \wedge \mathbf{b}$, we, as can be easily seen, obtain

$$\mathbf{a} \wedge \mathbf{b} = (a^2b^3 - a^3b^2)(\mathbf{e}_2 \wedge \mathbf{e}_3) + \\ + (a^3b^1 - a^1b^3)(\mathbf{e}_3 \wedge \mathbf{e}_1) + (a^1b^2 - a^2b^1)(\mathbf{e}_1 \wedge \mathbf{e}_2).$$

Using determinants we write this formula as

$$(7) \qquad \mathbf{a} \wedge \mathbf{b} = \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}(\mathbf{e}_2 \wedge \mathbf{e}_3) - \\ - \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}(\mathbf{e}_3 \wedge \mathbf{e}_1) + \begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}(\mathbf{e}_1 \wedge \mathbf{e}_2).$$

To memorise this formula it is convenient to write it in the following symbolic form (we assume that determinants have already been introduced in the parallel course in algebra).

$$(8) \qquad \mathbf{a} \wedge \mathbf{b} = \begin{vmatrix} \mathbf{e}_2 \wedge \mathbf{e}_3 & \mathbf{e}_3 \wedge \mathbf{e}_1 & \mathbf{e}_1 \wedge \mathbf{e}_2 \\ a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \end{vmatrix}.$$

Indeed, by formally expanding this determinant by the first row we obtain exactly formula (7).

Theorem 3. *A vector space* $\mathscr{V}^3 \wedge \mathscr{V}^3$ *has dimension* 3:

$$\dim(\mathscr{V}^3 \wedge \mathscr{V}^3) = 3.$$

Proof. Formula (7) means that the family consisting of three bivectors $\mathbf{e}_2 \wedge \mathbf{e}_3$, $\mathbf{e}_3 \wedge \mathbf{e}_1$, and $\mathbf{e}_1 \wedge \mathbf{e}_2$ is complete. So it is sufficient to prove that it is linearly independent (and hence is a basis).

Let k_1, k_2, and k_3 be numbers such that

$$k_1(\mathbf{e}_2 \wedge \mathbf{e}_3) + k_2(\mathbf{e}_3 \wedge \mathbf{e}_1) + k_3(\mathbf{e}_1 \wedge \mathbf{e}_2) = 0.$$

Suppose, for example, $k_1 \neq 0$ and consider the vectors

$$\mathbf{a} = -k_2\mathbf{e}_1 + k_1\mathbf{e}_2, \quad \mathbf{b} = -k_3\mathbf{e}_1 + k_1\mathbf{e}_3.$$

If $\lambda\mathbf{a} + \mu\mathbf{b} = 0$, then

$$-(\lambda k_2 + \mu k_3)\mathbf{e}_1 + (\lambda k_1)\mathbf{e}_2 + (\mu k_1)\mathbf{e}_3 = 0,$$

whence $\lambda k_1 = 0$ and $\mu k_1 = 0$, i.e. $\lambda = 0$ and $\mu = 0$. Therefore $\mathbf{a}$ and $\mathbf{b}$ are linearly independent and hence $\mathbf{a} \wedge \mathbf{b} \neq 0$.

But, on the other hand,

$$\begin{aligned}\mathbf{a} \wedge \mathbf{b} &= (-k_2\mathbf{e}_1 + k_1\mathbf{e}_2) \wedge (-k_3\mathbf{e}_1 + k_1\mathbf{e}_3) = \\ &= k_1^2(\mathbf{e}_2 \wedge \mathbf{e}_3) + k_1k_2(\mathbf{e}_3 \wedge \mathbf{e}_1) + k_1k_3(\mathbf{e}_1 \wedge \mathbf{e}_2) = \\ &= k_1[k_1(\mathbf{e}_2 \wedge \mathbf{e}_3) + k_2(\mathbf{e}_3 \wedge \mathbf{e}_1) + k_3(\mathbf{e}_1 \wedge \mathbf{e}_2)] = 0.\end{aligned}$$

The contradiction obtained shows that $k_1 = 0$ and hence

$$k_2(\mathbf{e}_3 \wedge \mathbf{e}_1) + k_3(\mathbf{e}_1 \wedge \mathbf{e}_2) = (k_2\mathbf{e}_3 - k_3\mathbf{e}_2) \wedge \mathbf{e}_1 = 0,$$

and this is possible, by virtue of the linear independence of the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, only for $k_2 = 0$ and $k_3 = 0$.

Thus any zero linear combination of the bivectors $\mathbf{e}_2 \wedge \mathbf{e}_3$, $\mathbf{e}_3 \wedge \mathbf{e}_1$, and $\mathbf{e}_1 \wedge \mathbf{e}_2$ is necessarily a trivial combination, which, by definition, means that these bivectors are linearly independent. □

At the same time it has been proved that *for any basis* $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ *of a vector space* $\mathscr{V}^3$ *the bivectors* $\mathbf{e}_2 \wedge \mathbf{e}_3$, $\mathbf{e}_3 \wedge \mathbf{e}_1$, *and* $\mathbf{e}_1 \wedge \mathbf{e}_2$ *constitute the basis for a vector space* $\mathscr{V}^3 \wedge \mathscr{V}^3$. The expansion of any bivector with respect to this basis is given by formula (7) (or (8)).

It follows, in particular that *a vector* $\mathbf{l}$ *with the coordinates* l, m, n *(in the basis* $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$*) is parallel to a bivector* $\mathfrak{a} \neq 0$ *with the coordinates* A, B, C *(in the basis* $\mathbf{e}_2 \wedge \mathbf{e}_3$, $\mathbf{e}_3 \wedge \mathbf{e}_1$, $\mathbf{e}_1 \wedge \mathbf{e}_2$*) if and only if*

$$Al + Bm + Cn = 0. \tag{9}$$

Indeed, if $\mathfrak{a} = \mathbf{a} \wedge \mathbf{b}$, then writing the relation $\mathbf{l} \parallel \mathfrak{a}$ is equivalent to saying that the vector $\mathbf{l}$ is linearly expressed in terms of $\mathbf{a}$ and $\mathbf{b}$. Therefore, if a^1, a^2, a^3, and b^1, b^2, b^3 are the coordinates of $\mathbf{a}$ and $\mathbf{b}$, then in the determinant

$$\begin{vmatrix} l & m & n \\ a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \end{vmatrix} \tag{10}$$

the first row is a linear combination of the other two rows and so the determinant is zero.

Conversely, it is known from the theory of determinants that if a determinant is zero, then its rows are linearly dependent. But in the determinant (10) the last rows are linearly independent (for $\mathbf{a} \wedge \mathbf{b} \neq 0$). Therefore, if this determinant is zero, then its first row is linearly expressible in terms of the last two rows, i.e. $\mathbf{l} \parallel \mathfrak{a}$.

This proves that the vector $\mathbf{l}\,(l, m, n)$ is parallel to the bivector $\mathfrak{a} = \mathbf{a} \wedge \mathbf{b}$ if and only if

$$\begin{vmatrix} l & m & n \\ a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \end{vmatrix} = 0. \quad \square \tag{11}$$

Formula (9) follows immediately: it is enough to expand the determinant (10) by the first row and take advantage of the fact that according to formula (7)

$$A = \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}, \quad B = -\begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}, \quad C = \begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}. \tag{12}$$

Lecture 9

Planes in space • Parametric equations of a plane • The general equation of a plane • A plane passing through three noncollinear points

We are now in a position to construct a theory of planes in space by repeating almost word for word the theory of straight line in a plane.

Let $\mathscr{A} = \mathscr{A}^3$ be an affine (three-dimensional) space and let $\mathscr{V} = \mathscr{V}^3$ be a vector space associated with $\mathscr{A}$.

Definition 1. For any point $M_0 \in \mathscr{A}$ and any nonzero bivector $\alpha \in \mathscr{V} \wedge \mathscr{V}$ the *plane* of the space $\mathscr{A}$ given by the point M_0 and the bivector α is the set of all points $M \in \mathscr{A}$ such that

$$\overrightarrow{M_0M} \parallel \alpha. \tag{1}$$

For $\alpha = \mathbf{e} \wedge \mathbf{b}$ this condition means that there are numbers u, v for which we have

$$\overrightarrow{M_0M} = u\mathbf{a} + v\mathbf{b}. \tag{2}$$

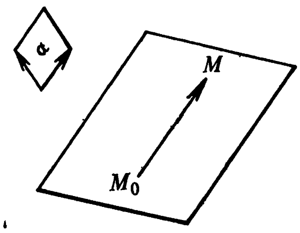
The bivector α is called the *direction bivector* of a plane. A vector $\mathbf{a} \in \mathscr{V}$ or bivector $\alpha' \in \mathscr{V} \wedge \mathscr{V}$ is said to be *parallel* to the plane if $\mathbf{a} \parallel \alpha$ or $\alpha' = k\alpha$ respectively.

All vectors parallel to a plane are easily seen to form a vector space of dimension 2. It follows literally as in the case of straight lines (see Lecture 5), that any plane is provided in a natural way with the structure of an affine space of dimension 2 (with an associated vector space consisting of all vectors parallel to the plane).

Since $\mathbf{0} \parallel \alpha$, the point M_0 is in the plane. Therefore the plane given by the point M_0 and a bivector $\alpha \neq 0$ is also called the *plane passing through the point* M_0 *and parallel to the bivector* $\mathfrak{a}$.

A point M_0 and linearly independent vectors $\mathbf{a}$ and $\mathbf{b}$ form an affine coordinate system in a plane. In this system the coordinates of a point M are the numbers u and v appearing in relation (2).

Proposition 1. *Let* N_0 *be an arbitrary point of the plane passing through a point* M_0 *and parallel to a bivector* $\mathfrak{a}$, *and let* $\mathfrak{b}$ *be an arbitrary nonzero bivector parallel to that plane. The point* N_0 *and the bivector* $\mathfrak{b}$ *then give the same plane.*



A plane given by a point M_0 *and a bivector* $\mathfrak{a} \neq 0$

Proof. Under the hypothe${}_0$is there are numbers u_0, vs, and $h_0 \neq 0$ such that

$$\overrightarrow{M_0N_0} = u_0\mathbf{a} + v_0\mathbf{b} \text{ and } \mathfrak{b} = h_0\mathbf{a} \wedge \mathbf{b}.$$

Therefore, if

$$\overrightarrow{M_0M} = u\mathbf{a} + v\mathbf{b}$$

then

$$\overrightarrow{N_0M} = \overrightarrow{M_0M} - \overrightarrow{M_0N_0} = (u - u_0)\,\mathbf{a} + (v - v_0)\,\mathbf{b} = u'\,(h_0\mathbf{a}) + v'\mathbf{b},$$

where $u' = \frac{u - u_0}{h_0}$ and $v' = v - v_0$.

Conversely, if

$$\overrightarrow{N_0M} = u'\,(h_0\mathbf{a}) + v'\mathbf{b}$$

then

$$\overrightarrow{M_0M} = \overrightarrow{N_0M} - \overrightarrow{N_0M_0} = u\mathbf{a} + v\mathbf{b},$$

where $u = u'h_0 + u_0$ and $v = v' + v_0$. □

Thus a plane can be given by any of its points and any nonzero bivector that is parallel to it.

Let an initial point O be chosen in $\mathcal{A}$ and let

$$\mathbf{r}_0 = \overrightarrow{OM_0}, \quad \mathbf{r} = \overrightarrow{OM}.$$

Formula (2) then takes the form

$$\mathbf{r} = \mathbf{r}_0 + u\mathbf{a} + v\mathbf{b}. \tag{3}$$

As the parameters u and v vary independently from $-\infty$ to $+\infty$, the point M with the radius vector given by formula (3) runs over the entire plane. This formula is therefore called the *parametric vector equation* of a plane.

In an arbitrary affine coordinate system $O\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$ equation (3) is equivalent to the three equations for the coordinates.

$$\begin{aligned} x &= x_0 + ua^1 + vb^1, \\ y &= y_0 + ua^2 + vb^2, \\ z &= z_0 + ua^3 + vb^3. \end{aligned} \tag{4}$$

These equations are called *coordinate parametric equations* of a plane.

Condition (1), i.e. the condition $(\mathbf{r} - \mathbf{r}_0) \parallel a$ can be expressed analytically in some other, fundamentally different ways. Thus, for example, it follows immediately from relation (11) of the preceding lecture that this condition is equivalent to the vanishing of the determinant made up of the coordinates of the vectors $\mathbf{r} - \mathbf{r}_0$, $\mathbf{a}$, and $\mathbf{b}$. Thus the point $M\ (x, y, z)$ is in the plane if and only if

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \end{vmatrix} = 0. \tag{5}$$

Briefly speaking, (5) is the equation of a plane passing through the point $M_0\ (x_0, y_0, z_0)$ and parallel to the bivector $\mathbf{a} \wedge \mathbf{b}$.

Expanding (5) by the first row and setting

$$A = \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}, \quad B = \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}, \quad C = \begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix} \tag{6}$$

we obtain the *general equation of a plane passing through* *the point* $M_0 (x_0, y_0, z_0)$.

$$A (x - x_0) + B (y - y_0) + C (z - z_0) = 0.$$

Setting $D = -Ax_0 - By_0 - Cz_0$, we can write the equation in the form

$$Ax + By + Cz + D = 0. \tag{7}$$

Comparing formulas (6) with formulas (12) of the preceding lecture shows that the coefficients A, B, C *of equation* (7) *are the coordinates of the direction bivector* $\mathfrak{a}$ (in the basis $\mathbf{e}_2 \wedge \mathbf{e}_3$, $\mathbf{e}_3 \wedge \mathbf{e}_1$, $\mathbf{e}_1 \wedge \mathbf{e}_2$).

It is now easy to see that *any equation of the form* (7), *where* $(A, B, C) \neq (0, 0, 0)$, *defines some plane*. Indeed, solving equation (7) we find some triple of numbers (x_0, y_0, z_0) satisfying it (for example, if $A \neq 0$, then it is possible to set $x_0 = -\frac{D}{A}$, $y_0 = 0$, $z_0 = 0$) and construct a plane passing through the point $M_0 (x_0, y_0, z_0)$ and parallel to the bivector $\mathfrak{a}$ (A, B, C). It follows directly from what has been said above that the equation of that plane is equation (7). □

We shall say that a vector $\mathbf{l}$ (l, m, n) *lies* in plane (7) (or, what is the same, is *parallel* to that plane) if it is parallel to its direction bivector $\mathfrak{a}$ (A, B, C), i.e. (see formula (8) of the preceding lecture) if

$$Al + Bm + Cn = 0. \tag{8}$$

It is easy to see that *for any two points* $M_0 (x_0, y_0, z_0)$, $M_1 (x_1, y_1, z_1)$ *of plane* (7) *the vector* $\overrightarrow{M_0M_1}$ (*with the coordinates* $x_1 - x_0$, $y_1 - y_0$, and $z_1 - z_0$) *lies in that plane*. Indeed, if $Ax_0 + By_0 + Cz_0 + D = 0$ and $Ax_1 + By_1 + Cz_1 + D = 0$, then

$$A (x_1 - x_0) + B (y_1 - y_0) + C (z_1 - z_0) = 0. \quad \square$$

If vectors $\mathbf{a}$ and $\mathbf{b}$ lie in a plane (are parallel to its direction bivector $\mathfrak{a}$), then their external product $\mathbf{a} \wedge \mathbf{b}$ is proportional to the bivector $\mathfrak{a}$ and therefore it is also a direction bivector (provided it is nonzero). It follows from this and the preceding assertion that *for any three points* M_0, M_1, M_2 *of a plane that are not on the same straight* *line, i.e. are noncollinear, the bivector* $\overrightarrow{M_0M_1} \wedge \overrightarrow{M_0M_2}$ *is the direction bivector of that plane.*

Setting in (5) $\mathbf{a} = \overrightarrow{M_0M_1}$ and $\mathbf{b} = \overrightarrow{M_0M_1}$ we see that the equation of a plane passing through the noncollinear points $M_0(x_0, y_0, z_0)$, $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$ may be written as

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ x_1-x_0 & y_1-y_0 & z_1-z_0 \\ x_2-x_0 & y_2-y_0 & z_2-z_0 \end{vmatrix} = 0. \tag{9}$$

This proves, in particular, the uniqueness of such a plane.

On the other hand, equation (9) has meaning for any three noncollinear points M_0, M_1, M_2 and, being (after expanding the determinant) an equation of the form (7), defines some plane obviously containing the given points.

This proves the following proposition.

Proposition 2. *There is one and only one plane passing through any three noncollinear points* M_0, M_1, M_2 *of an affine space.* □

That plane is designated by the symbol $M_0M_1M_2$.

Lecture 10

The half-spaces into which a plane divides space • Relative positions of two planes in space • Straight lines in space • A plane containing a given straight line and passing through a given point • Relative positions of a straight line and a plane in space • Relative positions of two straight lines in space • Change from one basis for a vector space to another

Let

$$Ax + By + Cz + D = 0 \tag{1}$$

be an arbitrary plane in space and let the field $\mathbb{R}$ of real numbers be the ground field $\mathbb{K}$.

Definition 1. Two points M_1 and M_2 not in the plane (1) are said to be *nonseparated* (by the plane (1)) if they either coincide or the segment $\overrightarrow{M_1M_2}$ and the plane (1) have no points in common.

To contract the formulas we set

$$F(x, y, z) = Ax + By + Cz + D.$$

Proposition 1. *Points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ not in the plane (1) are nonseparated if and only if the (nonzero) numbers $F_1 = F(x_1, y_1, z_1)$ and $F_2 = F(x_2, y_2, z_2)$ have the same signs.*

Proof. We may assume without loss of generality that $M_1 \neq M_2$. We then have the straight line M_1M_2 whose coordinate parametric equations are of the form

$$x = (1-t)x_1 + tx_2,$$
$$y = (1-t)y_1 + ty_2,$$
$$z = (1-t)z_1 + tz_2.$$

On substituting these expressions in the equation $F(x, y, z) = 0$ we get for t the equation

$$(1 - t) F_1 + tF_2 = 0,$$

hence

$$t = \frac{F_1}{F_1 - F_2}$$

(if $F_1 = F_2$, then there is no solution, i.e. the straight line M_1M_2 is parallel to the plane (1)). We derive from this, reasoning as before (see Lecture 6), that $0 < t < 1$ if and only if the numbers F_1 and F_2 have different signs. Therefore the segment does not intersect the plane (1) if and only if these numbers have the same signs. □

It follows directly from Proposition 1 that the relation of nonseparatedness is an equivalence relation and that there are exactly two corresponding equivalence classes.

Definition 2. These equivalence classes are called *half-spaces* determined by the plane (1).

Two distinct points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ not in the plane (1) thus lie in the same half-plane if and only if the segment $\overrightarrow{M_1M_2}$ does not intersect that plane, i.e. if and only if the numbers $F(x_1, y_1, z_1)$ and $F(x_2, y_2, z_2)$ have the same signs.

Let the ground field $\mathbb{K}$ be again an arbitrary field.

Definition 3. Two planes are said to be *parallel* if their direction bivectors are proportional (and can therefore be chosen equal).

If parallel planes have at least one point in common, then they coincide (for a plane is uniquely determined by a point and a direction bivector). Thus distinct parallel planes have no points in common (do not intersect).

Consider now two planes having a point in common. Let $\mathbf{r}_0$ be its radius vector and $\mathfrak{a}$ and $\mathfrak{b}$ the bivectors of the given planes.

According to Propositions 2 and 3 in Lecture 7 there exists for the bivectors $\mathfrak{a}$ and $\mathfrak{b}$ a vector $\mathbf{e} \neq 0$ such that

$$\mathfrak{a} = \mathbf{e} \wedge \mathbf{a}, \quad \mathfrak{b} = \mathbf{e} \wedge \mathbf{b}.$$

The parametric vector equations of the planes under consideration can therefore be written as

$$\mathbf{r} = \mathbf{r}_0 + u\mathbf{e} + v\mathbf{a}, \quad \mathbf{r} = \mathbf{r}_0 + u\mathbf{e} + v\mathbf{b}.$$

It follows that the straight line

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{e} \tag{2}$$

is contained in either of the planes (its points are obtained when $u = t$, $v = 0$). This proves that if the intersection of the two planes is not empty, then it contains a whole straight line.

Suppose that this intersection contains at least a point M_1 not lying on a straight line (2). The vector $\mathbf{c} = \overrightarrow{M_0M_1}$ then lies in both planes, i.e. we have the relations $\mathbf{c} \parallel \mathfrak{a}$ and $\mathbf{c} \parallel \mathfrak{b}$. Since by construction $\mathbf{e} \parallel \mathfrak{a}$ and $\mathbf{e} \parallel \mathfrak{b}$, the bivector $\mathbf{e} \wedge \mathbf{c}$ is proportional to both bivectors $\mathfrak{a}$ and $\mathfrak{b}$. Since under the hypothesis $\mathbf{e} \wedge \mathbf{c} \neq 0$, it follows that $\mathfrak{a}$ and $\mathfrak{b}$ are proportional, i.e. the given planes are parallel and so, having points in common, coincide.

Thus *the intersection of two distinct planes, when it is not empty, is a straight line.* □

A case is also possible a priori where the planes are not parallel but their intersection is empty. It turns out, however, that this case is impossible, i.e. *nonparallel planes necessarily have at least one point in common* (and therefore, according to what has been proved, intersect in a straight line). Indeed, let

$$\begin{aligned} Ax + By + Cz + D &= 0, \\ A_1x + B_1y + C_1z + D_1 &= 0 \end{aligned} \tag{3}$$

be the equations of two nonparallel planes. Since the coordinates of their direction bivectors are, respectively, the numbers A, B, C and A_1, B_1, C_1 we have that (by virtue of nonparallelism) the triple (A, B, C) is not proportional to the triple (A_1, B_1, C_1), i.e. either $\frac{A}{A_1} \neq \frac{B}{B_1}$ or $\frac{A}{A_1} \neq \frac{C}{C_1}$ or $\frac{B}{B_1} \neq \frac{C}{C_1}$. Let for definiteness $\frac{A}{A_1} \neq \frac{B}{B_1}$. Then the system of two equations

$$Ax + By + D = 0 \text{ and } A_1x + B_1y + D_1 = 0$$

will have a unique solution $x = x_0$, $y = y_0$. This means that the point with the coordinates $(x_0, y_0, 0)$ is in both planes. □

If the planes (3) are parallel, then the triples (A, B, C) and (A_1, B_1, C_1) must be proportional, i.e. we must have planes,

$$\frac{A}{A_1} = \frac{B}{B_1} = \frac{C}{C_1}.$$

If

$$\frac{A}{A_1} \neq \frac{D}{D_1},$$

then the planes have obviously no points in common (cf. the corresponding reasoning for straight lines in Lecture 6) and if

$$\frac{A}{A_1} = \frac{D}{D_1},$$

then the planes coincide. Since the converse statements are obviously also true, this proves the following theorem.

Theorem 1 (*relative positions of two planes in space*). *Two planes in space*

(a) *either have no point in common; or*

(b) *have one and only one straight line in common; or*

(c) *coincide.*

Case (a) *is characterized by the fact that*

$$\frac{A}{A_1} = \frac{B}{B_1} = \frac{C}{C_1} \neq \frac{D}{D_1},$$

case (b) *by the fact that there holds at least one of the inequalities*

$$\frac{A}{A_1} \neq \frac{B}{B_1}, \quad \frac{A}{A_1} \neq \frac{C}{C_1}, \quad \frac{B}{B_1} \neq \frac{C}{C_1},$$

and the case (c) *by the fact that*

$$\frac{A}{A_1} = \frac{B}{B_1} = \frac{C}{C_1} = \frac{D}{D_1}.$$

In cases (a) *and* (c) *the planes are parallel and in case* (b) *nonparallel.* □

As a consequence we derive (**the uniqueness theorem**) the fact that equations (3) determine the same plane if and only if they are proportional.

In addition we see that straight lines in space may be characterized as intersections of two planes, i.e. may be specified by *two* equations of the form

$$\text{(4)} \qquad \begin{cases} Ax+By+Cz+D=0, \\ A_1x+B_1y+C_1z+D_1=0. \end{cases}$$

It is of course required that the coefficients A, B, C and A_1, B_1, C_1 of these equations should not be proportional.

So it is easy to see that *any straight line in space is an intersection of two planes*, i.e. can be given by equations of the form (4).

Indeed, by definition (see Lecture 5), each straight line consists of all points $M(x, y, z)$ for which the vector $\overrightarrow{M_0M}$, where $M_0(x_0, y_0, z_0)$ is a given point, is collinear with a given vector $\mathbf{a}(l, m, n)$. This means that we must have for the coordinates of the vectors $\overrightarrow{M_0M}$ and $\mathbf{a}$ the proportion

$$\text{(5)} \qquad \frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n},$$

i.e. the coordinates x, y, z of the point M must satisfy the two equations:

$$\begin{cases} \dfrac{x-x_0}{l}=\dfrac{y-y_0}{m}, \\ \dfrac{x-x_0}{l}=\dfrac{z-z_0}{n}. \end{cases}$$

On writing these equations as

$$\begin{cases} m(x-x_0)-l(y-y_0)=0, \\ n(x-x_0)-l(z-z_0)=0, \end{cases}$$

i.e. as

$$\begin{cases} mx-ly-(mx_0-ly_0)=0, \\ nx-lz-(nx_0-lz_0)=0, \end{cases}$$

we obtain exactly equations of the form (4) (with nonproportional coefficients $A=m$, $B=-l$, $C=0$ and $A_1=n$, $B_1=0$, $C_1=-l$). □

Equations (5) are called *canonical equations* of a straight line in space.

As an example of using the results obtained we shall prove the following proposition.

Proposition 2. *For any straight line and any point not lying on it there exists a unique plane containing that line and that point.*

Proof. Let $M_1(x_1, y_1, z_1)$ be a given point and let

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n}$$

be the canonical equations of a given straight line. Suppose the required plane exists. Then the vectors $\mathbf{a}\ (l, m, n)$ and $\overrightarrow{M_0M_1}$ are in that plane and therefore their external product $\overrightarrow{M_0M_1} \wedge \mathbf{a}$ (nonzero under the hypothesis) is its direction bivector. This proves that the plane under consideration is specified by the point M_1 and the bivector $\overrightarrow{M_0M_1} \wedge \mathbf{a}$. It is therefore uniquely determined.

That the required plane exists is now obvious: it is the plane given by the point M_0 and bivector $\overrightarrow{M_0M_1} \wedge \mathbf{a}$. □

The equation of that plane is of the form

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ x_1-x_0 & y_1-y_0 & z_1-z_0 \\ l & m & n \end{vmatrix}=0.$$

We now study the relative positions of a straight line and a plane in space.

Definition 4. A straight line is said to be *parallel* to a plane if its direction vector $\mathbf{a}\ (l, m, n)$ is parallel to the direction bivector $\alpha\ (A, B, C)$ of the plane.

We know (see formula (9) of the preceding lecture) this to be the case if and only if

$$Al + Bm + Cn = 0. \tag{6}$$

According to Proposition 2 of Lecture 7, there exists for $\mathbf{a} \parallel \alpha$ a vector $\mathbf{b}$ such that $\alpha = \mathbf{a} \wedge \mathbf{b}$. Therefore, if a plane and a line parallel to it have a common point M_0 with radius vector $\mathbf{r}_0$, then the parametric vector equation of the line can be written as

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{a} \tag{7}$$

and the parametric vector equation of the plane as

(8) $$\mathbf{r} = \mathbf{r}_0 + u\mathbf{a} + v\mathbf{b}.$$

Since each vector (7) is of the form (8) (when $u = t$ and $v = 0$), this proves that the plane and the line parallel to it either have no point in common (do not intersect) or the line is entirely contained in the plane.

Alternatively this fact can be established by substituting the parametric equations of a straight line

$$x = x_0 + tl,$$
$$y = y_0 + tm,$$
$$z = z_0 + tn$$

in the general equation of a plane

$$Ax + By + Cz + D = 0.$$

The result of the substitution has the form

(9) $$(Al + Bm + Cn)\,t + (Ax_0 + By_0 + Cz_0 + D) = 0.$$

Therefore, if a line is parallel to a plane, i.e. condition (6) holds, then equation (9) for t (giving values of the parameter t for the common points of the line and the plane) is either satisfied identically (when the point M_0 is in the plane, i.e. when $Ax_0 + By_0 + Cz_0 + D = 0$) or (otherwise) has no solutions.

Moreover, since equation (9) has one and only one solution when $Al + Bm + Cn \neq 0$, we also find that a plane and a line not parallel to it have always one and only one point in common.

We can sum up all the above results by a single theorem.

Theorem 2 (relative positions of a straight line and a plane in space). *In space a plane and a straight line*

(a) *either do not intersect*; *or*

(b) *have a single point in common*; *or*

(c) *the straight line is contained entirely in the plane.*

Case (a) *is characterized by the fact that*

$$Al + Bm + Cn = 0 \text{ and } Ax_0 + By_0 + Cz_0 + D \neq 0,$$

case (b) *by the fact that*

$$Al + Bm + Cn \neq 0,$$

and case (c) *by the fact that*

$$Al + Bm + Cn = 0 \text{ and } Ax_0 + By_0 + Cz_0 + D = 0.$$

In cases (a) *and* (c) *the straight line is parallel to the plane and in case* (b) *it is not.* □

Similarly we investigate the possible positions of two straight lines in space.

We have already known (see Lecture 6) that parallel lines either do not intersect or coincide.

It can now be added that *noncoincident parallel lines lie in one and only one plane.*

Indeed, on choosing an arbitrary point $M_1(x_1, y_1, z_1)$ on the first line and an arbitrary point $M_2(x_2, y_2, z_2)$ on the second, consider a plane with the parametric vector equation

$$\mathbf{r} = \mathbf{r}_1 + u(\mathbf{r}_2 - \mathbf{r}_1) + v\mathbf{a}, \tag{10}$$

where $\mathbf{a}(l, m, n)$ is the direction vector of the given straight lines and $\mathbf{r}_1$ and $\mathbf{r}_2$ are the radius vectors of the points M_1 and M_2. Since the vector $\mathbf{r}_2 - \mathbf{r}_1$ is not collinear with the vector $\mathbf{a}$ (otherwise the straight lines coincide), formula (10) does give a certain plane. Since with $u = 0$ we obtain points $\mathbf{r} = r_1 + v\mathbf{a}$ of the first straight line and with $u = 1$ we obtain points $\mathbf{r} = r_2 + v\mathbf{a}$ of the second, this proves that the plane containing the given straight lines does exist. The equation of that plane in the coordinate form is

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ l & m & n \end{vmatrix} = 0. \tag{11}$$

To prove the uniqueness of the plane containing the given straight lines, it is sufficient to note that since any such plane contains the points M_1 and M_2 and is parallel to the vector $\mathbf{a}$ (and hence has a direction bivector of the form $\overrightarrow{M_1M_2} \wedge \mathbf{a}$), its parametric vector equation can be written in the form (10). □

Now consider two nonparallel straight lines. Such lines have at most one point in common (if straight lines have two points in common, then they coincide and therefore are parallel). It is easy to see that just as parallel lines *nonparal-*

lel straight lines having a point in common (intersecting lines) lie in one and only one plane.

Indeed, the plane with the parametric vector equation

$$\mathbf{r} = \mathbf{r}_0 + u\mathbf{a}_1 + v\mathbf{a}_2,$$

where $\mathbf{r}_0$ is the radius vector of a common point M_0 of given straight lines and $\mathbf{a}_1$ and $\mathbf{a}_2$ are their direction vectors, obviously contains both straight lines (they are obtained respectively when $u = 0$ and $v = 0$). The uniqueness of that plane is ensured by the fact that it passes through a given point M_0 and has a given (to within the proportionality) direction bivector $\mathbf{a}_1 \wedge \mathbf{a}_2$. □

In coordinate notation the equation of this plane has the form

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix} = 0,$$

where l_1, m_1, n_1 and l_2, m_2, n_2 are the coordinates of the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ and x_0, y_0, z_0 are the coordinates of the point M_0.

If nonparallel straight lines with the direction vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ do not intersect (such straight lines are called *skew lines*), then they cannot lie in one plane. On the other hand, either of the lines is parallel to any plane with the direction bivector $\mathbf{a}_1 \wedge \mathbf{a}_2$, i.e. either it is contained in such a plane or does not intersect it. By choosing on one of the straight lines a point M_1 and drawing through it a plane with the direction bivector $\mathbf{a}_1 \wedge \mathbf{a}_2$ we obtain a plane containing that straight line and therefore not intersecting the other straight line. So for no point M_2 of the second straight line the vector $\overrightarrow{M_1M_2}$ is linearly expressed in terms of the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and hence the three vectors $\overrightarrow{M_1M_2}$, $\mathbf{a}_1$, $\mathbf{a}_2$ are linearly independent. Consequently, the determinant made up of their coordinates is nonzero.

$$\begin{vmatrix} x_2-x_1 & y_2-y_1 & z_2-z_1 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix} \neq 0.$$

Thus we have proved the following theorem.

Theorem 3 (relative positions of two straight lines in space). *In space two straight lines*

(a) *either do not lie in one plane* (*are skew*) *and then do not intersect*; *or*

(b) *lie in one plane and do not intersect*; *or*

(c) *lie in one plane and intersect in a single point*; *or*

(d) *coincide.*

In case (a) *the determinant of the matrix*

$$\begin{pmatrix} x_2-x_1 & y_2-y_1 & z_2-z_1 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{pmatrix} \tag{12}$$

is nonzero; *in case* (b) *the last two rows of the matrix* (12) *are proportional to each other but not to the first row*; *in case* (c) *the last two rows of the matrix* (12) *are not proportional but the first row is their linear combination*; *and at last in case* (d) *all the three rows of the matrix* (12) *are proportional.*

In cases (b) *and* (d) *the given straight lines are parallel and in cases* (a) *and* (c) *they are not.*

Remark. Since a straight line is given by two planes, Theorem 3 is a special case of a certain general theorem on the relative positions of four planes in space. Similarly, Theorem 2 is a special case of a general theorem on the relative positions of three planes in space. The various possible positions of the planes prove to be conveniently characterized by setting equal to zero (or nonzero) some minors of a matrix composed of the coefficients of the general equations of planes. This is nothing else but a geometric reformulation (for the case of three unknowns and three or four equations) of the general algebraic statements about solutions to systems of linear equations. We cannot busy ourselves with this here but we strongly recommend that the reader should elaborate the matter on his own.

We have noted in Lecture 5 that coordinates (in vector or affine space) can be chosen in many different ways. Let us study the arbitrariness existing here. Strictly speaking, this should have been done already in Lecture 5, but then the algebra course still lacked the necessary concepts (those of matrices and operations on matrices).

We shall begin with the case of vector spaces. For the time being dimension n will be assumed to be arbitrary, although in fact we need only $n = 1, 2, 3$.

Let the following two bases be given in a vector space $\mathscr{V}$:

$$\mathbf{e}_1, \ldots, \mathbf{e}_n \text{ and } \mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$$

(notice that we prime the indices; this seemingly strange notation proves very convenient; however, this relates only to considerations of a general character, similar to the ones we are dealing with now; it is better to prime the letters in particular calculations). Expanding the vectors $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ with respect to the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ we obtain formulas of the form

$$\begin{aligned} &\mathbf{e}_{1'} = c_{1'}^1\mathbf{e}_1 + \ldots + c_{1'}^n\mathbf{e}_n, \\ (13) \qquad &\ldots\ldots\ldots\ldots\ldots \\ &\mathbf{e}_{n'} = c_{n'}^1\mathbf{e}_1 + \ldots + c_{n'}^n\mathbf{e}_n, \end{aligned}$$

which in the Einstein notation have the form

$$\mathbf{e}_{i'} = c_{i'}^i\mathbf{e}_i.$$

It is possible to write these formulas in matrix notation.

A matrix

$$C = \begin{pmatrix} c_{1'}^1 & \ldots & c_{n'}^1 \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ c_{1'}^n & \ldots & c_{n'}^n \end{pmatrix}$$

is called a *matrix of transition* from a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to the basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$. (Note that the coordinates of the vectors $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ are arranged in columns in the matrix.) By introducing row matrices consisting of vectors

$$\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_n) \text{ and } \mathbf{e}' = (\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'})$$

we can write formulas (13) as a single matrix equation:

$$(14) \qquad \mathbf{e}' = \mathbf{e}C.$$

Here we extend operations known for number matrices to matrices of vector elements, assuming formally that for any vector $\mathbf{a}$ and any number k the symbol $\mathbf{a}k$ designates $k\mathbf{a}$. It is easy to see that all the usual operations on matrices remain valid. In particular, for any two matrices C and C'

we have the equation $(\mathbf{e}C)\,C' = \mathbf{e}\,(CC')$.

If now

$$\mathbf{e}_{1''}, \ldots, \mathbf{e}_{n''}$$

is a third basis and

$$\mathbf{e}_{i''} = c_{i''}^{i'} \mathbf{e}_{i'},$$

i.e.

$$\mathbf{e}'' = \mathbf{e}'C',$$

where $C' = (c_{i''}^{i'})$ is a matrix of transition from the basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ to the basis $\mathbf{e}_{1''}, \ldots, \mathbf{e}_{n''}$ and $\mathbf{e}'' = (\mathbf{e}_{1''}, \ldots, \mathbf{e}_{n''})$, then by virtue of the formula $(\mathbf{e}C)\,C' = \mathbf{e}\,(CC')$ we have

$$\mathbf{e}'' = \mathbf{e}\,(CC').$$

This proves that *if C is a matrix of transition from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to the basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ and C' is a matrix of transition from the basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ to the basis $\mathbf{e}_{1''}, \ldots, \mathbf{e}_{n''}$, then the matrix CC' is a matrix of transition from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to the basis $\mathbf{e}_{1''}, \ldots, \mathbf{e}_{n''}$.* □

This can also be very easy proved in the Einstein notation: since $\mathbf{e}_{i'} = c_{i'}^{i} \mathbf{e}_i$ and $\mathbf{e}_{i''} = c_{i''}^{i'} \mathbf{e}_{i'}$, then we have

$$\mathbf{e}_{i''} = c_{i''}^{i'} (c_{i'}^{i} \mathbf{e}_i) = (c_{i''}^{i'} c_{i'}^{i})\, \mathbf{e}_i,$$

i.e. $c_{i''}^{i} = c_{i''}^{i'} c_{i'}^{i}$.

In particular, if $\mathbf{e}'' = \mathbf{e}$ we see that $CC' = E$, where E is a unit matrix, i.e. that $C' = C^{-1}$. Thus, *if C is a matrix of transition from a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to a basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$, then the inverse matrix C^{-1} is a matrix of inverse transition from the basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ to the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$.* □

Note that in our notation elements of the matrix C^{-1} have the form $c_i^{i'}$ (i.e. differ from the elements of the matrix C only in the position of the prime).

The fact that the matrix C^{-1} exists shows that *any transition matrix C is nonsingular.* □

It is easy to see that, conversely, *any nonsingular matrix C is a matrix of transition from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ (which may be arbitrary) to some basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$*, i.e., in other words, that for any nonsingular matrix C the vectors $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ given by formulas (13) constitute a basis. Indeed, they are n in number and linearly independent (otherwise the columns of the matrix C would be linearly dependent and therefore, contrary to the assumption, its determinant would be equal to zero). □

Lecture 11

Formulas for the transformation of vector coordinates•Formulas for the transformation of the affine coordinates of points•Orientation•Induced orientation of a straight line•Orientation of a straight line given by an equation•Orientation of a plane in space

We shall now find the formulas connecting the coordinates in two different bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$.

By definition the coordinates $x^1, \ldots, x^n$ of an arbitrary vector $\mathbf{x}$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are found from the formula

(1) $$\mathbf{x} = x^i \mathbf{e}_i,$$

and the coordinates $x^{1'}, \ldots, x^{n'}$ of the same vector $\mathbf{x}$, but in the basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ are found from the formula

(2) $$\mathbf{x} = x^{i'} \mathbf{e}_{i'}$$

therefore, if

(3) $$\mathbf{e}_{i'} = c_{i'}^{i} \mathbf{e}_i,$$

then

$$\mathbf{x} = x^{i'} (c_{i'}^{i} \mathbf{e}_i) = (x^{i'} c_{i'}^{i})\, \mathbf{e}_i = (c_{i'}^{i} x^{i'})\, \mathbf{e}_i$$

and hence

(4) $$x^i = c_{i'}^{i} x^{i'},$$

i.e.

$$\begin{aligned} x^1 &= c_{1'}^{1} x^{1'} + \ldots + c_{n'}^{1} x^{n'}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ x^n &= c_{1'}^{n} x_{1'} + \ldots + c_{n'}^{n} x^{n'}. \end{aligned}$$

Note that whereas in formulas (3) the "new" basis $\mathbf{e}_{1'}, \ldots$ $\ldots, \mathbf{e}_{n'}$ is expressed in terms of the "old" basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, in formulas (4) the "old" coordinates $x^1, \ldots, x^n$ are, on the contrary, expressed in terms of the "new" coordinates $x^{1'}, \ldots, x^{n'}$. Besides, the corresponding matrix is not a matrix C, but a transposed matrix $C^\top$. Formally it is expressed by summation being performed over the unprimed indices in formula (3) and over the primed ones in formula (4).

Of course, it is also easy to write expressions for the "new" coordinates in terms of the "old" coordinates. They are of the form

$$x^{i'} = c_i^{i'} x^i.$$

The change in the position of the prime implies, as we know, a transition to the inverse matrix.

Pay attention to the structure of the resulting formulas: on the left and on the right of all of them the same indices over which no summation is carried out occupy the same places (above or below). Very often this formal sign, together with the Einstein convention (of the two indices over which summation is carried out one is above and the other is below), makes it possible to write almost automatically the right formulas (and to detect the wrong ones).

Formulas (4) (called the *formulas for the transformation of coordinates*) are also easy to obtain in matrix notation.

Introducing along with the row matrix $\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ (see the preceding lecture) the column matrix

$$x = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix},$$

we can write formula (1) in the following form:

$$\mathbf{x} = \mathbf{e}x$$

(the order of multipliers being essential here as ever in matrix multiplication). Similarly, formula (2) in matrix notation becomes

$$\mathbf{x} = \mathbf{e}'x',$$

where $\mathbf{e}' = (\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'})$ and

$$x' = \begin{pmatrix} x^{1'} \\ \vdots \\ x^{n'} \end{pmatrix}.$$

Therefore, if

$$\mathbf{e}' = \mathbf{e}C$$

(see formula (14) of the preceding lecture), then

$$\mathbf{x} = (\mathbf{e}C)\, x' = \mathbf{e}\,(Cx')$$

and hence

$$x = Cx', \tag{5}$$

which is exactly equivalent to formula (4) (an actual transition to the transposed matrix is reflected in the fact that formula (5) contains left multiplication by the matrix C.

We shall now consider a similar question for affine coordinates in an affine space.

The coordinates $x^1, \ldots, x^n$ of a point M in an affine coordinate system $O\mathbf{e}_1 \ldots \mathbf{e}_n$ are determined in two steps: the radius vector $\overrightarrow{OM}$ is first constructed and then its coordinates calculated in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Therefore, if two coordinate systems $O\mathbf{e}_1 \ldots \mathbf{e}_n$ and $O'\mathbf{e}_{1'} \ldots \mathbf{e}_{n'}$ have the same origin $O = O'$, then a transition from coordinates in the system $O\mathbf{e}_1 \ldots \mathbf{e}_n$ to coordinates in the system $O\mathbf{e}_{1'} \ldots \mathbf{e}_{n'}$ is effected using the already familiar formulas for the coordinates of vectors.

The other extreme case arises when two coordinate systems differ only in the initial points, i.e. have the forms $O\mathbf{e}_1 \ldots \mathbf{e}_n$ and $O'\mathbf{e}_1 \ldots \mathbf{e}_n$. Since

$$\overrightarrow{OM} = \overrightarrow{O'M} + \overrightarrow{OO'},$$

the coordinates $x^1, \ldots, x^n$ and $x^1, \ldots, x^{n'}$ of the vectors $\overrightarrow{OM}$ and $\overrightarrow{O'M}$ are connected by the formulas

$$x^i = x^{i'} + b^i, \quad i = 1, \ldots, n,$$

where $b^1, \ldots, b^n$ are the coordinates (in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$) of the vector $\overrightarrow{OO'}$.

In the general case there naturally arises a combination of these transformations. Thus, if $x^1, \ldots, x^n$ are the coordinates of a point M in the coordinate system $O\mathbf{e}_1 \ldots \mathbf{e}_n$ and $x^{1'}, \ldots, x^{n'}$ are the coordinates of the same point in the coordinate system $O'\mathbf{e}_{1'} \ldots \mathbf{e}_{n'}$, then

$$x^i = c^i_{i'} x^{i'} + b^i, \tag{6}$$

where $(c^i_{i'})$ is the matrix of transition from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to the basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$, and $b^1, \ldots, b^n$ are the coordinates of the vector $\overrightarrow{OO'}$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

In the expanded form formula (5) is written as

$$\begin{aligned} x^1 &= c^1_{1'} x^{1'} + \ldots + c^1_{n'} x^{n'} + b^1, \\ &\cdots\cdots\cdots\cdots\cdots \\ x^n &= c^n_{1'} x^{1'} + \ldots + c^n_{n'} x^{n'} + b^n, \end{aligned}$$

and in matrix form as

$$x = Cx' + b, \tag{7}$$

where b is the column matrix

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b^n \end{pmatrix}.$$

Closely related to the question of the transformation of coordinates is the very important concept of *orientation* explicating formally the intuitive idea of the direction of rectilinear motion (to the right or left), of plane rotational motion (clockwise or counterclockwise) and of spatial screw motion (right-handed or left-handed).

Here in contrast to the preceding reasoning the essential assumption is that the ground field $\mathbb{K}$ is the field of real number $\mathbb{R}$.

Definition 1. Two bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ for a vector space $\mathscr{V}$ are said to be *of the same sign* if the determinant of the matrix of transition, C, from the first basis to the second is positive:

$$\det C > 0.$$

If $\det C < 0$, then the bases are said to be *of opposite signs.*

Two affine coordinate systems $Oe_1 \ldots e_n$ and $O'e_{1'} \ldots$ $\ldots e_{n'}$ of an affine space $\mathcal{A}$ are said to be *of the same* (*opposite*) *signs* if so are the bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ of the associated vector space $\mathcal{V}$.

Since a set of matrices with a positive determinant is a *group* (i.e. is closed under multiplication and finding the inverse matrix), the sameness-of-the-sign relation between bases (coordinate systems) is an equivalence relation.

Definition 2. The classes of bases (coordinate systems) of the same sign are called *orientations* of a vector space $\mathcal{V}$ (of an affine space $\mathcal{A}$).

It is clear that *there are exactly two orientations*: bases (coordinate systems) having the same orientation are of the same sign and those that have different orientation are of opposite signs. Orientation different from orientation o is designated by the symbol $-o$ and called *opposite orientation.*

Bases (coordinate systems) of the same sign are also called bases (coordinate systems) *of the same orientation.*

Not infrequently one has to determine where two bases $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and $\mathbf{b}_1, \ldots, \mathbf{b}_n$ whose vectors are given by their coordinates in some third basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are of the same or opposite signs. Let Δ_a (Δ_b respectively) be a determinant whose columns are the columns of the coordinates of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ (of $\mathbf{b}_1, \ldots, \mathbf{b}_n$ respectively). It is clear that, say, Δ_a is nothing else than the determinant of the matrix of transition from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to the basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Therefore *bases* $\mathbf{a}_1, \ldots, \mathbf{a}_n$ *and* $\mathbf{b}_1, \ldots, \mathbf{b}_n$ *are of the same sign if and only if the determinants* Δ_a *and* Δ_b *have the same sign.* $\square$

Definition 3. A vector (or affine) space is said to be *oriented* if a certain orientation has been chosen in it. This orientation is called *positive* and the opposite orientation is called *negative*. Accordingly, a basis (coordinate system) determining a chosen orientation is called *positively oriented* and a basis (coordinate system) determining the opposite orientation is called *negatively oriented.*

Note that it is often impossible to select from the two possible orientations of a vector (affine) space any definite one in an intrinsically mathematical way. To do this one has to resort to consideration extraneous to mathematics (say, to the earth's rotation or human anatomy).

For example, choosing an orientation on a straight line means choosing on it one of the two possible directions. If a straight line is represented in a drawing by a horizontal line, it is common to select on it the direction "from left to right". This direction is the more customary (if only because this is the direction of writing with most civilized peoples) and therefore it is given the name of the "positive" direction ("positive" orientation). It should be understood, however, that fixation of this direction has no invariant (drawing-independent) meaning: turn the drawing upside down and the "positive" direction becomes "negative".

Similarly, the direction from below upwards is considered as "positive" on vertical straight lines.

Giving an orientation in a plane is equivalent to giving a certain (clockwise or counterclockwise) "sense of rotation", namely the sense in which the first vector of the basis determining the given orientation should be rotated to make its direction coincide with that of the second vector in the shortest way. It is customary to consider as the "positive" orientation of the drawing the "counterclockwise" direction (the abscissa axis being directed to the right and the ordinate axis upwards). This convention is certainly not invariant either: look at the plane from the other side and you have a change in the orientation.

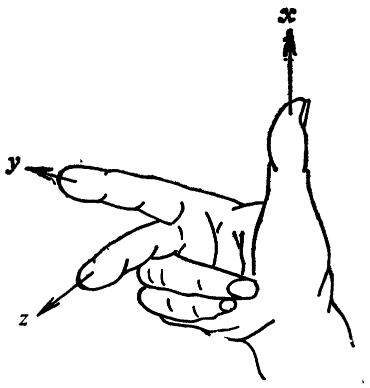
If you look at the palm of your right hand, its thumb and forefinger will form a basis having a counterclockwise orientation. For this reason a "counterclockwise" orientation of a plane is generally referred to as the "right-handed" orientation and a "clockwise" orientation as the "left-handed".

Similarly, in space a "right-handed" orientation is the orientation determined by the "basis" consisting of the thumb, index, and middle fingers of the right hand. In modern text-books it is the right-handed orientation of space that is generally regarded as its "positive" orientation; in many older works, however, that orientation was regarded as "negative".

Note that the terms "right-handed" and "left-handed" have an invariant meaning with respect to the orientation of space, since we cannot look at space "from the other side" (nor can we make our left hand coincide with our right hand). Nevertheless they cannot be defined in an intrinsically mathematical way.

Without additional data, the orientation of a plane (an affine space of dimension 2) and that of straight lines (affine spaces of dimension 1) located in that plane are in no way connected. To connect them, one may give, for example, a "side" of a straight line.

Definition 4. A *side* of a straight line d in a plane is said to be given if one of the two half-planes into which d



The right-handed orientation of space

divides the plane has been chosen. The chosen side is referred to as *positive* and the opposite side as *negative*.

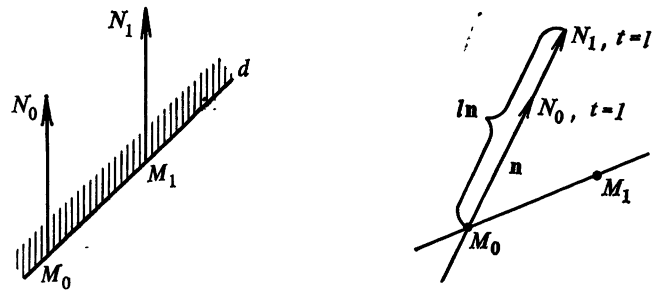
Let $\mathbf{n}$ be an arbitrary vector not parallel to a straight line d (and, in particular, different from zero). Having chosen on d an arbitrary point M_0, mark off from it the vector $\mathbf{n}$, i.e. find a point N_0 such that $\mathbf{n} = \overrightarrow{M_0N_0}$.

Definition 5. It is said that a side containing a point N_0 is *determined by the vector* $\mathbf{n}$ and that the vector $\mathbf{n}$ is *directed* toward that side.

It is certainly necessary here to check the correctness of the definition, i.e. the independence of the side containing the point N_0 on the choice of a point M_0. But it is easily done. Indeed, if M_1 is another point of d and $\mathbf{n} = \overrightarrow{M_1N_1}$, then

$$\overrightarrow{N_0N_1} = \overrightarrow{N_0M_0} + \overrightarrow{M_0M_1} + \overrightarrow{M_1N_1} = \overrightarrow{M_0M_1}$$

and hence the straight line N_0N_1 is parallel to the straight line $d = M_0M_1$. Consequently, the points N_0 and N_1 are on the same side of d. □



A side of a straight line

Let $\mathbf{a}$ be a direction vector of a straight line d. Together with a vector $\mathbf{n}$ it constitutes the basis $\mathbf{a}, \mathbf{n}$ of a plane and so any third vector $\mathbf{n}_1$ is decomposed with respect to these vectors:

$$\mathbf{n}_1 = l_1\mathbf{a} + l\mathbf{n}.$$

Suppose the vector $\mathbf{n}_1$ is not parallel to d (and hence $l \neq 0$). It is easy to see that in this case the *vector* $\mathbf{n}_1$ *is directed toward the same side of d as the vector* $\mathbf{n}$ *if and only if* $l > 0$. Indeed, on marking off from a point $M_0 \in d$ with radius vector $\mathbf{r}_0$ a vector $\mathbf{n}_1$ we obtain the same point as we do on marking from a point $M_1 \in d$ with radius vector $\mathbf{r}_0 + l_1\mathbf{a}$ a vector $l\mathbf{n}$. Therefore we may assume without loss of generality that $\mathbf{n}_1 = l\mathbf{n}$. But then the points N_0 and N_1 for which $\mathbf{n} = \overrightarrow{M_0N_0}$ and $\mathbf{n}_1 = \overrightarrow{M_0N_1}$ correspond on the straight line $\mathbf{r} = r_0 + t\mathbf{n}$ to the values $t = 1$ and $t = l$ of the parameter t and hence the segment $\overline{N_0N_1}$ is determined by the inequalities $1 \leqslant t \leqslant l$. On the other hand, the point M_0 of this straight line (which is its point of intersection with d) corresponds to the value $t = 0$.

Consequently, the points N_0 and N_1 are on the same side of d if and only if $l > 0$. □

The orientation of the straight line d is given by a direction vector **a** (constituting a basis on the straight line), the two vectors **a** and $\mathbf{a}_1 = k\mathbf{a}$ giving the same orientation if and only if $k > 0$.

Suppose we are given the orientation of a straight line d (by a direction vector **a**) and some side of that straight line (by a vector **n**). Then there arises a basis **a**, **n** in the plane. It is easy to see that the orientation of the plane given by that basis does not depend on the choice of **a** and **n** and is determined exclusively the given orientation and the given side of d. Indeed, if vectors $\mathbf{a}_1$ and $\mathbf{n}_1$ give the same orientation and side, then, according to what was said above,

$$\mathbf{a}_1 = k\mathbf{a},$$

$$\mathbf{n}_1 = l_1\mathbf{a} + l\mathbf{n},$$

where $k > 0$ and $l > 0$ and therefore

$$\begin{vmatrix} k & l_1 \\ 0 & l \end{vmatrix} = kl > 0. \quad \square$$

Thus of the three objects,
(a) the orientation of a straight line d,
(b) the side of the straight line d,
(c) the orientation of a plane,
the first two uniquely determine the third.

It is easy to see, however, that *any two of these objects uniquely determine a third.* Given an orientation of the plane and a side (or orientation) of the straight line by a vector **n** (a vector **a**), the corresponding orientation (side) of the straight line is given by the vector **a** (**n**) possessing the property that the basis **a**, **n** is positively oriented. The correctness of the construction is checked on the basis of the same formulas as above. □

All this may be considered in a slightly different way if the concept of the orientation of a half-plane is introduced.

Definition 6. *The orientation of a half-plane* is the orientation of the plane containing it. A vector in a plane is said to be *directed into* the given half-plane if it is directed toward that half-plane in the sense of Definition 5.

Every orientation of a half-plane determines the orientation of the straight line bounding that half-plane. This orientation is given by a direction vector **a** constituting, together with an arbitrary vector **n** directed into the half-plane, a positively oriented basis for the plane. This orientation is said to be *induced* by the given orientation of the half-plane.

Although this reiterates what was said above in slightly different terms, we shall see in due course that this reformulation is useful.

Since the orientation of a straight line is essentially the orientation of the associated vector space, the assertion that orientations on two different but parallel lines coincide and differ has meaning. Therefore it is also possible to compare the sides of parallel lines.

Note that the orientations (and sides) of non-parallel lines cannot be compared.

Giving in a plane a coordinate system $O\mathbf{e}_1\mathbf{e}_2$ defines a certain orientation of it (namely that in which the basis $\mathbf{e}_1$, $\mathbf{e}_2$ is positively oriented). On the other hand, giving the equation

$$Ax + By + C = 0 \tag{8}$$

by a straight line defines a certain side of it, namely that side for the points $M(x, y)$ of which $Ax + By + C > 0$. Thus once an equation of the straight line has been given (and hence a coordinate system written), that straight line automatically turns out to be oriented.

This orientation (and side) remains unchanged when equation (8) is multiplied by a positive number and becomes opposite when the equation is multiplied by a negative number.

It is easy to see that the vector **n** with the coordinates (A, B) is directed toward the positive side of the straight line (8). Indeed, if $\mathbf{n} = \overrightarrow{M_0N_0}$, where $M_0(x_0, y_0)$ is a point of the straight line (8), then for the coordinates $x_0 + A$, $y_0 + B$ of the point N_0 we have the inequality

$$A(x_0+A)+B(y_0+B)+C=$$
$$=A^2+B^2+(Ax_0+By_0+C=A^2+B^2>0. \quad \square$$

It follows that *the orientation of the straight line determined by equation* (8) *is given by the direction vector* **a** $(B, -A)$. Indeed,

$$\begin{vmatrix} B & A \\ -A & B \end{vmatrix} = A^2+B^2>0.$$

Quite similar results hold for planes in space.

Definition 7. A *side* of a plane in space is said to be given if one of the two half-spaces into which the plane divides the space has been chosen.

It can be proved by duplicating word for word the case of the straight line that

(i) for any vector **n** not parallel to a plane and for any point M_0 of the plane the side of the plane determined by a point N_0 for which $\mathbf{n} = \overrightarrow{M_0N_0}$ does not depend on the choice of the point M_0, i.e. is determined by the vector **n**; the vector **n** is said to be *directed toward that side*;

(ii) a vector of the form $\mathbf{a} + l\mathbf{n}$, where **a** is an arbitrary vector parallel to the plane, is directed toward the same side as the vector **n** if and only if $l > 0$;

(iii) of the three objects,

(a) the orientation of the plane,

(b) the side of the plane,

(c) the orientation of the space,

any two determine a third.

For example, if the orientation of a plane is given by a basis **a**, **b** and the side of the plane is given by a vector **n**, then the orientation of the space is by definition given by the basis **a**, **b**, **n**. If

$$\mathbf{a}_1 = k_1\mathbf{a} + k_2\mathbf{b},$$
$$\mathbf{b}_1 = k_1'\mathbf{a} + k_2'\mathbf{b}$$

is a basis for the plane, of the same sign as the basis **a**, **b**, and

$$\mathbf{n}_1 = l_1\mathbf{a} + l_2\mathbf{b} + l\mathbf{n}$$

is a vector directed toward the same side as the vector **n**, then the bases $\mathbf{a}$, $\mathbf{b}$, $\mathbf{n}$ and $\mathbf{a}_1$, $\mathbf{b}_1$, $\mathbf{n}_1$ are of the same sign, since the determinant of their transition matrix is

$$\begin{vmatrix} k_1 & k_1' & l_1 \\ k_2 & k_2' & l_2 \\ 0 & 0 & l \end{vmatrix} = \begin{vmatrix} k_1 & k_1' \\ k_2 & k_2' \end{vmatrix} l$$

and hence positive, for under the hypothesis

$$\begin{vmatrix} k_1 & k_1' \\ k_2 & k_2' \end{vmatrix} > 0 \text{ and } l > 0.$$

Consequently, the orientation of a space given by the basis $\mathbf{a}$, $\mathbf{b}$, $\mathbf{n}$ has been defined correctly.

A slightly different account of this construction may be given by introducing (in an obvious way) the *orientation of a half-space*. Then it turns out that every orientation of a half-space will naturally determine the orientation (called the *induced orientation*) of the plane bounding that half-space. This orientation is given by a basis $\mathbf{a}$, $\mathbf{b}$ constituting together with an arbitrary vector $\mathbf{n}$ directed into the half-space a positively oriented basis $\mathbf{a}$, $\mathbf{b}$, $\mathbf{n}$ for the space.

If in a space an affine coordinate system $O\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$ is given so is an orientation. And the equation of a plane

$$Ax + By + Cz + D = 0 \tag{9}$$

gives its side (by the condition $Ax + By + Cz + D > 0$). Therefore the equation of a plane gives its orientation, which remains unchanged when equation (9) is multiplied by a positive number and transforms into the opposite orientation when it is multiplied by a negative number.

As in the case of the straight line a vector $\mathbf{n}$ with the coordinates A, B, C is directed toward the positive side of the plane (2). As to the orientation of a plane defined by equation (9), it is given by an arbitrary pair $\mathbf{a}$, $\mathbf{b}$ of linearly independent vectors parallel to the plane and possessing the property that

$$\begin{vmatrix} a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \\ A & B & C \end{vmatrix} > 0, \tag{10}$$

where a^1, a^2, a^3 and b^1, b^2, b^3 are the coordinates of the vectors **a** and **b**. (By tradition the transposed determinant is used here.)

However, instead of vectors **a** and **b** it is convenient to consider their external product $\mathbf{a} \wedge \mathbf{b}$. Indeed, it follows immediately from Proposition 1 of Lecture 7 that if $\mathbf{a} \wedge \mathbf{b} = \mathbf{a}_1 \wedge \mathbf{b}_1 = \alpha$, then the bases $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{a}_1$, $\mathbf{b}_1$ for a vector space of dimension 2 determined by a bivector α are of the same sign. Therefore the orientation of a two-dimensional vector space $\mathscr{V}$ may be given by an arbitrary nonzero bivector of $\mathscr{V} \wedge \mathscr{V}$. In other words, just as the orientation of a straight line is nothing else than a class of its positively proportional direction vectors so the orientation of a plane may be regarded as a class of its positively proportional direction bivectors.

According to formulas (12) of Lecture 8, there exist linearly independent vectors **a** and **b**, parallel to the plane (9), such that the coordinates of the bivector $\mathbf{a}' \wedge \mathbf{b}$, i.e. the cofactors of elements in the last column of the determinant (10) are just the elements A, B, C of that column. Therefore for such vectors the determinant (10) equals $A^2 + B^2 + C^2$ and hence is positive. This proves that the orientation of the plane defined by equation (9) is given by the direction bivector $\alpha\,(A, B, C)$. □

Lecture 12

Deformation of bases • Sameness of the sign bases • Equivalent bases and matrices • The coincidence of deformability with the sameness of sign • Equivalence of linearly independent systems of vectors • Trivectors • The product of a trivector by a number • The external product of three vectors.

The sameness of the sign relation between bases (and hence the concept of orientation) may be described in a quite different way not connected directly with transition matrices. To do this we shall need some general definitions concerning the continuity of vector-valued functions.

Let $t \mapsto \mathbf{a}(t)$ be a function whose values are the vectors of a vector space $\mathscr{V}$. We shall assume for definiteness that this function is given for $0 \leqslant t \leqslant 1$.

For a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ the function $t \mapsto \mathbf{a}(t)$ is uniquely determined by the coordinates

$$a^1(t), \ldots, a^n(t) \tag{1}$$

of the vector $\mathbf{a}(t)$ which are n number functions.

Definition 1. A function $t \mapsto \mathbf{a}(t)$ is said to be *continuous* if all functions (1) are continuous.

When a basis is changed functions (1) are replaced by some of their linear combinations with constant coefficients and so remain continuous. This shows that Definition 1 is correct.

Definition 2. A *deformation of bases* for a vector space $\mathscr{V}$ is a family of n continuous functions

$$\mathbf{a}_1(t), \ldots, \mathbf{a}_n(t) \tag{2}$$

determined for $0 \leqslant t \leqslant 1$ such that for every t the vectors (2) constitute a basis for the vector space $\mathscr{V}$.

A basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is said to be *deformable* into a basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ if there exists a deformation of bases

$$\mathbf{a}_1(t), \ldots, \mathbf{a}_n(t)$$

such that

$$\mathbf{a}_1(0) = \mathbf{a}_1, \ldots, \mathbf{a}_n(0) = \mathbf{a}_n$$

and

$$\mathbf{a}_1(1) = \mathbf{b}_1, \ldots, \mathbf{a}_n(1) = \mathbf{b}_n.$$

This deformation is said to *connect* the first basis with the second.

Proposition 1. *The deformability relation between bases is an equivalence relation.*

Proof. This relation is reflexive since for any basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ the formulas

$$\mathbf{a}_1(t) = \mathbf{a}_1, \ldots, \mathbf{a}_n(t) = \mathbf{a}_n$$

obviously determine the deformation connecting that basis with itself. It is symmetrical since for any deformation of the bases $\mathbf{a}_1(t), \ldots, \mathbf{a}_n(t)$ the formulas

$$\mathbf{b}_1(t) = \mathbf{a}_1(1-t), \ldots, \mathbf{b}_n(t) = \mathbf{a}_n(1-t)$$

determine the deformation connecting the basis $\mathbf{b}_1 = \mathbf{a}_1(1), \ldots, \mathbf{b}_n = \mathbf{a}_n(1)$ with the basis $\mathbf{a}_1 = \mathbf{a}_1(0), \ldots, \mathbf{a}_n = \mathbf{a}_n(0)$. It is transitive since for any two deformations of the bases $\mathbf{a}_1(t), \ldots, \mathbf{a}_n(t)$ and $\mathbf{b}_1(t), \ldots, \mathbf{b}_n(t)$ having the property that $\mathbf{a}_1(1) = \mathbf{b}_1(0), \ldots, \mathbf{a}_n(1) = \mathbf{b}_n(0)$ the formulas

$$\mathbf{c}_i(t) = \begin{cases} \mathbf{a}_i(2t) & \text{if } 0 \leqslant t \leqslant \frac{1}{2} \\ \mathbf{b}_i(2t-1) & \text{if } \frac{1}{2} \leqslant t \leqslant 1 \end{cases} \qquad i = 1, \ldots, n$$

determine the deformation connecting the basis $\mathbf{a}_1 = \mathbf{a}_1(0), \ldots, \mathbf{a}_n = \mathbf{a}_n(0)$ with the basis $\mathbf{c}_1 = \mathbf{b}_1(1), \ldots, \mathbf{c}_n = \mathbf{b}_n(1)$. □

It is easy to see that *bases deformable into each other are of the same sign.* Indeed, let $\Delta(t)$ be the determinant of a matrix of transition from a basis $\mathbf{a}_1 = \mathbf{a}_1(0), \ldots, \mathbf{a}_n = \mathbf{a}_n(0)$ to a basis $\mathbf{a}_1(t), \ldots, \mathbf{a}_n(t)$. It is clear that $\Delta(t)$

is continuously dependent on t and that $\Delta(0) = 1$. Besides, $\Delta(t) \neq 0$ for every t, $0 \leqslant t \leqslant 1$. But it is known from calculus that a continuous function nonvanishing in an interval maintains its sign. Therefore $\Delta(1) > 0$ and hence the basis $\mathbf{b}_1 = \mathbf{a}_1(1), \ldots, \mathbf{b}_n = \mathbf{a}_n(1)$ is of the same sign as the basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$. □

The converse is also true and easy to prove for $n = 1$ (i.e. for the case of a straight line).

Indeed, for $n = 1$ the bases are one-member families consisting each of a nonzero vector. Any two such bases $\mathbf{a}$ and $\mathbf{b}$ are proportional, i.e. there is $\delta \in \mathbb{R}$ such that $\mathbf{b} = \delta\mathbf{a}$. If the bases are of the same sign, then $\delta > 0$. Therefore the formula

$$\text{(3)} \qquad \mathbf{a}(t) = (1 + t(\delta - 1))\,\mathbf{a}$$

determines the deformation of the basis $\mathbf{a} = \mathbf{a}(0)$ into the basis $\mathbf{b} = \mathbf{a}(1)$, for if $\delta > 0$, then the function $1 + t(\delta - 1)$ does not vanish for $0 \leqslant t \leqslant 1$.

Thus *for $n = 1$ bases of the same sign are deformable into each other.* □

As far as the case $n > 1$ is concerned, we are compelled to begin its study from afar, from questions which on the face of it have nothing to do with the matter at hand.

Definition 3. Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be an arbitrary basis (or more generally, an arbitrary family of vectors). An operation consisting in adding to some element of the basis another of its elements multiplied by an arbitrary number k will be called an *elementary transformation of type* 1 and an operation consisting in multiplying some element of the basis by an arbitrary number $\lambda \neq 0$ and simultaneously multiplying another of its elements by $1/\lambda$ will be called an *elementary transformation of type* 2.

An example of elementary transformation of type 1 is

$$(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) \Rightarrow (\mathbf{a}_1 + k\mathbf{a}_2, \mathbf{a}_2, \ldots, \mathbf{a}_n)$$

an example of type 2 being

$$(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) \Rightarrow \left(\lambda\mathbf{a}_1, \frac{1}{\lambda}\mathbf{a}_2, \ldots, \mathbf{a}_n\right).$$

For $n = 2$ these are exactly the elementary transformations in the sense of Definition 1 of Lecture 7.

Let C be a matrix of transition from a basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ to a basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$. If we subject the basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ to an elementary transformation, then in the matrix C one column multiplied by k will be added to the other or, respectively, one column will be multiplied by λ and the other by $1/\lambda$. The matrix C is subjected to a similar transformation (but with respect to the rows and with k and λ changed) if we make an elementary transformation of the basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

Matrix transformations of this kind will also be called elementary transformations (of type 1 and type 2 respectively).

Definition 4. Two bases (or matrices) are said to be *equivalent* if it is possible to connect them by a chain of elementary transformations.

Lemma 1. *Any square matrix is equivalent to a diagonal matrix.*

Proof. Let some matrix row have two nonzero elements c and c'. Then, on adding to the column containing c' the column containing c and multiplied by $k = -c'/c$ we obtain a matrix in which the number of nonzero elements in the row considered is decreased by one. Repeating this step we shall obtain a matrix the given row of which contains exactly one nonzero element c and applying then the same process to the column containing the element c we obtain a matrix which has no nonzero elements, except the element c, in that column either.

We again apply the same transformations to the constructed matrix (and to its arbitrary nonzero element c_1) to cause in that row and that column whose intersection contains the element c_1 the vanishing of all elements except the element c_1 itself. It is clear that the elements of the row and column containing the element c remain zero in the process.

After at most n steps (where n is the order of the matrix) we obtain a matrix in each row and each column of which there is at most one nonzero element, there being exactly one such element if the original matrix is nonsingular (for in elementary transformations the determinant of a matrix remains unchanged and hence the matrix remains nonsingular). In other words, we arrive at a matrix that is obtained from a certain diagonal matrix by interchanging its rows and columns.

On the other hand, in Lecture 7 (see formula (4) there) we have proved (true, in connection with pairs of vectors and not pairs of matrix rows or columns, but clearly this is of no importance) that by a chain of elementary transformations one can interchange any two rows (any two columns) in any matrix and simultaneously change the sign of one of them. This obviously proves Lemma 1. □

Note that we employed only elementary transformations of type 1 in this proof.

Lemma 2. *Any nonsingular matrix C is equivalent to a matrix of the form*

$$\begin{pmatrix} \delta & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

where $\delta = \det C$.

Proof. According to Lemma 1 the matrix C is equivalent to a matrix of the form

$$\begin{vmatrix} c_1 & & & 0 \\ & c_2 & & \\ & & \ddots & \\ 0 & & & c_n \end{vmatrix},$$

where $c_1 c_2 \dots c_n = \det C \neq 0$. (As was already noted above, in elementary transformations the determinant of a matrix remains unchanged.) We subject the last matrix to $n - 1$ elementary transformations of type 2 which consists in multiplying the first column by $c_2, c_3, \dots, c_n$ in turn and in dividing the remaining columns respectively by the same numbers. It is clear that as a result we obtain a matrix of the required form. □

We restate Lemma 2 for bases.

Lemma 3. *Two arbitrary bases for which the determinant of a transition matrix is equal to δ can be changed by elementary* *transformations into bases* $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and $\mathbf{b}_1, \ldots, \mathbf{b}_n$ *connected by the relations*

$$(4) \qquad \mathbf{b}_1 = \delta\mathbf{a}_1, \quad \mathbf{b}_2 = \mathbf{a}_2, \ldots, \mathbf{b}_n = \mathbf{a}_n. \ \square$$

To apply these results to the problem of the deformability of bases, it is enough to note that *equivalent bases are deformable into each other*. Indeed, in view of Proposition 1 it is enough to prove this for the bases $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and $\mathbf{b}_1, \ldots, \mathbf{b}_n$ connected by elementary transformations. But if, for example,

$$\mathbf{b}_1 = \mathbf{a} + k\mathbf{a}_2, \quad \mathbf{b}_2 = \mathbf{a}_2, \ldots, \mathbf{b}_n = \mathbf{a}_n,$$

then the corresponding deformation can be given by the formulas

$$\mathbf{a}_1(t) = \mathbf{a}_1 + kt\mathbf{a}_2, \ \mathbf{a}_2(t) = \mathbf{a}_2, \ldots, \mathbf{a}_n(t) = \mathbf{a}_n$$

and if

$$\mathbf{b}_1 = \lambda\mathbf{a}_1, \ \mathbf{b}_2 = \frac{1}{\lambda}\mathbf{a}_2, \ \mathbf{b}_3 = \mathbf{a}_3, \ \ldots, \ \mathbf{b}_n = \mathbf{a}_n,$$

where $\lambda > 0$, then it is possible to apply the deformation described above for $n = 1$ (see formula (3)) to each of the vectors $\mathbf{b}_1$ and $\mathbf{b}_2$. Finally, in the case $\lambda = -1$ (the only one left) it is possible to use the deformation given by the formulas

$$\mathbf{a}_1(t) = (\cos \pi t)\,\mathbf{a}_1 - (\sin \pi t)\,\mathbf{a}_2,$$
$$\mathbf{a}_2(t) = (\sin \pi t)\,\mathbf{a}_1 + (\cos \pi t)\,\mathbf{a}_2,$$
$$\mathbf{a}_3(t) = \mathbf{a}_3, \ \ldots, \ \mathbf{a}_n(t) = \mathbf{a}_n. \ \square$$

We can now prove the basic theorem.

Theorem 1. *Two bases are of the same sign if and only if they are deformable into each other.*

Proof. The fact that deformable bases are of the same sign was proved above. By the remark just made and Lemma 3 any two bases of the same sign can be deformed into bases $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and $\mathbf{b}_1, \ldots, \mathbf{b}_n$ connected by relations (4), where $\delta > 0$. Since $\delta > 0$, we can apply to the vectors $\mathbf{a}_1$ and $\mathbf{b}_1$ the deformation (3). Consequently, the original bases are deformable into each other. $\square$

Thus we have the right to define orientations as classes of bases deformable into each other.

We now return to the investigation of the equivalence relation introduced by Definition 4. As we already observed in passing, the concept of elementary transformation, and hence the corresponding equivalence relation, makes sense for any families of vectors. It turns out, however, that it is appropriate to investigate it only for linearly independent families. Therefore we restrict ourselves to such families only. To emphasize their possible incompleteness we shall denote the number of vectors in these families by the letter m (reserving as before the letter n for denoting the dimension of a vector space $\mathscr{V}$ that is considered).

When $m = 2$ Definition 4 (for linearly independent, i.e. noncollinear, vectors) coincides with Definition 2 of Lecture 7. It is natural therefore, in the general case as well, to call classes of equivalent m-member linearly independent families of vectors *m-vectors* adding to them the *zero m-vector* consisting by definition of all m-member linearly dependent families of vectors.

To formulate the conditions for the equality of two m-vectors (i.e. to find the analogue of Proposition 1, Lecture 7) it is convenient to introduce the following definition:

Definition 5. Two families of vectors are said to be *linearly equivalent* if any vector of either family can be linearly expressed in terms of the vectors of the other.

Obviously linearly equivalent linearly independent families consist of the same number of vectors. For any two of such families $\mathbf{a}_1, \ldots, \mathbf{a}_m$ and $\mathbf{b}_1, \ldots, \mathbf{b}_m$ we have formulas of the form

$$\begin{gathered}\mathbf{b}_1 = c_1^1 \mathbf{a}_1 + \ldots + c_1^m \mathbf{a}_m, \\ \cdots\cdots\cdots\cdots \\ \mathbf{b}_m = c_m^1 \mathbf{a}_1 + \ldots + c_m^m \mathbf{a}_m,\end{gathered}$$

whose coefficients make up a nonsingular matrix

$$C = \begin{pmatrix} c_1^1 & \ldots & c_m^1 \\ \cdot & \cdot\,\cdot\,\cdot & \cdot \\ c_1^m & \ldots & c_m^m \end{pmatrix}$$

called a *matrix of transition* from the family $\mathbf{a}_1, \ldots, \mathbf{a}_m$ to the family $\mathbf{b}_1, \ldots, \mathbf{b}_m$.

If the determinant of that matrix is equal to unity,

$$\det C = 1,$$

then the families $\mathbf{a}_1, \ldots, \mathbf{a}_m$ and $\mathbf{b}_1, \ldots, \mathbf{b}_m$ are said to be *unimodularly equivalent*.

It is easy to see that *Lemma* 3 *remains valid for any linearly equivalent linearly independent families of vectors.* This fact plays a key role in proving the following proposition turning for $m = 2$ into Proposition 1 of Lecture 7.

Proposition 2. *Two linearly independent families of vectors are unimodularly equivalent if and only if they are equivalent in the sense of Definition* 4 (*are converted into each other by a chain of elementary transformations*), *i.e. if they determine the same m-vector.*

Proof. Two linearly independent families connected by an elementary transformation are obviously unimodularly equivalent. Since, on the other hand, the relation of unimodular equivalence is transitive (for the product of two unimodular matrices is again a unimodular matrix), this proves that equivalent families are unimodularly equivalent. Conversely, according to Lemma 3 unimodularly equivalent families are equivalent to the same family (for $\delta = 1$) and hence equivalent. □

The case of 2-vectors (bivectors) was studied in detail in Lectures 7 and 8. We shall therefore consider the case of 3-vectors (also called "trivectors").

For clarity we shall repeat the basic definitions for this case.

An *elementary transformation* of a triple of vectors $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ consists either in adding to one of the vectors another vector multiplied by an arbitrary number k or in multiplying two of the three vectors by λ and $1/\lambda$ respectively.

Two triples of vectors are said to be *equivalent* if either they both consist of coplanar vectors or one of them is obtained from the other by a chain of elementary transformations.

Classes of equivalent triples are called *trivectors*.

A trivector given by a triple $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is designated by the symbol $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$.

A trivector consisting of triples of coplanar vectors is called the *zero* trivector and designated by the symbol 0. Thus $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c} = 0$ *if and only if the vectors* $\mathbf{a}$, $\mathbf{b}$ *and* $\mathbf{c}$ *are coplanar.*

Proposition 2 now assumes the following form (completely similar to Proposition 1 of Lecture 7).

Proposition 3. *If* $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c} \neq 0$, *then the equality*

$$\mathbf{a}_1 \wedge \mathbf{b}_1 \wedge \mathbf{c}_1 = \mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$$

holds if and only if

$$\mathbf{a}_1 = k_1\mathbf{a} + l_1\mathbf{b} + m_1\mathbf{c},$$

$$\mathbf{b}_1 = k_2\mathbf{a} + l_2\mathbf{b} + m_2\mathbf{c},$$

$$\mathbf{c}_1 = k_3\mathbf{a} + l_3\mathbf{b} + m_3\mathbf{c},$$

with

$$\begin{vmatrix} k_1 & k_2 & k_3 \\ l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{vmatrix} = 1. \square$$

It should be expected by analogy with bivectors (and vectors) that in an intuitive-geometrical situation trivectors are given by oriented solids of given volume "floating freely" in space. We now show that this is really the case.

Since we are only interested in volume, instead of considering solids of arbitrary shapes we may concentrate on parallelepipeds, i.e. on triples of (noncoplanar) vectors. We must assume that two of such triples determine the same "geometrical trivector" if

(a) the parallelepipeds constructed on them have the same volume;

(b) their orientations coincide.

But it is clear that elementary transformations leave unchanged the volume of the parallelepiped constructed on the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ (this is obvious for transformations of type 2, and for transformations of type 1 it follows from the fact that they leave unchanged the area of the base of the parallelepiped and the length of its altitude) and do not change the orientation of the triple $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ (for the determinant of the matrix of transition from the triple $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ to an equivalent triple is equal to unity and hence positive). Therefore triples that are equivalent "algebraically" are also equivalent "geometrically".

Conversely, let two noncoplanar triples of vectors satisfy conditions (a) and (b). Since we work in three-dimensional "geometric" space, both triples are bases and so we can apply to them Lemma 3. According to this lemma the given triples are equivalent ("algebraically" and hence, by what has already been proved, also "geometrically") to two triples of the form $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ and $(\delta\mathbf{a}, \mathbf{b}, \mathbf{c})$ which thus also satisfy conditions (a) and (b). But for such triples it is obviously possible only when $\delta = 1$. Consequently, the original triples are equivalent "algebraically" (can be converted into each other by elementary transformations). □

It follows immediately from Proposition 3 that for any trivector $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c} \neq 0$ and any number $k \neq 0$ we have

$$k\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c} = \mathbf{a} \wedge k\mathbf{b} \wedge \mathbf{c} = \mathbf{a} \wedge \mathbf{b} \wedge k\mathbf{c} \tag{5}$$

and that if $\mathbf{a}_1 \wedge \mathbf{b}_1 \wedge \mathbf{c}_1 = \mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$, then

$$k\mathbf{a}_1 \wedge \mathbf{b}_1 \wedge \mathbf{c}_1 = k\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}.$$

Besides, it is clear that both assertions hold also when $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c} = 0$ or $k = 0$.

We are justified in giving the following definition.

Definition 6. The *product* of a trivector $\mathfrak{A} = \mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$ by a number k is the trivector (5). This trivector is designated by the symbol $k\mathfrak{A}$.

Thus by the definition

$$k(\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}) = k\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c} = \mathbf{a} \wedge k\mathbf{b} \wedge \mathbf{c} = \mathbf{a} \wedge \mathbf{b} \wedge k\mathbf{c}.$$

In the notation $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$ both symbols $\wedge$ are nonseparable and constitute together a single sign of the operation (which by analogy with bivectors may also be called *external multiplication*) associating with the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ a trivector $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$.

This symbol can be "split", however, if the operation of external multiplication of a bivector by a vector is introduced.

Definition 7. *The external product* $\alpha \wedge \mathbf{c}$ of a bivector $\alpha = \mathbf{a} \wedge \mathbf{b}$ by a vector $\mathbf{c}$ is a trivector $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$. Thus by definition

$$(\mathbf{a} \wedge \mathbf{b}) \wedge \mathbf{c} = \mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}.$$

Similarly defined is the external product of a vector by a bivector

$$\mathbf{a} \wedge (\mathbf{b} \wedge \mathbf{c}) = \mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}.$$

Hence we see that by virtue of these definitions the external multiplication of three vectors satisfies the associative law

$$(\mathbf{a} \wedge \mathbf{b}) \wedge \mathbf{c} = \mathbf{a} \wedge (\mathbf{b} \wedge \mathbf{c}).$$

Of course, Definition 7 also needs to be checked for correctness.

Let $\mathbf{a} \wedge \mathbf{b} = \mathbf{a}_1 \wedge \mathbf{b}_1$, i.e. (see Proposition 1 of Lecture 7; it is obviously enough to consider only the case where $\mathbf{a} \wedge \mathbf{b} \neq 0$) let

$$\mathbf{a}_1 = k\mathbf{a} + l\mathbf{b},$$
$$\mathbf{b}_1 = k_1\mathbf{a} + l_1\mathbf{b},$$

where

$$\begin{vmatrix} k & k_1 \\ l & l_1 \end{vmatrix} = 1.$$

Then

$$\begin{vmatrix} k & k_1 & 0 \\ l & l_1 & 0 \\ 0 & 0 & l \end{vmatrix} = 1,$$

and so

$$\mathbf{a}_1 \wedge \mathbf{b}_1 \wedge \mathbf{c} = \mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}.$$

It can be proved in a similar way that $\mathbf{a} \wedge \mathbf{b}_1 \wedge \mathbf{c}_1 = \mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$ if $\mathbf{b}_1 \wedge \mathbf{c}_1 = \mathbf{b} \wedge \mathbf{c}$. □

It is now easy to see that interchanging two vectors in the trivector $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$ results in changing the sign of the trivector (the anticommutative property). For example,

$$\mathbf{b} \wedge \mathbf{a} \wedge \mathbf{c} = (\mathbf{b} \wedge \mathbf{a}) \wedge \mathbf{c} = (-\mathbf{a} \wedge \mathbf{b}) \wedge \mathbf{c} = -\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}.$$

Lecture 13

Trivectors in three-dimensional vector space • Addition of trivectors • The formula for the volume of a parallelepiped • Scalar product • Axioms of scalar multiplication • Euclidean spaces • The length of a vector and the angle between vectors • The Cauchy-Buniakowski inequality • The triangle inequality • Theorem on the diagonals of a parallelogram • Orthogonal vectors and the Pythagorean theorem

---

Thus far the dimension n of a vector space $\mathscr{V}$ has played no role in our theory of trivectors. To construct the operation of trivector addition, however, it is necessary to suppose that $n \leqslant 4$ (just as for bivector addition it was necessary to assume that $n \leqslant 3$). In accordance with our general purpose we shall not consider the case $n = 4$, restricting ourselves to $n \leqslant 3$.

Since when $n \leqslant 2$ and triple of vectors is linearly dependent, there are no trivectors, besides the zero one, when $n \leqslant 2$. (For a similar reason there are no non-zero m-vectors, with $m > 3$, when $n \leqslant 3$). Of interest is therefore only the "three-dimensional" case $n = 3$.

If two bases $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{a}_1$, $\mathbf{b}_1$, $\mathbf{c}_1$ for a three-dimensional vector space $\mathscr{V}$ are connected by the relations

$$\mathbf{a}_1 = \delta\mathbf{a}, \quad \mathbf{b}_1 = \mathbf{b}, \quad \mathbf{c}_1 = \mathbf{c}, \tag{1}$$

then by definition (see Definition 6 of the preceding lecture)

$$\mathbf{a}_1 \wedge \mathbf{b}_1 \wedge \mathbf{c}_1 = \delta\,(\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}). \tag{2}$$

On the other hand, according to Lemma 3 of the preceding lecture any two bases for a vector space $\mathscr{V}$ are equivalent to bases connected by relations (1), with δ equalling the determinant of the corresponding transition matrix. Therefore formula (2) holds for any two bases (noncoplanar triples of vectors) $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{a}_1$, $\mathbf{b}_1$, $\mathbf{c}_1$ for a vector space $\mathscr{V}$.

Moreover, formula (2) obviously holds for coplanar vectors $\mathbf{a}_1$, $\mathbf{b}_1$, $\mathbf{c}_1$ as well.

Thus we see (changing the symbols) that *for any basis* $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ *of a vector space* $\mathscr{V}$ *and any vectors* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *we have*

$$\mathbf{a}\wedge\mathbf{b}\wedge\mathbf{c}=\begin{vmatrix} a^1 & a^2 & a^3\\ b^1 & b^2 & b^3\\ c^1 & c^2 & c^3\end{vmatrix}(\mathbf{e}_1\wedge\mathbf{e}_2\wedge\mathbf{e}_3), \tag{3}$$

where a^1, a^2, a^3, b^1, b^2, b^3, c^1, c^2, c^3 are the coordinates of the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ in the basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ (cf. formula (6) of Lecture 8).

This in particular means that by an arbitrary choice of some trivector $\mathfrak{A}_0\neq 0$ we can represent any other trivector $\mathfrak{A}$ as

$$\mathfrak{A}=k\mathfrak{A}_0$$

where k is some (obviously, uniquely determined) number.

Definition 1. The *sum* of two trivectors $\mathfrak{A}=k\mathfrak{A}_0$ and $\mathfrak{B}=l\mathfrak{A}_0$ is the trivector

$$\mathfrak{A}+\mathfrak{B}=(k+l)\,\mathfrak{A}_0.$$

If $\mathfrak{A}_0'=\lambda\mathfrak{A}_0$, where $\lambda\neq 0$, then $\mathfrak{A}=\frac{k}{\lambda}\mathfrak{A}_0'$ and $\mathfrak{B}=\frac{l}{\lambda}=\mathfrak{A}_0'$. Since the trivector $\left(\frac{k}{\lambda}+\frac{l}{\lambda}\right)\mathfrak{A}_0'$ is obviously equal to the trivector $(k+l)\,\mathfrak{A}_0$, this proves that Definition 1 is correct.

Obvious calculations now show that the set $\mathscr{V}\wedge\mathscr{V}\wedge\mathscr{V}$ of all trivectors is a vector space of dimension 1 with respect to the introduced operations of addition and multiplication by a number.

From a descriptive-geometric viewpoint it is clear that if the volume of a trivector $\mathfrak{A}_0$ is equal to one, then that of a trivector $k\mathfrak{A}_0$, with $k\neq 0$, is equal to $|k|$. According to formula (3), therefore, if the volume of the parallelepiped constructed on the vectors of a basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ is equal to one, then that of an arbitrary parallelepiped $OABC$ is equal to the absolute value of the determinant

$$\begin{vmatrix} a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \\ c^1 & c^2 & c^3 \end{vmatrix},$$

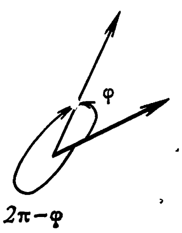
whose rows are the coordinates of the vectors $\overrightarrow{OA}$, $\overrightarrow{OB}$ and $\overrightarrow{OC}$ in the basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$.

In a formal-axiomatic theory some trivector $\mathfrak{A}_0 \neq 0$ should be chosen and the absolute value $|k|$ of the number k satisfying the relation

$$\overrightarrow{OA} \wedge \overrightarrow{OB} \wedge \overrightarrow{OC} = k\mathfrak{A}_0$$

declared to be the volume of the arbitrary parallelepiped $OABC$. The number k itself is here the so-called *oriented volume* of the parallelepiped.

At this point we shall temporarily discontinue the study of affine geometry to turn to the question of what axioms must be added to the affine axioms 1° to 11° to obtain Euclidean geometry, familiar from school, with its mensurations of lengths and angles.



An angle between two vectors

As ever, we first adopt an intuitive point of view.

Let **a** and **b** be two ("geometrical") vectors. Supposing them to be nonzero, consider the angle between them. This angle has two values, φ and $2\pi - \varphi$, whose cosines are equal. Therefore the formula

$$\mathbf{ab} = ab \cos \varphi, \tag{4}$$

where a and b are the lengths of **a** and **b**, uniquely determines some number **ab**.

Definition 2. A number **ab** defined by formula (4) is called a *scalar product* of vectors **a** and **b**. If $\mathbf{a} = 0$ or $\mathbf{b} = 0$, then by definition $\mathbf{ab} = 0$. The operation $\mathbf{a}, \mathbf{b} \mapsto \mathbf{ab}$ is called *scalar multiplication*.

It is clear that scalar multiplication is **commutative**:

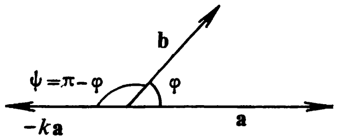
$$\mathbf{ab} = \mathbf{ba}$$

for any vectors $\mathbf{a}$ and $\mathbf{b}$.

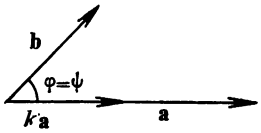
Besides, it is easy to see that it is **homogeneous**, i.e.

$$k(\mathbf{ab}) = (k\mathbf{a})\mathbf{b} = \mathbf{a}(k\mathbf{b}) \tag{5}$$

for any $\mathbf{a}$, $\mathbf{b}$ and any k. Indeed, for $\mathbf{a} = \mathbf{0}$ or $\mathbf{b} = \mathbf{0}$, as well as for $k = 0$, formula (5) is obvious. But if $\mathbf{a} \neq \mathbf{0}$, $\mathbf{b} \neq \mathbf{0}$ and $k \neq 0$, then the length of the vector $k\mathbf{a}$ is equal to



Case $k<0$



Case $k>0$

$|k|a$, the angle ψ between the vectors $k\mathbf{a}$ and $\mathbf{b}$ equalling the angle φ between $\mathbf{a}$ and $\mathbf{b}$ when $k > 0$ and $\pi - \varphi$ when $k < 0$. Therefore $k\cos\varphi = |k|\cos\psi$ and hence

$$(k\mathbf{a})\mathbf{b} = (|k|a)b\cos\psi = ab(|k|\cos\psi) =$$
$$= abk\cos\varphi = k(ab\cos\varphi) = k(\mathbf{ab}).$$

The formula $\mathbf{a}(k\mathbf{b}) = k(\mathbf{ab})$ is proved in a similar way (it is also possible to take advantage of commutativity). □

The significance of a scalar product resides in the fact that both lengths and angles are expressible in terms of it. Indeed, since $\varphi = 0$ when $\mathbf{a} = \mathbf{b}$, for the length a of any vector $\mathbf{a}$ we have the formula

$$a = \sqrt{\mathbf{a}^2}$$

where $\mathbf{a}^2 = \mathbf{aa}$ ("the scalar square of a vector is equal to the square of its length") and, hence, for the angle φ between nonzero vectors we have the formula

$$\cos\varphi = \frac{\mathbf{ab}}{\sqrt{\mathbf{a}^2}\sqrt{\mathbf{b}^2}}.$$

In particular we see that
$\mathbf{a}^2 \geqslant 0$ *for any vector* $\mathbf{a}$, *with* $\mathbf{a}^2 = 0$ *if and only if* $\mathbf{a} = \mathbf{0}$.
This property of scalar multiplication is called its **positivity**.

The well-known trigonometric cosine theorem

$$c^2 = a^2 + b^2 - 2ab\cos\varphi$$

can be written in the following vector form:

$$(\mathbf{a} - \mathbf{b})^2 = \mathbf{a}^2 + \mathbf{b}^2 - 2\mathbf{ab}$$

holding obviously for any vectors **a** and **b**.

Replacing here **b** by $-\mathbf{b}$ and considering that $(-\mathbf{a})\,\mathbf{b} = -\mathbf{ab}$ we have

$$(\mathbf{a} + \mathbf{b})^2 = \mathbf{a}^2 + \mathbf{b}^2 + 2\mathbf{ab},$$

from which it follows that for any three vectors **a**, **b** and **c** we have

$$(\mathbf{a} + \mathbf{b} + \mathbf{c})^2 = (\mathbf{a} + (\mathbf{b} + \mathbf{c}))^2 = \mathbf{a}^2 + (\mathbf{b} + \mathbf{c})^2 + 2\mathbf{a}\,(\mathbf{b} + \mathbf{c}) = \mathbf{a}^2 + \mathbf{b}^2 + \mathbf{c}^2 + 2\mathbf{bc} + 2\mathbf{a}\,(\mathbf{b} + \mathbf{c}),$$

$$(\mathbf{a} + \mathbf{b} + \mathbf{c})^2 = ((\mathbf{a} + \mathbf{b}) + \mathbf{c})^2 = (\mathbf{a} + \mathbf{b})^2 + \mathbf{c}^2 + 2\,(\mathbf{a} + \mathbf{b})\,\mathbf{c} = \mathbf{a}^2 + \mathbf{b}^2 + \mathbf{c}^2 + 2\mathbf{ab} + 2\,(\mathbf{a} + \mathbf{b})\,\mathbf{c},$$

and hence

$$\mathbf{a}\,(\mathbf{b} + \mathbf{c}) - (\mathbf{a} + \mathbf{b})\,\mathbf{c} = \mathbf{ab} - \mathbf{bc}.$$

Replacing here **c** by $-\mathbf{c}$ and adding we get

$$\mathbf{a}\,(\mathbf{b} + \mathbf{c}) + \mathbf{a}\,(\mathbf{b} - \mathbf{c}) = 2\mathbf{ab}.$$

Finally, applying the last formula to the vectors **a**, $\frac{\mathbf{b}+\mathbf{c}}{2}$ (instead of **b**) and $\frac{\mathbf{b}-\mathbf{c}}{2}$ (instead of **c**) and considering that

$$\frac{\mathbf{b}+\mathbf{c}}{2} + \frac{\mathbf{b}-\mathbf{c}}{2} = \mathbf{b}, \quad \frac{\mathbf{b}+\mathbf{c}}{2} - \frac{\mathbf{b}-\mathbf{c}}{2} = \mathbf{c}$$

and

$$2\mathbf{a}\left(\frac{\mathbf{b}+\mathbf{c}}{2}\right) = \mathbf{a}\,(\mathbf{b}+\mathbf{c}),$$

we ultimately find that

$$\mathbf{a}(\mathbf{b}+\mathbf{c}) = \mathbf{ab} + \mathbf{ac},$$

i.e. that scalar multiplication is distributive with respect to addition (and that is why the operation $\mathbf{a}, \mathbf{b} \mapsto \mathbf{ab}$ is termed multiplication).

In axiomatic theory we must as ever "reverse" the results obtained and take them as axioms.

Let $\mathscr{V}$ be an arbitrary vector space (over the field $\mathbb{R}$ of real numbers).

Definition 3. *Scalar multiplication* in a space $\mathscr{V}$ is an arbitrary function $\mathbf{a}, \mathbf{b} \mapsto \mathbf{ab}$ of a pair of vectors which takes numerical values and has the following properties (we continue a single numbering of axioms):

12° (*distributivity*). For any vectors $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathscr{V}$

$$\mathbf{a}(\mathbf{b}+\mathbf{c}) = \mathbf{ab} + \mathbf{ac}.$$

13° (*homogeneity*). For any vectors $\mathbf{a}, \mathbf{b} \in \mathscr{V}$ and any number k

$$k(\mathbf{ab}) = (k\mathbf{a})\mathbf{b} = \mathbf{a}(k\mathbf{b}).$$

14° (*commutativity*). For any vectors $\mathbf{a}, \mathbf{b} \in \mathscr{V}$

$$\mathbf{ab} = \mathbf{ba}.$$

15° (*positivity*). For any nonzero vector $\mathbf{a} \in \mathscr{V}$

$$\mathbf{a}^2 > 0.$$

A number $\mathbf{ab}$ here is called a scalar product of vectors $\mathbf{a}$ and $\mathbf{b}$.

Remark. The symbol $\mathbf{ab}$ is inconvenient when the elements of a space $\mathscr{V}$ are, say, functions. Therefore, instead of $\mathbf{ab}$ one often writes $(\mathbf{a}, \mathbf{b})$. One should get accustomed to this notation too.

Definition 4. A vector space $\mathscr{V}$ for which some scalar multiplication $\mathbf{a}, \mathbf{b} \mapsto \mathbf{ab}$ has been chosen and fixed is called a *Euclidean vector space.* (Some writers call these spaces also *real unitary spaces* or even, without further ado, *spaces with scalar multiplication.*)

A *Euclidean* (*point*) *space* is an affine space $\mathscr{A}$ in the associated vector space of which a scalar multiplication is chosen (i.e. which is turned into a Euclidean vector space).

The branch of mathematics which studies Euclidean spaces is called *Euclidean geometry*. In comparison with affine geometry Euclidean geometry has one additional primary notion ("scalar multiplication") and four additional axioms 12° to 15°.

In a Euclidean vector space the *length of a vector* **a** (usually denoted by $|\mathbf{a}|$ and sometimes by $\|\mathbf{a}\|$) is *defined* by the formula

$$|\mathbf{a}| = \sqrt{\mathbf{a}^2} \tag{6}$$

(or, in other notation, $|\mathbf{a}| = \sqrt{(\mathbf{a}, \mathbf{a})}$) and the *angle* φ between two nonzero vectors **a** and **b** by the formula

$$\cos\varphi = \frac{\mathbf{ab}}{|\mathbf{a}|\cdot|\mathbf{b}|} \tag{7}$$

which gives for φ two values (summing to 2π).

In a Euclidean point space a *distance* $|M_0M_1|$ between two points is *defined* by the formula

$$|M_0M_1| = |\overrightarrow{M_0M_1}| = \sqrt{\overrightarrow{M_0M}^2}$$

and the angle φ between two straight lines by formula (7) in which **a** and **b** are arbitrary direction vectors of the given straight lines. (This formula gives four values for φ, depending on the choice of vectors **a** and **b**, and two if we impose on φ the constraint $0 \leqslant \varphi \leqslant \pi$).

Axiom 15° on positivity ensures that for any vector **a** formula (6) uniquely defines its length $|\mathbf{a}|$, which is a real nonnegative number, with $|\mathbf{a}| > 0$ if $\mathbf{a} \neq \mathbf{0}$ and $|\mathbf{a}| = 0$ if $\mathbf{a} = \mathbf{0}$ (the latter follows from axiom 13° on homogeneity for $k = 0$).

The situation is worse with formula (7). Here it must be specially proved that its right-hand side is in the domain of definition of arc cosine, i.e. that

$$|\mathbf{ab}| \leqslant |\mathbf{a}|\cdot|\mathbf{b}|,$$

or, in other words, that

$$(\mathbf{ab})^2 \leqslant \mathbf{a}^2\mathbf{b}^2. \tag{8}$$

This is the **Cauchy-Buniakowski inequality**. To prove it we shall consider the function

$$f(t) = (\mathbf{a} + t\mathbf{b})^2$$

of a number variable t. According to axiom 15°, for every t we have $f(t) \geqslant 0$. On the other hand,

$$f(t) = \mathbf{a}^2 + 2t(\mathbf{ab}) + t^2\mathbf{b}^2,$$

and it is well known from elementary algebra that if a second-degree trinomial takes only nonnegative values, its discriminant (the radicand in the formula for the roots of a quadratic equation) is nonpositive. In our case this discriminant is equal to $(\mathbf{ab})^2 - \mathbf{a}^2\mathbf{b}^2$, which proves inequality (8). □

If vectors $\mathbf{a}$ and $\mathbf{b}$ are collinear, we have equality in (8) Conversely, if $(\mathbf{ab})^2 = \mathbf{a}^2\mathbf{b}^2$, then for $\mathbf{b} \neq \mathbf{0}$, according to the formula for the roots of a quadratic equation, $f(t_0) = 0$, where $t_0 = -\frac{\mathbf{ab}}{\mathbf{b}^2}$. Therefore $\mathbf{a} + t_0\mathbf{b} = \mathbf{0}$. Thus in the *Cauchy-Buniakowski inequality equality holds only for collinear vectors* $\mathbf{a}$ *and* $\mathbf{b}$. □

When $\mathbf{a} \neq \mathbf{0}$ and $\mathbf{b} \neq \mathbf{0}$ this means that the *angle* φ *between noncollinear vectors is different from zero and from* π, whereas between collinear vectors it is equal either to zero (if the vectors are positively proportional) or to π (otherwise).

It follows from the Cauchy-Buniakowski inequality that

$$|\mathbf{a} + \mathbf{b}|^2 = (\mathbf{a} + \mathbf{b})^2 = \mathbf{a}^2 + 2\mathbf{ab} + \mathbf{b}^2 \leqslant$$
$$\leqslant |\mathbf{a}|^2 + 2|\mathbf{a}| \cdot |\mathbf{b}| + |\mathbf{b}|^2 = (|\mathbf{a}| + |\mathbf{b}|)^2,$$

and, similarly, that

$$|\mathbf{a} + \mathbf{b}|^2 = (\mathbf{a} + \mathbf{b})^2 = \mathbf{a}^2 + 2\mathbf{ab} + \mathbf{b}^2 \geqslant$$
$$\geqslant |\mathbf{a}|^2 - 2|\mathbf{a}| \cdot |\mathbf{b}| + |\mathbf{b}|^2 = (|\mathbf{a}| - |\mathbf{b}|)^2.$$

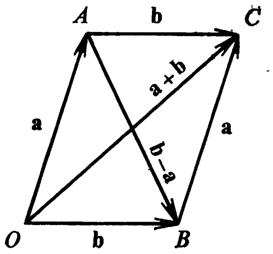
Hence

(9) $$\big||\mathbf{a}| - |\mathbf{b}|\big| \leqslant |\mathbf{a} + \mathbf{b}| \leqslant |\mathbf{a}| + |\mathbf{b}|.$$

If ABC is an arbitrary triangle and $\mathbf{a} = \overrightarrow{AB}$, $\mathbf{b} = \overrightarrow{BC}$, then $\mathbf{a} + \mathbf{b} = \overrightarrow{AC}$ and we obtain from (9) the familiar inequalities

$$||AB| - |BC|| \leqslant |AC| \leqslant |AB| + |BC|$$

("any side of a triangle is not greater than the sum and not less than the difference of the other two sides"). For this reason inequalities (9) are generally called the **triangle inequalities.**



Scalar multiplication is a powerful tool for proving geometrical theorems. We have already seen that the cosine theorem is simply a formula for the square of a sum. Here is another example.

Let $OACB$ be a parallelogram. Setting $\mathbf{a} = \overrightarrow{OA}$ and $\mathbf{b} = \overrightarrow{OB}$ yields $\overrightarrow{OC} = \mathbf{a} + \mathbf{b}$ and $\overrightarrow{AB} = \mathbf{b} - \mathbf{a}$. Therefore

$$\begin{aligned} |OC|^2 + |AB|^2 &= (\mathbf{a}+\mathbf{b})^2 + (\mathbf{b}-\mathbf{a})^2 = \\ &= 2(\mathbf{a}^2 + \mathbf{b}^2) = 2(|OA|^2 + |OB|^2) = \\ &= |OA|^2 + |AC|^2 + |BC|^2 + |OB|^2, \end{aligned}$$

i.e. *the sum of the squares of the lengths of the diagonals of a parallelogram is equal to the sum of the squares of the lengths of its four sides.*

Definition 5. Vectors $\mathbf{a}$ and $\mathbf{b}$ are said to be *orthogonal* if their scalar product $\mathbf{ab}$ is zero

$$\mathbf{ab} = 0.$$

For $\mathbf{a} \neq \mathbf{0}$ and $\mathbf{b} \neq \mathbf{0}$ this is equivalent to saying that the angle between $\mathbf{a}$ and $\mathbf{b}$ is equal to $\pi/2$.

If vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal, then

$$(10) \qquad (\mathbf{a} + \mathbf{b})^2 = \mathbf{a}^2 + \mathbf{b}^2.$$

For $\mathbf{a} = \overrightarrow{OA}$, $\mathbf{b} = \overrightarrow{OB}$, when $\mathbf{b} - \mathbf{a} = \overrightarrow{AB}$, we obtain from this the theorem of Pythagoras

$$|AB|^2 = |OA|^2 + |OB|^2$$

("the square of the hypotenuse is equal to the sum of the squares of the legs"). For this reason equation (10) is also called the *Pythagorean theorem*.

Lecture 14

Metric form and metric coefficients • The condition of positive definiteness • Formulas for the transformation of metric coefficients when changing a basis • Orthonormal families of vectors and Fourier coefficients • Orthonormal bases and rectangular coordinates • Decomposition of positive definite matrices • The Gram-Schmidt orthogonalization process • Isomorphism of Euclidean spaces • Orthogonal matrices • Second-order orthogonal matrices • Formulas for the transformation of rectangular coordinates

We shall find the expression for a scalar product in coordinates.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a basis for a Euclidean vector space $\mathscr{V}$ and let

$$\mathbf{x} = x^1\mathbf{e}_1 + \ldots + x^n\mathbf{e}_n = x^i\mathbf{e}_i,$$
$$\mathbf{y} = y^1\mathbf{e}_1 + \ldots + y^n\mathbf{e}_n = y^j\mathbf{e}_j$$

be two arbitrary vectors of that space.

In the Einstein notation

$$\mathbf{xy} = (x^i\mathbf{e}_i)(y^j\mathbf{e}_j) = (\mathbf{e}_i\mathbf{e}_j)\, x^i y^j,$$

i.e.

(1) $$\mathbf{xy} = g_{ij}x^i y^j,$$

where

(2) $$g_{ij} = \mathbf{e}_i\mathbf{e}_j \quad i, j = 1, \ldots, n.$$

Note that

(3) $$g_{ij} = g_{ji} \quad \text{for any } i, j = 1, \ldots, n,$$

For $n = 1$ formula (1) has the form

$$\mathbf{xy} = g_{11}x^1y^1,$$

where $g_{11} = \mathbf{e}_1\mathbf{e}_1$;
for $n = 2$ it has the form

$$\mathbf{xy} = g_{11}x^1y^1 + g_{12}x^1y^2 + g_{21}x^2y^1 + g_{22}x^2y^2 = \\ = g_{11}x^1y^1 + g_{12}(x^1y^2 + x^2y^1) + g_{22}x^2y^2$$

where

$$g_{11} = \mathbf{e}_1\mathbf{e}_1,$$
$$g_{12} = g_{21} = \mathbf{e}_1\mathbf{e}_2,$$
$$g_{22} = \mathbf{e}_2\mathbf{e}_2;$$

for $n = 3$ it has the form

$$\mathbf{xy} = g_{11}x^1y^1 + g_{12}x^1y^2 + g_{13}x^1y^3 + g_{21}x^2y^1 + g_{22}x^2y^2 + \\ + g_{23}x^2y^3 + g_{31}x^3y^1 + g_{32}x^3y^2 + g_{33}x^3y^3 = \\ = g_{11}x^1y^1 + g_{12}(x^1y^2 + x^2y^1) + g_{13}(x^1y^3 + x^3y^1) + \\ + g_{22}x^2y^2 + g_{23}(x^2y^3 + x^3y^2) + g_{33}x^3y^3,$$

where

$$g_{11} = \mathbf{e}_1\mathbf{e}_1, \quad g_{12} = g_{21} = \mathbf{e}_1\mathbf{e}_2, \quad g_{13} = g_{31} = \mathbf{e}_1\mathbf{e}_3,$$
$$g_{22} = \mathbf{e}_2\mathbf{e}_2, \quad g_{23} = g_{32} = \mathbf{e}_2\mathbf{e}_3, \quad g_{23} = \mathbf{e}_3\mathbf{e}_3.$$

In the general case we can also reduce the "similar" terms:

$$\mathbf{xy} = \sum_i g_{ii}x^iy^i + \sum_{i<j} g_{ij}(x^iy^j + x^jy^i).$$

In the first sum summation is performed from 1 to n over i and in the second from 1 to n over i and j provided that $i < j$.

Formula (1) can be written in matrix form. Let, as in Lecture 11,

$$x = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}, \quad y = \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix}$$

and let G be a square $n \times n$ matrix

$$G = \begin{pmatrix} g_{11} & \dots & g_{1n} \\ g_{21} & \dots & g_{2n} \\ \cdot & \cdot \cdot \cdot & \cdot \\ g_{n1} & \dots & g_{nn} \end{pmatrix}. \tag{4}$$

Consider a matrix

$$x^\top G y,$$

where $y^\top = (y^1, \dots, y^n)$ (the index $^\top$ is the transposition sign). By the rule of rectangular matrix multiplication this matrix is of size 1×1, i.e. represents a number. Computing the latter we immediately find that it is equal to

$$\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} x^i y^j, \tag{5}$$

i.e. to (1). This proves that

$$\mathbf{xy} = x^\top G y. \tag{6}$$

For $\mathbf{x} = \mathbf{y}$ we get

$$\mathbf{x}^2 = g_{ij} x^i x^j = \sum_i g_{ii} (x^i)^2 + 2 \sum_{i<j} g_{ij} x^i x^j = x^\top G x \tag{7}$$

and in particular we get

$$\mathbf{x}^2 = g_{11} (x^1)^2$$

for $n = 1$,

$$\mathbf{x}^2 = g_{11} (x^1)^2 + 2 g_{12} x^1 x^2 + g_{22} (x^2)^2$$

for $n = 2$ and

$$\mathbf{x}^2 = g_{11} (x^1)^2 + 2 g_{12} x^1 x^2 + 2 g_{13} x^1 x^3 + \\ + g_{22} (x^2)^2 + 2 g_{23} x^2 x^3 + g_{33} (x^3)^2.$$

for $n = 3$.

The algebraic expression on the right of formula (1) (written in a more explicit form as (5)) is a polynomial homogeneous in two sets of variables $x^1, \dots, x^n$ and $y^1, \dots, y^n$ and linear in each of them. Homogeneous polynomials are generally called *forms* and polynomials of the kind (5) *bilinear forms*. Forms (5) whose coefficients have the property (3) are called *symmetrical forms*.

When $y^1 = x^1, \ldots, y^n = x^n$ a symmetrical bilinear form turns (see formulas (7)) into a homogeneous second-degree polynomial in variables $x^1, \ldots, x^n$, i.e. into what is said to be a *quadratic form*. It is clear that this form uniquely determines the corresponding symmetrical bilinear form (i.e. the coefficients g_{ij}). According to axiom 15° form (7) has the property

$$g_{ij}x^ix^j > 0$$

if $(x^1, \ldots, x^n) \neq (0, \ldots, 0)$. Such quadratic forms (and the corresponding bilinear symmetrical forms) are called *positively definite*.

Thus in this terminology formula (1) (and the equivalent formula (6)) states that a *scalar product of two vectors is a bilinear symmetrical and positively definite form of their coordinates*.

The converse turns out to be true too.

Proposition 1. *Let $\mathscr{V}$ be an arbitrary vector space, let* $\mathbf{e}_1, \ldots, \mathbf{e}_n$ *be some basis for it and let $g_{ij}x^iy^j$ be an arbitrary bilinear symmetrical and positively definite form in variables $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$. Then formula* (1) *defines in $\mathscr{V}$ some scalar multiplication.*

Proof. Distributivity and homogeneity (axioms 12° and 13°) follow from bilinearity

$$g_{ij}(x_1^i + x_2^i)\,y^j = g_{ij}x_1^iy^j + g_{ij}x_2^iy^j,$$
$$g_{ij}(kx^i)\,y^j = g_{ij}x^i(ky^j) = kg_{ij}x^iy^j.$$

Commutativity (axiom 14°) ensues from symmetry

$$g_{ij}y^ix^j = g_{ji}y^ix^j = g_{ji}x^jy^i = g_{ij}x^iy^j.$$

This last transformation uses the fact that *the index ove- which summation is performed may be denoted by any prer viously unused letter*; explicitly this transformation is done as follows:

$$\begin{aligned} g_{ji}x^jy^i &= g_{\alpha i}x^\alpha y^i \quad \text{(we replace } j \text{ by a new letter } \alpha\text{)},\\ &= g_{\alpha j}x^\alpha y^j \quad \text{(we replace } i \text{ by the now free letter } j\text{)},\\ &= g_{ij}x^iy^j \quad \text{(we replace } \alpha \text{ by the now free letter } i\text{)}. \end{aligned}$$

Finally, positivity (axiom 15°) is ensured, by definition, by the positive definiteness of the form $g_{ij}x^iy^j$. $\square$

Thus we see that there are *many* different scalar products in the same vector space, i.e. that it can be transformed into a Euclidean space in many different ways.

Definition 1. The form $g_{ij}x^iy^j$ (and the corresponding quadratic form $g_{ij}x^ix^j$) is called a *metric form* of a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and its coefficients g_{ij} are called the *metric coefficients* of that basis.

The condition of positive definiteness imposes certain constraints on the coefficients g_{ij}. For example, it is clear that

$$g_{ii} > 0 \text{ for any } i = 1, \ldots, n.$$

For $n = 1$, when there is one coefficient g_{11}, the condition $g_{11} > 0$ is not only necessary but, obviously, sufficient for the form to be positively definite.

Let $n = 2$ and let $g_{11} > 0$. Then

$$g_{11}(x^1)^2 + 2g_{12}x^1x^2 + g_{22}(x^2)^2 =$$
$$= \left(\sqrt{g_{11}}\,x^1 + \frac{g_{12}}{\sqrt{g_{11}}}\,x^2\right)^2 + \frac{g_{22}g_{11} - g_{12}^2}{g_{11}}(x^2)^2,$$

from which it follows that the form considered is positively definite if and only if $g_{22}g_{11} - g_{12}^2 > 0$. This proves that the *symmetrical matrix*

$$G = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}, \quad g_{12} = g_{21}$$

is a matrix of metric coefficients of some basis for a two-dimensional Euclidean vector space if and only if

$$g_{11} > 0, \quad \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} > 0.$$

For $n = 3$ similar conditions are of the form

$$g_{11} > 0, \quad \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} > 0.$$

We shall prove this in the next semester's lectures.

By construction, a metric form depends on the choice of a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Let us find how it changes when this basis is replaced by another. We shall make our calculation in matrix notation.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ be two bases;

$$x = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}, \quad x' = \begin{pmatrix} x^{1'} \\ \vdots \\ x^{n'} \end{pmatrix}$$

be the columns of the coordinates of the same vector $\mathbf{x}$ in these bases, let $C = (c^i_{i'})$ be a matrix of transition from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to the basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ and let $G = (g_{ij})$ be a matrix of metric coefficients of the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Then

$$x = Cx' \text{ and } \mathbf{x}^2 = x^{\mathsf{T}}Gx.$$

Therefore

$$\mathbf{x}^2 = (x'^{\mathsf{T}}C^{\mathsf{T}})\, G\, (Cx') = x'^{\mathsf{T}}\, (C^{\mathsf{T}}GC)\, x'.$$

This formula implies that the matrix $G' = (g_{i'j'})$ of the metric coefficients of the basis $\mathbf{e}_{1'}, \ldots, \mathbf{e}_{n'}$ is expressed by the formula

$$G' = C^{\mathsf{T}}GC. \tag{8}$$

In other words

$$g_{i'j'} = c^i_{i'}c^j_{j'}g_{ij}. \tag{9}$$

Definition 2. A family $\mathbf{e}_1, \ldots, \mathbf{e}_m$ of vectors of a Euclidean space is said to be *orthonormal* if

(a) the length of each vector is equal to one;

(b) any two different vectors are orthogonal, i.e. if

$$\mathbf{e}_i\mathbf{e}_j = \begin{cases} 1, \text{ when } i = j, \\ 0, \text{ when } i \neq j. \end{cases}$$

Definition 3. *Fourier coefficients* of a vector $\mathbf{x}$ with respect to a family $\mathbf{e}_1, \ldots, \mathbf{e}_m$ are the scalar products $x_1 = \mathbf{x}\mathbf{e}_1, \ldots, x_m = \mathbf{x}\mathbf{e}_m$.

All Fourier coefficients of the zero vector are equal to zero, but the inverse is, generally speaking, incorrect. If it follows from $x_1 = 0, \ldots, x_m = 0$ that $\mathbf{x} = \mathbf{0}$, then the family $\mathbf{e}_1, \ldots, \mathbf{e}_m$ is said to be *closed*.

Proposition 2 (the Bessel inequality). *If a family* $\mathbf{e}_1, \ldots, \mathbf{e}_m$ *is orthonormal, then for any vector* $\mathbf{x}$

$$x_1^2 + \ldots + x_m^2 \leqslant |\mathbf{x}|^2,$$

the vector $\mathbf{x}' = \mathbf{x} - x_1\mathbf{e}_1 - \ldots - x_m\mathbf{e}_m$ *being orthogonal to all vectors* $\mathbf{e}_1, \ldots, \mathbf{e}_m$ *(i.e. all its Fourier coefficients with respect to the family* $\mathbf{e}_1, \ldots, \mathbf{e}_m$ *are equal to zero).*

Proof. It is enough to note that

$$0 \leqslant |\mathbf{x}'|^2 = \mathbf{x}'^2 = \Big(\mathbf{x} - \sum_{i=1}^{m} x_i\mathbf{e}_i\Big)\Big(\mathbf{x} - \sum_{j=1}^{m} x_j\mathbf{e}_j\Big) =$$

$$= \mathbf{x}^2 - 2\sum_{i=1}^{m} x_i(\mathbf{x}\mathbf{e}_i) + \sum_{i=1}^{m}\sum_{j=1}^{m} x_ix_j(\mathbf{e}_i\mathbf{e}_j) =$$

$$= \mathbf{x}^2 - 2\sum_{i=1}^{m} x_i^2 + \sum_{i=1}^{m} x_i^2 = \mathbf{x}^2 - \sum_{i=1}^{m} x_i^2$$

and that

$$\mathbf{e}_i(\mathbf{x} - x_1\mathbf{e}_1 - \ldots - x_m\mathbf{e}_m) = \mathbf{e}_i\mathbf{x} - x_i\mathbf{e}_i\mathbf{e}_i = x_i - x_i = 0$$

for any $i = 1, \ldots, m$. □

Recall that a family of vectors (not necessarily orthonormal) is said to be complete if any vector of the space is linearly expressible in terms of it.

Definition 4. An orthonormal family of vectors is said to be *maximal* if adding to it any vector makes it stop being orthonormal.

Proposition 3. *The following properties of an orthonormal family of vectors* $\mathbf{e}_1, \ldots, \mathbf{e}_n$ *are equivalent.*

(i) *it is maximal;*
(ii) *it is closed;*
(iii) *it is complete;*
(iv) *for any vector* $\mathbf{x}$

$$\mathbf{x} = x_1\mathbf{e}_1 + \ldots + x_n\mathbf{e}_n,$$

where $x_1, \ldots, x_n$ *are the Fourier coefficients of the vector* $\mathbf{x}$;

(v) *for any vectors* $\mathbf{x}$ *and* $\mathbf{y}$

$$\mathbf{x}\mathbf{y} = x_1y_1 + \ldots + x_ny_n; \tag{10}$$

(vi) *for any vector* $\mathbf{x}$

$$\mathbf{x}^2 = x_1^2 + \ldots + x_n^2,$$

i.e. *there holds equality in the Bessel inequality.*

Proof.

(i) $\Rightarrow$ (ii). If for some vector $\mathbf{x} \neq \mathbf{0}$ we have the equations $x_1 = 0, \ldots, x_n = 0$, then by adding to the family $\mathbf{e}_1, \ldots, \mathbf{e}_n$ a vector $\frac{\mathbf{x}}{|\mathbf{x}|}$ we obtain an orthonormal family.

(ii) $\Rightarrow$ (iii). If a vector $\mathbf{x}$ is not linearly expressible in terms of vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$, the vector $\mathbf{x}'$ constructed in Proposition 1 is nonzero but all its Fourier coefficients are zero.

(iii) $\Rightarrow$ (iv). If $\mathbf{x} = k_1\mathbf{e}_1 + \ldots + k_n\mathbf{e}_n$, then $x_i = \mathbf{x}\mathbf{e}_i = k_i(\mathbf{e}_i\mathbf{e}_i) = k_i$.

(iv) $\Rightarrow$ (v). If $\mathbf{x} = \sum\limits_i x_i\mathbf{e}_i$ and $\mathbf{y} = \sum\limits_j y_j\mathbf{e}_j$, then

$$\mathbf{x}\mathbf{y} = \Big(\sum_i x_i\mathbf{e}_i\Big)\Big(\sum_j y_j\mathbf{e}_j\Big) = \sum_i\sum_j x_iy_j(\mathbf{e}_i\mathbf{e}_j) = \sum_i x_iy_i.$$

(v) $\Rightarrow$ (vi). It is sufficient to set $\mathbf{x} = \mathbf{y}$ in (10).

(vi) $\Rightarrow$ (i). If it were possible to add to the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$, while preserving orthonormality, another vector $\mathbf{x}$, then we should have

$$1 = \mathbf{x}^2 = x_1^2 + \ldots + x_n^2 = 0,$$

for being orthogonal to the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ the vector $\mathbf{x}$ would have only zero Fourier coefficients $x_1, \ldots, x_n$. $\square$

Equation (10) is called *Parseval's formula.*

Remark. The conclusion that $x_i = k_i$ (see the proof of the implication (iii) $\Rightarrow$ (iv)) uses only the orthonormality of the family $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and can be applied (without any additional assumptions about that family) to any vector of the form $\mathbf{x} = k_1\mathbf{e}_1 + \ldots + k_n\mathbf{e}_n$. But if $\mathbf{x} = \mathbf{0}$, then $x_i = 0$ and therefore $k_i = 0$. This shows that *any orthonormal family of vectors is linearly independent.*

Definition 5. The orthonormal family of vectors which forms a basis is called an *orthonormal basis.*

According to Proposition 3, for an orthonormal family of vectors to be a basis it is necessary and sufficient for it to be either maximal, or closed, or complete (or to consist of $n = \dim \mathscr{V}$ vectors). The coordinates of an arbitrary vector in an orthonormal basis are its Fourier coefficients (statement (iv) of Proposition 3).

Therefore in what follows we shall not as a rule use the term 'Fourier coefficient' in connection with orthonormal bases. At the same time for the coordinates in an orthonormal basis we shall preserve the notation $x_1, \ldots, x_n$ with lowered indices.

Vectors of an orthonormal basis will ordinarily be denoted by $\mathbf{i}_1, \ldots, \mathbf{i}_n$ (by $\mathbf{i}$, $\mathbf{j}$ for $n = 2$ and by $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ for $n = 3$).

Definition 6. A coordinate system $O\mathbf{e}_1 \ldots \mathbf{e}_n$ in a point Euclidean space, for which the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is orthonormal, is called a *system of rectangular* (or *Euclidean* or, sometimes, *Cartesian*, which is absolutely unfounded from the viewpoint of the history of mathematics) *coordinates*. The same term is used with respect to the coordinates of vectors of a Euclidean vector space in an orthonormal basis.

An orthonormal basis is characterized by the fact that the matrix G of its metric coefficients is a unit matrix E. Therefore all metric formulas are substantially simplified in rectangular coordinates:

$$\mathbf{xy} = x_1y_1 + \ldots + x_ny_n,$$

$$\mathbf{x}^2 = x_1^2 + \ldots + x_n^2,$$

$$|\mathbf{x}| = \sqrt{x_1^2 + \ldots + x_n^2},$$

$$|AB| = \sqrt{(y_1 - x_1)^2 + \ldots + (y_n - x_n)^2}$$

($x_1, \ldots, x_n$ are the coordinates of the point A and $y_1, \ldots, y_n$ the coordinates of the point B) and so on.

For this reason only rectangular coordinates are commonly used in the study of Euclidean spaces.

The fact of the existence (for $\dim \mathscr{V} > 0$) of orthonormal bases and rectangular coordinates follows directly from Proposition 3. Indeed, on taking an arbitrary nonzero vector $\mathbf{a}$ and multiplying it by a number $1/|\mathbf{a}|$, we obtain a vector $\mathbf{e} = \mathbf{a}/|\mathbf{a}|$ of unit length (incidentally, such vectors are sometimes called *unit vectors*), i.e. an orthonormal family consisting of a single vector. This proves that orthonormal families of vectors do exist. Since no such family can (by virtue of linear independence) contain more than $n = \dim \mathscr{V}$ vectors, among orthonormal families there are maximal ones, i.e. bases.

It is interesting that some nontrivial algebraic statements follow from the mere fact of the existence of orthonormal bases. For example, let G be an arbitrary symmetrical *positive definite matrix* (i.e. such that the quadratic form $x^{\top}Gx$ is positively definite). On changing arbitrarily a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in some vector space $\mathscr{V}$ (for example, in $\mathbb{R}^n$) we convert that vector space into a Euclidean one, setting $\mathbf{xy} = x^{\top}Gy$ for any vectors $\mathbf{x}, \mathbf{y} \in \mathscr{V}$ (see Proposition 1), i.e. taking the matrix G as the matrix of metric coefficients of the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. According to what has been said, there is an orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ in $\mathscr{V}$. Let C be a matrix of transition from $\mathbf{i}_1, \ldots, \mathbf{i}_n$ to $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Then, as we know, the matrix G will be equal to $C^{\top}EC = C^{\top}C$, where E is the matrix of metric coefficients of the orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$, i.e. a unit matrix. Conversely, let C be an arbitrary nonsingular matrix, let $\mathbf{i}_1, \ldots, \mathbf{i}_n$ be some orthonormal basis for an arbitrary Euclidean vector space $\mathscr{V}$ and let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a basis connected with the basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ by the transition matrix C. Then the matrix G of metric coefficients of that basis (known to be a symmetrical positive definite matrix) will be equal to $C^{\top}EC = C^{\top}C$. This proves that a *matrix G is a symmetrical positive definite matrix if and only if there is a nonsingular matrix C such that*

$$G = C^{\top}C. \quad \square \tag{11}$$

The practical significance of this result is narrowed, however, since it gives no recipe for finding the matrix C given a matrix G. To overcome this drawback, it is necessary to indicate an explicit way of constructing from any basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ some orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$. We shall describe one such method called the *Gram-Schmidt orthogonalization process.*

It consists in a gradual step-by-step transformation of a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ into an orthonormal one. Let $0 \leqslant k \leqslant n$. We say that the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ *is orthonormal up to a number k* if its subfamily $\mathbf{e}_1, \ldots, \mathbf{e}_k$ is orthonormal. In accordance with this definition any basis is orthonormal up to the number 0 and the basis orthonormal up to the number n is nothing else but an ordinary orthonormal basis. It is therefore sufficient to indicate a method for transforming an arbitrary basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ orthonormal up to a number $k < n$ into a basis orthonormal up to the number $k + 1$.

The method suggested by Gram and Schmidt consists in changing in a given basis only one vector $\mathbf{e}_{k+1}$ which is first replaced by the vector

$$\mathbf{e}'_{k+1} = \mathbf{e}_{k+1} - x_1\mathbf{e}_1 - \ldots - x_k\mathbf{e}_k$$

where $x_1, \ldots, x_k$ are the Fourier coefficients of the vector $\mathbf{x} = \mathbf{e}_{k+1}$ with respect to the orthonormal family $\mathbf{e}_1, \ldots$ $\ldots, \mathbf{e}_k$ (naturally, nothing happens when $k = 0$). It is clear that after such a replacement a basis remains a basis. In addition (see Proposition 2) the vector $\mathbf{e}'_{k+1}$ is orthogonal to all vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$. Therefore, to complete the construction it is only necessary to "normalize" that vector, i.e. to divide it by its length. As a result we obviously obtain a basis orthonormal up to the number $k + 1$. □

Note that the ultimately resulting orthonormal basis is connected with the original basis by a *triangular* transition matrix. Thus we have not only found a practical way of realizing the decomposition (11) but also proved that *for any symmetrical positive definite matrix G there is a decomposition* (11) *with a triangular matrix C.* □

As already noted above, the same vector space can be converted into many different Euclidean spaces. For example, it is always possible to turn any preassigned basis into an orthonormal one.

However, for any $n \geqslant 0$ there exists, in a certain exact sense, only one Euclidean (vector or point) space.

Definition 7. Let $\mathscr{V}$ and $\mathscr{V}'$ be Euclidean vector spaces. The mapping $\varphi\colon \mathscr{V} \to \mathscr{V}'$ is said to be an *isomorphism* if it is an isomorphism of vector spaces (see Definition 1 of Lecture 5) and, besides, leaves the scalar product unchanged, i.e.

$$\varphi(\mathbf{x})\,\varphi(\mathbf{y}) = \mathbf{xy}$$

for any two vectors $\mathbf{x}, \mathbf{y} \in \mathscr{V}$.

Euclidean vector spaces $\mathscr{V}$ and $\mathscr{V}'$ are said to be *isomorphic* if there exists at least one isomorphism $\mathscr{V} \to \mathscr{V}'$.

Theorem 1. *Any two Euclidean vector spaces of the same dimension n are isomorphic. Isomorphism is realized by equating coordinates in any two orthonormal bases.*

Proof. It is sufficient to observe that in every orthonormal basis scalar product is expressed by the same formula

$$\mathbf{xy} = x_1y_1 + \ldots + x_ny_n. \quad \square$$

The natural way to convert the space $\mathbb{R}^n$ into a Euclidean one is to declare the standard basis orthonormal, i.e. to set for any vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, by definition,

$$\mathbf{xy} = x_1y_1 + \ldots + x_ny_n.$$

Then for any orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ for an arbitrary Euclidean vector space $\mathscr{V}$ the corresponding coordinate isomorphism $\mathscr{V} \to \mathbb{R}^n$ is an isomorphism of Euclidean spaces.

Similar definitions and results hold, of course, for Euclidean point spaces as well.

In conclusion we shall consider the question of matrices connecting two orthonormal bases.

Definition 8. A matrix C of transition from one orthonormal basis to another is called an *orthogonal matrix.*

Since it is the unit matrix E that is the matrix of coefficients in both bases, the matrix C is orthogonal if and only if

$$C^\top C = E. \tag{12}$$

Setting

$$C = \begin{pmatrix} c_{11} & \ldots & c_{1n} \\ \cdot & \cdot\,\cdot\,\cdot & \cdot \\ c_{n1} & \ldots & c_{nn} \end{pmatrix}$$

we see that equation (12) is equivalent to $\frac{n(n+1)}{2}$ relations of the form

$$c_{1i}c_{1j} + c_{2i}c_{2j} + \ldots + c_{ni}c_{nj} = \begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{if} \quad i \neq j, \end{cases}$$

where $i, j = 1, \ldots, n$ and $i \leqslant j$, expressing the fact that the columns of the matrix C constitute an orthonormal family (basis) of vectors of $\mathbb{R}^n$ (as it must be expected, since these columns consist of the coordinates of orthonormal vectors in the orthonormal basis).

Equation (12) is equivalent to the equation

$$C^{-1} = C^{\top}$$

which in turn is equivalent to the relation

$$CC^{\top} = E.$$

The latter implies that

$$c_{i1}c_{j1} + \ldots + c_{in}c_{jn} = \begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{if} \quad i \neq j \end{cases}$$

for any $i, j = 1, \ldots, n$, i.e. that the rows of the matrix C also constitute an orthonormal family.

On summing up all these statements we obtain the following proposition.

Proposition 4. *A matrix C is orthogonal if and only if it has one (and therefore any other) of the following five properties.*

(a) $C^{\top}C = E$;

(b) *the columns of the matrix C are orthonormal;*

(c) $C^{-1} = C^{\top}$;

(d) $CC^{\top} = E$;

(e) *the rows of the matrix C are orthonormal.* □

Going over in relation (a) to the determinants and taking into account the fact that $\det C^{\top} = \det C$ we obtain the equation

$$(\det C)^2 = 1,$$

i.e. the equation

$$\det C = \pm 1.$$

The determinant of an orthogonal matrix is thus *equal to* ± 1.

Definition 9. An orthogonal matrix C is said to be *proper* (or *unimodular*) if $\det C = 1$.

It is clear that the inverse of an orthogonal (proper orthogonal) matrix, as well as a product of two orthogonal (proper orthogonal) matrices, is an orthogonal (proper orthogonal) matrix (if say $C_1^{\top}C_1 = E$ or $C_2^{\top}C_2 = E$, then $(C_1C_2)^{\top}(C_1C_2) = C_2^{\top}(C_1^{\top}C_1)C_2 = C_2^{\top}C_2 = E$). In algebraic terms this means that the set of all orthogonal (proper orthogonal) matrices of a given order n forms a group. This group is designated by the symbol O (n) (respectively by the symbol SO (n)).

When $n = 1$ an orthogonal matrix $C = (c_{11})$ is simply a number (c_{11}) satisfying the relation $c_{11}^2 = 1$. This means that the group O (1) consists of two elements ± 1 and that the group SO (1) consists of one element

$$O(1) = \{1, -1\}, \quad SO(1) = \{1\}.$$

When $n = 2$ there are three conditions of orthogonality:

$$c_{11}^2 + c_{21}^2 = 1,$$
$$c_{11}c_{12} + c_{21}c_{22} = 0,$$
$$c_{12}^2 + c_{22}^2 = 1.$$

The first and the last imply that there are angles α and β such that

$$c_{11} = \cos\alpha, \quad c_{21} = \sin\alpha,$$
$$c_{12} = \cos\beta, \quad c_{22} = \sin\beta,$$

and the second condition means that these angles are connected by the relation

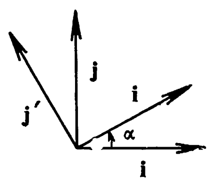
$$\cos\alpha\cos\beta + \sin\alpha\sin\beta = 0,$$

i.e. by the relation

$$\cos(\beta - \alpha) = 0.$$

Thus, either $\beta = \alpha + \frac{\pi}{2}$ or $\beta = \alpha + \frac{3\pi}{2}$, i.e. the matrix C is either of the form

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$



The angle α

or of the form

$$\begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}.$$

In the former case the matrix is proper and in the latter it is improper.

In particular we see that *any proper orthogonal matrix of the second order is of the form*

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

This means that two orthonormal plane bases of the same sign $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{i}'$, $\mathbf{j}'$ are connected by the formulas

$$\mathbf{i}' = \cos\alpha\cdot\mathbf{i} + \sin\alpha\cdot\mathbf{j},$$
$$\mathbf{j}' = -\sin\alpha\cdot\mathbf{i} + \cos\alpha\cdot\mathbf{j}$$

and that the corresponding coordinates x, y and x', y' are connected by the formulas

$$x = \cos\alpha\cdot x' - \sin\alpha\cdot y'$$
$$y = \sin\alpha\cdot x' + \cos\alpha\cdot y'.$$

Since $\mathbf{i}'\mathbf{i} = \cos\alpha$, $\mathbf{j}'\mathbf{i} = -\sin\alpha$ and $|\,\mathbf{i}\,| = |\,\mathbf{i}'\,| = 1$, to obtain the basis $\mathbf{i}'$, $\mathbf{j}'$ it is necessary to rotate the basis $\mathbf{i}$, $\mathbf{j}$ through the angle α.

A similar description of orthogonal matrices of the third order is rather complicated, and we shall not deal with them.

In a point Euclidean space formulas for the transformation of rectangular coordinates are of the form (in matrix notation; see formula (7) of Lecture 11):

$$x = Cx' + b,$$

where C is an orthogonal matrix (and b is an arbitrary column). In particular (denoting the coordinates by x and y) we obtain for $n = 2$ the following *formulas for the transformation of rectangular coordinates in the plane*

$$x = \cos\alpha\cdot x' - \sin\alpha\cdot y' + x_0,$$
$$y = \sin\alpha\cdot x' + \cos\alpha\cdot y' + y_0,$$

where α is the angle between positive directions of the abscissa axes of the old and new coordinate systems and (x_0, y_0) are the coordinates of a "new" origin O' in the "old" coordinate system.

Lecture 15

Trivectors in oriented Euclidean space • Triple product of three vectors • The area of a bivector in Euclidean space • A vector complementary to a bivector in oriented Euclidean space • Vector multiplication • Isomorphism of spaces of vectors and bivectors • Expressing a vector product in terms of coordinates • The normal equation of a straight line in the Euclidean plane and the distance between a point and a straight line • Angles between two straight lines in the Euclidean plane

In Euclidean (three-dimensional) space the theory of bivectors and trivectors is substantially simplified and essentially reduces to the theory of vectors. We shall first deal with the simpler case of trivectors.

Let $\mathscr{V}$ be an arbitrary Euclidean (three-dimensional) space. Assuming it to be oriented, consider in it an arbitrary orthonormal positively oriented basis $\mathbf{i}, \mathbf{j}, \mathbf{k}$. With any other such basis $\mathbf{i}', \mathbf{j}', \mathbf{k}'$ the basis $\mathbf{i}, \mathbf{j}, \mathbf{k}$ is connected by a proper orthogonal transition matrix. Since the determinant of the matrix is equal to $+1$, we deduce (see formula (2) of Lecture 13) that

$$\mathbf{i}' \wedge \mathbf{j}' \wedge \mathbf{k}' = \mathbf{i} \wedge \mathbf{j} \wedge \mathbf{k}.$$

Thus the *trivector* $\mathfrak{A}_0 = \mathbf{i} \wedge \mathbf{j} \wedge \mathbf{k}$ *does nct depend on the choice of the basis* $\mathbf{i}, \mathbf{j}, \mathbf{k}$, i.e. it is the same for all (orthonormal and positively oriented) bases.

Since $\mathfrak{A}_0 \neq 0$, any trivector $\mathfrak{A} \in \mathscr{V} \wedge \mathscr{V} \wedge \mathscr{V}$ is of the form

$$\mathfrak{A} = a\mathfrak{A}_0,$$

where a is a number.

Definition 1. A number a is called the *magnitude* (or *oriented volume*) of a trivector $\mathfrak{A}$, and its absolute value $|a|$ is called the *volume* of the trivector $\mathfrak{A}$.

In accordance with the general theory of volume Definition 1 means that we take the trivector $\mathfrak{A}_0$ as the standard of volume in space. In other words, the volume of a unit cube (i.e. a cube with an edge of unit length) is assumed equal to one. We thus see that the *Euclidean structure of a vector space uniquely determines the mensuration of volumes in it.*

Definition 2. The magnitude of a trivector $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$ is called a *mixed* (or *triple*) product of the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and designated by the symbol $\mathbf{abc}$.

According to formula (3) of Lecture 13, if $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$, $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$ and $\mathbf{c} = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$, then

$$\mathbf{abc} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}. \tag{1}$$

Note that a triple product *is dependent* on the orientation of the space and changes sign when the orientation is reversed.

The algebraic properties of a triple product are naturally the same as those of an external product: it is distributive (with respect to addition), homogeneous (with respect to the multiplication of vectors by numbers), anticommutative (changes sign when the factors are transposed) and free (equal to zero if and only if the vectors to be multiplied are coplanar).

Note that these are exactly the properties of determinants known from algebra. We may say that a triple product is simply a geometrical interpretation of third-order determinants.

Thus in a Euclidean (and oriented!) space trivectors can be replaced in a natural way (without any arbitrariness) by numbers (their values) and the external products of triples of vectors by their triple products.

As it was to be expected, the situation proves to be more complicated for bivectors.

Let $\mathfrak{a} = \mathbf{a} \wedge \mathbf{b}$ be an arbitrary nonzero bivector (in some Euclidean space $\mathscr{V}$). When the vector $\mathbf{b}$ is replaced by the vector $\mathbf{b} - \frac{\mathbf{ab}}{\mathbf{a}^2}\mathbf{a}$ orthogonal to the vector $\mathbf{a}$, the bivector $\mathfrak{a}$ obviously remains unchanged. Therefore from the very outset we can assume without loss of generality that the vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal. Normalizing these vectors, i.e. dividing them by their lengths, and taking out the common multiplier, we obtain for the bivector $\mathfrak{a}$ a representation of the form

(2) $$\mathfrak{a} = a(\mathbf{e}_1 \wedge \mathbf{e}_2)$$

where the vectors $\mathbf{e}_1$, $\mathbf{e}_2$ are orthonormal and $a > 0$.

Representation (2) (with $a = 0$) holds for $\mathfrak{a} = 0$ as well.

It is now easy to see that *in representation* (2) *the number* $a \geqslant 0$ *is uniquely determined by the bivector* $\mathfrak{a}$. Indeed, the equation

$$a(\mathbf{e}_1 \wedge \mathbf{e}_2) = a'(\mathbf{e}_1' \wedge \mathbf{e}_2'),$$

where $a > 0$, $a' > 0$ and the pairs of vectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_1'$, $\mathbf{e}_2'$ are orthonormal, means that we have

$$a'\mathbf{e}_1' = k(a\mathbf{e}_1) + l\mathbf{e}_2,$$
$$\mathbf{e}_2' = k'(a\mathbf{e}_1) + l'\mathbf{e}_2$$

with

$$\begin{vmatrix} k & l \\ k' & l' \end{vmatrix} = 1.$$

But then

$$\mathbf{e}_1' = \frac{ka}{a'}\mathbf{e}_1 + \frac{l}{a'}\mathbf{e}_2,$$
$$\mathbf{e}_2' = (k'a)\mathbf{e}_1 + l'\mathbf{e}_2,$$

from which it follows that the matrix

$$\begin{pmatrix} \frac{ka}{a'} & k'a \\ \frac{l}{a'} & l' \end{pmatrix}$$

is orthogonal and so its determinant a/a' is equal to ± 1. This proves that $a' = \pm a$. But under the hypothesis the numbers a and a' are positive. Therefore $a' = a$. □

Definition 3. The number a appearing in representation (2) is the *area* of a bivector $\mathfrak{a}$.

By what has been proved the definition is correct.

Thus the Euclidean structure of a space allows an area standard to be introduced in every plane, these standards being in some sense (which can be made more precise, if desired) equal in different planes.

Note that there is no natural way to compare area standards in different (nonparallel) planes in an affine space.

Suppose now that the Euclidean space under consideration is *three-dimensional* (bivectors in a two-dimensional space are of little interest since they are completely similar to trivectors in a three-dimensional space and represent a geometrical equivalent of second-order determinants) and, in addition, oriented.

Note in the first place that *for any two orthonormal vectors* $\mathbf{e}_1$, $\mathbf{e}_2$ *there exists one and only one vector* $\mathbf{e}_3$ *forming together with these vectors a positively oriented orthonormal basis* $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$.

Indeed, since the orthonormal family consisting of the vectors $\mathbf{e}_1$, $\mathbf{e}_2$ is not complete and hence (see Proposition 3 in the preceding lecture) not maximal, there is a vector $\mathbf{e}_3$ such that the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ constitute an orthonormal basis. If this basis is positively oriented, then this proves the existence of the vector $\mathbf{e}_3$. In case the orientation of the basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ has turned out to be negative, it is enough to change the sign of the vector $\mathbf{e}_3$.

Let there exist another vector

$$\mathbf{e}_3' = a\mathbf{e}_1 + b\mathbf{e}_2 + c\mathbf{e}_3$$

forming together with the vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ a positively oriented orthonormal basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3'$. Then the matrix

$$\begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & c \end{pmatrix}$$

as a matrix of transition from the basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ to the basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3'$ is orthogonal and proper which obviously is possible only for $a = b = 0$ and $c = 1$. This proves the uniqueness of the vetor $\mathbf{e}_3$. $\square$

Definition 4. A vector *complementary to a bivector* (2) is designated by the symbol $\mathfrak{a}^{\perp}$ and defined by the formula

$$\mathfrak{a}^{\perp} = a\mathbf{e}_3$$

where $\mathbf{e}_3$ is a vector such that the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ constitute a positively oriented orthonormal basis.

Here we must first check the correctness of the definition; to do this it is obviously enough to prove that *if* $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ *and* $\mathbf{e}'_1$, $\mathbf{e}'_2$, $\mathbf{e}'_3$ *are orthonormal bases of the same sign such that* $\mathbf{e}'_1 \wedge \mathbf{e}'_2 = \mathbf{e}_1 \wedge \mathbf{e}_2$, *then* $\mathbf{e}'_3 = \mathbf{e}_3$. Let

$$\begin{pmatrix} k_1 & k_2 & k_3 \\ l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

be a matrix of transition from the basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ to the basis $\mathbf{e}'_1$, $\mathbf{e}'_2$, $\mathbf{e}'_3$. This matrix is orthogonal and proper, so that

$$\begin{aligned} k_1^2+k_2^2+k_3^2&=1 \\ l_1^2+l_2^2+l_3^2&=1, \\ m_1^2+m_2^2+m_3^2&=1 \end{aligned} \quad \text{and} \quad \begin{vmatrix} k_1 & k_2 & k_3 \\ l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{vmatrix} = 1.$$

On the other hand, since $\mathbf{e}'_1 \wedge \mathbf{e}'_2 = \mathbf{e}_1 \wedge \mathbf{e}_2$, we have

$$m_1 = 0, \quad m_2 = 0$$

and the matrix

$$\begin{pmatrix} k_1 & k_2 \\ l_1 & l_2 \end{pmatrix}$$

(obviously, orthogonal) is unimodular. Thus

$$\begin{aligned} k_1^2+k_2^2&=1 \\ l_1^2+l_2^2&=1 \end{aligned} \quad \text{and} \quad \begin{vmatrix} k_1 & k_2 \\ l_1 & l_2 \end{vmatrix} = 1.$$

From the above relations it follows directly that $k_3 = 0$, $l_3 = 0$, and $m_3 = 1$. Thus, $\mathbf{e}'_3 = \mathbf{e}_3$. □

Definition 5. *The vector product* $\mathbf{a} \times \mathbf{b}$ of vectors $\mathbf{a}$ and $\mathbf{b}$ is a vector complementary to their external product $\mathbf{a} \wedge \mathbf{b}$

$$\mathbf{a} \times \mathbf{b} = (\mathbf{a} \wedge \mathbf{b})^{\perp}.$$

The following proposition explains why the product **abc** is called "mixed".

Proposition 1. *For any three vectors* **a**, **b**, **c** *we have the formula*

$$\mathbf{abc} = (\mathbf{a} \times \mathbf{b})\,\mathbf{c}.$$

Proof. For $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ the formula is obvious. Let $\mathbf{a} \times \mathbf{b} \neq = \mathbf{0}$. To construct the vector $\mathbf{a} \times \mathbf{b}$ we must first represent the bivector $\mathbf{a} \wedge \mathbf{b} \neq 0$ in the form

$$\mathbf{a} \wedge \mathbf{b} = a\,(\mathbf{e}_1 \wedge \mathbf{e}_2)$$

where $a > 0$ and $\mathbf{e}_1$, $\mathbf{e}_2$ are orthonormal vectors. Following the above method of constructing such a representation we shall obviously have formulas of the form

$$\mathbf{a} = a_1\mathbf{e}_1,$$
$$\mathbf{b} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2.$$

From these it follows that $\mathbf{a} \wedge \mathbf{b} = a_1b_2\,(\mathbf{e}_1 \wedge \mathbf{e}_2)$. Therefore $a = a_1b_2$ and hence $\mathbf{a} \times \mathbf{b} = a_1b_2\mathbf{e}_3$, where $\mathbf{e}_3$ is a vector complementing the vectors $\mathbf{e}_1$, $\mathbf{e}_2$ to form a positively oriented orthonormal basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$.

Consequently, if

$$\mathbf{c} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3$$

then $(\mathbf{a} \times \mathbf{b})\,\mathbf{c} = a_1b_2c_3$. On the other hand,

$$\mathbf{abc} = \begin{vmatrix} a_1 & 0 & 0 \\ b_1 & b_2 & 0 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1b_2c_3.$$

Thus, $(\mathbf{a} \times \mathbf{b})\,\mathbf{c} = \mathbf{abc}$. $\square$

Proposition 2. *Vector multiplication is distributive over addition, i.e.*

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$$

for any three vectors **a**, **b**, **c**.

Proof. Let **x** be an arbitrary vector. Then, by virtue of the distributivity of mixed and scalar multiplications,

$$(\mathbf{a} \times (\mathbf{b} + \mathbf{c}))\,\mathbf{x} = \mathbf{a}\,(\mathbf{b} + \mathbf{c})\,\mathbf{x} = \mathbf{abx} + \mathbf{acx} =$$
$$= (\mathbf{a} \times \mathbf{b})\,\mathbf{x} + (\mathbf{a} \times \mathbf{c})\,\mathbf{x} = (\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c})\,\mathbf{x}.$$

Taking as $\mathbf{x}$ the vectors of some basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ we find from here that all Fourier coefficients of the vectors $\mathbf{a} \times (\mathbf{b} + \mathbf{c})$ and $\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$ are equal. Hence these vectors are also equal. □

Corollary. *The complementary vector of a sum of bivectors is equal to the sum of the complementary vectors of the summands, i.e.*

$$(\mathfrak{a} + \mathfrak{b})^{\perp} = \mathfrak{a}^{\perp} + \mathfrak{b}^{\perp}$$

for any bivectors $\mathfrak{a}$ and $\mathfrak{b}$. □

This corollary is merely a reformulation of Proposition 2.

Proposition 3. *Vector multiplication is homogeneous and anticommutative, i.e.*

$$k(\mathbf{a} \times \mathbf{b}) = k\mathbf{a} \times \mathbf{b} = \mathbf{a} \times k\mathbf{b},$$

$$\mathbf{b} \times \mathbf{a} = -\mathbf{a} \times \mathbf{b}$$

for any number k and any vectors $\mathbf{a}$ *and* $\mathbf{b}$.

Proof. The first statement follows from the homogeneity of mixed and scalar multiplications (cf. the proof of Proposition 2). The second follows from the first plus the anticommutativity of external multiplication. □

Corollary. *For any number k and any bivector $\mathfrak{a}$ we have*

$$(k\mathfrak{a})^{\perp} = k\mathfrak{a}^{\perp}. \ \square$$

Note that this corollary can easily be derived directly. Indeed, if $\mathfrak{a} = a(\mathbf{e}_1 \wedge \mathbf{e}_2)$ and $\mathfrak{a}^{\perp} = a\mathbf{e}_3$, then $k\mathfrak{a} = ka(\mathbf{e}_1 \wedge \mathbf{e}_2)$ for $k \geqslant 0$ and therefore $(k\mathfrak{a})^{\perp} = ka\mathbf{e}_3 = k\mathfrak{a}^{\perp}$. But if $k < 0$, then $k\mathfrak{a} = |k| a(\mathbf{e}_2 \wedge \mathbf{e}_1)$ and therefore $(k\mathfrak{a})^{\perp} = |k| a(-\mathbf{e}_3) = ka\mathbf{e}_3 = k\mathfrak{a}^{\perp}$, since the basis $\mathbf{e}_2, \mathbf{e}_1, -\mathbf{e}_3$ is orthonormal and positively oriented. □

This gives a direct proof also for Proposition 3 (there seems to be no similar sufficiently simple proof of Proposition 2).

Theorem 1. *The correspondence $\mathfrak{a} \mapsto \mathfrak{a}^{\perp}$ is an isomorphism of a vector space $\mathscr{V} \wedge \mathscr{V}$ onto a vector space $\mathscr{V}$.*

Proof. In view of the corollaries of Propositions 2 and 3 one only needs to prove that the mapping $\mathfrak{a} \mapsto \mathfrak{a}^{\perp}$ is bijective, i.e. that for any vector $\mathbf{a}$ there exists a unique bivector $\mathfrak{a}$ such that $\mathfrak{a}^{\perp} = \mathbf{a}$.

Let a be the length of the vector $\mathbf{a}$, i.e. let $\mathbf{a} = a\mathbf{e}_3$, where $\mathbf{e}_3$ is a vector of unit length (uniquely determined when $a \neq 0$ and arbitrary when $a = 0$). Further let $\mathbf{e}_1$ and $\mathbf{e}_2$ be vectors such that the vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ form a positively oriented orthonormal basis (that such vectors exist is obvious) and let $\alpha = a(\mathbf{e}_1 \wedge \mathbf{e}_2)$. Then, by definition, $\alpha^\perp = \mathbf{a}$. This proves the existence of the bivector α.

To prove its uniqueness it is obviously enough to establish that for the orthonormal bases of the same sign $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and $\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'$ the equation $\mathbf{e}_3' = \mathbf{e}_3$ yields the equation $\mathbf{e}_1' \wedge \mathbf{e}_2' = \mathbf{e}_1 \wedge \mathbf{e}_2$. But this is nearly obvious. Indeed, since $\mathbf{e}_3' = \mathbf{e}_3$, the matrix of transition from the basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ to the basis $\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'$ is of the form

$$\begin{pmatrix} k_1 & k_2 & 0 \\ l_1 & l_2 & 0 \\ m_1 & m_2 & 1 \end{pmatrix}.$$

The condition of orthogonality for the last row gives

$$m_1^2 + m_2^2 + 1 = 1,$$

from which it follows that $m_1 = 0$ and $m_2 = 0$. Therefore

$$\mathbf{e}_1' = k_1\mathbf{e}_1 + l_1\mathbf{e}_2,$$
$$\mathbf{e}_2' = k_2\mathbf{e}_1 + l_2\mathbf{e}_2$$

with

$$\begin{vmatrix} k_1 & k_2 \\ l_1 & l_2 \end{vmatrix} = 1.$$

Hence $\mathbf{e}_1' \wedge \mathbf{e}_2' = \mathbf{e}_1 \wedge \mathbf{e}_2$. □

Corrollary. *Vector multiplication is free, i.e.* $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ *if and only if the vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *are collinear.* □

Note that complete reduction of bivectors to vectors, provided for by Theorem 1, is possible only in a three-dimensional space and only if a definite orientation of the space is chosen.

It is obvious that for any positively oriented orthonormal basis $\mathbf{i}, \mathbf{j}, \mathbf{k}$ we have the formulas

$$(\mathbf{j} \wedge \mathbf{k})^\perp = \mathbf{i}, \quad (\mathbf{k} \wedge \mathbf{i})^\perp = \mathbf{j}, \quad (\mathbf{i} \wedge \mathbf{j})^\perp = \mathbf{k},$$

i.e. the formulas

$$\mathbf{j} \times \mathbf{k} = \mathbf{i}, \quad \mathbf{k} \times \mathbf{i} = \mathbf{j}, \quad \mathbf{i} \times \mathbf{j} = \mathbf{k}.$$

It follows (cf. formula (8) of Lecture 8) that for any vectors

$$\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \text{ and } \mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$$

we have the formula

$$\mathbf{a}\times\mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}.$$

We stress that the basis **i**, **j**, **k** is here supposed to be not only orthonormal but also positively oriented (right-handed).

By construction, the length $|\mathbf{a}\times\mathbf{b}|$ of the vector product $\mathbf{a}\times\mathbf{b}$ is equal to the area S of the parallelogram constructed on the vectors **a** and **b**. Therefore

$$S = \sqrt{\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}^2}.$$

The existence in the plane or in space of a Euclidean structure (scalar multiplication) also adds substantially to the theory of straight lines and planes. We shall begin as ever with a discussion of straight lines in the plane.

Suppose an arbitrary nonzero vector **n** and some point M_0 are given in a Euclidean plane.

It is easy to see that the set of all points M of the plane, for which the vector **n** is orthogonal to the vector $\overrightarrow{M_0M}$, i.e. for which we have

$$\mathbf{n}\overrightarrow{M_0M} = 0, \tag{3}$$

is a straight line passing through the point M_0. Indeed, in rectangular coordinates condition (3) has the form

$$A(x - x_0) + B(y - y_0) = 0,$$

where A, B are the coordinates of the vector **n**, and x, y, and x_0, y_0 are the coordinates of the points M and M_0, respectively. □

The vector **n** is said to be *orthogonal* to a straight line. Up to collinearity it is uniquely determined by a straight line. Any equation of a straight line

$$Ax + By + C = 0 \tag{4}$$

(in rectangular coordinates) determines the vector $\mathbf{n}$ (A, B) orthogonal to the straight line. If the vector is normalized (is of unit length), i.e. if $A^2 + B^2 = 1$, and if $C \leqslant 0$, then equation (4) is said to be the *normal equation* of a straight line. A normal equation is ordinarily written as

(5) $$x \cos \alpha + y \sin \alpha - p = 0,$$

where α is the angle between the vector $\mathbf{n}$ and the unit vector $\mathbf{i}$ of the abscissa axis. To reduce equation (4) to the form (5), it is enough to divide it by $\sqrt{A^2 + B^2}$ when $C \leqslant 0$ and by $-\sqrt{A^2 + B^2}$ when $C \geqslant 0$.

Introducing the radius vector $\mathbf{r}$ (x, y) and the unit vector $\mathbf{n}$ $(\cos \alpha, \sin \alpha)$ we can write equation (5) in vector form:

(6) $$\mathbf{nr} - p = 0.$$

Defininion 6. Two straight lines are said to be *perpendicular* if their direction vectors are orthogonal. Let N be the point of intersection of the straight line passing through a given point M and perpendicular to the straight line (6) (it is easy to see that such a straight line always exists and is unique) with the straight line (6). The length $d = |NM|$ of the segment $\overrightarrow{NM}$ is called the *distance* from the point M to the straight line (6).

This distance is obviously equal to the absolute value of the scalar product $\mathbf{n}\overrightarrow{NM} = \mathbf{n}(\mathbf{r} - \mathbf{s})$, where $\mathbf{r}$ and $\mathbf{s}$ are the radius vectors of the points M and N. But the point N lies on the straight line (6) and therefore $\mathbf{ns} = p$. Hence

(7) $$d = |\mathbf{nr} - p| = |x \cos \alpha + y \sin \alpha - p|.$$

This formula explains why we need normal equations.

For $p \neq 0$ the value $\mathbf{nr} - p$ is positive (and hence equal to d) if and only if the point M is not in the halfplane containing the origin 0.

In general it is not so simple to define the angle φ between two straight lines in a Euclidean plane uniquely. As already noted (see Lecture 13), its unsophisticated definition as the angle between direction vectors of these straight lines, generally speaking, gives four different values. Even if we impose the restriction $0 \leqslant \varphi \leqslant \pi$ (such angles may be called "elementary-geometry" angles), there will all the same be a choice left between two adjacent angles (summing to π).

Let the straight lines considered be given (in rectangular coordinates) by the equations

(8) $$Ax + By + C = 0 \quad \text{and} \quad A_1x + B_1y + C = 0.$$

It is natural then to take as the angle φ the angle between the direction vectors $\mathbf{a}$ $(B, -A)$ and $\mathbf{a}_1$ $(B_1, -A_1)$ (giving the orientation of the straight lines defined by the equations; see Lecture 12). Hence

(9) $$\cos\varphi = \frac{AA_1 + BB_1}{\sqrt{A^2+B^2}\,\sqrt{A_1^2+B_1^2}}.$$

Together with the condition $0 \leqslant \varphi < \pi$ this uniquely determines the angle φ. We have an acute angle when $AA_1 + BB_1 > 0$, an obtuse angle when $AA_1 + BB_1 < 0$ and a right angle when $AA_1 + BB_1 = 0$.

In particular we see that the *straight lines* (8) *are perpendicular if and only if*

$$AA_1 + BB_1 = 0.$$

Instead of imposing on the angle φ the elementary-geometric condition $0 \leqslant \varphi < \pi$, it is possible to require that the sum $\varphi + \frac{\pi}{2}$ should be one of the angles between the vectors $\mathbf{a}$ $(B, -A)$ and $\mathbf{n}_1$ (A_1, B_1). Since $\cos\left(\varphi + \frac{\pi}{2}\right) = -\sin\varphi$, it is equivalent to the requirement that

(10) $$\sin\varphi = \frac{AB_1 - A_1B}{\sqrt{A^2+B^2}\,\sqrt{A_1^2+B_1^2}}.$$

Formulas (9) and (10) uniquely determine the angle φ satisfying the condition $-\pi < \varphi \leqslant \pi$. Its absolute value is equal to an elementary-geometry angle φ.

Another way of fixing the angle φ is to seek the elementary-geometry angle with the same tangent as the angle determined by the second method. In other words, this angle is uniquely determined by the formula

(11) $$\tan\varphi = \frac{AB_1 - A_1B}{AA_1 + BB_1}$$

and the condition $0 \leqslant \varphi < \pi$.

The last method is ordinarily employed when straight lines are given by equations of the form

(12) $$y = kx + b \quad \text{and} \quad y = k_1x + b_1.$$

For such straight lines formula (11) takes the form

$$\tan\varphi = \frac{k_1 - k}{1 + kk_1}.$$

In particular we see that *straight lines* (12) *are*

$$\left.\begin{array}{r}\textit{parallel}\\ \textit{perpendicular}\end{array}\right\}\ \begin{array}{l}\textit{if and}\\ \textit{only if}\end{array}\ \left\{\begin{array}{c}k = k_1\\ kk_1 = -1.\end{array}\right.$$

Lecture 16

The plane in Euclidean space • The distance from a point to a plane • The angle between two planes, between a straight line and a plane, between two straight lines • The distance from a point to a straight line in space • The distance between two straight lines in space • The equations of the common perpendicular of two skew lines in space

The situation is perfectly similar in the case of planes in Euclidean space.

For any nonzero vector **n** (A, B, C) the set of all points $M(x, y, z)$ of a space, for which

$$\mathbf{n}\overrightarrow{M_0M} = 0,$$

where $M_0(x_0, y_0, z_0)$ is some fixed point, is the plane

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

The vector **n** is said to be *orthogonal* to that plane. It is simply a vector complementary to the direction bivector of the plane. Therefore up to collinearity the vector **n** is uniquely determined by the plane.

Any equation of a plane

$$Ax + By + Cz + D = 0 \tag{1}$$

(in rectangular coordinates) determines the vector **n** (A, B, C) orthogonal to the plane. If the vector is normalized, i.e. $A^2 + B^2 + C^2 = 1$, and if $D \leqslant 0$, then equation (1) is said to be a *normal equation* of the plane. A normal equation is ordinarily written as

$$x \cos\alpha + y \cos\beta + z \cos\gamma - p = 0, \tag{2}$$

where α, β, γ are the angles between the vector $\mathbf{n}$ and the vectors of a basis $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ and $p \leqslant 0$. To reduce equation (1) to the form (2) it is sufficient to divide it by $\sqrt{A^2 + B^2 + C^2}$ when $D \leqslant 0$ and by $-\sqrt{A^2 + B^2 + C^2}$ when $D \geqslant 0$.

In "vector form" equation (2) becomes

$$\mathbf{nr} - p = 0. \tag{3}$$

Definition 1. A straight line is said to be *perpendicular* to the plane (3) if its direction vector is collinear with the vector $\mathbf{n}$. Let N be the point of intersection of the straight line passing through a given point M and perpendicular to the plane (3) with that plane. The length $d = |NM|$ of the segment $\overline{NM}$ is called the *distance from the point M to the plane* (3).

Just as for the distance from a point to a straight line it can be proved word for word that the *distance from the point M to the plane* (3) *is expressed by the formula*

$$d = |\mathbf{nr} - p|,$$

where $\mathbf{r}$ is the radius vector of the point M.

The *angle between two planes* is defined to be the angle between the vectors orthogonal to those planes. It is possible to fix it uniquely by the same three methods as those used to fix the angle between two straight lines in the plane, and obtain similar formulas. This case being quite trivial we shall waste neither time nor space on it. We shall only note that the *two planes*

$$Ax + By + Cz + D = 0 \text{ and } A_1x + B_1y + C_1z + D_1 = 0$$

are perpendicular (i.e. the angle between them is equal to $\pi/2$) *if and onlyifwe have*

$$AA_1 + BB_1 + CC_1 = 0.$$

Similarly the *angle between a straight line and a plane* is defined to be a complement to $\pi/2$ of the angle between the direction vector $\mathbf{a}$ (l, m, n) of the straight line and the vector $\mathbf{n}$ (A, B, C) orthogonal to the plane, with additional restrictions of some character which ensure uniqueness.

For this angle we have the formula

$$\sin\varphi = \frac{Al + Bm + Cn}{\sqrt{A^2+B^2+C^2}\,\sqrt{l^2+m^2+n^2}}.$$

In particular we see that *a straight line and a plane are perpendicular if and only if*

$$Al + Bm + Cn = 0.$$

Finally, the *angle between two straight lines in space* is defined according to the general rule (as the angle between their direction vectors **a** (l, m, n) and $\mathbf{a}_1$ (l_1, m_1, n_1)) and therefore we have for it the formula

$$\cos\varphi = \frac{ll_1 + mm_1 + nn_1}{\sqrt{l^2+m^2+n^2}\,\sqrt{l_1^2+m_1^2+n_1^2}}.$$

More interesting are the questions of the distance from a point to a straight line in space and that of the distance between straight lines.

An arbitrary straight line $\mathbf{r} = \mathbf{r}_0 + t\mathbf{a}$ and an arbitrary point M $(\mathbf{r})$ not lying on it are in one and only one plane (namely, in the plane passing through the point M_0 $(\mathbf{r}_0)$ and having the direction bivector $\overrightarrow{M_0M} \wedge \mathbf{a}$). The distance d from the point M to the straight line in that plane is precisely the *distance from the point M to the straight line* in space.

By definition, for it to be computed it is necessary to find on the straight line a point N such that the straight line NM is perpendicular to the given straight line, i.e. that the vector $\overrightarrow{NM}$ is orthogonal to the vector **a**. Then d will be equal to the length of the vector $\overrightarrow{NM}$, i.e. we shall have

$$\overrightarrow{NM} = d\mathbf{e}_1,$$

where $\mathbf{e}_1$ is some unit vector, and hence

$$\overrightarrow{NM} \wedge \mathbf{a} = ad\,(\mathbf{e}_1 \wedge \mathbf{e}_2),$$

where a is the length of the vector **a** and $\mathbf{e}_2 = \frac{\mathbf{a}}{a}$. Since $\mathbf{e}_1$ and $\mathbf{e}_2$ are orthogonal unit vectors (i.e. form an orthonormal family) it follows that $|\overrightarrow{NM} \wedge \mathbf{a}| = ad$, i.e. that

$$d = \frac{1}{a} |\overrightarrow{NM} \times \mathbf{a}|.$$

On the other hand, since the point N is in the given straight line, its radius vector $\mathbf{s}$ has the form $\mathbf{r}_0 + t_1\mathbf{a}$, where t_1 is a number. Therefore $\overrightarrow{NM} = \mathbf{r} - (\mathbf{r}_0 + t_1\mathbf{a}) = (\mathbf{r} - \mathbf{r}_0) - t_1\mathbf{a}$ and hence

$$\overrightarrow{NM} \times \mathbf{a} = (\mathbf{r} - \mathbf{r}_0) \times \mathbf{a}.$$

This proves that

$$d = \frac{1}{a} |(\mathbf{r} - \mathbf{r}_0) \times \mathbf{a}|,$$

or, in coordinates, that

$$d^2 = \frac{\begin{vmatrix} y-y_0 & z-z_0 \\ m & n \end{vmatrix}^2 + \begin{vmatrix} x-x_0 & z-z_0 \\ l & n \end{vmatrix}^2 + \begin{vmatrix} x-x_0 & y-y_0 \\ l & m \end{vmatrix}^2}{l^2+m^2+n^2}.$$

Now let

(4) $$\mathbf{r} = \mathbf{r}_0 + s\mathbf{a}, \quad \mathbf{r} = \mathbf{r}_1 + t\mathbf{b}$$

be two nonparallel straight lines in space not lying in the same plane (skew lines). It turns out that *there exists one and only one pair of numbers* (s_0, t_0) *such that the straight line* N_0N_1 *passing through the points* N_0 $(\mathbf{n}_0)$ *and* N_1 $(\mathbf{n}_1)$, *where* $\mathbf{n}_0 = \mathbf{r}_0 + s_0\mathbf{a}$, $\mathbf{n}_1 = \mathbf{r}_1 + t_0\mathbf{b}$, *is perpendicular to both straight lines* (4).

Indeed, the straight line N_0N_1 has the direction vector $\mathbf{n}_1 - \mathbf{n}_0 = (\mathbf{r}_1 - \mathbf{r}_0) + t_0\mathbf{b} - s_0\mathbf{a}$ and therefore it is perpendicular to the straight lines (4) if and only if $(\mathbf{n}_1 - \mathbf{n}_0)\,\mathbf{a} = 0$ and $(\mathbf{n}_1 - \mathbf{n}_0)\,\mathbf{b} = 0$, i.e. if

$$(\mathbf{r}_1 - \mathbf{r}_0)\,\mathbf{a} + t_0\mathbf{ba} - s_0\mathbf{a}^2 = 0,$$
$$(\mathbf{r}_1 - \mathbf{r}_0)\,\mathbf{b} + t_0\mathbf{b}^2 - s_0\mathbf{ab} = 0.$$

To complete the proof it remains to note that this system of linear equations has for s_0, t_0 the unique solution

$$s_0 = \frac{\begin{vmatrix} (\mathbf{r}_1-\mathbf{r}_0)\,\mathbf{a} & \mathbf{ab} \\ (\mathbf{r}_1-\mathbf{r}_0)\,\mathbf{b} & \mathbf{b}^2 \end{vmatrix}}{\begin{vmatrix} \mathbf{a}^2 & \mathbf{ab} \\ \mathbf{ab} & \mathbf{b}^2 \end{vmatrix}}, \quad t_0 = -\frac{\begin{vmatrix} \mathbf{a}^2 & (\mathbf{r}_1-\mathbf{r}_0)\,\mathbf{a} \\ \mathbf{ab} & (\mathbf{r}_1-\mathbf{r}_0)\,\mathbf{b} \end{vmatrix}}{\begin{vmatrix} \mathbf{a}^2 & \mathbf{ab} \\ \mathbf{ab} & \mathbf{b}^2 \end{vmatrix}}.$$

(The denominator is nonzero since the vectors **a** and **b** are not collinear.) □

The straight line N_0N_1 is called the *common perpendicular* of skew lines, and the length $d = |N_0N_1|$ of the segment $\overline{N_0N_1}$ is called the *distance* between those lines.

The name of the number d is accounted for by the fact that, as is easily seen, *for any point M_0 of the straight line* $\mathbf{r} = \mathbf{r}_0 + s\mathbf{a}$ *and any point M_1 of the straight line* $\mathbf{r} = \mathbf{r}_1 + t\mathbf{b}$ *we have the inequality*

$$|M_0M_1| \geqslant |N_0N_1|,$$

equality holding only for $M_0 = N_0$ and $M_1 = N_1$. Indeed

$$|M_0M_1|^2 = ((\mathbf{r}_1 - \mathbf{r}_0) + t\mathbf{b} - s\mathbf{a})^2 = \{[(\mathbf{r}_1 - \mathbf{r}_0) + t_0\mathbf{b} - s_0\mathbf{a}] + (t - t_0)\,\mathbf{b} - (s - s_0)\,\mathbf{a}\}^2 = ((\mathbf{n}_1 - \mathbf{n}_0) + (t - t_0)\,\mathbf{b} - (s - s_0)\,\mathbf{a})^2 = (\mathbf{n}_1 - \mathbf{n}_0)^2 + [(t - t_0)\,\mathbf{b} - (s - s_0)\,\mathbf{a}]^2,$$

for $(\mathbf{n}_1 - \mathbf{n}_0)\,\mathbf{a} = 0$ and $(\mathbf{n}_1 - \mathbf{n}_0)\,\mathbf{b} = 0$ (the Pythagorean theorem). Therefore

$$|M_0M_1|^2 \geqslant (\mathbf{n}_1 - \mathbf{n}_0)^2 = |N_0N_1|^2$$

equality holding only for $s = s_0$, $t = t_0$. □

We now note that the mixed product $\overrightarrow{M_0M_1}\mathbf{ab}$ does not depend on the choice of the points M_0 and M_1 on the straight lines (4) since when these points are changed we add to the vector $\overrightarrow{M_0M_1}$ a linear combination of vectors **a** and **b**. But for $s = 0$, $t = 0$ this product is equal to $(\mathbf{r}_1 - \mathbf{r}_0)\,\mathbf{ab}$ and for $s = s_0$ and $t = t_0$ it is equal to $\overrightarrow{N_0N_1}\,(\mathbf{a} \times \mathbf{b}) = \pm |N_0N_1| \cdot |\mathbf{a} \times \mathbf{b}|$ (for the vectors $\overrightarrow{N_0N_1}$ and $\mathbf{a} \times \mathbf{b}$ are collinear). Hence

$$d = \frac{|(\mathbf{r}_1 - \mathbf{r}_0)\,\mathbf{ab}|}{|\mathbf{a} \times \mathbf{b}|}.$$

In coordinates

$$d = \left| \frac{\begin{vmatrix} x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ l & m & n \\ l_1 & m_1 & n_1 \end{vmatrix}}{\sqrt{\begin{vmatrix} m & n \\ m_1 & n_1 \end{vmatrix}^2 + \begin{vmatrix} l & n \\ l_1 & n_1 \end{vmatrix}^2 + \begin{vmatrix} l & m \\ l_1 & m_1 \end{vmatrix}^2}} \right|, \tag{5}$$

where

l, m, n are the coordinates of the direction vector $\mathbf{a}$;
l_1, m_1, n_1 are the coordinates of the direction vector $\mathbf{b}$;
x_0, y_0, z_0 are the coordinates of the radius vector $\mathbf{r}_0$;
x_1, y_1, z_1 are the coordinates of the radius vector $\mathbf{r}_1$.

Thus formula (5) gives the distance between two nonparallel straight lines

$$(6)\quad \frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n} \text{ and } \frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1}.$$

For the straight line N_0N_1 we know the radius vector of its point N_0 and the direction vector $\overrightarrow{N_0N_1}$. Therefore we can write immediately the parametric equation of that straight line. After obvious identical transformations it becomes

$$\mathbf{r}=\left|\frac{\begin{vmatrix} (1-t)\,\mathbf{r}_0+t\mathbf{r}_1 & -(1-t)\,\mathbf{a} & t\mathbf{b} \\ (\mathbf{r}_1-\mathbf{r}_0)\,\mathbf{a} & \mathbf{a}^2 & \mathbf{ab} \\ (\mathbf{r}_1-\mathbf{r}_0)\,\mathbf{b} & \mathbf{ab} & \mathbf{b}^2 \end{vmatrix}}{\begin{vmatrix} \mathbf{a}^2 & \mathbf{ab} \\ \mathbf{ab} & \mathbf{b}^2 \end{vmatrix}}\right|.$$

This straight line can be characterized as the line of intersection of two planes, one of them passing through the point $M_0(x_0, y_0, z_0)$ and having the direction bivector $\mathbf{a} \wedge \overrightarrow{N_0N_1}$ (or equivalently the direction bivector $\mathbf{a} \wedge (\mathbf{a} \times \mathbf{b})$) and the other passing through the point $M_1(x_1, y_1, z_1)$ and having the direction bivector $\mathbf{b} \wedge \wedge (\mathbf{a} \times \mathbf{b})$, i.e. by the equations

$$\left\{\begin{aligned} &\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ l & m & n \\ \begin{vmatrix} m & n \\ m_1 & n_1 \end{vmatrix} & -\begin{vmatrix} l & n \\ l_1 & n_1 \end{vmatrix} & \begin{vmatrix} l & m \\ l_1 & m_1 \end{vmatrix} \end{vmatrix}=0, \\ &\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ l_1 & m_1 & n_1 \\ \begin{vmatrix} m & n \\ m_1 & n_1 \end{vmatrix} & -\begin{vmatrix} l & n \\ l_1 & n_1 \end{vmatrix} & \begin{vmatrix} l & m \\ l_1 & m_1 \end{vmatrix} \end{vmatrix}=0. \end{aligned}\right.$$

Lecture 17

The parabola • The ellipse • The focal and directorial properties of the ellipse • The hyperbola • The focal and directorial properties of the hyperbola

In this lecture we begin the study of three remarkable curves in the Euclidean plane: the parabola, ellipse, and hyperbola. In spite of their outward dissimilarity we shall see that they may be combined in a natural way into one group.

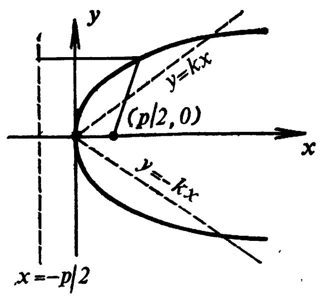
Definition 1. A curve in a Euclidean plane is said to be a *parabola* if there exists a system of rectangular coordinates x, y in which the equation of that curve is of the form

$$y^2 = 2px, \quad \text{where} \quad p > 0. \tag{1}$$

Note that we are not giving any general definition of what a curve is since we are not going to prove any theorems on "curves at large". We shall use the term "curve" as a synonym of a "subset of a plane" but in a narrower sense determined solely by tradition.

The "equation" of a curve is an arbitrary relation between x- and y-coordinates which holds if and only if the point $M(x, y)$ having these coordinates is on the curve. This is not a rigorous definition but only a verbal description of a rather vague notion. Therefore we state nothing about "equations" and "curves": neither that any curve has an equation nor that any equation gives a curve. In practice, defining a particular curve (or a class of curves) we shall always indicate its (their) equation.

In principle it would be possible to do without the concepts of a curve and its equation, but this would result in unusual and cumbersome statements. For example, Definition 1 would then take the following form: a set of points of a plane is said to be a parabola if there exists a system of rectangular coordinates x, y and a number $p > 0$ such that the point $M(x, y)$ belongs to that set if and only if $y^2 = 2px$.



A parabola

Rectangular coordinates x, y specified by Definition 1 are called the *canonical coordinates* for the given parabola and equation (1) is called its *canonical equation*.

The abscissa axis of a system of canonical coordinates is the axis of symmetry of the parabola (for when y changes sign equation (1) remains unchanged). For this reason this straight line is called the *axis* of the parabola (or sometimes the *focal axis*).

When $x < 0$ there are no points satisfying equation (1). This means that the whole parabola is situated in the half-plane $x \geqslant 0$.

The ordinate axis $x = 0$ is intersected by the parabola (1) only in the point $O(0, 0)$ which is called the *vertex* of the parabola.

From these two remarks it immediately follows that the *axis of parabola is its only axis of symmetry*. The ratio

$$\frac{|y|}{x} = \frac{\sqrt{2p}}{\sqrt{x}}$$

tends to zero as $x \to +\infty$. This means that beginning with a sufficiently large x the parabola is contained in any symmetrical angle containing the positive semiaxis of the abscissa axis. Visually, this means that if we look along that semiaxis the parabola will appear to be convergent, although in fact it is arbitrarily far from the abscissa axis ($|y| \to +\infty$ as $x \to +\infty$).

We stress that both the axis and the vertex of the parabola are uniquely characterized, in a purely geometrical way, without resorting to any coordinates: the axis is the axis of symmetry and the vertex is the common point of the axis and the parabola. This means that the axes of a system of canonical coordinates are also uniquely characterized by the parabola: the abscissa axis is characterized as the axis of the parabola and the ordinate axis as the straight line passing through the vertex and perpendicular to the axis of the parabola. The positive direction of the abscissa axis is also determined by the parabola (as the direction giving the half-plane containing the parabola).

This proves that up to a change in the orientation of the ordinate axis (i.e. in the sign of the y-coordinate) the canonical coordinates are uniquely determined by the parabola.

Therefore all the objects determined with the aid of canonical coordinates but independent of the orientation of the ordinate axis are invariantly (i.e. without any arbitrariness) related to the parabola. Among them are:

the number p called the *focal parameter*,

the number $\frac{p}{2}$ called the *focal distance*,

the point $\left(\frac{p}{2}, 0\right)$ called the *focus*,

the straight line $x = -\frac{p}{2}$ called the *directrix*.

It is easy to see that the *parabola is a set* (or locus, as people prefer to say after the old fashion) *of all points equidistant from the focus and directrix*. Indeed, after squaring and reducing the similar terms the condition of equidistance

$$\left|x+\frac{p}{2}\right|=\sqrt{\left(x-\frac{p}{2}\right)^2+y^2}$$

turns into equation (1) and, conversely, if $y^2 = 2px$, then this condition obviously holds. □

Definition 2. A curve in the Euclidean plane is said to be an *ellipse* if there exists a system of rectangular coordinates x, y in which the equation of that curve is of the form

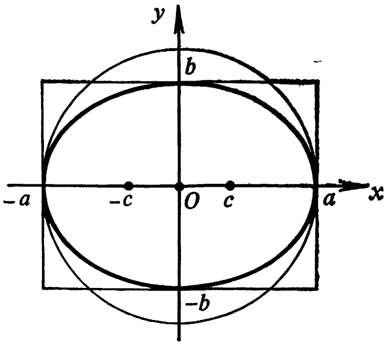
$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \text{ where } a \geqslant b > 0, \tag{2}$$

The coordinates in which the equation of the ellipse is of the form (2) are called *canonical coordinates* (for the given ellipse) and equation (2) is called the *canonical equation* of the ellipse.

When $b = a$ the ellipse has the equation

(3) $$x^2 + y^2 = a^2$$

which is obviously the equation of a circle of radius a with centre at the point O (0, 0). Hence the circle is a particular case of the ellipse.



An ellipse

Compare for $b < a$ the ellipse (2) and the circle (3). Let $k = \frac{b}{a}$. If a point with coordinates x, y is in the circle (3), then a point with coordinates x, ky is in the ellipse (2) $\left(\text{for } \frac{x^2}{a^2} + \frac{(ky^2)}{b^2} = \frac{x^2+y^2}{a^2} = 1\right)$ and vice versa. This means that the ellipse (2) is obtained from the circle (3) by means of the transformation $(x, y) \mapsto (x, ky)$ geometrically representing a *compression* of the plane toward the abscissa axis in the ratio k. This not only gives a quite satisfactory idea of the shape of the ellipse, but also proves (since $k < 1$) that all points of the ellipse (2), except the points $\pm(a, 0)$, are inside the circle (3).

When $b = a$ any straight line passing through the point O (0, 0) is an exis of symmetry of the ellipse. Since equation (2) contains only the squares of coordinates, the coordinate axes are axes of symmetry of the ellipse for any a and b. Being the point of intersection of the axes of symmetry, the point O (0, 0) is the centre of symmetry of the ellipse.

It turns out that when $b < a$ the ellipse has no other axes of symmetry. Indeed, any axis of symmetry of the ellipse passes through its centre of symmetry $O(0, 0)$ and hence is the axis of symmetry of the circle (3). Therefore symmetry in this axis must preserve the intersection of the circle (3) and the ellipse (2) consisting, as we have determined, of two points $(\pm a, 0)$. Hence this symmetry either leaves both points $(a, 0)$ and $(-a, 0)$ where they are or interchanges them. In the former case the abscissa axis of the system of canonical coordinates is the axis of symmetry and in the latter case its ordinate axis.

This proves that when $b < a$ the axes of the system of canonical coordinates are uniquely characterized by the ellipse, i.e. up to a sign the canonical coordinates are unique.

Therefore when $b < a$ all the objects determined with the aid of canonical coordinates but independent of the orientation of the coordinates of the axes (remaining unchanged when arbitrary changes are made of the signs of the coordinates) are invariantly related to the ellipse.

Among them are:

the number a called the *major semiaxis*;

the number b called the *minor semiaxis*;

the number $c = \sqrt{a^2 - b^2}$ called the *linear eccentricity*;

the number $2c$ called the *focal distance*;

the number $e = \frac{c}{a} = \sqrt{1 - \frac{b^2}{a^2}}$ called the (numerical) *eccentricity* (it is obvious that $0 \leqslant e < 1$);

the number $p = \frac{b^2}{a}$ called the *focal parameter* (or simply *parameter*);

the abscissa axis called the *major* (or *focal*) *axis*;

the ordinate axis called the *minor axis*;

the point $O(0, 0)$ called the *centre*;

the points $(\pm a, 0)$ and $(0, \pm b)$ called *vertices*;

the points $(\pm c, 0)$ called *foci*;

when $e \neq 0$ the straight lines $x = \pm \frac{a}{e}$ called *directrices*.

The focus $(c, 0)$ and directrix $x = \frac{a}{e}$ are called *right-hand* and the focus $(-c, 0)$ and directrix $x = -\frac{a}{e}$ are called *left-hand*. A focus and directrix are said to be *on the same* *side* if they are both right-hand or left-hand. It is clear that this relation between a focus and a directrix is geometrically invariant, whereas their property of being right- or left-hand depends on the orientation of the abscissa axis.

For a circle $b = a$, $c = 0$, $e = 0$, $p = a$, the foci coincide with the centre and the directrices are indeterminate.

A segment connecting a point $M(x, y)$ of an ellipse to a focus is called the *focal radius* of that point. There are two focal radii, the *right-hand* and the *left-hand* radius.

For a length r_1 of the left-hand focal radius we have the formula

$$r_1^2 = (x+c)^2 + y^2 = (x+c)^2 + b^2\left(1 - \frac{x^2}{a^2}\right) =$$
$$= \left(1 - \frac{b^2}{a^2}\right) x^2 + 2xc + c^2 + b^2 = \frac{c^2}{a^2} x^2 + 2xc + a^2 =$$
$$= e^2x^2 + 2xea + a^2 = (ex + a)^2.$$

Since $|x| \leqslant a$ and hence $|ex| < a$, it follows that

$$r_1 = a + ex.$$

It is proved in a similar way that for a length r_2 of the right-hand focal radius we have

$$r_2 = a - ex.$$

Hence

$$r_1 + r_2 = 2a.$$

Conversely, let $M(x, y)$ be a point of the plane such that the sum of its distances from the foci of the ellipse is equal to $2a$:

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a.$$

Isolating one root, squaring and reducing the similar terms, again isolating the root and squaring we obtain after obvious transformations equation (2). This proves that the *ellipse* (2) *is a locus of points the sum of whose distances from the foci is equal to* $2a$.

This property of the ellipse is called its *focal property*.

The distance from a point $M(x, y)$ of the ellipse (2) to the left-hand directrix $x = -\frac{a}{e}$ is equal to

$$\left|x+\frac{a}{e}\right|=\frac{|ex+a|}{e}=\frac{r_1}{e}$$

and that to the right-hand directrix is equal to

$$\left|x-\frac{a}{e}\right|=\frac{|ex-a|}{e}=\frac{r_2}{e}.$$

Conversely, if

$$\sqrt{(x\pm c)^2+y^2}-e\left|x\pm\frac{a}{e}\right|,$$

then

$$(x\pm c)^2+y^2=(ex\pm a)^2,$$

and so

$$(1-e^2)\,x^2+y^2=a^2-c^2,$$

which is obviously equivalent to equation (2). This proves that the *ellipse* (2) *is a locus of points the ratio of whose distances from the focus to the directrix on the same side is equal to e.*

This property of the ellipse is called its *directorial property*. It is quite similar to the corresponding property of the parabola into which it turns when $e = 1$. For this reason it is convenient to regard the parabola as a kind of ellipse with the eccentricity $e = 1$.

Definition 3. A curve in a Euclidean plane is said to be a *hyperbola* if there exists a system of rectangular coordinates x, y in which the equation of that curve has the form

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1,\ \text{where } a>0,\ b>0. \tag{4}$$

The coordinates in which the equation of the hyperbola has the form (4) are called *canonical* (for this hyperbola) and equation (4) itself is called the *canonical equation* of the hyperbola.

When $b = a$ the hyperbola is called an *equilateral* hyperbola. In the coordinates

$$u=\frac{\sqrt{2}}{2}(x-y),\quad v=\frac{\sqrt{2}}{2}(x+y)$$

(which are obviously also rectangular) its equation

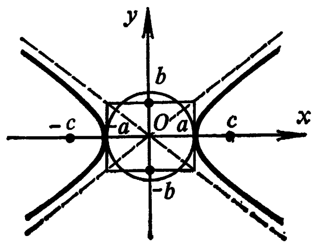
(5) $$x^2 - y^2 = a^2$$

takes the form

$$uv = 2a^2,$$

from which it follows that in reference to the coordinates u and v the equilateral hyperbola is the *graph of inverse proportionality* known from school. In the coordinates x and y we obtain therefore the same graph turned, however, by $\pi/4$.

As $u \to \pm\infty$ (and also as $v \to \pm\infty$) the graph of inverse proportionality approaches more and more closely the abscissa axis $v = 0$ (respectively the ordinate axis $u = 0$), that is to say, has these axes as its (two-sided) *asymptotes*. In the canonical coordinates x, y these asymptotes are the bisectrices $y = x$ and $y = -x$ of the coordinate angles.



A hyperbola

To go over from the equilateral hyperbola (5) to an arbitrary hyperbola (4) it is sufficient to perform a compression $(x, y) \mapsto (x, ky)$ toward the abscissa axis with the coefficient $k = \frac{b}{a}$ (note that, in contrast to the case of the ellipse, this coefficient may well be greater than one, so that our "compression" may in fact be an expansion). This gives a quite satisfactory idea of the shape of the hyperbola.

In particular we see that the hyperbola consists of two connected parts obtained respectively for $x > a$ and for $x < -a$ and has two asymptotes with the equations $y = \frac{b}{a} x$ and $y = -\frac{b}{a} x$, being situated as it is in two vertical angles formed by them.

These parts are called the *branches* of the hyperbola, the *left-hand* and the *right-hand* branch respectively.

Since equation (4) contains only the squares of coordinates, the coordinate axes are the axes of symmetry of the hyperbola and the point O (0, 0) is its centre of symmetry.

We show that the hyperbola has no other axes of symmetry (not excepting the case $b = a$). Indeed, any axis of symmetry of the hyperbola passes through its centre of symmetry and so is the axis of symmetry of the circle

$$x^2 + y^2 = a^2.$$

But it follows directly from what was said above that the hyperbola (4) intersects that circle in two points $(\pm a, 0)$. Therefore the symmetry considered either leaves each of these points where it is (and hence is a symmetry about the axis of abscissae) or interchanges the points (and hence is a symmetry about the axis of ordinates).

This proves that the axes of the system of canonical coordinates are uniquely determined by the hyperbola, i.e. that up to a sign the canonical coordinates are unique. Therefore all the objects determined with the aid of the canonical coordinates but remaining unchanged when their signs are changed are invariantly related to the hyperbola. Among them are:

the number a called the *real semiaxis*;
the number b called the *imaginary semiaxis*;
the number $c = \sqrt{a^2 + b^2}$ called the *linear eccentricity*;
the number $2c$ called the *focal distance*;
the number $e = \frac{c}{a} = \sqrt{1 + \frac{b^2}{a^2}}$ called (numerical) *eccentricity* (it is obvious that $1 < e < \infty$);
the number $p = \frac{b^2}{a}$ called the *focal parameter*;
the abscissa axis called the *real* (or *focal*) axis;
the ordinate axis called the *imaginary axis;*
the point $O\,(0, 0)$ called the *centre*;
the points $(\pm a, 0)$ called *vertices*;
the points $(\pm c, 0)$ called *foci*;
the straight lines $x = \pm \frac{a}{e}$ called *directrices*.

Left- and right-hand foci, directrices and focal radii, as well as foci, directrices and focal radii on the same side, are defined for the hyperbola in the same way as they are for the ellipse. The formulas

$$r_1^2 = (ex + a)^2, \quad r_2^2 = (ex - a)^2$$

for the squares of the lengths of focal radii also remain (together with the proof). But now taking the roots should be carried out with the caution, since for the hyperbola $|ex| > |x| \geqslant a$ and so

$$r_1 = \begin{cases} a + ex \text{ when } x > 0, \\ -a - ex \text{ when } x < 0 \end{cases}$$

and

$$r_2 = \begin{cases} -a + ex \text{ when } x > 0, \\ a - ex \text{ when } x < 0. \end{cases}$$

Hence

$$r_1 - r_2 = \begin{cases} 2a \text{ when } x > 0, \\ -2a \text{ when } x < 0, \end{cases}$$

i.e. for all x

$$|r_1 - r_2| = 2a.$$

Conversely, if the absolute value of the difference between the distances of some point $M(x, y)$ from the foci of the hyperbola is equal to $2a$, i.e. if

$$\left| \sqrt{(x-c)^2 + y^2} - \sqrt{(x+c)^2 + y^2} \right| = 2a,$$

then, after practically the same transformations as in the case of the ellipse, we obtain for x and y equation (4). This proves that the *hyperbola* (4) *is a locus of points the absolute value of the difference between whose distances from the foci is equal to* $2a$ (*the focal property of the hyperbola*). □

From the formulas for r_1 and r_2, just as for the ellipse, we can derive word for word the *directorial property of the hyperbola*, i.e. that the *hyperbola is a locus of points the ratio of whose distances from the focus to the directrix on the same side is equal to* e.

The ellipse, parabola, and hyperbola can thus be obtained by the same "directrix-focus" construction. All the difference will lie in the value of the eccentricity e.

Lecture 18

The equations of ellipses, parabolas and hyperbolas referred to a vertex • Polar coordinates • The equations of ellipses, parabolas and hyperbolas in polar coordinates • Affine ellipses, parabolas, hyperbolas • Algebraic curves • Second-degree curves and associated difficulties • Complex affine geometry and its insufficiency

The unity of ellipses, parabolas and hyperbolas can be made clear from a different viewpoint.

Consider for an arbitrary hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad a > 0, \quad b > 0$$

its equation in the coordinates

$$x' = x - a, \qquad y' = y$$

resulting from the translation of the origin of the canonical coordinates to the right-hand vertex of the hyperbola. The equation is of the form

$$\frac{(x'+a)^2}{a^2} - \frac{y'^2}{b^2} = 1,$$

i.e. of the form (we omit the primes)

(1) $$y^2 = 2px + qx^2,$$

where $q = \frac{b^2}{a^2} = e^2 - 1 > 0$ $\left(\text{and } p = \frac{b^2}{a} \text{ as before}\right)$.

It is easy to see that a similar translation of the origin to the left-hand vertex $(-a, 0)$ of the ellipse will reduce its equation to the same form (1), with q expressed by the formula

$$q=-\frac{b^2}{a^2}=e^2-1$$

and hence satisfying the inequalities $-1 \leqslant q < 0$.

Finally, for $q = 0$ we obtain from (1) the canonical equation of a parabola.

Thus *all hyperbolas, parabolas and ellipses can be given by the same equation* (1), where $p > 0$ and $q > 0$ for hyperbolas, $q = 0$ for parabolas and $-1 \leqslant q < 0$ for ellipses (for $q = -1$ we obtain a circle).

If we change q in equation (1), leaving p unchanged, then with q very large, we obtain a wide "open" hyperbola pressed to the ordinate axis. As q decreases, the branches of the hyperbola gradually close, the left-hand branch moving farther and farther to the left $\left(\text{the abscissa of the left-hand vertex is equal to } -\frac{2p}{q}\right)$ until at last, when $q = 0$, the left-hand branch vanishes ("goes into infinity") and the right-hand branch becomes a parabola. As q continues to decrease the parabola closes to form a strongly elongated ellipse $\left(\text{with the right-hand vertex at the point with the abscissa } -\frac{2p}{q}\right)$ which, gradually rounding, becomes a circle when $q = -1$.

It is interesting to see what happens if we take $q < -1$ in equation (1). To this end we introduce new coordinates

$$x'=y, \quad y'=x+\frac{p}{q}.$$

In these coordinates curve (1) has the equation (we again discard the primes)

$$x^2=2p\left(y-\frac{p}{q}\right)+q\left(y-\frac{p}{q}\right)^2,$$

i.e. the equation

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,$$

where $a^2 = -\frac{p^2}{q}$, $b^2 = \frac{p^2}{q^2}$. Since $b < a$, this is the canonical equation of an ellipse. But now the ellipse is elongated in the vertical direction (along the old axis of ordinates). Its focal parameter is equal to $\frac{p}{\sqrt{-q}}$ and its eccentricity is equal to $\sqrt{1+q^{-1}}$.

We thus see that as q changes from -1 to $-\infty$ the circle changes its size gradually elongating in the vertical direction and at the same time pressing itself to the (old) axis of ordinates. As $q \to -\infty$ a point is obtained "in the limit".

Note that as q changes from -1 to $-\infty$ so does the focal parameter which when $q < -1$ is equal not to p but to $\frac{p}{\sqrt{-q}}$. Therefore no parabola is obtained as $q \to -\infty$ (although $e \to 1$), since the focal parameter tends to zero.

According to Kepler's first law, the planets move about the Sun in ellipses (and the comets do in parabolas and hyperbolas) at one focus of which the Sun is situated. It is therefore convenient to give ellipses, parabolas and hyperbolas in the so-called *polar coordinates* in astronomic applications. We shall first describe these coordinates.

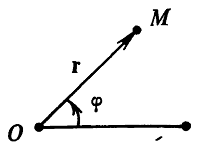
Polar coordinates are given by

(a) some (usually "counterclockwise") orientation of the plane;

(b) some point O called the *pole*;

(c) some oriented straight line passing through the point O and called the *polar axis* (the polar axis is usually chosen to be horizontal and orientated "from left to right").

The orientation of the plane and that of the polar axis allow us to uniquely associate with every vector $\mathbf{r} \neq 0$ the so-called "oriented angle" φ from the axis to the vector which takes values from $-\pi$ (exclusively) to π (inclusively). If $\mathbf{r}$ is directed along the axis, then $\varphi = 0$; if $\mathbf{r}$ has the opposite direction, then $\varphi = \pi$; if $\mathbf{r}$ is directed toward the positive side of the polar axis ("upwards"), then $0 < \varphi < \pi$, otherwise $-\pi < \varphi < 0$.



Together with the length r of the vector $\mathbf{r}$ the angle φ uniquely determines the vector $\mathbf{r}$. The numbers r and φ

for the vector $\mathbf{r} = \overrightarrow{OM}$ are exactly the *polar coordinates* of the point M (the number r being the *polar radius* and the angle φ the *polar angle*). For $M = 0$ the angle φ is assumed to have any value (and $r = 0$).

If x and y are rectangular coordinates *concordant* with the polar coordinates r, φ, i.e.

(a) defining the given orientation of the plane;
(b) having the point O as the origin;
(c) having the polar axis as abscissa axis, then

$$x = r\cos\varphi, \quad y = r\sin\varphi$$

and

$$r = \sqrt{x^2+y^2},$$

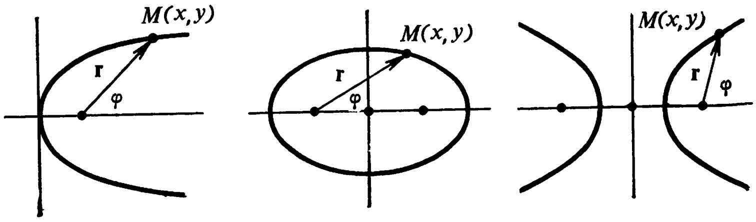
$$\cos\varphi = \frac{x}{\sqrt{x^2+y^2}}, \quad \sin\varphi = \frac{y}{\sqrt{x^2+y^2}}$$

(it is clear that the last two equations uniquely define the angle φ).

Now let

(2) $$y^2 = 2px$$

be an arbitrary parabola. According to the directorial property of the parabola the distance r from its arbitrary point



to the focus is equal to $x + \frac{p}{2}$. But this distance is the polar radius of the polar coordinate system concordant with the rectangular coordinates $x - \frac{p}{2}$, y obtained from the canonical coordinates by the translation of the origin to the focus. Therefore

$$x - \frac{p}{2} = r\cos\varphi$$

which together with the equation $r = x + \frac{p}{2}$ gives the relation

$$r - p = r \cos \varphi,$$

i.e. the relation

$$r = \frac{p}{1 - \cos \varphi}. \tag{3}$$

Conversely, if (3) holds and $x - \frac{p}{2} = r \cos \varphi$, $y = r \sin \varphi$, then

$$x + \frac{p}{2} = r \cos \varphi + p = r$$

and hence (by the directorial property) the point $M(x, y)$ is in the parabola (2). Thus (3) is the equation of the parabola (2) in the polar coordinates r, φ.

For the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad a \geqslant b > 0 \tag{4}$$

we place the pole of polar coordinates at the left-hand focus and direct the polar axis as before along the abscissa axis. Then the rectangular coordinates concordant with r and φ are the coordinates $x + c$ and y, so that in particular we have the equation $x + c = r \cos \varphi$. On the other hand, the polar radius is the left-hand focal radius $r_1 = a + ex$. Therefore

$$r = a + e(-c + r \cos \varphi) = p + er \cos \varphi$$

(since $a - ec = p$), i.e.

$$r = \frac{p}{1 - e \cos \varphi}. \tag{5}$$

Conversely, if (5) holds and $x + c = r \cos \varphi$, $y = r \sin \varphi$ then

$$x + \frac{a}{e} = -c + r \cos \varphi + \frac{a}{e} = \frac{p + er \cos \varphi}{e} = \frac{r}{e}$$

and therefore (by the directorial property of the ellipse) the point $M(x, y)$ is in the ellipse (4). Thus (5) is the equation of the ellipse (4) in the polar coordinates r, φ.

Finally, for the hyperbola

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1, \quad a>0, \quad b>0 \tag{6}$$

we place the pole at the right-hand focus $(c, 0)$ and restrict ourselves only to the right-hand branch $(x>0)$. Then we have the equations

$$r = ex - a$$

$$x - c = r \cos \varphi$$

from which it follows that

$$r=\frac{p}{1-e\cos\varphi}. \tag{7}$$

Conversely, by the directorial property of the hyperbola (7) gives (6) (for $x = c + r \cos \varphi > 0$ and $y = r \sin \varphi$). Thus (7) is the equation of the right-hand branch of the hyperbola (6) in the polar coordinates r, φ.

We see that the ellipse, parabola and hyperbola (more precisely, one of its branches) *may be given in polar coordinates by the same equation*

$$r=\frac{p}{1-e\cos\varphi}, \tag{8}$$

which again emphasizes the unity of all these curves.

For the parabola the pole of the coordinates r, φ is located at its only focus for a branch of the hyperbola it is at the focus on the same side, for the ellipse it is at any of its foci. In each of the three cases the polar axis is the focal axis of the curve considered, and the orientation of the polar axis is the canonical orientation for the parabola, and the orientation in which the given focus follows the vertex on the same side, for a branch of the hyperbola and for the ellipse. The orientation of the plane is of no consequence. For the parabola and hyperbola we thus have two systems of polar coordinates (differing in the orientation of the plane, i.e. in the sign of the angle φ) in which equation (8) holds, whereas for the ellipse we have a total of four systems of polar coordinates differing, in addition, in the choice of the focus to be the pole). That is because the ellipse has two axes of symmetry and the parabola and the branch of the hyperbola have one.

We now go over from the Euclidean to the affine plane, i.e. return again to affine geometry.

Recall that any Euclidean plane is affine and that there are many different ways to turn an affine plane into a Euclidean one, i.e., as we say, to introduce a Euclidean structure in it. In particular, it is possible to have any preassigned affine coordinate system turn into a Euclidean system of rectangular coordinates.

Definition 1. A curve in an affine plane is said to be an *affine ellipse* if it is possible to introduce into the plane a Euclidean structure with respect to which the given curve will be an ellipse in the sense of Definition 2 of Lecture 17. Similarly defined are the *affine parabola* and the *affine hyperbola*.

A curve in an affine plane is thus an ellipse if there exist coordinates x, y such that the equation of the curve is of the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$$

But then in the coordinates

$$x'=\frac{x}{a},\quad y'=\frac{y}{b}$$

the equation of that curve takes the form (we omit the primes)

$$x^2+y^2=1. \tag{9}$$

This proves that *for any affine ellipse there exists a system of affine coordinates x, y in which its equation has the form* (9). We shall call this equation and the corresponding coordinates the *canonical affine equation* and *canonical affine coordinates.*

Introducing into an affine plane a Euclidean structure in which canonical affine coordinates are rectangular we obtain a circle of radius 1 (a "unit circle"). Thus a curve in an affine plane is an ellipse if and only if it is possible to introduce into the plane a Euclidean structure with respect to which that curve is a unit circle.

Similarly, for any hyperbola there exist affine coordinates in which it is a "unit hyperbola", i.e. has the equation

$$x^2-y^2=1$$

and for any parabola there exist affine coordinates in which it is a "unit parabola"

$$y^2 = 2x.$$

These equations and the corresponding coordinates are called canonical affine equations and canonical affine coordinates.

Recall that a *polynomial* $F(x, y)$ in two variables x and y is a sum of *monomials* of the form ax^py^q, where a is some number. If $a \neq 0$, then the monomial ax^py^q is said to be *contained* in the given polynomial. The *degree of the monomial* ax^py^q is the number $p + q$ and the *degree of the polynomial* $F(x, y)$ is the highest degree of the monomials it contains.

It is clear that any reversible linear change of variables, i.e. any change of the form

$$x' = c_{11}x + c_{12}y + c_{13},$$
$$y' = c_{21}x + c_{22}y + c_{23},$$

where

$$\begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} \neq 0$$

transforms the polynomial in x, y into a polynomial in x', y' *of the same degree* (the degree cannot grow, nor can it decrease, by virtue of reversibility).

Definition 2. An *algebraic curve of degree n* in an affine (or Euclidean) plane is a set of points whose coordinates (x, y) in some (and hence any) affine coordinate system satisfy a relation of the form

$$F(x, y) = 0, \tag{10}$$

where $F(x, y)$ is a polynomial of degree n. This relation is called the *equation of an algebraic curve*.

A certain awkwardness of this definition is accounted for by the fact that for reasons indicated on pp. 166-167 we avoided using the terms "curve" and "equation of a curve" in it.

It should be stressed that there arise substantial difficulties in connection with Definition 2. For example, it is quite possible that the same curve should have two nonproportional equations, even possibly of different degrees, in the same coordinate system. Therefore one may speak of a degree of an algebraic curve only if its equation (10) has been chosen.

First-degree curves are given by equations of the form

$$Ax + By + C = 0$$

where at least one of the coefficients A and B is non-zero. As we know (see Lecture 6), these are all exactly straight lines and only straight lines.

Second-degree curves must have an equation of the form

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0 \tag{11}$$

where at least one of the coefficients a_{11}, a_{12}, a_{22} is nonzero.

Examples of second-degree curves.

[1] A curve with the equation

$$x^2 + y^2 - 1 = 0.$$

This is the *ellipse* as we know it.

[2] A curve with the equation

$$x^2 + y^2 + 1 = 0.$$

This curve is called an *imaginary ellipse.*

[3] A curve with the equation

$$x^2 + y^2 = 0.$$

This is a *pair of imaginary intersecting straight lines.*

[4] A curve with the equation

$$x^2 - y^2 - 1 = 0.$$

This is a *hyperbola.*

[5] A curve with the equation

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

This is a *pair of intersecting straight lines.*

[6] A curve with the equation

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

This is a *parabola.*

[7] A curve with the equation

$$y^2 - 1 = 0.$$

This is a *pair of distinct parallel lines.*

[8] A curve with the equation

$$y^2 + 1 = 0.$$

This is a *pair of imaginary parallel lines.*

[9] A curve with the equation

$$y^2 = 0.$$

This is a *pair of coincident straight lines.*

Definition 3. The nine enumerated equations are called *canonical affine equations* of second-degree curves.

For equations [1], [4], [5], [6], [7] the names of the corresponding curves are quite clear (although one may doubt whether a pair of intersecting straight lines, for example, corresponds to the intuitive notion of a curve). But the name of curve [9] is rather relative and serves only to remind that its equation is of degree 2; in fact (as a set of points) this curve is a straight line. Just as much relative are the names of the other "curves" which only serve to show that their equations are similar to those of the corresponding "real" curves. Curve [2] is an empty set, curve [3] consists of a single point, and curve [8] is again an empty set.

We thus see that for second-degree curves algebraic consideration is not adequate to geometrical one; when going from equations to curves algebraic analogues are lost and the general picture is slurred (which manifests itself, say, in different equations giving, as the corresponding curve, the same empty set).

One would like therefore to perfect "geometry" so that it should exactly reflect "algebra".

To do this it is necessary, of course, to introduce "imaginary" points with complex (generally speaking, nonreal) coordinates. On the face of it it seems that we have at our disposal all the necessary concepts for this purpose. Indeed, as we have repeatedly noted, the concept of a vector space has meaning over any field $\mathbb{K}$ and in particular over the field of complex numbers, $\mathbb{C}$. A ground field figures in axioms 10°, 11° of the affine space only indirectly, and to obtain the concept of an *affine space over a field* $\mathbb{K}$ (and in particular the concept of a *complex affine space*) one should suppose the associated vector space $\mathscr{V}$ in these axioms to be a vector space over the field $\mathbb{K}$ (respectively, over the field $\mathbb{C}$). Affine theory as a whole (except for the questions connected with orientation) remains quite valid for the case of any field $\mathbb{K}$.' Euclidean geometry is not transferred to this case (at least not directly), because axiom 15° of positivity has meaning only over the field $\mathbb{R}$.

However, going from the theory of second-degree curves to complex affine geometry has its disadvantages. Consider, for example, curves [1] and [2]. They are different both geometrically and algebraically. Nevertheless a complex change of coordinates $x' = x$, $y' = iy$ transforms one of them into the other.

The thing is that within the framework of complex geometry there is no room for a "real" point: all the points in it are absolutely equivalent. But if we are to keep contact with geometry, we must be able to distinguish between ordinary ("real") and new ("imaginary") points which we are forced to add in view of an algebraic necessity. Therefore, although coordinates may now take any complex values, formulas for the transition from one coordinate system to the other must have real coefficients (otherwise a point with real coordinates may become "imaginary", and vice versa). The corresponding axiomatic construction will be dealt with in the next lecture.

Lecture 19

Real-complex vector spaces • Their dimensionality • Isomorphism of real-complex vector spaces • Complexification • Real-complex affine spaces • The complexification of affine spaces • Real-complex Euclidean spaces • Real and imaginary curves of second degree

Definition 1. *A real-complex* vector space is a complex (i.e. over the field $\mathbb{C}$) vector space in which a subset $\mathscr{V}^{\mathrm{R}}$ is given whose elements are called *real vectors* and which has the following properties (we continue our general numbering of axioms):

16°. If $\mathbf{a}, \mathbf{b} \in \mathscr{V}^{\mathrm{R}}$, then $\mathbf{a} + \mathbf{b} \in \mathscr{V}^{\mathrm{R}}$.

17°. If $a \in \mathscr{V}^{\mathrm{R}}$, then $k\mathbf{a} \in \mathscr{V}^{\mathrm{R}}$ for any real number k.

18°. For any vector $\mathbf{c} \in \mathscr{V}$ there exist vectors $\mathbf{a}, \mathbf{b} \in \mathscr{V}^{\mathrm{R}}$ such that

$$\mathbf{c} = \mathbf{a} + i\mathbf{b}. \tag{1}$$

19°. If $\mathbf{a} = i\mathbf{b} = 0$, where $\mathbf{a}, \mathbf{b} \in \mathscr{V}^{\mathrm{R}}$, then $\mathbf{a} = 0$ and $\mathbf{b} = 0$.

It follows from axiom 19° that the decomposition (1) provided by axiom 18° is unique. The vector $\mathbf{a}$ involved in the decomposition is called the *real part* of the vector $\mathbf{c}$ and designated by the symbol $\operatorname{Re} \mathbf{c}$. Similarly, the vector $\mathbf{b}$ is called the *imaginary part* of the vector $\mathbf{c}$ and designated by $\operatorname{Im} \mathbf{c}$. Axiom 18° may loosely be written as the mnemonic formula

$$\mathscr{V} = \mathscr{V}^{\mathrm{R}} + i\mathscr{V}^{\mathrm{R}}.$$

Axioms 16° and 17° mean that the subset $\mathscr{V}^{\mathrm{R}}$ is a vector space over the field of real numbers $\mathbb{R}$. For this reason $\mathscr{V}^{\mathrm{R}}$ is called the *real subspace* of the space $\mathscr{V}$.

Let n be the dimension of the vector space $\mathscr{V}^{\mathbb{R}}$. To em phasize that the dimension of a vector space over $\mathbb{R}$ is involved we shall designate it by the symbol $\dim_{\mathbb{R}} \mathscr{V}^{\mathbb{R}}$.

In a similar way the dimension of a vector space $\mathscr{V}$ as a vector space over $\mathbb{C}$ will be designated by $\dim_{\mathbb{C}} \mathscr{V}$.

Proposition 1. *We have the equation*

$$\dim_{\mathbb{C}} \mathscr{V} = \dim_{\mathbb{R}} \mathscr{V}^{\mathbb{R}}.$$

Any basis for the vector space $\mathscr{V}^{\mathbb{R}}$ is a basis for a vector space $\mathscr{V}$.

Proof. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a basis for the vector space $\mathscr{V}^{\mathbb{R}}$ and let $\mathbf{c} = \mathbf{a} + i\mathbf{b} \in \mathscr{V}$. Then, if $\mathbf{a} = a^1\mathbf{e}_1 + \ldots + a^n\mathbf{e}_n$ and $\mathbf{b} = b^1\mathbf{e}_1 + \ldots + b^n\mathbf{e}_n$, then $\mathbf{c} = (a^1 + ib^1)\mathbf{e}_1 + \ldots + (a^n + ib^n)\mathbf{e}_n$. Consequently, the family of vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is complete in $\mathscr{V}$.

On the other hand, if $c^1\mathbf{e}_1 + \ldots + c^n\mathbf{e}_n = \mathbf{0}$, where $c^1 = a^1 + ib^1, \ldots, c^n = a^n + ib^n$, then $(a^1\mathbf{e}_1 + \ldots + a^n\mathbf{e}_n) + i(b^1\mathbf{e}_1 + \ldots + b^n\mathbf{e}_n) = \mathbf{0}$ and by axiom 19° $a^1\mathbf{e}_1 + \ldots + a^n\mathbf{e}_n = \mathbf{0}$ and $b^1\mathbf{e}_1 + \ldots + b^n\mathbf{e}_n = 0$. Therefore $a^1 = 0, \ldots, a^n = 0$, $b^1 = 0$, $b^n = 0$ and hence $c^1 = 0, \ldots, c^n = 0$. Consequently, the family $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is linearly independent in $\mathscr{V}$. □

Bases in $\mathscr{V}$ that are bases in $\mathscr{V}^{\mathbb{R}}$ will be called *real* bases. They have the property that the coordinates of the vector $\mathbf{c} \in \mathscr{V}$ are real in them if and only if so is the vector $\mathbf{c}$.

As an example of a real-complex vector space we may cite a vector space $\mathbb{C}^n$ in which sequences consisting of real numbers are assumed to be real vectors. For this vector space

$$(\mathbb{C}^n)^{\mathbb{R}} = \mathbb{R}^n.$$

Definition 2. Two real-complex vector spaces $\mathscr{V}$ and $\mathscr{V}_1$ are said to be *isomorphic* if there exists their isomorphism $\mathscr{V} \to \mathscr{V}_1$ as vector spaces over $\mathbb{C}$ mapping the real subspace $\mathscr{V}^{\mathbb{R}}$ onto the real subspace $\mathscr{V}_1^{\mathbb{R}}$ (and therefore inducing the isomorphism $\mathscr{V}^{\mathbb{R}} \to \mathscr{V}_1{}^{\mathbb{R}}$).

According to what has been said above each real basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ determines the isomorphism of a real-complex vector space $\mathscr{V}$ onto a real-complex vector space $\mathbb{C}^n$. Thus, as was to be expected, *all real-complex vector spaces of the same dimension are isomorphic.* □

Any real-complex vector space $\mathscr{V}$ may be thought of as simply a vector space over $\mathbb{C}$, whether its vectors are real or not. The operation of transition from a real-complex to a complex vector space will be called the *ignoring of reality*.

Conversely, we can "introduce reality" in every complex vector space by choosing an arbitrary basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and declaring real the vectors having real coordinates in that basis. It is clear that axioms 16° to 19° will hold here.

It is obvious that this construction applied to two different bases gives the same real-complex space if and only if the matrix of transition from one basis to the other is real (consists of real numbers).

It is more interesting that *for any real vector space* $\mathscr{W}$ *there exists a real-complex space* $\mathscr{V}$ *whose real subspace* $\mathscr{V}^{\mathbb{R}}$ *coincides with the space* $\mathscr{W}$ (or, more exactly, is isomorphic to it in a natural way). The construction of this space $\mathscr{V}$ completely parallels the well-known construction of the field of complex numbers (and turns into it when $\mathscr{W} = \mathbb{R}$).

We take as the vectors of the space $\mathscr{V}$ pairs of the form $(\mathbf{a}, \mathbf{b})$, where $\mathbf{a}, \mathbf{b} \in \mathscr{W}$. We define the addition of such pairs by the formula

$$(\mathbf{a}, \mathbf{b}) + (\mathbf{a}_1, \mathbf{b}_1) = (\mathbf{a} + \mathbf{a}_1, \mathbf{b} + \mathbf{b}_1),$$

their multiplication by complex numbers being defined by the formula

$$(k + il)(\mathbf{a}, \mathbf{b}) = (k\mathbf{a} - l\mathbf{b}, k\mathbf{b} + l\mathbf{a}).$$

An automatic check shows that as a result we obtain the vector space $\mathscr{V}$ over the field $\mathbb{C}$.

To turn it into a real-complex vector space, it is enough to declare pairs of the form $(\mathbf{a}, \mathbf{0})$ to be real vectors. (Axioms 16° to 19° are checked automatically; for example, axiom 17° follows from the fact that $k\mathbf{b} + l\mathbf{a} = \mathbf{0}$ for $\mathbf{b} = \mathbf{0}$ and $l = 0$ and axiom 18° follows from the fact that $(\mathbf{0}, \mathbf{b}) = i(\mathbf{b}, \mathbf{0})$. On identifying each pair $(\mathbf{a}, \mathbf{0})$ with the corresponding vector $\mathbf{a} \in \mathscr{W}$ we find that $\mathscr{W} = \mathscr{V}^{\mathbb{R}}$. $\square$

The constructed space $\mathscr{V}$ is called the *complexification* of the space $\mathscr{W}$. It is designated by the symbol $\mathscr{W}^{\mathbb{C}}$.

We now pass on to spaces of points.

Definition 3. A *real-complex affine space is a complex affine space* $\mathscr{A}$ in which

(a) some subset $\mathscr{A}^{\mathrm{R}}$ is fixed whose elements are called *real points*;

(b) the associated vector space $\mathscr{V}$ is a real-complex space, with the following axioms holding.

20°. For any two real points A, $B \in \mathscr{A}^{\mathrm{R}}$ the vector $\overrightarrow{AB}$ is real:

$$A,\ B \in \mathscr{A}^{\mathrm{R}} \Rightarrow \overrightarrow{AB} \in \mathscr{V}^{\mathrm{R}}.$$

21°. If a point A and a vector $\overrightarrow{AB}$ are real, then the point B is also real:

$$A \in \mathscr{A}^{\mathrm{R}},\ \overrightarrow{AB} \in \mathscr{V}^{\mathrm{R}} \Rightarrow B \in \mathscr{A}^{\mathrm{R}}.$$

It follows from axiom 21° (and axiom 17°) that if a point B and a vector $\overrightarrow{AB}$ are real, then the point A is also real (it is enough to go over to the vector $\overrightarrow{BA} = -\overrightarrow{AB}$).

Axioms 20° and 21° mean that the subset $\mathscr{A}^{\mathrm{R}}$ is a real affine space with an associated vector space $\mathscr{V}^{\mathrm{R}}$. It is called the *real subspace* of the real-complex space $\mathscr{A}$. Its ("real") dimensionality is equal to the ("complex") dimensionality of the space $\mathscr{A}$.

An affine coordinate system $O\mathbf{e}_1 \ldots \mathbf{e}_n$ of the space $\mathscr{A}$ is said to be *real* if the point O and the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are real. Such coordinate systems are characterized by the coordinates of a point of $\mathscr{A}$ being real in them if and only if that point is real.

Two points of a real-complex space $\mathscr{A}$ are said to be *complex conjugate* if they have complex conjugate coordinates in some (and hence any) real coordinate system.

The isomorphism of real-complex affine spaces is defined in an obvious way. As was expected, *all-real-complex affine spaces of the same dimension are isomorphic.*

Any real-complex affine space may be thought of, "ignoring reality", as simply a complex affine space. Conversely, any complex affine space can be turned into a real-complex space by choosing an arbitrary affine coordinate system $O\mathbf{e}_1 \ldots \mathbf{e}_n$ and declaring the points and vectors to be real if they have real coordinates in that system. Two coordinate systems lead to the same result if and only if the formulas for the transition from one system to the other have real coefficients.

It is proved essentially in the same way as for vector spaces that *for any real affine space $\mathscr{B}$ there is a real-complex affine space $\mathscr{A} = \mathscr{B}^{\mathbb{C}}$* (called the *complexification* of the space $\mathscr{B}$) *such that* $\mathscr{A}^{\mathrm{R}} = \mathscr{B}$. That is, we shall assume pairs (A_1, A_2) of points of the space $\mathscr{B}$ to be points of the space $\mathscr{A}$ and, taking the complexification $\mathscr{W}^{\mathbb{C}}$ of a vector space $\mathscr{W}$ associated with the affine space $\mathscr{B}$ as an associated vector space $\mathscr{V}$, define the vector $\overrightarrow{(A_1, A_2)(B_1, B_2)}$ by the formula

$$\overrightarrow{(A_1, A_2)(B_1, B_2)} = (\overrightarrow{A_1, B_1}, \overrightarrow{A_2, B_2}).$$

It is clear that axioms 10° and 11° of affine space will then hold. To introduce into the spaces thus obtained the structure of a real-complex space, we arbitrarily choose in $\mathscr{B}$ a point 0 and declare pairs of the form $(A, 0)$, where $A \in \mathscr{B}$, to be real points in $\mathscr{A}$. Then axioms 20° and 21° will also hold, and by virtue of the identity $(A, 0) = \mathscr{A}$ we shall have the equation $\mathscr{A}^{\mathrm{R}} = \mathscr{B}$. □

For completeness we introduce yet another real-complex space, one with a Euclidean structure.

Definition 4. A real-complex vector or affine space is said to be *Euclidean* if a Euclidean structure is introduced into its real subspace.

In such a space formulas defining lengths (distances) and angles may be also applied to nonreal objects. No good theory (one with the customary properties) results, however, outside the real domain (for example, the length of a nonzero vector may be zero). Nevertheless it is sometimes useful to go beyond this domain.

We now return to the case $n = 2$, i.e. consider a real-complex affine (or Euclidean) plane.

Definition 5. An algebraic curve in a real-complex plane is said to be *real* if it can be given in some (and hence any) real coordinate system by an equation whose left-hand side is a polynomial with real coefficients.

For example, a *real curve of second degree* must have an equation of the form (11), Lecture 18, in which all the coefficients a_{11}, a_{12}, a_{22}, a_{13}, a_{23}, a_{33} are real. The nine canonical affine equations [1] to [9] enumerated in the same lecture are all real.

The names of the corresponding curves (except curve [9], of course) now assume their full meaning. The epithet "imaginary" simply signifies that the curve in question does not intersect the real subspace (or, as curve [3], has only one point in common with it). "Nonimaginary" curves will be called *real curves*. Thus the term "real" will have double meaning*; the curves [1] to [9] are all real in the former sense, but only some of them are so in the new sense. Real curves in the former sense are also called curves *defined over* $\mathbb{R}$.

It should be stressed that, for example, real ellipses [1] and affine ellipses in the sense of Definition 1 of the preceding lecture strictly speaking represent different objects since the former have nonreal points and the latter (being by definition curves in the real plane) cannot have such points. The connection between them is that affine ellipses are intersections of real ellipses with the real plane. That is why while affine ellipses, for example, have obviously no asymptotes, real ellipses, as we shall see in due course, do have asymptotes (for the ellipse $x^2 + y^2 = 1$ these are imaginary straight lines $y = \pm ix$).

The situation is similar of course for the other real curves of the list [1] to [9] (for parabolas, hyperbolas, and pairs of real straight lines).

Our immediate aim will be to prove that any real curve of second degree in a real-complex plane is one (and only one) curve of the list [1] to [9]. To do this we shall have to extend the general theory of these curves sufficiently far in the next lectures.

---

* Two different words, "veshchestvenny" and "deistvitelny", are used in the original. *Translator's note.*

Lecture 20

Introductory remarks • The centre of a second-degree curve • Centres of symmetry • Central and noncentral curves of second degree • Straight lines of non-asymptotic direction • Tangents • Straight lines of asymptotic direction

The initial stages in the theory of second-degree curves are independent of the ground field $\mathbb{K}$ (it is only required that the characteristic of this field should be other than two). In more delicate questions it is necessary to assume this field to be the field of complex numbers, $\mathbb{C}$, i.e. to work in the complex affine plane. And in order not to lose contact entirely with intuitive notions this plane should be considered a real-complex plane. To distinguish between the last two cases the first will be referred to as *situation* $\mathbb{C}$, and the second as *situation* $(\mathbb{C}, \mathbb{R})$. Situation $(\mathbb{C}, \mathbb{R})$ will sometimes be restricted to the real plane alone (*situation* $\mathbb{R}$).

Throughout this lecture x and y denote arbitrary but fixed affine coordinates. In situation $(\mathbb{C}, \mathbb{R})$ it is naturally assumed that these coordinates are real, i.e. they are real numbers for real points.

All our statements refer to some fixed curve of second degree (unless otherwise indicated).

(1) $$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0.$$

In the general theory (over an arbitrary field $\mathbb{K}$ of a characteristic other than two) the coefficients of that equation may be arbitrary numbers of the field $\mathbb{K}$ but of course at least one of the coefficients a_{11}, a_{12}, a_{22} must be nonzero. In particular, in situation $\mathbb{C}$ they may be any complex numbers. But in situation $(\mathbb{C}, \mathbb{R})$ they will be assumed to be real.

For the symmetry of the formulas we shall also introduce the coefficients a_{21}, a_{31}, and a_{32}, considering by definition that

$$a_{21} = a_{12}, \quad a_{31} = a_{13}, \quad a_{32} = a_{23}.$$

We begin by investigating the change of the coefficients in equation (1) after the translation of the origin to an arbitrary point $M_0(x_0, y_0)$, i.e. after the transition to the coordinates

$$x' = x - x_0, \quad y' = y - y_0.$$

The ground field $\mathbb{K}$ is for the time being assumed to be an arbitrary field (of a characteristic other than two).

On denoting the left-hand side of equation (1) by $F(x, y)$ we find that in the coordinates x', y', the equation of the urve considered is of the form

$$F(x' + x_0, y' + y_0) = 0.$$

Let

$$F(x' + x_0, y' + y_0) = \\ = a'_{11}(x')^2 + 2a'_{12}x'y' + a'_{22}(y')^2 + 2a'_{13}x' + 2a'_{23}y' + a'_{33}.$$

A trivial computation shows that

$$\begin{aligned} &a'_{11} = a_{11}, \ a'_{12} = a_{12}, \ a'_{22} = a_{22}, \\ (2) \quad &a'_{13} = a_{11}x_0 + a_{12}y_0 + a_{13}, \quad a'_{23} = a_{21}x_0 + a_{22}y_0 + a_{23} \\ &a'_{33} = F(x_0, y_0). \end{aligned}$$

In particular we see that after the translation of the origin the coefficients of the second-order terms remain unchanged.

Definition 1. A point $M_0(x_0, y_0)$ is a *centre* of the curve (1) if

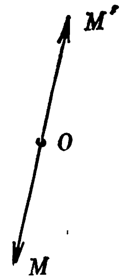
$$(3) \qquad \begin{aligned} a_{11}x_0 + a_{12}y_0 + a_{13} = 0, \\ a_{21}x_0 + a_{22}y_0 + a_{23} = 0, \end{aligned}$$

i.e. if after the translation of the origin to that point equation (1) loses the first-order terms.

This definition seems to be unsatisfactory since it is bound to the given equation (1). We have as yet no assurance that for a different equation of the same curve (in the same system of coordinates) we shall not get quite different equations (3), i.e. that Definition 1 will turn out to be correct

In fact these fears are groundless, but we cannot yet prove this in full generality.

The only case accessible to direct investigation arises when the curve (1) "does not lie on a straight line", i.e. there are at least three noncollinear points on it. It turns out that in this case the centres of the curve (1) may be characterized in a purely geometrical fashion, without any appeal to the equation of the curve. Hence for such curves Definition 1 is correct.



We begin with the remark that the concept of central symmetry (unlike, for example, that of axial symmetry) may be correctly defined in any affine space (over an arbitrary field). That is, we say that a point M' is obtained from a point M *by symmetry in a centre* O (or about O) if $\overrightarrow{OM'} = -\overrightarrow{OM}$. The point O is said to be the *centre of symmetry* of some set (for example, that of a curve) if together with each point M to that set belongs a symmetrical point M'. The origin O is the centre of symmetry of the curve $F(x, y) = 0$ if and only if the equation $F(x, y) = 0$ yields the equation $F(-x, -y) = 0$.

It follows immediately that *any centre of the curve* (1) *is its centre of symmetry* (this explains in particular the origin of the term "centre"). Indeed, if a centre of the curve (1) is at the origin, i.e. if there are no linear terms in equation (1), then we have the identity $F(-x, -y) = F(x, y)$ and so the above condition holds. □

Conversely, if the centre of symmetry of the curve (1) coincides with the origin, then for any point $M(x, y)$ of the curve (1) we have $F(-x, -y) = 0$, together with $F(x, y) = 0$, and therefore $F(x, y) - F(-x, -y) = 0$, i.e.

$$a_{13}x + a_{23}y = 0. \tag{4}$$

For $a_{13} \neq 0$ or $a_{23} \neq 0$ this means that the point $M(x, y)$ is on the straight line (4), so that the curve (1) lies entirely on that straight line. This proves that *if the curve* (1) *is not on any straight line, then any of its centres of symmetry is its centre.* □

Thus for a curve of the form (1) not lying on any straight line its centres are exactly the centres of symmetry. Therefore for such curves Definition 1 is correct.

In the general case it must be remembered, however, that this definition is formally dependent on the choice of equation (1).

In general relations (3) imply that the centres of the curve (1) are common points of two straight lines:

$$(5) \qquad a_{11}x + a_{12}y + a_{13} = 0 \text{ and } a_{21}x + a_{22}y + a_{23} = 0,$$

and so are either absent (the straight lines (5) are parallel) or there is only one centre (the straight lines (5) intersect) or centres fill a whole straight line (the straight lines (5) coincide). It may of course happen that one of the equations (5) is satisfied identically or that, conversely, the equation is incompatible, i.e. there is not a point in the plane that satisfies it (for the first equation this is the case when $a_{11} = 0$, $a_{12} = 0$, $a_{13} \neq 0$). Nothing new arises, however (if one of the equations (5) is satisfied identically, then there is a straight line of centres, and if it is incompatible, then there are no centres at all). The following three cases are thus possible for the curve (1).

I. There is only one centre.

II. There is no centre.

III. There is a whole straight line of centres.

Definition 2. In case I the curve (1) is called a second-degree *central curve* and in cases II and III it is called a *noncentral curve.*

According to the theorem of relative positions of two straight lines in the plane, proved in Lecture 6, case I holds when $\frac{a_{11}}{a_{21}} \neq \frac{a_{12}}{a_{22}}$, i.e. when $a_{11}a_{22} - a_{21}^2 \neq 0$; case II holds when $\frac{a_{11}}{a_{21}} = \frac{a_{12}}{a_{22}} \neq \frac{a_{13}}{a_{23}}$ and case III, when $\frac{a_{11}}{a_{21}} = \frac{a_{12}}{a_{22}} = \frac{a_{13}}{a_{23}}$. It is convenient to write these conditions in a somewhat different form by introducing into consideration the determinants

$$\delta=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad \text{and} \quad \Delta=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

called respectively the *minor* and the *major determinants* of equation (1).

Since $\delta = a_{11}a_{22} - a_{12}^2$, we see that $\delta \neq 0$ in case I and $\delta = 0$ in cases II and III. In addition, $\Delta = 0$ in case III (the first two rows of this determinant are proportional). Conversely, let $\Delta = 0$ and $\delta = 0$. On expanding the determinant Δ by the last column we find that

$$\Delta = a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} - a_{23}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} + a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

The last determinant is the zero determinant δ. In addition, it follows from the equation $\delta = 0$ that there is a number k such that $a_{11} = ka_{21}$, $a_{12} = ka_{22}$ and so

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} = k\begin{vmatrix} a_{12} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

This proves that for $\delta = 0$

$$\Delta = (a_{13} - ka_{23})\begin{vmatrix} a_{12} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Therefore, if $\Delta = 0$, then either $a_{13} = ka_{23}$ (and hence the first two rows of the determinant Δ are proportional) or $\begin{vmatrix} a_{12} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = 0$ and so there is a number l such that $a_{31} = la_{12}$, $a_{32} = la_{22}$ and hence $a_{31} = la_{12} = lka_{22} = ka_{32} = ka_{23}$, so that in this case too the first two rows of the determinant Δ are proportional. This proves that case III holds if and only if $\delta = 0$ and $\Delta = 0$. The remaining possibility $\delta = 0$ and $\Delta \neq 0$ characterizes therefore case II.

Summing up all that has been proved we obtain the following theorem.

Theorem 1 (on centres). *A second-degree curve is a central curve if and only if* $\delta \neq 0$.

A second-degree curve has no centre if and only if $\delta = 0$ *and* $\Delta \neq 0$.

A second-degree curve has a straight line of centres if and only if $\delta = 0$ and $\Delta = 0$. □

As an extension to this theorem it is of interest to clarify the geometrical meaning of the condition $\Delta = 0$. Since for $\delta = 0$ the answer is provided by Theorem 1 (the equation $\Delta = 0$ is equivalent to the existence of a straight line of centres), it is enough to restrict ourselves to the case $\delta \neq 0$.

Proposition 1. *For $\delta \neq 0$ the equation $\Delta = 0$ holds if and only if a centre of the curve* (1) *lies on the curve.*

Proof. Let a centre $M_0(x_0, y_0)$ of the curve (1) lie on it. Then for the coordinates of the centre we have equations (3) as well as equation (1) in which we suppose $x = x_0$ and $y = y_0$. On multiplying the first of the equations (3) by x_0 and the second by y_0 and subtracting them from (1) we obviously obtain the relation

$$a_{31}x_0 + a_{32}y_0 + a_{33} = 0.$$

Together with equations (3) it means that the three numbers x_0, y_0, 1 make up a nontrivial solution of the system of homogeneous equations

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= 0, \\ a_{21}x + a_{22}y + a_{23}z &= 0, \\ a_{31}x + a_{32}y + a_{33}z &= 0. \end{aligned}$$

Hence the determinant Δ of this system is zero. Note that the assumption $\delta \neq 0$ has not been used in our argument.

Conversely, if $\Delta = 0$, then the system above has a nontrivial solution x_0, y_0, z_0. Here $z_0 \neq 0$, for $\delta \neq 0$. Therefore we may assume without loss of generality that $z_0 = 1$. Then the first two equations will show that the point $M_0(x_0, y_0)$ is a centre of the curve (1). On the other hand, on multiplying the first of these equations (which assume that $x = x_0$, $y = y_0$ and $z = 1$) by x_0 and the second by y_0, and adding to the third equation, we obviously obtain equation (1) (which assumes $x = x_0$, $y = y_0$). Hence the centre $M_0(x_0, y_0)$ is on the curve (1). □

Examples. We find the centres of the curves [1] to [9].

[1] The real ellipse: $\delta = 1$ and $\Delta = -1$. There is one centre (0, 0).

[2] The imaginary ellipse: $\delta = 1$ and $\Delta = 1$. There is one centre (0, 0).

[3] The pair of imaginary intersecting straight lines: $\delta = 1$ and $\Delta = 0$. There is one centre (0, 0).

[4] The hyperbola: $\delta = -1$ and $\Delta = 1$. There is one centre (0, 0).

[5] The pair of real intersecting straight lines: $\delta = -1$ and $\Delta = 0$. There is one centre (0, 0).

[6] The parabola: $\delta = 0$ and $\Delta = -1$. There are no centres.

[7] The pair of real parallel lines: $\delta = 0$ and $\Delta = 0$. There is a straight line of centres $y = 0$.

[8] The pair of imaginary parallel lines: $\delta = 0$ and $\Delta = 0$. There is a straight line of centres $y = 0$.

[9] The pair of coincident straight lines: $\delta = 0$ and $\Delta = 0$. There is a straight line of centres $y = 0$.

We see that central curves (in the list [1] to [9]) are the two ellipses, the hyperbola and the pairs of intersecting straight lines (real or imaginary). In all the cases the centre is, as supposed, the centre of symmetry.

A noncentral curve without a centre is only the parabola.

Noncentral curves with a straight line of centres are the pairs of parallel lines (real or imaginary, distinct or coincident).

Note that for the curve [9], the only curve among the curves [1] to [9] that lies (in situations $\mathbb{C}$ and $(\mathbb{C}, \mathbb{R})$) on a straight line, the centres of symmetry and the centres nevertheless coincide.

We now study the relative positions of the curve (1) and an arbitrary straight line

$$\begin{aligned} x &= x_0 + tl, \\ y &= y_0 + tm. \end{aligned} \tag{6}$$

The points the curve (1) and the straight line (6) have in common are determined from the equation $F(x_0 + tl, y_0 + tm) = 0$, i.e. from the equation

$$\begin{aligned} &(a_{11}l^2 + 2a_{12}lm + a_{22}m^2)\, t^2 + 2\,(a_{11}lx_0 + a_{12}(ly_0 + mx_0) + \\ &\qquad + a_{22}my_0 + a_{13}l + a_{23}m)\, t + F(x_0, y_0) = 0. \end{aligned} \tag{7}$$

Definition 3. When $a_{11}l^2 + 2a_{12}lm + a_{22}m^2 \neq 0$ the straight line (6) is said to have *nonasymptotic direction* (with respect to the curve (1)) and when

(8) $$a_{11}l^2 + 2a_{12}lm + a_{22}m^2 = 0$$

this straight line is said to have *asymptotic direction.*

For a straight line of a nonasymptotic direction equation (7) is a second-degree equation. Therefore it has (in the ground field $\mathbb{K}$) either two roots or one (double) root or has no root. In the first case the straight line (6) intersects the curve (1) in two different points, in the second it has one point in common with the curve (1), and in the third case the straight line (6) and the curve (1) do not intersect.

In situations $\mathbb{C}$ and ($\mathbb{C}$, $\mathbb{R}$) the third case is impossible, and in situation ($\mathbb{C}$, $\mathbb{R}$) it is possible in the first case for the common points to be either real (then in situation $\mathbb{R}$ the first case will also hold) or complex conjugate, nonreal (in situation $\mathbb{R}$ the third case will then hold).

Definition 4. A straight line of nonasymptotic direction for which equation (7) has one (double) root is called a *tangent* to the curve (1) and its point corresponding to that root (the only point in common with the curve (1)) is called a *point of tangency.*

Example. Let the curve (1) be an ellipse $x^2 + y^2 = 1$ and let a point $M_0(x_0, y_0)$ of the straight line (6) lie in that ellipse. Equation (7) is in this case of the form

$$(l^2+m^2)\, t^2 + 2\,(lx_0 + my_0)\, t = 0$$

and has a double root if and only if $lx_0 + my_0 = 0$, i.e. if $l: m = -y_0: x_0$. Therefore the tangent to the ellipse $x^2 + y^2 = 1$ at its point $M_0(x_0, y_0)$ is the straight line

(9) $$x = x_0 - ty_0, \; y = y_0 + tx_0.$$

If now the plane is Euclidean and the coordinates x, y are rectangular (and so the ellipse under consideration is a unit circle), then

$$\mathbf{ar}_0 = (-y_0)\, x_0 + x_0 y_0 = 0,$$

where $\mathbf{r}_0$ is the radius vector of the point M_0. This means that the straight line (9) is perpendicular to the radius vector of the point M_0 and therefore is a tangent to the circle in the usual sense.

It can be checked in a similar way that in a Euclidean plane the tangent at any point of an arbitrary ellipse, parabola or hyperbola is the tangent defined in calculus as the limiting position of the secant. For pairs of distinct straight lines of the form (1) tangents are all possible straight lines different from those straight lines and passing through the point they have in common. For the curve $y^2 = 0$ the tangent is any straight line distinct from straight lines $y = k$.

For a straight line of an asymptotic direction equation (7) is linear and so has either one root (when the coefficient of t is nonzero) or has no root (when the coefficient of t is zero but the free term is nonzero) or is satisfied identically (when all coefficients are zero). In the first case the straight line (6) intersects the curve (1) in one point (but is not tangent to it!), in the second case it has no point in common with the curve (1), in the third case it lies entirely on the curve (1).

Since the undertaken investigation covers all the cases we see in particular that *a straight line having three points in common with the curve* (1) *lies entirely on that curve.*

Definition 5. A straight line of asymptotic direction having no points in common with the curve (1) is called its *asymptote.*

Examples.

For curves [1], [2], and [3] equation (8) defining asymptotic directions is of the form $l^2 + m^2 = 0$ and so has no real solutions other than zero solution $(l, m) = (0, 0)$ which we do not need. Thus, for these curves there are no straight lines of asymptotic direction in situation $\mathbb{R}$ and there are two asymptotic directions, $1 : i$ and $1 : -i$, in situations $\mathbb{C}$ and $(\mathbb{C}, \mathbb{R})$. For, curves [1] and [2] asymptotes are straight lines $y = ix$ and $y = -ix$. Curve [3] (a pair of straight lines) has no asymptotes.

For curves [4] and [5] asymptotic directions are defined by the equation $l^2 - m^2 = 0$. Therefore (over any field of a characteristic other than two) there are two asymptotic directions ($1 : 1$ and $-1 : 1$). For curve [4] (the hyperbola) asymptotes are presumably the straight lines $y = x$ and $y = -x$. Curve [5] (a pair of straight lines) has no asymptotes.

For curves [6], [7], [8] the equation of asymptotic directions is of the form $m^2 = 0$, so that there is one (double) asymptotic direction $1 : 0$. Curve [6] (the parabola) has no asymptotes, and curves [7] and [8] (pairs of parallel lines) have as their asymptote any line parallel to, but distinct from them.

In the general case of curve (1) (and an arbitrary field $\mathbb{K}$) equation (8) has 0, 1 or 2 solutions. The case of 0 solutions (there are no asymptotic directions) occurs if and only if there is no element $\sqrt{-\delta}$ in $\mathbb{K}$, where as ever $\delta = a_{11}a_{22} - a_{12}^2$. If, however, $\sqrt{-\delta} \in \mathbb{K}$, then solutions are given by any of the two formulas:

$$\begin{aligned} l : m &= (-a_{12} \pm \sqrt{-\delta}) : a_{11}, \\ l : m &= a_{22} : (-a_{12} \mp \sqrt{-\delta}). \end{aligned} \tag{10}$$

When $a_{11} \neq 0$ and $a_{22} \neq 0$ both of these formulas are equivalent and when $a_{11} = 0$ and $a_{22} = 0$ we should use the one that has meaning. There is a unique asymptotic direction if and only if $\delta = 0$.

Suppose now that we have situation $(\mathbb{C}, \mathbb{R})$ (or $\mathbb{R}$). Then all the coefficients of equation (1) are real and so the number δ is real.

Definition 6. Curve (1) is called the curve of

$$\left.\begin{matrix} \textit{elliptic type} \\ \textit{hyperbolic type} \\ \textit{parabolic type} \end{matrix}\right\} \text{ if } \left\{\begin{matrix} \delta > 0, \\ \delta < 0, \\ \delta = 0. \end{matrix}\right.$$

An elliptic-type curve has two imaginary asymptotic directions, a hyperbolic-type curve has two real asymptotic directions, a parabolic-type curve has one (real) asymptotic direction.

Note that parabolic-type curves are precisely noncentral curves.

Lecture 21

Singular and nonsingular directions • Diameters • Diameters and centres • Conjugate directions and conjugate diameters • Simplification of the equation of the second-degree central curve • Necessary refinements • Simplification of the equation of the second-degree noncentral curve

Let

(1) $$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0$$

be as before an arbitrary second-degree curve in an affine plane.

For any direction $l:m$ the coordinates x_0, y_0 of points $M_0(x_0, y_0)$, for which the coefficient of t in equation (7) of the preceding lecture is zero, satisfy the equation

(2) $$(a_{11}l + a_{12}m)\,x + (a_{12}l + a_{22}m)\,y + \\ + (a_{13}l + a_{23}m) = 0.$$

Definition 1. A direction $l:m$ is said to be *singular* if

(3) $$\begin{aligned} a_{11}l + a_{12}m &= 0, \\ a_{12}l + a_{22}m &= 0; \end{aligned}$$

otherwise it is called a *nonsingular* direction.

For a nonsingular direction equation (2) defines a certain straight line. This line is called a *diameter* conjugate to a nonsingular direction $l:m$.

Since

$$a_{11}l^2 + 2a_{12}lm + a_{22}m^2 = (a_{11}l + a_{12}m)\,l + (a_{12}l + a_{22}m)\,m,$$

every singular direction is asymptotic.

The converse is in general untrue. For example, when $\delta \neq 0$ equations (3) have for l and m only a trivial solution $(l, m) = (0, 0)$ (for δ is the determinant of this system of homogeneous equations), so that *for every central curve all directions are nonsingular* (including both asymptotic directions when they exist). □

If, on the other hand, $\delta = 0$, then equations (3) have a unique (up to proportionality) nontrivial solution $(l, m) \neq \neq (0, 0)$. For this solution $l : m = -a_{12} : a_{11} = -a_{22} : a_{12}$. Thus *for noncentral curves there exists a unique singular direction* $-a_{12} : a_{11} = -a_{22} : a_{12}$. □

But it follows from formulas (10) of the preceding lecture that for $\delta = 0$ this is a (unique) asymptotic direction of the curve. Thus *for noncentral curves an asymptotic direction is singular.* □

In the case of the parabola it is shown by direct computation to represent the direction of the parabola axis, and in the case of a pair of parallel lines, the direction of these lines.

Consider now the diameter (2) conjugate to a nonsingular direction $l : m$. Clearly, it is correctly defined by that direction, i.e. on giving the same direction by a different (but proportional) pair of numbers l and m and constructing for that pair equation (2) we obtain the same straight line (for equation (2) is replaced by a proportional equation).

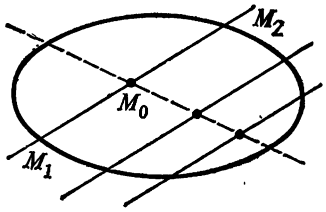
The question as to how the straight line (2) depends on the choice of equation (1) of the second-degree curve under consideration is much more difficult. Since at this point we do not know what degree of arbitrariness is possible in choosing this equation, the only way to establish the correctness of the definition of the diameter (2) is by giving it a straightforward geometrical description not appealing to equation (1).

For a nonsingular direction $l : m$ the diameter (2) has the direction $-(a_{11}l + a_{22}m) : (a_{11}l + a_{12}m)$. Therefore, if the direction $l : m$ is asymptotic (recall that this is possible only when $\delta \neq 0$) and hence

$$(a_{11}l + a_{12}m)\, l + (a_{12}l + a_{22}m)\, m = 0,$$

then it will coincide with the direction of the diameter (2), In addition, for any point $M_0(x_0, y_0)$ of this diameter the coefficient of t in equation (7) of the preceding lecture is by definition equal to zero. Therefore the diameter (2) is either the asymptote of the direction $l : m$ (which uniquely characterizes it) or the straight line of the direction $l : m$ entirely on the curve (1) (which in view of centrality of curve (1) also, as we shall later see, uniquely characterizes it). Thus for an asymptotic direction $l : m$ the diameter (2) is defined correctly.

Now let the direction $l : m$ be nonasymptotic. Consider an arbitrary straight line of this direction (i.e. one having a direction vector $\mathbf{a}(l, m)$. Suppose that the straight line



intersects curve (1) in two distinct points M_1 and M_2, and take the centre of the segment $\overline{M_1M_2}$ as the point $M_0(x_0, y_0)$ giving together with the vector $\mathbf{a}(l, m)$ this straight line. Let

$$x = x_0 + tl, \quad y = y_0 + tm \tag{4}$$

be the corresponding parametric equations and let t_1 and t_2 be the values of the parameter t corresponding to the points M_1 and M_2. Then $\overrightarrow{M_0M_1} = t_1\mathbf{a}$ and $\overrightarrow{M_0M_2} = t_2\mathbf{a}$. But under the hypothesis $\overrightarrow{M_0M_1} = -\overrightarrow{M_0M_2}$ and so $t_1 + t_2 = 0$.

Since according to Viète's formulas the sum $t_1 + t_2$ is the coefficient of t in equation (7) of the preceding lecture, taken with the opposite sign and divided by the coefficient of t^2, this proves that the indicated coefficient is zero. This means that the coordinates (x_0, y_0) of the centre of the segment cut out by the curve (1) in each straight line of the direction $l : m$ satisfy the equation of the diameter conjugate to the direction $l : m$, i.e. that that centre is on the diameter (2).

Conversely, let $M_0 (x_0, y_0)$ be an arbitrary point of the diameter (2). Suppose that the straight line (4) of a direction $l : m$ passing through the point intersects curve (1) in two points M_1 and M_2. Then the values t_1 and t_2 of the parameter t corresponding to these points are the roots of equation (7) of the preceding lecture. But since the point $M_0 (x_0, y_0)$ is on the diameter (2) the coefficient of t in this equation is zero. Therefore $t_1 + t_2 = 0$ and hence the point M_0 is the centre of the segment $\overline{M_1M_2}$.

Definition 2. A direction $l : m$ is said to be *chordwise* (for curve (1)) if there exist at least two straight lines of that direction, each intersecting curve (1) in two distinct points.

Any chordwise direction is nonasymptotic.

According to what has just been proved the diameter of curve (1) conjugate to the chordwise direction $l : m$ is uniquely characterized as a straight line passing through the centres of the segments cut out by the curve (1) in the straight lines of that direction.

Thus the diameters conjugate to chordwise directions are defined correctly.

As to the diameters conjugate to nonchordwise directions nothing remains for us but to reconcile ourselves to the idea of possible incorrectness of their definition and never forget about it.

Note that for curves [1] to [8] all (nonasymptotic) directions are chordwise. For these curves Definition 2 is therefore correct. A trivial computation shows that for a pair of coincident straight lines [9] the diameter conjugate to any nonasymptotic direction is either of these straight lines. For this curve Definition 2 is thus also correct.

Equation (2) may be rewritten as

$$(a_{11}x + a_{12}y + a_{13})\, l + (a_{21}x + a_{22}y + a_{23})\, m = 0,$$

from which it is immediate (cf. equation (3) of the preceding lecture) that it is satisfied by the coordinates of any centre of curve (1). Thus *all diameters of the curve* (1) *pass through each of its centres.* □

It follows in particular that a noncentral curve possessing a straight line of centres has a unique diameter coinciding with that straight line. This diameter is conjugate to each nonsingular (i.e. nonasymptotic) direction.

A direction

$$l' : m' = -(a_{12}l + a_{22}m) : (a_{11}l + a_{12}m)$$

of the diameter (2) conjugate to the direction $l : m$ satisfies the relation

$$(a_{11}l + a_{12}m)\, l' + (a_{12}l + a_{22}m)\, m' = 0,$$

i.e. the relation

$$a_{11}ll' + a_{12}(lm' + ml') + a_{22}mm' = 0. \tag{5}$$

This justifies the following definition:

Definition 3. A direction $l' : m'$ is said to be *conjugate* (with respect to curve (1)) to a direction $l : m$ if condition (5) holds.

This condition remains valid after interchanging l, m with l', m'. Hence the *conjugacy relation is symmetrical*, i.e. the direction $l : m$ is conjugate to the direction $l' : m'$. One may therefore speak of *conjugate directions* $l : m$ and $l' : m'$.

For a nonsingular direction $l : m$ the conjugate direction $l' : m'$ is the direction of the conjugate diameter and is therefore uniquely defined. As for a singular direction $l : m$ it is conjugate to any direction.

Besides, a direction $l : m$ (singular or not) is asymptotic if and only if it is *self-conjugate*, i.e. conjugate to itself.

Definition 4. Two diameters of a central curve (1) are said to be *conjugate* if they have conjugate directions, i.e. either diameter is conjugate to the direction of the other.

For noncentral curves the concept of conjugate diameters is not defined.

A diameter is self-conjugate, i.e. conjugate to itself, if and only if it is an asymptote. Conversely, each asymptote (of a central curve) is a self-conjugate diameter.

We now investigate the question of possible simplification of the equation of curve (1) by an appropriate choice of affine coordinates.

Condition (5) shows that *the directions* $1:0$ *and* $0:1$ *of coordinate axes are conjugate if and only if*

$$a_{12} = 0.$$

Indeed, condition (5) for the directions $l:m = 1:0$ and $l':m' = 0:1$ holds if and only if $a_{12} = 0$. □

Therefore, by choosing coordinate axes having conjugate directions we eliminate in equation (1) the term with the product xy of the coordinates x and y.

By placing, however, the origin of coordinates at the centre we eliminate linear terms as well. This proves the following proposition:

Proposition 1. *For any second-degree central curve there exists a system of affine coordinates* x, y *in which its equation is of the form*

$$\text{(I)} \qquad a_{11}x^2 + a_{22}y^2 + a_{33} = 0, \quad \textit{where } a_{11} \neq 0,\ a_{22} \neq 0.$$

A system of affine coordinates has this property if and only if the origin is at the centre of the given curve and the coordinate axes are conjugate diameters. □

The condition $a_{11} \neq 0$, $a_{22} \neq 0$ holds because otherwise equation (I) would not be the equation of a central curve.

Unfortunately, the presented proof of Proposition 1 contains a serious logical error. As discussed in detail above, the definition of a centre and conjugate directions is bound in the general case to the fixed, preassigned equation (1) of the second-degree curve under consideration. Therefore, if we are to be absolutely strict, we must say "a centre with respect to equation (1)" rather than "a centre", "directions conjugate with respect to equation (1)" rather than "conjugate directions", etc. Keeping this in mind we shall restate the proof of Proposition 1. To clarify the logical structure of the proof beyond a shadow of doubt we shall divide our argument into elementary steps.

Step 1. We choose an arbitrary nonasymptotic (with respect to equation (1)) direction and take it as the direction of the ordinate axis of a new coordinate system.

Step 2. The conjugate direction (with respect to equation (1)) is taken as the direction of the abscissa axis.

Step 3. Under the hypothesis the curve under consideration has a unique centre (with respect to equation (1)). We take it as the origin.

Step 4. We finally fix the coordinate system by choosing arbitrarily direction vectors of the coordinate axes.

Step 5. Denoting the coordinates with respect to the coordinate system thus constructed by the symbols x', y' we write the equation of our curve in these coordinates:

$$a'_{11}(x')^2+2a'_{12}x'y'+a'_{22}(y')^2+2a'_{13}x'+2a'_{23}y'+a'_{33}=0. \tag{6}$$

Step 6. Since the coordinate axes are conjugate and the centre is at the origin we have

$$a'_{12}=0, \quad a'_{13}=0, \quad a'_{23}=0.$$

Hence equation (6) is of the form (I) (up to the primes).

The error is now evident: at the last step conjugacy and centre are understood with respect to equation (6) rather than with respect to equation (1)!

To save the proof we must therefore show that nothing is spoilt by going over to equation (6). It turns out that this is really so if equation (6) is derived from equation (1) by replacing the coordinates x and y with their expressions

$$\begin{aligned} x &= c_{11}x'+c_{12}y'+x_0, \\ y &= c_{21}x'+c_{22}y'+y_0, \end{aligned} \tag{7}$$

in terms of the coordinates x', y', rather than chosen anyhow. However, appropriate computations, if carried out "amateurishly", without invention, are rather tedious. In order to simplify them we shall use the notation of matrix calculus (there are also other ways, the Einstein notation for example, to do this).

The left-hand side of the conjugacy condition

$$a_{11}ll_1+a_{12}(lm_1+ml_1)+a_{22}mm_1=0$$

(with respect to equation (1)) is a bilinear form with the matrix

$$A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

and so it may be written as

$$(l_1, m_1) A \begin{pmatrix} l \\ m \end{pmatrix}.$$

Similarly, the left-hand side of equation (1) is expressed by the formula

$$(x, y) A \begin{pmatrix} x \\ y \end{pmatrix} + 2(a_{13}, a_{23}) \begin{pmatrix} x \\ y \end{pmatrix} + a_{33}.$$

On the other hand, formulas (7) have in matrix notation the form

$$\begin{pmatrix} x \\ y \end{pmatrix} = C \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix},$$

where

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

is the transition matrix. After replacing on the left of equation (1) the coordinates x, y by their expressions (7) we thus obtain the polynomial

$$\left[C\begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right]^{\top} A \left[C\begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right] +$$

$$+ 2(a_{13}, a_{23}) \left[C\begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right] + a_{33} =$$

$$= (x', y') C^{\top} A C \begin{pmatrix} x' \\ y' \end{pmatrix} + (x', y') C^{\top} A \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} +$$

$$+ (x_0, y_0) A C \begin{pmatrix} x' \\ y' \end{pmatrix} + (x_0, y_0) A \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + 2(a_{13}, a_{23}) C \begin{pmatrix} x' \\ y' \end{pmatrix} +$$

$$+ 2(a_{13}, a_{23}) \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + a_{33} = (x', y') C^{\top} A C \begin{pmatrix} x' \\ y' \end{pmatrix} +$$

$$+ 2\left[(a_{13}, a_{23}) + (x_0, y_0) A\right] C \begin{pmatrix} x' \\ y' \end{pmatrix} +$$

$$+ (x_0, y_0) A \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + 2(a_{13}, a_{23}) \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + a_{33}$$

(in the transformations we have used the symmetry, $A^{\top} = A$, of the matrix A and the identity $(a, b)\begin{pmatrix} c \\ d \end{pmatrix} = (c, d)\begin{pmatrix} a \\ b \end{pmatrix}$). This proves that if equation (6) is derived by substituting in equation (1) expressions (7), then the matrix

$$A' = \begin{pmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{pmatrix}$$

is expressed by the formula

$$A' = C^{\top}AC$$

and the row (a'_{13}, a'_{23}) and the free term a'_{33} are expressed by the formulas

$$(a'_{13}, a'_{23}) = [(a_{13}, a_{23}) + (x_0, y_0) A] C,$$

$$a'_{33} = (x_0, y_0) A \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + 2(a_{13}, a_{23}) \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + a_{33}.$$

If now $\mathbf{a}$ is an arbitrary vector and l, m and l', m' are its coordinates in the old and the new coordinate systems, then

$$\begin{pmatrix} l \\ m \end{pmatrix} = C \begin{pmatrix} l' \\ m' \end{pmatrix}.$$

Hence for any two vectors $\mathbf{a}$ and $\mathbf{a}_1$

$$(l_1, m_1) A \begin{pmatrix} l \\ m \end{pmatrix} = (l'_1 m'_1) C^{\top}AC \begin{pmatrix} l' \\ m' \end{pmatrix} = (l'_1 m'_1) A' \begin{pmatrix} l' \\ m' \end{pmatrix}$$

and so the relations

$$(l_1, m_1) A \begin{pmatrix} l \\ m \end{pmatrix} = 0 \quad \text{and} \quad (l'_1, m'_1) A' \begin{pmatrix} l' \\ m' \end{pmatrix} = 0$$

are equivalent to each other.

This proves that the directions of the straight lines with the direction vectors $\mathbf{a}$ and $\mathbf{a}_1$ are conjugate with respect to equation (1) in the coordinates x, y if and only if they are conjugate with respect to equation (6) in the coordinates x', y'. At step 6 we thus have the right to understand conjugacy in the required sense (when equation (6) is chosen as described, we stress).

The situation is similar for the centre as well. The coordinates x, y of the centre of the curve under consideration with respect to equation (1) are found from the relations

$$a_{11}x + a_{12}y + a_{13} = 0,$$
$$a_{21}x + a_{22}y + a_{23} = 0,$$

equivalent to a single matrix relation

(8) $$(x,\ y)\,A + (a_{13},\ a_{23}) = (0,\ 0).$$

Therefore the coordinates x', y' of the centre of the same curve but with respect to equation (6) satisfy the relation

(9) $$(x',\ y')\,A' + (a'_{13},\ a'_{23}) = (0,\ 0).$$

But in view of the formulas proved above

$$(x',\ y')\,A' + (a'_{13},\ a'_{23}) =$$
$$= \left[C^{-1}\left[\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right]\right]^{\mathsf{T}} C^{\mathsf{T}}AC + [(a_{13},\ a_{23}) + (x_0,\ y_0)\,A]\,C =$$
$$= [(x,\ y) - (x_0,\ y_0)]\,(C^{-1})^{\mathsf{T}}\,C^{\mathsf{T}}AC + [(a_{13},\ a_{23}) + (x_0,\ y_0)\,A]\,C =$$
$$= [(x,\ y)\,A + (a_{13},\ a_{23})]\,C.$$

Therefore (the matrix C is invertible) relations (8) and (9) are equivalent to each other.

Step 6 in the proof of Proposition 1 is thus fully justified. □

For a noncentral curve we take as the direction $0:1$ of the ordinate axis an arbitrary nonsingular (i.e. nonasymptotic in this case) direction and as the abscissa axis the diameter conjugate to this direction (thereby we fix not only the direction but also the position of the abscissa axis). According to the general formula (2) the diameter conjugate to the direction $0:1$ has the equation

$$a_{12}x + a_{22}y + a_{23} = 0$$

(here we mean of course the coefficients of the corresponding equation (6) and the coordinates x', y' with the primes discarded). It therefore follows from the condition that this diameter is the abscissa axis $y = 0$ that $a_{12} = 0$, $a_{22} \neq 0$ and $a_{23} = 0$, from which it further ensues that $a_{11} = 0$ (for $a_{11}a_{22} - a_{12}^2 = 0$). Thus if the abscissa axis and the direction of the ordinate axis are chosen in this way, the equation of a noncentral curve becomes

(10) $$a_{22}y^2+2a_{13}x+a_{33}=0 \text{ where } a_{22}\neq 0.$$

Since for this equation

$$\Delta=\begin{vmatrix}0 & 0 & a_{13}\\ 0 & a_{22} & 0\\ a_{31} & 0 & a_{33}\end{vmatrix}=-a_{22}a_{13}^2$$

curve (10) is of type II (has no centres) when $a_{13}\neq 0$ and of type III (has a straight line of centres) when $a_{13}=0$.

For $a_{13}=0$ we thus obtain the following proposition:

Proposition 2. *For any noncentral second-degree curve of type* III, *i.e. having a straight line of centres, there exists a system of affine coordinates x, y in which the equation of the curve has the form*

(III) $$a_{22}y^2+a_{33}=0 \text{ where } a_{22}\neq 0.$$

The system of affine coordinates has this property if and only if the axis of ordinates has a nonsingular direction and the axis of abscissas is the diameter conjugate to the direction of the ordinate axis. □

Now let $a_{13}\neq 0$. Then the abscissa axis intersects curve (10) in the point with the abscissa $x=-\frac{a_{33}}{2a_{13}}$. By choosing the origin at that point (which fixes the position of the ordinate axis) we have the free term a_{33} in equation (10) go to zero. This proves the following proposition:

Proposition 3. *For any noncentral second-degree curve of type* II, *i.e. having no centres, there exists a system of affine coordinates x, y in which the equation of the curve has the form*

(II) $$a_{22}y^2+2a_{13}x=0 \text{ where } a_{22}\neq 0,\ a_{13}\neq 0.$$

The system of affine coordinates has this property if and only if the axis of ordinates has a nonsingular direction, the axis of abscissas is the diameter conjugate to the direction of the ordinate axis and the origin of coordinates is on the curve. □

It is advisable to combine the propositions we have proved into a single theorem.

Theorem 1. *For any second-degree curve in an affine plane there exists a system of affine coordinates x, y in which the* *equation of the curve belongs to one of the following three types:*

(I) $a_{11}x^2 + a_{22}y^2 + a_{33} = 0$ *where* $a_{11} \neq 0$, $a_{22} \neq 0$;

(II) $a_{22}y^2 + 2a_{13}x = 0$ *where* $a_{22} \neq 0$, $a_{13} \neq 0$;

(III) $a_{22}y^2 + a_{33} = 0$ *where* $a_{22} \neq 0$.

For no curve there are two systems of coordinates in which the equation of the given curve would belong to different types. □

The last statement follows from the fact that each of the three types is characterized geometrically as: central curves, curves without a centre, and curves with a straight line of centres.

We emphasize that Theorem 1 holds in an affine plane over an arbitrary field $\mathbb{K}$ (with a characteristic $\neq 2$).

Lecture 22

Second-degree curves in the complex affine plane • Second-degree curves in the real-complex affine plane • The uniqueness of the equation of a second-degree curve • Second-degree curves in the Euclidean plane • Circles

Depending on whether the individual coefficients are zero or nonzero and after dividing by a nonzero coefficient we obtain from Theorem 1 of the preceding lecture the following types of second-degree equations:

(I_1) $Ax^2+By^2=1$,

(I_0) $Ax^2+By^2=0$,

(II) $y^2=2Ax$,

(III_1) $y^2+A=0$,

(III_0) $y^2=0$,

where $A\neq 0$, $B\neq 0$.

The possibility of further simplification on these equations depends on the arithmetic properties of the ground field $\mathbb{K}$. For example, if in the field $\mathbb{K}$ there is a square root for any element (this property is peculiar, in particular, to the field of complex numbers, $\mathbb{C}$), then one may introduce new coordinates x', y' by the formulas

$$x'=\sqrt{A}x, \quad y'=\sqrt{B}y$$

in cases (I_1) and (I_0) or by the formulas

$$x'=x, \quad y'=\frac{y}{\sqrt{A}}$$

in cases (II) and (III_1). This provides us with the following theorem:

Theorem 1 (reduction to canonical form in situation $\mathbb{C}$). *For any second-degree curve in a complex affine plane there exists a system of affine coordinates x, y in which the equation of the given curve is one of the following five equations:*

[1] $x^2 + y^2 = 1$,

[2] $x^2 + y^2 = 0$,

[3] $y^2 = 2x$,

[4] $y^2 + 1 = 0$,

[5] $y^2 = 0$.

For no curve there are two systems of coordinates in which the curve would have different equations of this list.

The last statement follows from the fact that curve [2] is a pair of intersecting straight lines, curve [4] is a pair of parallel lines, curve [5] is a pair of coincident straight lines, curve [3] has no centre (of symmetry), and curve [1] is a central curve (with a centre of symmetry) and contains no straight line (if we have identically for t an equation of the form

$$(x_0 + tl)^2 + (y_0 + tm)^2 = 1,$$

then

$$l^2 + m^2 = 0, \quad x_0 l + y_0 m = 0, \quad x_0^2 + y_0^2 = 1,$$

which is possible only for $(l, m) = (0, 0)$). □

Corollary (theorem of the classification of second-degree curves in situation $\mathbb{C}$). *In a complex affine plane any second-degree curve is either*

(a) *a central curve* [1] *on which there is no straight line or*

(b) *a curve without a centre* [3] *or*

(c) *a pair of* (*intersecting* [2], *parallel and distinct* [4] *coincident* [5]) *straight lines.* □

One is equally justified in calling curve (a) either an ellipse or a hyperbola. Some writers call it an *oval.*

Curve (b) may well be called a parabola.

In situation ($\mathbb{C}$, $\mathbb{R}$) if we are not to violate the reality of the coefficients we should introduce new coordinates by the formulas

$$x' = \sqrt{|A|}\, x, \quad y' = \sqrt{|A|}\, y$$

in cases (I_1) and (I_0) or by the formulas

$$x' = x, \qquad y' = \frac{y}{\sqrt{|A|}}$$

in cases (II) and (III_1). In view of the possibility of multiplying equations by -1 this yields the following theorem:

Theorem 2 (reduction to canonical form in situation ($\mathbb{C}$, $\mathbb{R}$)). *For any real second-degree curve in a real-complex affine plane there exists a real affine coordinate system in which the equation of the given curve is one of the following nine canonical affine equations*:

[1] $x^2 + y^2 - 1 = 0$,

[2] $x^2 + y^2 + 1 = 0$,

[3] $x^2 + y^2 = 0$,

[4] $x^2 - y^2 - 1 = 0$,

[5] $x^2 - y^2 = 0$,

[6] $y^2 - 2x = 0$,

[7] $y^2 - 1 = 0$,

[8] $y^2 + 1 = 0$,

[9] $y^2 = 0$.

For no curve there are two systems of coordinates in which the curve would have different equations of this list.

The last statement follows from the fact that each of the curves [1] to [9] can be characterized geometrically:

[1] A central curve cutting out an ellipse in the real plane.

[2] A central curve intersecting no real plane.

[3] A pair of complex conjugate intersecting straight lines.

[4] A central curve cutting out a hyperbola in the real plane.

[5] A pair of real intersecting straight lines.

[6] A curve without centres (cutting out a parabola in the real plane).

[7] A pair of real parallel and distinct lines.

[8] A pair of complex conjugate parallel and distinct lines.

[9] A pair of coincident straight lines (automatically real). □

Affine coordinates in which the equation of a given second-degree curve has the form indicated in Theorem 2 are called *canonical* (for the given curve).

As already said, curve [1] is called a real ellipse, curve [2] is called an imaginary ellipse and curves [4] and [6] are respectively called a hyperbola and a parabola (with or without adding the epithet "real").

Corollary (theorem of the classification of second-degree curves in situation ($\mathbb{C}$, $\mathbb{R}$)). *In a real-complex affine plane any real second-degree curve is either*

(a) *an ellipse (real* [1] *or imaginary* [2]) *or*

(b) *a hyperbola* [4] *or*

(c) *a parabola* [6] *or*

(d) *a pair of (intersecting complex conjugate* [3], *intersecting real* [5], *parallel distinct and real* [7], *parallel distinct and complex conjugate* [8], *coincident* [9]) *straight lines.*

In order to complete the theory of second-degree curves (in situations $\mathbb{C}$ and ($\mathbb{C}$, $\mathbb{R}$)) it remains to consider the question of the uniqueness of their equations.

It is clear that if a second-degree curve is given by some equation of the form

$$F(x, y) = 0,$$

where $F(x, y)$ is a second-degree polynomial, then the proportional equation

$$kF(x, y) = 0,$$

where k is an arbitrary nonzero number, gives the same curve. It is natural to consider such equations as being equal.

It turns out that it is to this that (in situations $\mathbb{C}$ and ($\mathbb{C}$, $\mathbb{R}$)) all the arbitrariness is confined.

Theorem 3 (uniqueness of the equation of a second-degree curve). *If in a complex or real-complex affine plane two second-degree equations*

$$F(x, y) = 0, \quad G(x, y) = 0$$

give (in the same system of affine coordinates x, y) the same curve, then the equations are proportional

$$G(x, y) = kF(x, y), \ k \neq 0.$$

This theorem shows that the scrupulousness which we exercised in proving Theorem 1 of the preceding lecture was really unnecessary. In order to avoid the vicious circle, however, we could not use Theorem 3 there, for its proof rests on the classification theorem (and hence on the results of the preceding lecture).

We shall preface the proof of Theorem 3 with a number of general remarks which are of interest in their own right as well. The ground field $\mathbb{K}$ is for the time being taken to be arbitrary.

The fact that the second-degree curve

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0$$

passes through a point $M_0(x_0, y_0)$ means that we have the relation

(1) $\quad a_{11}x_0^2 + 2a_{12}x_0y_0 + a_{22}y_0^2 + 2a_{13}x_0 + 2a_{23}y_0 + a_{33} = 0,$

which is a homogeneous linear condition on the six coefficients

(2) $\quad a_{11}, a_{12}, a_{22}, a_{13}, a_{23}, a_{33}$

of the equation of the curve. But it is known from algebra (this is a special case of the general theorem on the number of fundamental solutions of a system of homogeneous equations for which, incidentally, we shall provide a geometrical interpretation and a new proof in the next semester) that a system of n homogeneous linear equations in $n + 1$ unknowns has a unique up to proportionality nontrivial solution if the equations of that system are linearly independent. Therefore, if five distinct points

(3) $\quad M_1, M_2, M_3, M_4, M_5$

have the property that the corresponding five conditions of the form (1) on the coefficients (2) are linearly independent (throughout the proof of Theorem 3 these points will be referred to as *independent*) and if there exists a second-degree curve passing through these points (the theorem of equations does not imply the existence of the curve since all the three components a_{11}, a_{12}, a_{22} may turn out to be zero in the solution), then any two equations of the curve are proportional. This means that the following proposition is true:

Proposition 1. *If there are five independent points on a second-degree curve, then any two equations of the curve (of the form* $F(x, y) = 0$, *where* $F(x, y)$ *is a second-degree polynomial) are proportional.* □

An important extension of the proposition is the following lemma:

Lemma. *The points* (3) *are independent if no four of them are on the same straight line.*

Proof. Suppose that the points (3) are dependent, i.e. that a relation of the form (1) corresponding to one of the points (say to the point M_5) is a linear combination of relations corresponding to the other points. This means that any second-degree curve passing through the first four points also passes through the point M_5. But one of such curves is a pair of straight lines M_1M_2 and M_3M_4. Therefore the point M_5 lies on one of these straight lines. For similar reasons this point lies on one of the curves M_1M_3 or M_2M_4 as well. But it is obvious that this is possible only if at least two of these four straight lines coincide, i.e. if at least three points of the quadruple M_1, M_2, M_3, M_4 are on the same straight line.

Let these be the points M_1, M_2 and M_3. Then the second-degree curve consisting of the straight line M_1M_2 and an arbitrary straight line passing through the point M_4 will contain all the four points M_1, M_2, M_3, M_4 and hence the point M_5. It is clear that this is possible only if the point M_5 is on the straight line M_1M_2. But then, contrary to the assumption, the points M_1, M_2, M_3 and M_5 will lie on the same straight line. The contradiction obtained proves the lemma. □

Now we are already in a position to pass on directly to the proof of Theorem 3.

Proof of Theorem 3. According to Proposition 1 it is enough to find five independent points on each second-degree curve. In the case of situation ($\mathbb{C}$, $\mathbb{R}$) there is no need for these points to be real (since we are not constructing the equation of a curve from these points but are only proving the uniqueness of the equation). We may therefore restrict ourselves to situation $\mathbb{C}$ without loss of generality.

If the second-degree curve under consideration does not contain any straight line (i.e. is an oval or a parabola), then any three (and all the more four) of its points are not on the same straight line. Therefore, according to the lemma any five of its points are independent. This proves Theorem 3 for such a curve.

Five points satisfying the condition of the lemma and hence independent can clearly be found on any pair of distinct straight lines as well. Therefore for a pair of distinct straight lines Theorem 3 is also true.

The remaining case of coincident straight lines requires a separate study.

Let $F(x, y) = 0$, where $F(x, y)$ is a second-degree polynomial, be an arbitrary equation of a pair of coincident straight lines. According to the reduction theorem (in its precise statement) it is possible to transform from the coordinates x, y to other coordinates

$$x' = c_{11}x + c_{12}y + x_0',$$
$$y' = c_{21}x + c_{22}y + y_0',$$

having the property that after substituting in the polynomial $F(x, y)$ the expressions for the coordinates x, y in terms of the coordinates x', y' we obtain a polynomial proportional to the polynomial y'^2. But then the polynomial $F(x, y)$ itself is clearly proportional to the polynomial $(c_{21}x + c_{22}y + y_0')^2$. This proves that any equation of a pair of coincident straight lines is of the form $(Ax + By + C)^2 = 0$, where $Ax + By + C = 0$ is the equation of either of the straight lines.

The proportionality of any two equations of a pair of coincident straight lines now follows from the proportionality of any two equations of a straight line.

This completes the proof of Theorem 3. □

Remark. It is easy to show in the case of an arbitrary ground field (of a characteristic $\neq 2$) that curves of types (I_1) and (II) (see the beginning of this lecture) contain no straight lines, that a curve of type (I_0) is either a point or a pair of intersecting straight lines and that a curve of type (III_1) is either an empty set or a pair of parallel lines. Besides, the set of points in every curve of type (II) is equivalent to the set of elements in a field $\mathbb{K}$ and the set of points in every nonempty curve of type (I_1) is equivalent to the set of elements in a field $\mathbb{K}$ plus one more element (prove this!) Thus, for the case of an arbitrary field $\mathbb{K}$ the situation is the same, formally, as that in the case of the field $\mathbb{R}$.

Therefore, if the number of elements in a field $\mathbb{K}$ is greater than three, then Theorem 3 holds for any second-degree curves in an affine plane over $\mathbb{K}$ that contain more than one point. But it cannot hold in principle for empty and single-point curves.

We now turn to second-degree curves in a Euclidean (real or real-complex) plane. The coordinates x, y will now be assumed to be rectangular.

Definition 1. A direction $l : m$ is said to be *principal* with respect to a given second-degree curve

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0$$

if the perpendicular direction $-m : l$ is conjugate to it (and in particular is not singular).

This condition holds of course for the singular direction $-a_{12} : a_{11} = -a_{22} : a_{12}$. Thus, for every noncentral curve the singular direction (i.e. the direction of the axis for a parabola and their own direction for parallel lines) is the principal direction.

Also principal is the direction perpendicular to the singular direction. A second-degree noncentral curve has no other principal directions. Thus, *for any noncentral curve there are two and only two perpendicular principal directions.* □

On recalling now the method presented in the preceding lecture for reducing the equation of a second-degree curve to the simplest form (II) or (III) we immediately discover that in the Euclidean plane this can be done in the class of rectangular coordinates: it is enough to choose in the initial step of construction the direction of the ordinate axis as principal and nonsingular. Thus, *for any second-degree noncentral curve in a real or real-complex Euclidean plane there exist rectangular coordinates x, y in which the equation of the curve has the form* (II) *or* (III). □

A similar statement for central curves will of course be also proved should we prove the existence of at least one pair of conjugate perpendicular (and hence principal) directions. But this is easily done. Indeed, the conjugacy of a nonsingular direction $l : m$ and the perpendicular direction $-m : l$ implies that we have

$$l\,(a_{21}l + a_{22}m) - m\,(a_{11}l + a_{12}m) = 0,$$

i.e.

$$a_{21}l^2 + (a_{22} - a_{11})\,lm - a_{12}m^2 = 0. \tag{4}$$

Thus *a direction $l : m$ is a principal direction if and only if we have relation* (4). □

(A direct check shows that incidentally this is also true when the direction $l : m$ is singular).

For $a_{12} = 0$ and $a_{11} = a_{22}$ relation (4) is satisfied identically, i.e. in this case any direction is a principal direction. For $a_{12} \neq 0$ relation (4) is a quadratic equation in $l : m$ whose solutions can be written in one of the equivalent forms:

$$l : m = (a_{11} - a_{22} \pm \sqrt{(a_{11} - a_{22})^2 + 4a_{12}^2}) : 2a_{12},$$

$$l : m = -2a_{12} : (a_{11} - a_{22} \mp \sqrt{(a_{11} - a_{22})^2 + 4a_{12}^2}).$$

These same formulas are also suitable for $a_{12} = 0$, $a_{11} \neq \neq a_{22}$; it is only necessary to use the formula that makes sense, i.e. the upper one when the root is taken with the positive sign and the lower one otherwise.

Thus when $(a_{11} - a_{22})^2 + 4a_{12}^2 > 0$ these formulas give for $l : m$ two real values. This means that *for a second-degree central curve either there are two and only two principal directions* (automatically conjugate and perpendicular) *or any direction is principal.* □

We have thus proved in particular that for any such curve there exist rectangular coordinates in which its equation has the form (I), i.e. (I_1) or (I_0).

Multiplying (if necessary) by -1 we may assume that in this equation $A > 0$. We set

$$a = \frac{1}{\sqrt{|A|}}, \quad b = \frac{1}{\sqrt{|B|}}.$$

Moreover, in case (I_1) for $B > 0$, as well as in case (I_0) we may in addition assume that $a \geqslant b$ (otherwise it is necessary to interchange the coordinates). Besides, in case (I_0) we can also obtain

$$\frac{1}{a^2} + \frac{1}{b^2} = 1.$$

In a similar way we can assume in case (II) (by changing if necessary, the orientation of the abscissa axis) that $A > 0$. We set $p = A$.

In case (III_1) however, we set

$$b = \sqrt{|A|}.$$

This yields the following theorem:

Theorem 4 (reduction to canonical form in a Euclidean plane). *For any (real) second-degree curve in a real or real-complex Euclidean plane there exists a (real) system of rectangular coordinates in which the equation of the given curve belongs to one of the following nine types:*

[1] $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ *where* $a \geqslant b > 0$;

[2] $\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$ *where* $a \geqslant b > 0$;

[3] $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$ *where* $a \geqslant b > 0$ *and* $\frac{1}{a^2} + \frac{1}{b^2} = 1$;

[4] $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ *where* $a > 0$, $b > 0$;

[5] $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$ *where* $a > 0$, $b > 0$, *and* $\frac{1}{a^2} + \frac{1}{b^2} = 1$;

[6] $y^2 = 2px$ *where* $p > 0$;

[7] $y^2 - b^2 = 0$ *where* $b > 0$;

[8] $y^2 + b^2 = 0$ *where* $b > 0$;

[9] $y^2 = 0$.

In the real plane these are respectively

[1] *an ellipse,*

[2] *an empty set,*

[3] *a point,*

[4] *a hyperbola,*

[5] *a pair of intersecting straight lines,*
[6] *a parabola,*
[7] *a pair of parallel but distinct lines,*
[8] *an empty set,*
[9] *a pair of coincident straight lines, and in the real-complex plane they are*
[1] *a real ellipse,*
[2] *an imaginary ellipse,*
[3] *a pair of complex conjugate intersecting straight lines,*
[4] *a hyperbola,*
[5] *a pair of real intersecting straight lines,*
[6] *a parabola,*
[7] *a pair of real parallel distinct lines,*
[8] *a pair of complex conjugate parallel distinct lines,*
[9] *a pair of coincident straight lines.*

For no curve containing in the real case more than one point there exist two systems of coordinates in which that curve would have different equations of the list.

The last statement follows from the corresponding statement for the affine case with respect to equations of different types, and with respect to equations of one type it follows from the fact that (in cases [1] to [8]) the coefficients of these equations can be characterized geometrically:

[1] The numbers a and b are the lengths of segments cut out by the curve in the axes of symmetry.

[2] (Relates only to the real-complex case). The numbers ia and ib are the lengths of segments cut out by the curve in the axes of symmetry.

[3] (Relates only to the real-complex case). For $a = b$ the length of any segment in each of the straight lines making up the curve is zero (such straight lines are called *isotropic*) and for $a \neq b$ the cosine of the angle between the straight lines is equal to $\frac{a^2+b^2}{a^2-b^2}$. Together with the normalizing condition $\frac{1}{a^2}+\frac{1}{b^2}=1$ this uniquely determines the numbers a and b obeying the inequalities $a \geqslant b > 0$.

[4] The number $2a$ is the length of a segment cut out by the curve in the axis of symmetry and the number $\frac{a^2-b^2}{a^2+b^2}$ is the cosine of the angle between the asymptotes of the curve.

[5] The number $\frac{a^2 - b^2}{a^2 + b^2}$ is the cosine of the angle between the straight lines that make up the curve.

[6] The number $\frac{p}{2}$ is the distance from the focus to the vertex.

[7] The number $2b$ is the distance between the straight lines that make up the curve.

[8] (Relates to the real-complex case only). The number $2ib$ is the distance between the straight lines that make up the curve. □

Corollary (theorem of the classification of second-degree curves in the Euclidean plane). *Any real second-degree curve in a real-complex Euclidean plane (any real second-degree curve in a real Euclidean plane containing more than one point) is either*

(a) *an ellipse (real* [1] *or imaginary* [2]*) or*

(b) *a hyperbola* [4] *or*

(c) *a parabola* [6] *or*

(d) *a pair of (intersecting complex conjugate* [3], *intersecting real* [5], *parallel distinct and real* [7], *parallel distinct and complex conjugate* [8], *coincident* [9]*) straight lines.* □

In conclusion we consider the question of the geometrical description of central curves for which any direction is principal.

In a real Euclidean plane an example of such a curve is provided by the *circle*, i.e. a locus of points whose distance from a given point (the centre) is equal to a given number R. Indeed, the equation of a circle (in the rectangular coordinates x, y) has obviously the form

$$(x - a)^2 + (y - b)^2 = R^2, \tag{5}$$

where a, b are the coordinates of the centre, i.e. the form

$$x^2 + y^2 - 2ax - 2by + c = 0, \tag{6}$$

where $c = a^2 + b^2 - R^2$. Thus, indeed, for a circle $a_{12} = 0$ and $a_{11} = a_{22}$. □

In a real-complex plane we assume the last property as a definition:

Definition 2. A real second-degree curve in a real-complex Euclidean plane is said to be a *circle* if $a_{12} = 0$ and

$a_{11} = a_{22}$, i.e. if the equation of the curve may be written in the form (6). A (real) point with the coordinates a, b is called here the *centre* of the circle, and the number $R = \sqrt{a^2 + b^2 - c}$ is called the *radius* of the circle.

It can be directly verified that these definitions are correct, i.e. do not depend on the choice of rectangular coordinates x, y.

Three cases are possible:

The case $a^2 + b^2 - c > 0$. Here the radius R of the circle is a positive real number. The intersection of the circle with a real plane is an ordinary circle of radius R with centre (a, b). As a second-degree curve such a circle is of type [1], i.e. is a real ellipse. It is usually called a *real circle.*

The case $a^2 + b^2 - c = 0$. The radius R of the circle is equal to zero. The circle has a single point (0, 0), the centre, in common with a real plane. As a second-degree curve this circle belongs to type [3] and is a pair of intersecting (at the point (0, 0)) complex conjugate isotropic straight lines.

The case $a^2 + b^2 - c < 0$. Here it is usually assumed that $R = \sqrt{|a^2 + b^2 - c|}$, denoting thereby the radius of the circle by iR. Such a circle has no real points. It belongs to type [2] and is usually called an *imaginary circle.*

In all the three cases the circle is a locus of points in a real-complex Euclidean plane whose distance from the centre is equal to a given real or pure imaginary number. This may be accepted at will as a definition of the circle.

Lecture 23

Ellipsoids • Imaginary ellipsoids • Second-degree imaginary cones • Hyperboloids of two sheets • Hyperboloids of one sheet • Rectilinear generators of a hyperboloid of one sheet • Second-degree cones • Elliptical paraboloids • Hyperbolic paraboloids • Elliptical cylinders • Other second-degree surfaces • The statement of the classification theorem

The analogues of second-degree curves in space are *second-degree surfaces* having an equation of the form

$$F(x, y, z) = 0$$

where $F(x, y, z)$ is some second-degree polynomial in x, y, z. One can repeat for these surfaces, with appropriate modifications and complications, what was said above about second-degree curves. For lack of the necessary time we shall restrict ourselves to a brief description of all the possible types of surfaces. We shall work in the Euclidean space in rectangular coordinates.

Type [1]. Of this type are surfaces having in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1 \tag{1}$$

where $a \geqslant b \geqslant c > 0$. These surfaces are called *ellipsoids*.

When $a = b = c$ an ellipsoid is a *sphere* of radius a.

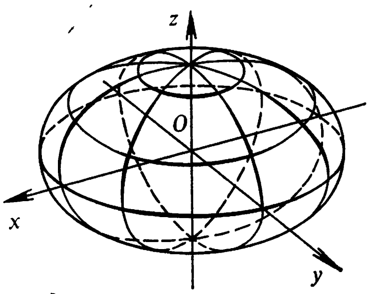
The easiest way to get an idea of what the shape of an arbitrary ellipsoid must be is through a study of its sections by planes parallel to the coordinate planes. Consider, for example, a plane $z = h$ parallel to the plane Oxy. In this plane the numbers x, y are coordinates (relative to a coordinate system H_{ij}, where $\mathbf{i}$, $\mathbf{j}$ are the basis vectors of the axes Ox and Oy and H is a point with the coordinates $(0, 0, h)$) and the equation of the curve cut out in the plane by ellipsoid (1) has in these coordinates the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1-\frac{h^2}{c^2}$$

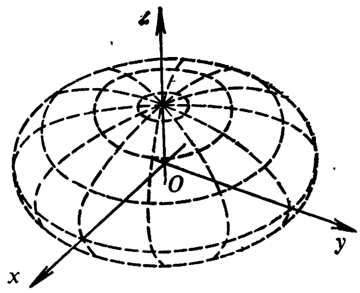
i.e. the form

$$\frac{x^2}{\left(a\sqrt{1-\frac{h^2}{c^2}}\right)^2}+\frac{y^2}{\left(b\sqrt{1-\frac{h^2}{c^2}}\right)^2}=1.$$

Hence the plane $z = h$ does not cut ellipsoid (1) when $|h| > c$, has a single point in common with the ellipsoid



An ellipsoid



An imaginary ellipsoid

(1) when $|h| = c$ (the point $(0, 0, c)$ when $h = c$ and $(0, 0, -c)$ when $h = -c$) and cuts ellipsoid (1) in an ellipse with the semiaxes

$$a\sqrt{1-\frac{h^2}{c^2}},$$

$$b\sqrt{1-\frac{h^2}{c^2}}$$

when $|h| < c$, the semiaxes being the largest (and equal to a and b) when $h = 0$ and monotonically decreasing to zero when $|h|$ increases from zero to c.

It can be shown in a similar way that the plane $y = h$ does not cut ellipsoid (1) when $|y| > b$, has a single point in common with ellipsoid (1) when $|y| = b$ (the point $(0, b, 0)$ when $y = b$ (and $(0, -b, 0)$ when $y = -b$) and cuts ellipsoid (1) in an ellipse with the semiaxes

$$a\sqrt{1-\frac{h^2}{b^2}}, \quad c\sqrt{1-\frac{h^2}{b^2}}$$

when $|y| < b$, the semiaxes being the largest (and equal to a and c) when $h = 0$ and monotonically decreasing to zero when $|h|$ increases from zero to b.

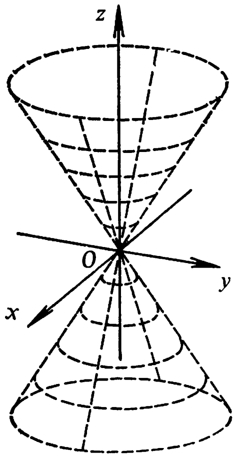
In just the same way the plane $x = h$ does not cut ellipsoid (1) when $|h| > a$, has a single point in common with the ellipsoid (1) when $|h| = a$ (the point $(a, 0, 0)$ when $h = a$ and $(-a, 0, 0)$ when $h = -a$) and cuts ellipsoid (1) in an ellipse with the semiaxes

$$b\sqrt{1-\frac{h^2}{a^2}}, \quad c\sqrt{1-\frac{h^2}{b^2}}$$

then $|h| < a$, the semiaxes being the largest (and equal to b and c) when $h = 0$ and monotonically decreasing to zero when $|h|$ increases from zero to a.

All this gives a quite satisfactory idea of the shape of ellipsoid (1). In particular we see that ellipsoid (1) is located wholly inside a rectangular parallelepiped (is inscribed in that parallelepiped) with centre at the point O (0, 0, 0), with faces parallel to the coordinate planes and with sides of lengths $2a$, $2b$, and $2c$.

It can be added that since equation (1) remains unchanged when the signs of the coordinates x, y, z are changed the coordinate planes are the planes of symmetry of ellipsoid (1) and the origin of coordinates is its centre of symmetry.



An imaginary cone of second degree

The ellipsoid has no other planes of symmetry when $a > b > c$.

Type [2]. Of this type are surfaces having in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=-1, \tag{2}$$

where $a \geqslant b \geqslant c > 0$. They have no real points and are called *imaginary ellipsoids.*

Type [3]. Of this type are surfaces that have in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0 \tag{3}$$

where $a \geqslant b \geqslant c > 0$ and $\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}=1$. They have a single real point O (0, 0, 0) and are called *imaginary cones of the second degree.*

Type [4]. Of this type are surfaces that have in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1 \tag{4}$$

where $a \geqslant b > 0$, $c > 0$. They are called *hyperboloids of two sheets.*

The plane $z = h$ does not cut hyperboloid (4) when $|h| < c$, has a single point in common with hyperboloid (4) when $|h| = c$ (a point $(0, 0, c)$ when $h = c$ and $(0, 0, -c)$ when $h = -c$) and cuts hyperboloid (4) in an ellipse with semiaxes

$$a\sqrt{\frac{h^2}{c^2}-1}, \quad b\sqrt{\frac{h^2}{c^2}-1}$$

when $|h| > c$, the semiaxes increasing monotonically (from 0 to $+\infty$) as $|h|$ increases from c to $+\infty$.

Any plane $y = h$ cuts hyperboloid (1) in a hyperbola with semiaxes

$$c\sqrt{1+\frac{h^2}{b^2}}, \quad a\sqrt{1+\frac{h^2}{b^2}}$$

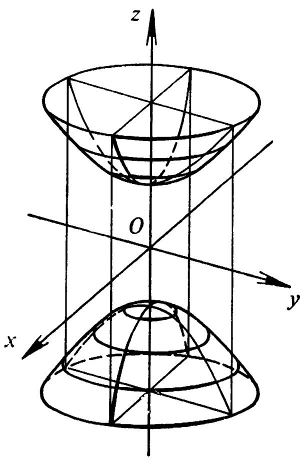
increasing monotonically (from c and a to $+\infty$) as $|h|$ increases from zero to $+\infty$.

Any plane $x = h$ cuts the hyperboloid (1) in a hyperbola with semiaxes

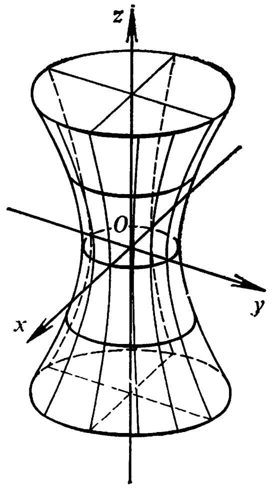
$$c\sqrt{1+\frac{h^2}{a^2}}, \quad b\sqrt{1+\frac{h^2}{a^2}}$$

increasing monotonically (from c and b to $+\infty$) as $|h|$ increases from zero to $+\infty$.

This fully explains the shape of the hyperboloid of two



A hyperboloid of two sheets



A hyperboloid of one sheet

sheets. In particular we see that this hyperboloid is made up of two symmetrical parts ("laps") located respectively in the half-spaces $z \geqslant c$ and $z \leqslant -c$.

By the same reasons as those for the ellipsoid the coordinate planes are the planes of symmetry of the hyperboloid of two sheets and the origin of coordinates is its centre of symmetry.

Type [5]. Of this type are surfaces having in some system of rectangular coordinates x, y, z an equation of the form

(5) $$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1,$$

where $a \geqslant b > 0$, $c > 0$. These surfaces are called *hyperboloids of one sheet.*

Any plane $z = h$ cuts hyperboloid (5) in an ellipse with semiaxes

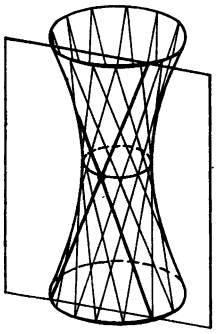
$$a\sqrt{1+\frac{h^2}{c^2}},$$

$$b\sqrt{1+\frac{h^2}{c^2}}$$

increasing monotonically (from a and b to $+\infty$) as $|h|$ increases from zero to $+\infty$.

The ellipse

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$



Rectilinear generators of a one-sheeted hyperboloid

we get when $h = 0$ is called the *striction ellipse* of hyperboloid (5).

As for the planes $y = h$ and $x = h$, when $|h| < b$ the plane $y = h$ cuts hyperboloid (5) in a hyperbola with semiaxes

$$a\sqrt{1-\frac{h^2}{b^2}}, \quad c\sqrt{1-\frac{h^2}{b^2}}$$

decreasing monotonically from a and c to zero as $|h|$ increases from zero to b, when $|h| = b$ it cuts hyperboloid (5) in a pair of straight lines which has in the coordinates x, z (that are obviously rectangular coordinates in this plane) an equation

$$\frac{x^2}{a^2}-\frac{z^2}{c^2}=0,$$

and when $|h| > b$ it cuts (5) in a hyperbola with semiaxes

$$c\sqrt{\frac{h^2}{b^2}-1}, \quad a\sqrt{\frac{h^2}{b^2}-1}$$

increasing monotonically (from zero to $+\infty$) as $|h|$ increases from b to $+\infty$. The imaginary (real) axes of hyperbolas obtained for $|h| > b$ are parallel to the real (imaginary) axes of hyperbolas obtained for $|h| < b$.

Similarly, the plane $x = h$ cuts hyperboloid (5) in a hyperbola with semiaxes

$$b\sqrt{1-\frac{h^2}{a^2}}, \quad c\sqrt{1-\frac{h^2}{a^2}}$$

decreasing monotonically (from b and c to zero) as $|h|$ increases from zero to a when $|h| < a$; when $|h| = a$ it cuts hyperboloid (5) in a pair of straight lines which has in the coordinates y, z (that are obviously rectangular coordinates in this plane) an equation

$$\frac{y^2}{b^2}-\frac{z^2}{c^2}=0,$$

and when $|h| > a$ it cuts (5) in a hyperbola with semiaxes

$$c\sqrt{\frac{h^2}{a^2}-1}, \quad b\sqrt{\frac{h^2}{a^2}-1}$$

increasing monotonically (from zero to $+\infty$) as $|h|$ increases from a to $+\infty$. The imaginary (real) axes of hyperbolas obtained for $|h| > a$ are parallel to the real (imaginary) axes of hyperbolas obtained for $|h| < a$.

Besides, like the preceding surfaces the hyperboloid has the coordinate planes as the planes of symmetry and the origin of coordinates as the centre of symmetry.

The most remarkable property of hyperboloid (5) is that it has straight lines lying wholly on it.

Definition 1. A surface in space is said to be an *l-fold ruled surface* if l (and only l) distinct straight lines lying wholly on it pass through any of its points. These straight lines are called *rectilinear generators* (or generatrices) of a ruled surface.

Proposition 1. *The one-sheeted hyperboloid* (5) *is a doubly ruled surface.*

Proof. Let $M_0(x_0, y_0, z_0)$ be an arbitrary point of the hyperboloid (5). A straight line passing through the point M_0 and having a direction vector $\mathbf{a}(l, m, n)$ wholly lies on the hyperboloid (5) if and only if we have identically for t

$$\frac{(x_0+tl)^2}{a^2}+\frac{(y_0+tm)^2}{b^2}-\frac{(z_0+tn)^2}{c^2}=1.$$

Multiplying out and taking into account the fact that under the hypothesis

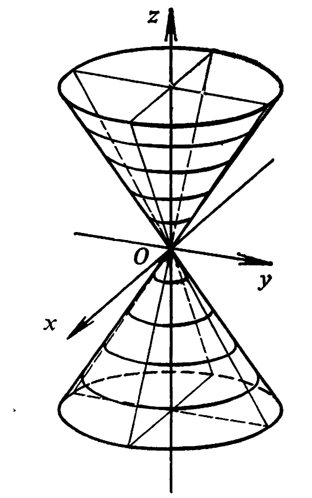
$$\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}-\frac{z_0^2}{c^2}=1$$

we obtain from this two relations:

$$\frac{l^2}{a^2}+\frac{m^2}{b^2}-\frac{n^2}{c^2}=0$$

and

$$\frac{lx_0}{a^2}+\frac{my_0}{b^2}-\frac{nz_0}{c^2}=0.$$



A cone of the second degree

It follows from the first relation that $n \neq 0$ (since otherwise we should have the equation $(l, m, n) = (0, 0, 0)$). Therefore we may assume without loss of generality that $n = c$.

For l and m we then have the equations

因 $n\neq 0$, 所以

$$\frac{l^2}{a^2}+\frac{m^2}{b^2}=1, \qquad \frac{lx_0}{a^2}+\frac{my_0}{b^2}=\frac{z_0}{c}.$$

(l,m,n) 不∥ xoy 面,

Setting in the second equation

因若 $(l,m,n)\parallel xoy$ 面,

$$x_0=x_1+l\frac{z_0}{c} \text{ and } y_0=y_1+m\frac{z_0}{c}$$

则 $(l,m,n)\perp(0,0,1)$

we immediately get (taking into account the first equation)

$$\frac{lx_1}{a^2}+\frac{my_1}{b^2}=0.$$

即 $n=0$ 矛盾.

所以直线与 xoy 面有交点, 设为 $(x_1,y_1,0)$

In addition, since

$$\frac{\left(x_1+l\frac{z_0}{c}\right)^2}{a^2}+\frac{\left(y_1+m\frac{z_0}{c}\right)^2}{b^2}-\frac{z_0^2}{c^2}=\frac{x_1^2}{a^2}+\frac{y_1^2}{b^2}+$$
$$+2\left(\frac{lx_1}{a^2}+\frac{my_1}{b^2}\right)\frac{z_0}{c}+\left(\frac{l^2}{a^2}+\frac{m^2}{b^2}-1\right)\frac{z_0^2}{c^2}=\frac{x_1^2}{a^2}+\frac{y_1^2}{b^2}$$

we get

$$\frac{x_1^2}{a^2}+\frac{y_1^2}{b^2}=1,$$

i.e. the point (x_1, y_1) lies on the striction ellipse of hyperboloid (5).

Therefore the numbers x_1 and y_1 are not both zero and hence the equation

$$\frac{lx_1}{a^2}+\frac{my_1}{b^2}=0$$

uniquely determines the relation $l:m=-a^2y_1:b^2x_1$. We set

$$l=-\frac{a}{b}y_1u, \quad m=\frac{b}{a}x_1u$$

where u is the factor of proportionality. Since

$$\frac{l^2}{a^2}+\frac{m^2}{b^2}=\left(\frac{y_1^2}{b^2}+\frac{x_1^2}{a^2}\right)u^2$$

we have $u^2=1$, i.e. $u=\pm1$. Besides,

$$x_0=x_1+l\frac{z_0}{c}=x_1-u\frac{a}{b}\frac{z_0}{c}y_1,$$

$$y_0=y_1+m\frac{z_0}{c}=y_1+u\frac{b}{a}\frac{z_0}{c}x_1$$

from which we find x_1 and y_1:

$$(*) \qquad x_1=\frac{x_0+u\frac{a}{b}\frac{z_0}{c}y_0}{\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}}, \quad y_1=\frac{y_0-u\frac{b}{a}\frac{z_0}{c}x_0}{\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}}.$$

Conversely, a direct check shows that for any t a point with the coordinates

$$x=x_0-u\frac{a}{b}y_1t, \quad y=y_0+u\frac{b}{a}x_1t, \quad z=z_0+ct,$$

where $u=\pm1$ and the numbers x_1 and y_1 are computed from formulas $(*)$, is on hyperboloid (5) (and the point (x_1, y_1) is on its striction ellipse). □

理解: $\forall (x_0,y_0)\in S$, 则直线

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n}$$

$$\begin{cases} l=-\frac{a}{b}y_1u, \quad m=\frac{b}{a}x_1u \\ x_1=\dfrac{x_0+u\frac{a}{b}\frac{z_0}{c}y_0}{\frac{z_0^2}{a^2}+\frac{y_0^2}{b^2}} \\ y_1=\dfrac{y_0-u\frac{b}{a}\frac{z_0}{c}x_0}{\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}}, \quad u=\pm1 \end{cases}$$

上所有点都在S上.

The obtained parametric equations

$$x = x_0 - u\frac{a}{b} y_1 t,$$
$$y = y_0 + u\frac{b}{a} x_1 t,$$
$$z = z_0 + ct$$

of the rectilinear generators of hyperboloid (5) can be conveniently written in a different way by noting that

$$x_1 = x_0 - u\frac{a}{b} y_1 \left(-\frac{z_0}{c}\right),$$
$$y_1 = y_0 + u\frac{b}{a} x_1 \left(-\frac{z_0}{c}\right),$$
$$0 = z_0 + c\left(-\frac{z_0}{c}\right),$$

i.e. that for $t = -\frac{z_0}{c}$ we have the point $(x_1, y_1, 0)$. Therefore, setting $t' = t + \frac{z_0}{c}$ and omitting the primes we obtain for the rectilinear generators the same equations but with replacement of x_0, y_0, and z_0 by x_1, y_1, and 0. This proves the following proposition:

Proposition 2. *Any rectilinear generator of a hyperboloid of one sheet cuts the striction ellipse of the hyperboloid. The parametric equations of rectilinear generators passing through a point (x_1, y_1) of the striction ellipse are of the form*

$$x = x_1 - u\frac{a}{b} y_1 t,$$
$$y = y_1 + u\frac{b}{a} x_1 t,$$
$$z = ct$$

where $u = \pm 1$. □

Two rectilinear generators of hyperboloid (5) are said to be *of the same sign* if they have the same value of u corresponding to them. Thus all generators are divided into two classes: generators of the same sign belong to the same class, those of opposite signs belong to different classes. These classes are usually called *families* of rectilinear generators of hyperboloid (5).

Theorem 1 (properties of rectilinear generators of a hyperboloid of one sheet). *The following statements hold:*

A. *One and only one rectilinear generator of each family passes through any point of a hyperboloid of one sheet.*

B. *Any two generators of a hyperboloid of one sheet that are of opposite signs lie in the same plane.*

C. *Any two noncoincident generators of a hyperboloid of one sheet that are of the same sign are skew.*

D. *No three mutually disjoint generators of a hyperboloid of one sheet that are of the same sign can be parallel to any plane.*

Proof. Property **A** has in fact been proved above. Let (x_1, y_1) and (x_2, y_2) be two points of the striction ellipse. Then according to Proposition 1 the direction vectors of two distinct rectilinear generators passing through these points are of the form

$$\left(-u_1\frac{a}{b}y_1,\ u_1\frac{b}{a}x_1,\ c\right),\quad \left(-u_2\frac{a}{b}y_2,\ u_2\frac{b}{a}x_2,\ c\right)$$

where $u_1 = \pm 1$ and $u_2 = \pm 1$, and therefore these generators are skew if and only if the determinant

$$\begin{vmatrix} x_2-x_1 & y_2-y_1 & 0 \\ -u_1\frac{a}{b}y_1 & u_1\frac{b}{a}x_1 & c \\ -u_2\frac{a}{b}y_2 & u_2\frac{b}{a}x_2 & c \end{vmatrix} =$$

$$= -u_2abc\left[\frac{(x_2-x_1)(x_2-ux_1)}{a^2}+\frac{(y_2-y_1)(y_2-uy_1)}{b^2}\right]$$

where $u=u_1u_2$, is nonzero. But if $u=1$, then

$$\frac{(x_2-x_1)(x_2-ux_1)}{a^2}+\frac{(y_2-y_1)(y_2-uy_1)}{b^2}=$$

$$=\frac{(x_2-x_1)^2}{a^2}+\frac{(y_2-y_1)^2}{b^2}>0$$

(when $u = 1$ both equations $x_1 = x_2$, $y_1 = y_2$ are impossible, for both generators would then be equal) while if $u = -1$, then

$$\frac{(x_2-x_1)(x_2-ux_1)}{a^2}+\frac{(y_2-y_1)(y_2-uy_1)}{b^2}=$$

$$=\frac{x_2^2-x_1^2}{a^2}+\frac{y_2^2-y_1^2}{b^2}=\left(\frac{x_2^2}{a^2}+\frac{y_2^2}{b^2}\right)-\left(\frac{x_1^2}{a^2}+\frac{y_1^2}{b^2}\right)=0.$$

This proves properties **B** and **C**.

To prove property **D** it is enough to establish that for any three mutually disjoint points (x_1, y_1), (x_2, y_2), and (x_3, y_3) of the striction ellipse the vectors

$$\left(-u\frac{a}{b}y_1,\ u\frac{b}{a}x_1,\ c\right),$$

$$\left(-u\frac{a}{b}y_2,\ u\frac{b}{a}x_2,\ c\right),$$

$$\left(-u\frac{a}{b}y_3,\ u\frac{b}{a}x_3,\ c\right)$$

where $u = \pm 1$, are not coplanar, i.e. that the determinant

$$\begin{vmatrix} -u\frac{a}{b}y_1 & u\frac{b}{a}x_1 & c \\ -u\frac{a}{b}y_2 & u\frac{b}{a}x_2 & c \\ -u\frac{a}{b}y_3 & u\frac{b}{a}x_3 & c \end{vmatrix} = c\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

is nonzero. But this is really so, since the points (x_1, y_1), (x_2, y_2), and (x_3, y_3) are not on the same straight line. □

Let us now resume the enumeration of all the possible types of second-degree surfaces.

Type [6]. Belonging to this type are surfaces having in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0 \tag{6}$$

where $a \geqslant b > 0$, $c > 0$ and $\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}=1$. These surfaces are called *real cones of the second degree*.

The coordinate planes are the planes of symmetry of cone (6) and the origin of coordinates is its centre of symmetry.

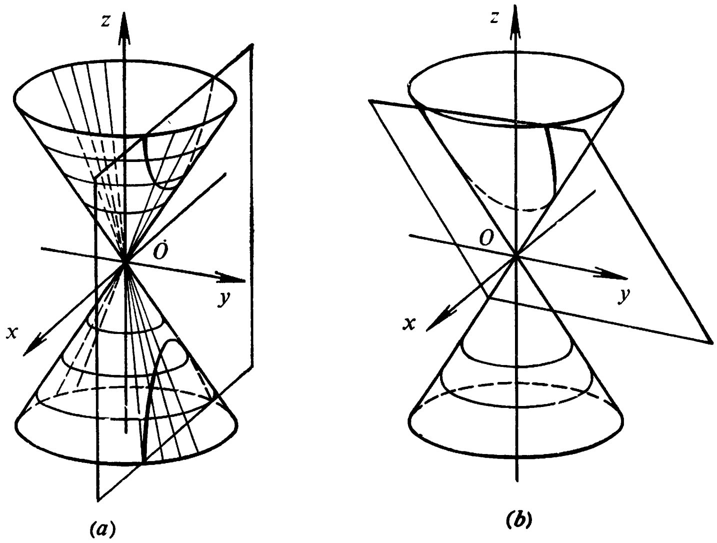
In general a *cone* is a ruled surface all rectilinear generators of which pass through the same point called the *vertex* of the cone.

For any point M_0 (x_0, y_0, z_0) of surface (6) other than the point O $(0, 0, 0)$ each point of the form (tx_0, ty_0, tz_0), i.e. each point of the straight line OM_0, lies on this surface:

$$\frac{(tx_0)^2}{a^2}+\frac{(ty_0)^2}{b^2}-\frac{(tz_0)^2}{c^2}=t^2\left(\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}-\frac{z_0^2}{c^2}\right)=0.$$

Surface (6) is thus indeed a cone.

A *directrix* of a cone is a curve arbitrarily located on the



Sections of a second-degree cone: (a) a hyperbola; (b) a parabola

cone and having the property that any rectilinear generator intersects it in one and only one point.

An example of a directrix of cone (6) is a section of the cone by an arbitrary plane of the form $z = h$, where $h \neq 0$. This section is an ellipse with semiaxes

$$\frac{a}{c}\,|\,h\,|,\quad \frac{b}{c}\,|\,h\,|$$

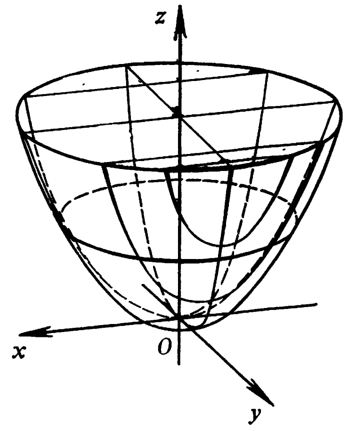
increasing monotonically together with $|\,h\,|$ from zero to $+\infty$. The straight line passing through the centres of ellipses obtained in this way for different h, i.e. the axis Oz, is called the *axis of the cone*. It is interesting that intersecting cone (6) with the planes not perpendicular to the axis we can obtain a circle (prove this!). Therefore cone (6) is also called a *circular cone* or *oblique circular cone*, when one desires to emphasize that generally speaking the axis of the cone does not pass through the centre of that circle. When the axis of the cone passes through the centre of the circle, which occurs if and only if the plane of the circle is perpendicular to the axis or, differently, if $a = b$, the cone (6) is called a *right circular cone*.

Directrices of cone (6), or, more exactly, of this cone without two generators, are also the sections of the cone by the planes $y = h$ and $x = h$, where $h \neq 0$, that are hyperbolas with semiaxes

$$\frac{c}{b}|h|, \quad \frac{a}{b}|h| \quad \text{and} \quad \frac{c}{a}|h|, \quad \frac{b}{a}|h|$$

also increasing monotonically together with $|h|$ from zero to $+\infty$.

The plane $z = 0$ intersects cone (6) in a point and the planes $y = 0$ and $x = 0$ do in a pair of straight lines (in two generators).



An elliptical paraboloid

Plane sections (and directrices) of cone (6) are not only ellipses and hyperbolas but also parabolas. Thus, for example, a parabola is a section of the cone (6) by any plane of the form $z = \frac{c}{a}x + h$, where $h \neq 0$. Indeed, in this plane the numbers x, y are affine (and not rectangular!) coordinates, and the equation of the curve cut out in the plane by the cone (6) has in these coordinates the form

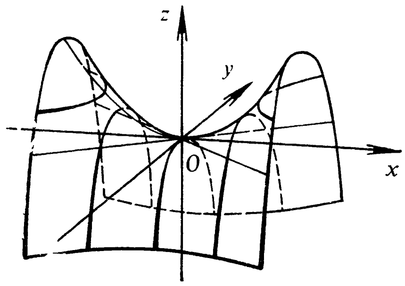
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{\left(\frac{c}{a}x + h\right)^2}{c^2} = 0.$$

Simple transformations reduce this equation to the form

$$y^2 = 2\frac{hb^2}{ac}\left(x + \frac{ha}{2c}\right)$$

which clearly shows that the equation defines a parabola.

We see that the ellipse, the hyperbola, and the parabola



A hyperbolic paraboloid

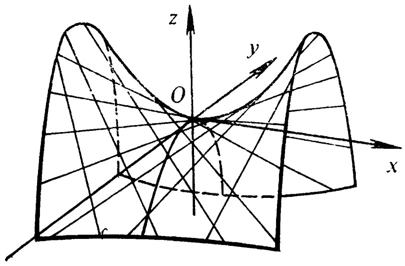
are plane sections of cone (6). For this reason these curves are usually called *conic sections*.

Type [7]. Of this type are surfaces that have in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{p}+\frac{y^2}{q}=2z \tag{7}$$

where $p \geqslant q > 0$. They are called *elliptical paraboloids*.

The plane $z = h$ does not intersect paraboloid (7) when



Rectilinear generators of a hyperbolic paraboloid

$h < 0$, has a single point (0, 0, 0) in common with it when $h = 0$ and intersects the paraboloid in an ellipse with semi-axes

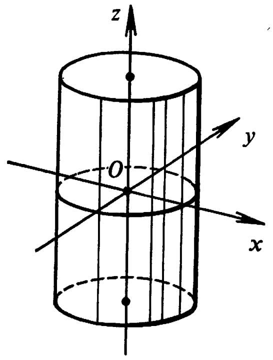
$$\sqrt{2hp}, \quad \sqrt{2hq}$$

increasing monotonically together with h from zero to $+\infty$ when $h > 0$.

The planes $y = h$ and $x = h$ intersect the paraboloid (7) in parabolas with focal parameters p and q, with vertices at the points $\left(0, h, \frac{h^2}{2q}\right)$ and $\left(h, 0, \frac{h^2}{2p}\right)$ and upturned "horns".

The planes $x = 0$ and $y = 0$ are the planes of symmetry of paraboloid (7). When $p \neq q$ it has no other planes of symmetry.

It is clear that an elliptical paraboloid (like an ellipsoid and a hyperboloid of two sheets) is not a ruled surface.



An elliptical cylinder

Type [8]. Of this type are surfaces having in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{p} - \frac{y^2}{q} = 2z \tag{8}$$

where $p > 0$, $q > 0$. They are called *hyperbolic paraboloids.*

Of all second-degree surfaces the shape of a hyperbolic paraboloid is the most difficult to imagine.

The plane $z = h$ intersects paraboloid (8), when $h < 0$, in a hyperbola with semiaxes

$$\sqrt{-2qh} \qquad \sqrt{-2ph}$$

decreasing monotonically from $+\infty$ to zero as h increases from $-\infty$ to zero. The real axis of the hyperbola is parallel to the axis Ox and the imaginary axis is parallel to the axis Oy.

The plane $z = 0$ intersects hyperboloid (8) in a pair of straight lines that has (in coordinates x, y) the equation

$$\frac{x^2}{p} - \frac{y^2}{q} = 0$$

and the plane $z = h$ intersects (8), when $h > 0$, in a hyperbola with semiaxes

$$\sqrt{2ph}, \quad \sqrt{2qh}$$

increasing monotonically together with h from zero to $+\infty$. Contrary to the case $h < 0$ the real axis of this hyperbola is parallel to the axis Oy and the imaginary axis is parallel to the axis Ox.

The planes $y = h$ and $x = h$ intersect the hyperbolic paraboloid (8) in parabolas with focal parameters p and q and with vertices respectively at the points $\left(0,\ h,\ -\frac{h^2}{2q}\right)$ and $\left(h,\ 0,\ \frac{h^2}{2p}\right)$. The "horns" of the first parabola are turned up and those of the second are turned down. The vertices of parabolas cut out by the planes $y = h$ lie on the parabola cut out by the plane $x = 0$ and those of parabolas cut out by the planes $x = h$ lie on the parabola cut out by the plane $y = 0$.

The planes $x = 0$ and $y = 0$ are the planes of symmetry of the hyperbolic paraboloid (8). This paraboloid has no other planes of symmetry.

Proposition 3. *A hyperbolic paraboloid is a doubly ruled surface.*

Proof. Let $M_0\ (x_0,\ y_0,\ z_0)$ be an arbitrary point of paraboloid (8). If a straight line

$$x = x_0 + tl,$$
$$y = y_0 + tm,$$
$$z = z_0 + tn$$

passing through the point M_0 lies wholly on the paraboloid (8), then we must have identically for t the equation

$$\frac{(x_0+tl)^2}{p} - \frac{(y_0+tm)^2}{q} = 2\,(z_0+tn),$$

i.e. the equation

$$t^2\left(\frac{l^2}{p}-\frac{m^2}{q}\right)+2t\left(\frac{lx_0}{p}-\frac{my_0}{q}-n\right)=0$$

which is possible only if

$$\frac{l^2}{p}-\frac{m^2}{q}=0$$

and

$$\frac{lx_0}{p}-\frac{my_0}{q}-n=0.$$

Up to proportionality these equations have two and only two solutions:

$$l:m:n=\sqrt{p}:u\sqrt{q}:\left(\frac{x_0}{\sqrt{p}}-u\frac{y_0}{\sqrt{q}}\right)$$

where $u = \pm 1$.

Conversely, a direct check shows that both straight lines

$$x=x_0+t\sqrt{p},$$
$$y=y_0+ut\sqrt{q},$$
$$z=z_0+t\left(\frac{x_0}{\sqrt{p}}-u\frac{y_0}{\sqrt{q}}\right)$$

lie wholly on the paraboloid (8). □

If $\frac{x_0}{\sqrt{p}} - u\frac{y_0}{\sqrt{q}} \neq 0$, then for $t = t_1$, wherej

$$t_1=-\frac{z_0}{\frac{x_0}{\sqrt{p}}-u\frac{y_0}{\sqrt{q}}}=-\frac{1}{2}\left(\frac{x_0}{\sqrt{p}}+u\frac{y_0}{\sqrt{q}}\right)$$

the found rectilinear generators intersect the plane $z = 0$ and, consequently, one of the straight lines

$$\frac{x}{\sqrt{p}}-\frac{y}{\sqrt{q}}=0 \quad \text{or} \quad \frac{x}{\sqrt{p}}+\frac{y}{\sqrt{q}}=0$$

of this plane that lie on the paraboloid. Since

$$\frac{x_0+t_1\sqrt{p}}{\sqrt{p}}+u\frac{y_0+ut_1\sqrt{q}}{\sqrt{q}}=\left(\frac{x_0}{\sqrt{p}}+u\frac{y_0}{\sqrt{q}}\right)+2t_1=0$$

that straight line is the line

$$\frac{x}{\sqrt{p}}+u\frac{y}{\sqrt{q}}=0$$

with the parametric equations

$$\begin{aligned} x &= \tau\sqrt{p}, \\ y &= -\tau u\sqrt{q}, \end{aligned} \quad -\infty < \tau < +\infty.$$

The point of intersection $(x_0 + t_1\sqrt{p},\ y_0 + ut_1\sqrt{q})$ has a nonzero value

$$\tau_1 = \frac{x_0 + t_1\sqrt{p}}{\sqrt{p}} = \frac{x_0}{\sqrt{p}} + t_1 = \frac{1}{2}\left(\frac{x_0}{\sqrt{p}} - u\frac{y_0}{\sqrt{q}}\right)$$

of the parameter τ. Geometrically the number τ_1 is a distance of the point of intersection from the origin divided by the length of the vector $(\sqrt{p}, -u\sqrt{q})$, i.e. by $\sqrt{p+q}$.

Setting $t' = t - t_1$ we derive from this that

$$x = x_0 + t\sqrt{p} = (x_0 + t_1\sqrt{p}) + (t - t_1)\sqrt{p} = (t' + \tau_1)\sqrt{p},$$

$$y = y_0 + tu\sqrt{q} = (y_0 + t_1 u\sqrt{q}) + u(t - t_1)\sqrt{q} = \\ = u(t' - \tau_1)\sqrt{q},$$

$$z = z_0 + t\left(\frac{x_0}{\sqrt{p}} - u\frac{y_0}{\sqrt{q}}\right) = (t - t_1)\left(\frac{x_0}{\sqrt{p}} - u\frac{y_0}{\sqrt{q}}\right) = 2t'\tau_1.$$

If, however, $\frac{x_0}{\sqrt{p}} - u\frac{y_0}{\sqrt{q}} = 0$ and hence (according to equation (8)) $z_0 = 0$, then the rectilinear generator under consideration lies wholly in the plane $z = 0$ and passes through the origin. Setting as before $t' = t - t_1$, we can write the parametric equation of that generator as

$$x = t'\sqrt{p}, \quad y = ut'\sqrt{q}, \quad z = 0,$$

i.e. in the same form as before but with $\tau_1 = 0$.

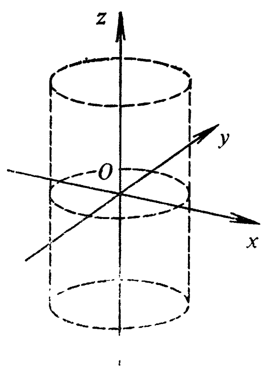
Thus we have proved the following proposition (we again denote t' by t and τ_1 by τ):

Proposition 4. *Each rectilinear generator of the hyperbolic paraboloid* (8) *intersects the plane* $z = 0$ *or is in this plane. Its parametric equations can be written as*

$$\begin{aligned} x &= (t + \tau)\sqrt{p}, \\ y &= u(t - \tau)\sqrt{q}, \\ z &= 2t\tau, \end{aligned}$$

where $u = \pm 1$ *and the number* τ *is either zero (if the generator is wholly in the plane* $z = 0$*) or equal to the distance from the origin to the point of intersection of the generator with the plane* $z = 0$ *divided by* $\sqrt{p+q}$. □

As in the case of a hyperboloid of one sheet two rectilinear generators are said to be *of the same sign* (to belong to the same *family* of generators) if the same value of u corresponds to them.



An imaginary elliptical cylinder

Theorem 2 (properties of rectilinear generators of a hyperbolic paraboloid). *The following statements hold:*

A. *One and only one rectilinear generator of each family passes through any point of a hyperbolic paraboloid.*

B. *Any two generators of opposite signs in a hyperbolic paraboloid intersect (and hence are in the same plane).*

C. *Any two noncoincident generators of the same sign in a hyperbolic paraboloid are skew.*

D. *Generators of the same family are all parallel to the same plane.*

We see that the properties of generators of a hyperbolic paraboloid closely correspond to those of generators of a hyperboloid of one sheet (property **B** being even somewhat stronger, because of the excluded parallelism of the generators). The only fundamental difference lies in property **D**.

Proof of Theorem 2. Property **A** has already been proved above in fact. For two generators

$$x = (t+\tau_1)\sqrt{p}, \qquad x = (t+\tau_2)\sqrt{p},$$

$$y = u_1(t-\tau_1)\sqrt{q}, \quad \text{and} \quad y = u_2(t-\tau_2)\sqrt{q},$$

$$z = 2t\tau_1 \qquad z = 2t\tau_2$$

not to be skew, it is necessary and sufficient that the determinant

$$\begin{vmatrix} \tau_2\sqrt{p}-\tau_1\sqrt{p} & -u_2\tau_2\sqrt{q}+u_1\tau_1\sqrt{q} & 0 \\ \sqrt{p} & u_1\sqrt{q} & 2\tau_1 \\ \sqrt{p} & u_2\sqrt{q} & 2\tau_2 \end{vmatrix} = $$
$$= 2\sqrt{pq}\,(u_1+u_2)\,(\tau_2-\tau_1)^2$$

should be zero. This proves property **C**. But to prove property **B** it is necessary in addition to note that since

$$\begin{vmatrix} \sqrt{p} & u_1\sqrt{q} \\ \sqrt{p} & u_2\sqrt{q} \end{vmatrix} = \sqrt{pq}\,(u_2-u_1)$$

generators of opposite signs cannot be parallel.

Finally, it is clear that vectors of the form $(\sqrt{p}, u\sqrt{q}, 2\tau)$ are all parallel to the plane

$$\frac{x}{\sqrt{p}} - u\frac{y}{\sqrt{q}} = 0.$$

This proves property **D**. □

Type [9]. Of this type are surfaces that have in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{9}$$

where

$$a \geqslant b > 0.$$

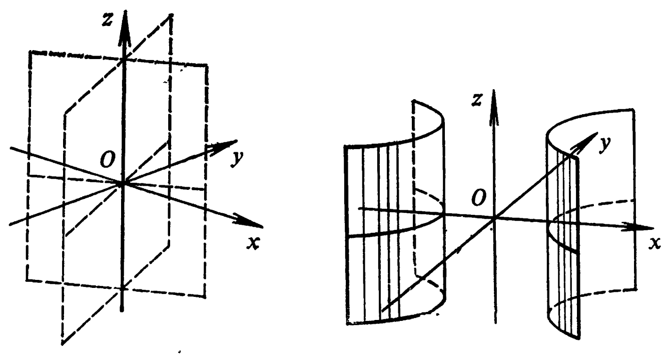
They are called *elliptical cylinders*.

The coordinate planes are planes of symmetry of cylinder (9) and the origin is its centre of symmetry. Cylinder (9) has no other planes of symmetry for $a \neq b$.

In general, a *cylinder* is a ruled surface whose rectilinear generators are all parallel to one another. If a cylinder has a centre of symmetry, the straight line passing through this centre and parallel to the generators is said to be the *axis of the cylinder* (it is its axis of symmetry).

It is clear that for any point $M_0\,(x_0, y_0, z_0)$ of surface (9) each point of the form (x_0, y_0, z), i.e. each point of the straight line passing through the point (x_0, y_0) and parallel to the axis Oz is on this surface. Surface (9) is thus indeed a cylinder. The axis of the cylinder is the axis Oz.

A *directrix* of a cylinder is a curve arbitrarily located on the cylinder and having the property that each generator intersects it in one and only one point. In particular, plane directrices are exactly the curves cut out on the cylinder by planes not parallel to the generators. For example, each plane $z = h$ (perpendicular to the generators of the cylinder



A pair of imaginary planes *A hyperbolic cylinder*

(9)) intersects the cylinder (9) in an ellipse having in the coordinates x, y defined in this plane equation (9). This accounts for the use of the adjective "elliptical" in the name of cylinder (9).

It is interesting that one of the plane directrices of cylinder (9) is a circle (prove it!). Therefore elliptical cylinders are also called *circular cylinders* or, more precisely, *oblique circular cylinders* when it should be emphasized that the plane cutting out the circle is not perpendicular to the axis. When the plane is perpendicular to the axis, i.e. when $a = b$, cylinder (9) is said to be a *right circular cylinder*. It is these cylinders that are considered in the school mathematics course.

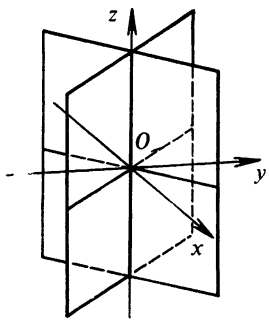
The remaining second-degree surfaces are all also cylinders (whose directrices are other second-degree curves).

Type [10]. Belonging to this type are surfaces having in some system of rectangular coordinates x, y, z an equation of the form

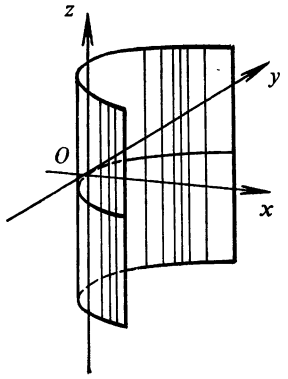
$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1 \tag{10}$$

where $a \geqslant b > 0$. They have no real points and are called *imaginary elliptical cylinders.*

Type [11]. Belonging to this type are surfaces that have



A pair of intersecting planes



A parabolic cylinder

in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=0 \tag{11}$$

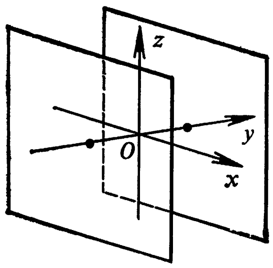
where $a > 0$, $b > 0$ and $\frac{1}{a^2}+\frac{1}{b^2}=1$.

The real points of every surface of this kind make up a straight line. In a real-complex space surface (11) is a pair of imaginary (complex conjugate) planes intersecting in a real straight line.

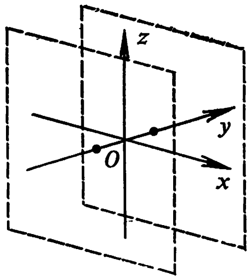
Type [12]. Belonging to this type are surfaces that have in some system of rectangular coordinates x, y, z an equation of the form

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1 \tag{12}$$

where $a > 0$, $b > 0$. They are called *hyperbolic cylinders*. Each plane $z = h$ intersects cylinder (12) in a hyperbola



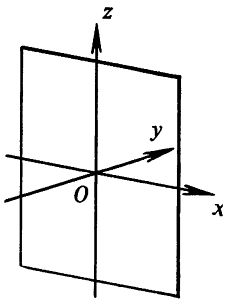
A pair of parallel distinct planes



A pair of imaginary distinct planes

having in the coordinates x, y in this plane equation (12).

Type [13]. Belonging to this type are surfaces that have



A pair of coincident planes

in some system of rectangular coordinates x, y, z an equation of the form

(13) $$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$$

where $a > 0$, $b > 0$ and $\frac{1}{a^2} + \frac{1}{b^2} = 1$. Every surface of this kind is a pair of intersecting planes.

Type [14]. Belonging to this type are surfaces that have in some system of rectangular coordinates x, y, z an equation of the form

$$(14) \qquad y^2 = 2px$$

where $p > 0$. They are called *parabolic cylinders.* Each plane $z = h$ intersects the cylinder (14) in a parabola having in the coordinates x, y in this plane equation (14).

Type [15]. Belonging to this type are surfaces that have in some system of rectangular coordinates x, y, z an equation of the form

$$(15) \qquad y^2 - b^2 = 0$$

where $b > 0$. They are pairs of parallel distinct planes.

Type [16]. Belonging to this type are surfaces that have in some system of rectangular coordinates x, y, z an equation of the form

$$(16) \qquad y^2 + b^2 = 0$$

where $b > 0$. They are pairs of imaginary (complex conjugate and distinct) planes.

Type [17]. Belonging to this type are surfaces having in some system of rectangular coordinates x, y, z an equation of the form

$$(17) \qquad y^2 = 0.$$

Every surface of this kind may be regarded as a pair of coincident planes.

The *theorem on the classification of second-degree surfaces* states that the enumerated seventeen types cover all second-degree surfaces in Euclidean space, no surface having two systems of rectangular coordinates in which it would have different equations (1) to (17). All second-degree surfaces are thus exhausted by:

(a) ellipsoids (real [1] and imaginary [2]),
(b) hyperboloids (of one sheet [5] and of two sheets [4]),
(c) paraboloids (elliptical [7] and hyperbolic [8]),
(d) second-degree cones (imaginary [3] and real [6]),
(e) second-degree cylinders (real elliptical [9], imaginary elliptical [10], parabolic [14] and hyperbolic [12]),

(f) pairs of (imaginary intersecting [11], real intersecting [13], real parallel and distinct [15], imaginary parallel [16], real coincident [17]) planes.

We shall not prove this theorem now, since in the next semester we are going to prove its extension to the case of spaces of any dimension.

Lecture 24

Coordinates of a straight line • Pencils of straight lines • Ordinary and ideal pencils • Extended planes • Models of projective-affine geometry

Thus far we have assumed that the basic element of geometry is the point. But this is not at all obligatory. It is possible, for example, to construct a geometry whose basic elements are straight lines. We shall first consider straight lines in the (affine) plane.

Assuming that some system of affine coordinates x, y is fixed in a plane we may characterize any straight line in the plane by three numbers A, B, C, the coefficients of its equation

$$Ax + By + C = 0. \tag{1}$$

Just as coordinates uniquely determine a point, these numbers uniquely determine a straight line. It is natural therefore to call them the *coordinates of a straight line.* However, while a point uniquely determines its coordinates, a straight line does not, its coordinates being determined only up to proportionality. This property of the coordinates of a straight line is called the *homogeneity* of the coordinates. For the purposes of notation we shall designate a straight line with coordinates A, B, C by the symbol $(A:B:C)$.

Another difference between the coordinates of straight lines and those of points is that *not every* triple of numbers A, B, C may be the triple of coordinates of a straight line: for this to be the case it is required that at least one of the numbers A or B should be nonzero, whereas the number C is not constrained in any way. This asymmetry of coordinates (a peculiar role played by the coordinate C) leads, as we shall later see, to numerous reservations and complications. But for the present we are forced to put up with it.

If for a straight line (1) the coefficient A is nonzero, then we may characterize this straight line by the *nonhomogeneous coordinates* $\frac{B}{A}$, $\frac{C}{A}$ that are already uniquely determined by the straight line. Straight lines may therefore be represented by points of a plane, associating with a straight line (1) the point $\left(\frac{B}{A}, \frac{C}{A}\right)$. However, in doing so one omits from consideration straight lines with $A = 0$. Homogeneous coordinates are good just because they presuppose no omissions.

We know that the coordinates of points $M(x, y)$ of a straight line passing through two distinct points $M_0(x_0, y_0)$ and $M_1(x_1, y_1)$ are expressed by the formulas

$$x = (1-t)x_0 + tx_1, \quad y = (1-t)y_0 + ty_1$$

by analogy it is possible to introduce into consideration for any two distinct straight lines $(A_0:B_0:C_0)$ and $(A_1:B_1:C_1)$ having nonhomogeneous coordinates $\left(\frac{B_0}{A_0}, \frac{C_0}{A_0}\right)$ and $\left(\frac{B_1}{A_1}, \frac{C_1}{A_1}\right)$ various straight lines $(A:B:C)$ for which

$$(2) \qquad \frac{B}{A} = (1-t)\frac{B_0}{A_0} + t\frac{B_1}{A_1}, \quad \frac{C}{A} = (1-t)\frac{C_0}{A_0} + t\frac{C_1}{A_1}.$$

The set of all these straight lines is then an analogue of the straight line passing through two points.

Writing now equation (2) as

$$\frac{B}{A} = \frac{(1-t)A_1B_0 + tA_0B_1}{A_0A_1} = \frac{[(1-t)A_1]B_0 + [tA_0]B_1}{[(1-t)A_1]A_0 + [tA_0]A_1},$$

$$\frac{C}{A} = \frac{(1-t)A_1C_0 + tA_0C_1}{A_0A_1} = \frac{[(1-t)A_1]C_0 + [tA_0]C_1}{[(1-t)A_1]A_0 + [tA_0]A_1}$$

we obtain for the coefficients A, B, C the following expressions:

$$\begin{aligned} A &= [(1-t)\rho A_1]A_0 + [t\rho A_0]A_1, \\ B &= [(1-t)\rho A_1]B_0 + [t\rho A_0]B_1, \\ C &= [(1-t)\rho A_1]C_0 + [t\rho A_0]C_1, \end{aligned}$$

where ρ is an arbitrary factor of proportionality. Setting

$$\mu = (1 - t)\,\rho A_1, \quad \nu = t\rho A_0$$

we can write these formulas in the following form:

$$\begin{aligned} A &= \mu A_0 + \nu A_1 \\ B &= \mu B_0 + \nu B_1, \\ C &= \mu C_0 + \nu C_1. \end{aligned} \tag{3}$$

In deriving these formulas we used nonhomogeneous coordinates, i.e. assumed all straight lines to be non-parallel to the axis of ordinates. However, formulas (3) make sense for arbitrary straight lines $(A_0:B_0:C_0)$, $(A_1:B_1:C_1)$ and determine for any permissible values of the parameters μ and ν (i.e. such that either $A \neq 0$ or $B \neq 0$) some straight line $(A:B:C)$. Of interest, however, is only the case where the straight lines $(A_0:B_0:C_0)$ and $(A_1:B_1:C_1)$ are distinct since if $(A_0:B_0:C_0) = (A_1:B_1:C_1)$, then we obtain the same straight line $(A_0:B_0:C_0)$ for any μ and ν.

All this motivates the following definition:

Definition 1. The set of all straight lines $(A:B:C)$ obtained from formulas (3) for various permissible values of parameters μ and ν is called a *pencil of straight lines* which is determined by straight lines $(A_0:B_0:C_0)$ and $(A_1:B_1:C_1)$ (assumed to be distinct).

Pencils of straight lines are thus analogues of straight lines in the geometry of points.

If we set $f = A_0x + B_0y + C_0$ and $g = A_1x + B_1y + C_1$ the straight lines of the pencil (3) will have an equation of the form

$$\mu f + \nu g = 0.$$

It is often convenient to use this notation for a pencil.

Straight lines in a plane may be given by equations of the form (1) rather than by parametric equations. The corresponding analogues in the geometry of straight lines are sets of straight lines $(A:B:C)$ defined by conditions of the kind

$$K\left(\frac{B}{A}\right) + L\left(\frac{C}{A}\right) + M = 0$$

or, in a more general form (not assuming that $A \neq 0$) of the kind

$$KB + LC + MA = 0$$

where K, L, M are fixed numbers. Somewhat changing the notation we shall write the last relation as

$$AX_0 + BY_0 + CZ_0 = 0 \tag{4}$$

where X_0, Y_0, Z_0 are some numbers. In close analogy with the "point" case it should be assumed that at least one of the numbers X_0, Y_0 is nonzero. However, we shall somewhat extend (and "symmetrize") the statement to include only the requirement that at least one of the numbers X_0, Y_0, Z_0 should be nonzero (otherwise relation (4) is satisfied identically).

It is remarkable that this approach to the concept of a pencil should lead to the same result:

Proposition 1. *A set of straight lines is a pencil of straight lines if and only if the straight lines $(A:B:C)$ it contains are characterized by a condition of the form* (4).

For this reason relation (4) is called the *equation of a pencil.*

We preface the proof of Proposition 1 with the following lemma:

Lemma. *If two distinct straight lines $(A_0:B_0:C_0)$ and $(A_1:B_1:C_1)$ satisfy relation* (4), *then the pencil* (3) *they determine coincides with the set of straight lines satisfying relation* (4).

Proof. If a straight line $(A:B:C)$ is in the pencil (3), then

$$\begin{aligned} AX_0+BY_0+CZ_0 &= \\ &= (\mu A_0 + \nu A_1) X_0 + (\mu B_0 + \nu B_1) Y_0 + (\mu C_0 + \nu C_1) Z_0 = \\ &= \mu (A_0X_0 + B_0Y_0 + C_0Z_0) + \nu (A_1X_0 + B_1Y_0 + C_1Z_0) = 0. \end{aligned}$$

Conversely, let a straight line $(A:B:C)$ satisfy relation (4). Then the columns of the matrix

$$\begin{pmatrix} A & B & C \\ A_0 & B_0 & C_0 \\ A_1 & B_1 & C_1 \end{pmatrix}$$

are linearly dependent and so are the rows. But under the assumption the last two rows are not proportional, i.e. are linearly independent. The first row is therefore linearly expressible in terms of the last two and, hence, the straight line $(A:B:C)$ is in the pencil (3). □

Proof of Proposition 1. Suppose we are given the pencil (3). Consider a system of two homogeneous linear equations in three unknowns

$$A_0X + B_0Y + C_0Z = 0,$$
$$A_1X + B_1Y + C_1Z = 0.$$

We know from algebra that such a system necessarily has a nontrivial solution $(X_0, Y_0, Z_0) \neq (0, 0, 0)$. This solution gives us relation (4) with respect to which the given straight lines satisfy the conditions of the lemma. Therefore the pencil (3) is characterized by this relation.

Conversely, suppose we are given relation (4). To apply the lemma and thus prove the proposition, it suffices to find two nonproportional triples (A_0, B_0, C_0) and (A_1, B_1, C_1) satisfying relation (4) and capable of serving as triples of coordinates of straight lines (i.e. such that at least one of the numbers A_0, B_0 or, respectively, A_1, B_1 is nonzero). But this is easily done. Indeed, if, for example $Z_0 \neq 0$, we may set

$$A_0 = Z_0,\ B_0 = 0,\quad C_0 = -X_0,$$
$$A_1 = 0,\quad B_1 = Z_0,\ C_1 = -Y_0. \ \square$$

In connection with Proposition 1 it is interesting to clarify the geometrical meaning of relation (4).

Suppose first that $Z_0 \neq 0$. Then setting $x_0 = \frac{X_0}{Z_0}$, $y_0 = \frac{Y_0}{Z_0}$ we may write the condition (4) as the relation

$$Ax_0 + By_0 + C = 0$$

implying that the straight line $(A:B:C)$ passes through a point M_0 (x_0, y_0). We therefore derive that *the totality of all straight lines passing through a given point M_0 is a pencil.* □

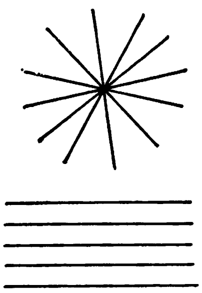
Definition 2. Pencils of such a form, i.e. pencils with equation (4) for which $Z_0 \neq 0$ are called *ordinary* pencils. A point M_0 through which the straight lines of an ordinary pencil pass are called the *centre* of the pencil.

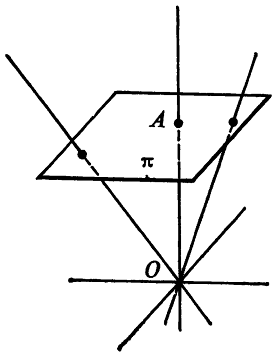
On the other hand, pencils having equation (4) with $Z_0 = 0$ are called *ideal* pencils.

The straight lines of an ideal pencil are thus characterized by a condition of the form

$$AX_0 + BY_0 = 0$$

implying, as we know, that the straight line $(A{:}B{:}C)$ is parallel to a vector with the coordinates X_0, Y_0. This proves





Pencils of straight lines

that *ideal pencils are exactly sets consisting of all straight lines parallel to one another* (i.e. having a common direction vector $\mathbf{a}\,(X_0,\ Y_0)$). □

Note that for an ordinary pencil (3) we have the relation

$$\frac{A_1}{A_0} \neq \frac{B_1}{B_0}$$

and therefore for any μ and ν we obtain some straight line $(A{:}B{:}C)$. But for an ideal pencil there are "prohibited" pairs $(\mu,\ \nu)$, viz. pairs for which $\mu{:}\nu = -A_1{:}A_0$. These pairs have no corresponding straight line.

The statements proved may be combined into the following proposition:

Proposition 2. *For any two intersecting straight lines* $f = 0$ *and* $g = 0$ *each straight line of the form*

$$\mu f + \nu g = 0 \tag{5}$$

passes through a point M_0 *they have in common, and, conversely, any straight line passing through the point* M_0 *is of the form* (5).

Similarly, for any two parallel lines $f = 0$ *and* $g = 0$ *each straight line* (5) *is parallel to them, and, conversely, any straight line parallel to them is of the form* (5).

In the first case (5) *is the equation of a straight line for any* μ *and* ν *and in the second we have a single* (*up to proportionality*) *exception* (*arising when* $\mu:\nu = -A_1:A_0$). □

Although pencils do not explicitly figure in this proposition (and it can be proved therefore without them), still it is only on the basis of the concept of a pencil that one can fully comprehend its geometrical meaning.

The numerous reservations we had to make above and the incomplete symmetry between points and straight lines cry for elimination. In the light of what has been said above it is sufficiently clear that after all this is due to ideal pencils having no centres. Let us try therefore to provide each ideal pencil with a centre.

Since all points of the plane are already occupied by the centres of ordinary pencils, to do this we must *extend* the plane and add new points to it which it is appropriate to call "ideal" points. It is necessary to add as many of these points as there are ideal pencils, i.e. classes of parallel lines. If we add fewer than necessary, some pencils will remain without centres, and if we add more than required some points will not be the centres of pencils. We thus arrive at the following definition:

Definition 3. A set $\mathscr{A}^+$ is said to be an *extended plane* if it contains an affine plane $\mathscr{A}$ and we are given a mapping $d \mapsto d^+$ of a set of all straight lines d of the plane $\mathscr{A}$ onto the complement $\mathscr{A}^+ \setminus \mathscr{A}$ having the following properties:

(a) any element of $\mathscr{A}^+ \setminus \mathscr{A}$ has the form d^+ for some straight line d;

(b) $d_1^+ = d_2^+$ if and only if the straight lines d_1 and d_2 are parallel.

The extended plane $\mathscr{A}$ is also called a *projective-affine plane*, and its geometry is called *projective-affine geometry*.

Elements of a set $\mathscr{A}^+ \setminus \mathscr{A}$ are called *ideal points* of the projective-affine plane $\mathscr{A}$ and the points of $\mathscr{A}$ are called *ordinary points*.

(Ordinary) straight lines of a plane $\mathscr{A}$ are straight lines d of the plane $\mathscr{A}$ to each of which an ideal point d^+ is added.

These straight lines are designated by the same symbols d, due to which the point d^+ has to be called an ideal point of a *straight line* d. It is sometimes convenient, however, to understand by a straight line in this phrase the initial straight line of the plane $\mathscr{A}$ as well.

It thus turns out that the "growth" $\mathscr{A}^+\setminus\mathscr{A}$ possesses a characteristic property of straight lines, having as it does a single point in common with each straight line d. For this reason it is appropriate to call it an *ideal straight line* of the projective-affine plane $\mathscr{A}^+$.

Pencils of a projective-affine plane $\mathscr{A}^+$ are ordinary pencils of a plane $\mathscr{A}$ (it is implied, of course, that to each straight line of a pencil its ideal point is added) and ideal pencils of the plane $\mathscr{A}$ to which an ideal straight line is added.

Accordingly, any pencil in a plane $\mathscr{A}^+$ consists of all straight lines passing through some fixed point (ordinary one for ordinary pencils and ideal one for ideal pencils). This point is called the *centre* of a pencil.

In projective-affine geometry, just as in affine geometry, a single straight line passes through any two distinct points but in addition (a property affine geometry lacks) any two distinct straight lines intersect in a single point.

In complete analogy with this, in projective-affine geometry two distinct straight lines determine a single pencil and (what is incorrect in affine geometry) two distinct pencils have a single straight line in common.

It is very important to note, and it is not specified in Definition 3, over what field the affine plane $\mathscr{A}$ is given. If this plane is an affine plane over a field $\mathbb{K}$, then the projective-affine plane $\mathscr{A}^+$ is also called a plane *over the field* $\mathbb{K}$. In particular, for $\mathbb{K} = \mathbb{C}$ we obtain a complex projective-affine plane. If $\mathscr{A}$ is a real-complex plane, then $\mathscr{A}^+$ is called a *real-complex projective-affine plane*.

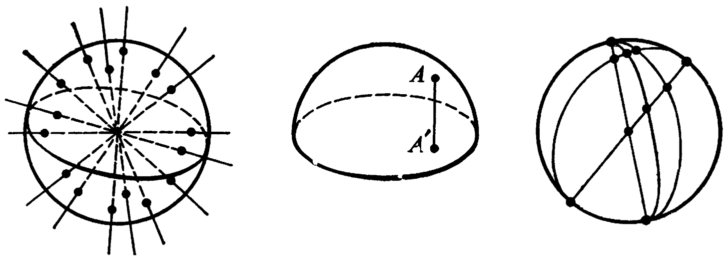
What is more, the plane $\mathscr{A}$ in Definition 3 may well be assumed to be a Euclidean plane. We then have a *projective Euclidean plane* $\mathscr{A}^+$, which even has two variants, a *real* and a *real-complex* one.

We thus have a whole bunch of different "planes" and hence different "geometries". In relation to each geometrical construction and theorem one should always have a clear idea of what geometry is, strictly speaking, involved, i.e.

what plane is used. Although this is nearly always clear from the context, still there are situations in which lack of proper attention to this point may result in an error.

To visualize a projective-affine plane consider in affine space an arbitrary plane π and some point O not lying in that plane. Let $\mathscr{A}$ be a set of all straight lines passing through the point O and intersecting the plane π. This set is in a bijective correspondence with the set of points of the plane π (a straight line OA corresponds to a point A of the plane) and therefore it may also be considered to be an affine plane (or a *model* or *interpretation* of affine geometry, as some people sometimes prefer to say). This model considers as points straight lines of the form OA and takes as straight lines sets of straight lines of the form OA that lie in the planes passing through the point O (and, certainly, not parallel to the plane π). For this reason straight lines are sometimes said to be represented in the model under consideration by planes (passing through the point O and not parallel to the plane π).

A natural extension of a set $\mathscr{A}$ is a set $\mathscr{A}_0$ of *all* straight lines in space passing through the point O (such a set is



(a) *(b)* *(c)*

(a) The sphere model, (b) a change to the circle model, (c) the circle model

called a *bundle of straight lines*). Let d be an arbitrary straight line in $\mathscr{A}$. It is represented by some plane in space passing through the point O. This plane has a single straight line passing through the point O and parallel to the plane π. We take this straight line as an ideal point of the straight line d. It is clear that in this way we transform the bundle $\mathcal{A}_O$ into a projective-affine plane. This plane is called a *bundle* model (interpretation) of projective-affine geometry.

By surrounding the point O with an arbitrary sphere (the surface of a ball) we may change from straight lines to pairs of points they cut out on the sphere. We obtain a new interpretation of projective-affine geometry. In this interpretation points are represented by pairs of antipodal (diametrically opposite) points of the sphere and straight lines are represented by great circles of the sphere (circles cut out on the sphere by the planes passing through the centre). This model is called the *sphere* model.

In this model an ideal straight line is represented by the great circle (equator) parallel to the plane π. Consider one of the hemispheres into which the equator divides the sphere. Each straight line passing through the point O and intersecting the plane π (the ordinary point of the model $\mathcal{A}_O$) has a single point in common with that hemisphere. By orthogonally projecting this point onto the equator plane we obtain some interior point of the circle $\mathcal{C}$ bounded by the equator (in the plane passing through the point O and parallel to the plane π). We may thus represent ordinary points of the plane $\mathcal{A}_O$ by interior points of the circle $\mathcal{C}$. As to the ideal points of the plane we are still forced to represent them by pairs of diametrically opposite points of the circumference of $\mathcal{C}$. The resulting model of projective-affine geometry is called the *circle* model. Keeping in mind precisely this model a *projective-affine plane is sometimes said to be obtained from a circle by identifying diametrically opposite points of its boundary.*

The circle model appears to be the most demonstrative of those presented, if only because ordinary points are represented by "real" points in it. Its disadvantage is that straight lines are represented in it in general by curved lines (semiellipses). These lines join diametrically opposite points of the circumference and hence any two of them have a point in common. The fact that in a projective-affine plane two distinct straight lines have a single point in common is thus visually palpable in the circle model.

Lecture 25

Homogeneous affine coordinates • Equations of straight lines in homogeneous coordinates • Second-degree curves in the projective-affine plane • Circles in the projective-Euclidean real-complex plane • Projective planes • Homogeneous affine coordinates in the bundle of straight lines • Formulas for the transformation of homogeneous affine coordinates • Projective coordinates • Second-degree curves in the projective plane

Let $\mathscr{A}^+$ be an arbitrary projective-affine plane, let $\mathscr{A} \subset \mathscr{A}^+$ be an affine plane of its ordinary points and let an affine coordinate system $O\mathbf{e}_1\mathbf{e}_2$ be given in the plane $\mathscr{A}$. What is the reasonable way of introducing the coordinates of points of the plane $\mathscr{A}^+$?

Each (ordinary or ideal) point of the plane is the centre of some (ordinary or ideal) pencil. According to Proposition 1 of the preceding lecture a pencil is uniquely characterized by the coefficients X, Y, Z of its equation (see equation (4) of the preceding lecture; the subscripts of X_0, Y_0, Z_0 are omitted). It is therefore natural to take them as the coordinates of the point under consideration. These coordinates are determined up to proportionality, i.e. they are *homogeneous coordinates.*

Definition 1. Coordinates X, Y, Z are called *homogeneous affine coordinates* determined by the affine coordinate system $O\mathbf{e}_1\mathbf{e}_2$.

A point with the coordinates X, Y, Z is designated by the symbol $(X{:}Y{:}Z)$.

From the description given in the preceding lecture of ordinary and ideal pencils in connection with their equations we immediately obtain the following proposition which gives rules for explicit calculation of homogeneous affine coordinates (and may therefore serve as another definition of them):

Proposition 1. *A point* $(X{:}Y{:}Z)$ *is an ordinary point if and only if* $Z \neq 0$. *In this case its ordinary (nonhomogeneous) coordinates* x, y *are expressed by the formulas*

$$x = \frac{X}{Z}, \quad y = \frac{Y}{Z}.$$

An ideal point $(X{:}Y{:}0)$ *is an ideal point of a straight line with a direction vector* $\mathbf{a}$ (l, m) *if and only if*

$$X{:}Y = l{:}m. \ \square$$

Note that *any triple* $(X, Y, Z) \neq (0, 0, 0)$ *is a triple of homogeneous affine coordinates of some point of a projective-affine plane.*

We now consider straight lines. According to Proposition 1 the homogeneous coordinates of points of an arbitrary straight line $(A{:}B{:}C)$ in the affine plane $\mathscr{A}$ satisfy the equation

$$A\left(\frac{X}{Z}\right) + B\left(\frac{Y}{Z}\right) + C = 0,$$

i.e. the equation

$$AX + BY + CZ = 0. \tag{1}$$

In the projective-affine plane $\mathscr{A}^+$ this means that the coordinates of ordinary points of an ordinary straight line $(A{:}B{:}C)$ satisfy equation (1). But the direction vector of the straight line $(A{:}B{:}C)$ is known to have the coordinates B, $-A$ whence it follows by virtue of Proposition 1 that the only ideal point of the straight line has homogeneous coordinates $(B{:}\ -A{:}0)$. Since these coordinates obviously satisfy equation (1) we derive that *in a projective-affine plane equations of the form* (1), *where either* $A \neq 0$ *or* $B \neq 0$, *are equations of ordinary straight lines.* $\square$

If, however, $A = 0$, $B = 0$ but $C \neq 0$ (otherwise relation (1) is satisfied identically), then equation (1) is satisfied by all ideal points (and only by ideal points). Hence *for* $A = 0$, $B = 0$ *equation* (1) *is the equation of an ideal straight line.*

Thus in a projective-affine plane any equation of the form (1), with $(A, B, C) \neq (0, 0, 0)$, is the equation of a straight line. In other words, just as any triple of numbers $(X, Y, Z) \neq \neq (0, 0, 0)$ is a triple of homogeneous coordinates of some point, so *any triple of numbers* $(A, B, C) \neq (0, 0, 0)$ *is a triple of homogeneous coordinates of some straight line.*

We see that the introduction of ideal points restores complete symmetry between points and straight lines. Analytically this symmetry manifests itself in one and the same relation (1) being (with A, B, C constant and X, Y, Z variable) the equation of a straight line and (with A, B, C variable and X, Y, Z constant) the equation of a pencil, i.e. a peculiar "equation" of the centre of the pencil.

Note that in a real-complex plane an ideal straight line is a real straight line.

By analogy we may now introduce the following definition:

Definition 2. A *second-degree curve* in a projective-affine (say, complex or real-complex) plane is a set of all points of the curve the homogeneous coordinates X, Y, Z of which satisfy an equation of the form

$$(2)\quad a_{11}X^2+2a_{12}XY+a_{22}Y^2+2a_{13}XZ+2a_{23}YZ+a_{33}Z^2=0.$$

Defined in a similar way in a projective-affine plane are *algebraic curves* of an arbitrary degree.

If at least one of the coefficients a_{11}, a_{12}, a_{22} is nonzero, then the ordinary points of curve (2) make up in an affine plane a second-degree curve

$$(3)\qquad a_{11}x^2+2a_{12}xy+a_{22}y^2+2a_{13}x+2a_{23}y+a_{33}=0.$$

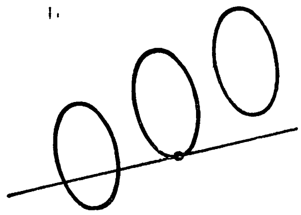
The ideal points $(X:Y:0)$ of the curve satisfy the equation

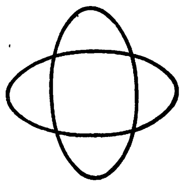
$$a_{11}X^2+2a_{12}XY+a_{22}Y^2=0$$

whence we see that they are ideal points of the asymptotic directions of curve (3). Thus in the case under consideration the second-degree curve (2) in a projective-affine plane is obtained from the second-degree curve (3) in an affine plane by adding ideal points of the asymptotic directions of the curve.

It follows that if the curve (3) is a pair of straight lines, then the curve (2) is a pair of the "same" straight lines (i.e. with an addition of ideal points).

The curve (2) is called an *ellipse*, a *parabola* or a *hyperbola* if the curve (3) is respectively an ellipse, a parabola or a hyperbola. Thus in a projective-affine (real or real-complex) plane the ellipse, having no real asymptotic directions, does not intersect the ideal straight line in real points, the parabola has one point in common with the ideal straight line, the ideal point of its axis (the parabola is said to be *tangent* to the ideal straight line at that point), and the hyperbola has two points in common with the ideal straight line.





The intersection of ellipses

When $a_{11} = a_{12} = a_{22} = 0$ equation (2) is of the form

$$2a_{13}XZ + 2a_{23}YZ + a_{33}Z^2 = 0$$

and is therefore satisfied by the points of the ideal straight line $Z = 0$ and the straight line

$$2a_{13}X + 2a_{23}Y + a_{33}Z = 0,$$

which may be ordinary (if $a_{13} \neq 0$ or $a_{23} \neq 0$) or ideal (if $a_{13} = 0$ and $a_{23} = 0$).

Summarizing, we obtain the following theorem (which for definiteness is stated for situation ($\mathbb{C}$, $\mathbb{R}$)):

Theorem 1 (theorem on the classification of real second-degree curves in the projective-affine real-complex plane). *Every real second-degree curve in a projective-affine real-complex plane is one of the following eleven curves:*

[1] *A real ellipse* ($X^2 + Y^2 - Z^2 = 0$).

[2] *An imaginary ellipse* ($X^2 + Y^2 + Z^2 = 0$).

[3] *A pair of distinct imaginary (complex conjugate) straight lines intersecting in an ordinary point* ($X^2 + Y^2 = 0$).

[4] *A hyperbola* ($X^2 - Y^2 - Z^2 = 0$).

[5] *A pair of distinct real ordinary straight lines intersecting in an ordinary point* ($X^2 - Y^2 = 0$).

[6] *A parabola* ($Y^2 - 2XZ = 0$).

[7] *A pair of distinct real ordinary straight lines intersecting in an ideal point* ($Y^2 - Z^2 = 0$).

[8] *A pair of distinct imaginary* (*complex conjugate*) *straight lines intersecting in an ideal point* ($Y^2 + Z^2 = 0$).

[9] *A pair of coincident real ordinary straight lines* ($Y^2 = 0$).

[10] *A pair of straight lines that is made up of a real ordinary straight line and an ideal straight line* ($YZ = 0$).

[11] *A pair of coincident straight lines each of which is an ideal straight line* ($Z^2 = 0$). □

Written out in parentheses are the equations resulting from an appropriate (canonical) choice of homogeneous affine coordinates.

Similar results hold of course in the projective-Euclidean real-complex plane as well. We shall not state them (this would be an unnecessary repetition) and instead consider a more interesting question of the characterization of circles in projective-Euclidean geometry.

It is natural that in the projective-Euclidean plane we shall take homogeneous coordinates X, Y, Z to be *homogeneous rectangular coordinates*, i.e. to be determined by a coordinate system $O\mathbf{e}_1\mathbf{e}_2$ with orthonormal vectors $\mathbf{e}_1$, $\mathbf{e}_2$.

Definition 3. A *circle* in a projective-Euclidean real-complex plane is a second-degree curve having in homogeneous rectangular coordinates X, Y, Z equation (2) in which

$$a_{11} = a_{22} \neq 0, \quad a_{12} = 0.$$

Ideal points $(X{:}Y{:}Z)$ of such a circle are defined by the equation $X^2 + Y^2 = 0$, i.e. have the form $(\pm 1{:}i{:}0)$. They are called *cyclical points* of a projective-Euclidean plane.

We thus see that a change from Euclidean to projective-Euclidean geometry contributes to circles only two new points, simultaneously ideal and imaginary.

Since cyclical points lie in *every* circle, it is impossible to introduce in a reasonable way the concept of their distance from ordinary points. At any rate there are no grounds to consider this distance infinite. This reveals the imperfection of terminology, widely used (especially in the past), in which ideal points are called points "at infinity" (not to mention the fact that within the framework of affine theory this term has in general no right to exist).

It is appropriate to rank with circles also pairs of straight lines at least one of which is an ideal straight line (such circles are characterized by the conditions $a_{11} = a_{22} = 0$ and $a_{12} = 0$). *Among all real second-degree curves circles will then be characterized by the fact that they pass through cyclical points.* Indeed, if equation (2) is satisfied for $X = \pm 1$, $Y = i$ and $Z = 0$, then

$$(a_{11} - a_{22}) \pm 2ia_{12} = 0$$

from which it follows that $a_{11} = a_{22}$ and $a_{12} = 0$.

The fact that all circles pass through cyclical points explains many features of these curves. For example, generally speaking two ellipses intersect in four points. Why should circles have only two points in common? The answer is: two other points in common are cyclical points.

We have already stressed above the complete symmetry (or duality, as is the usual way to put it) of points and straight lines in projective-affine geometry. However, in the harmonious melody of duality there is in fact a most unpleasant discord connected with the difference between ordinary and ideal points and straight lines. Indeed, there are many ideal points, but there is only one ideal straight line! To get rid of this latter roughness it is necessary to ignore the difference between ordinary and ideal points or straight lines and consider them fully equivalent and interchangeable.

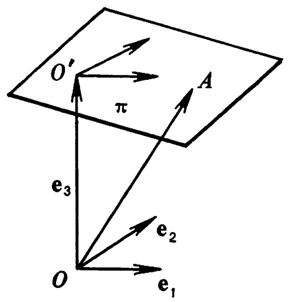
Definition 4. A projective-affine plane in which we ignore the difference between ordinary and ideal points and straight lines is called a *projective plane* and its geometry is called *projective geometry*.

This definition is equally suitable for planes over any field $\mathbb{K}$. We thus obtain in particular a *real projective plane* and a *complex projective plane*, as well as, a *real-complex projective plane*.

Each projective plane can be transformed into a projective-affine plane by choosing arbitrarily in it a straight line and declaring the points of that straight line to be ideal. Thus the same projective plane can be turned into a projective-affine plane in an infinite number of ways (if the ground field $\mathbb{K}$ is infinite), whereas each projective-affine plane uniquely determines the corresponding projective plane. Therefore, by way of a formal decoding of Definition 4 the projective plane may be defined as a class in the obvious sense of equivalent projective-affine planes. But we shall not engage in this rather fruitless play of definitions and prefer a different, more interesting way.

A model of projective-affine geometry which is the easiest to transform into a projective geometry model (a projective plane) is the bundle model $\mathscr{A}_O$ (rather than, say, the circle model). Indeed, ideal points of that model are straight lines in space parallel to the plane π, and therefore, in order to turn it into a projective geometry model it suffices to forget about that plane. The model $\mathscr{A}_O$ considered as a projective geometry model will be designated by the symbol $\mathscr{P}_O$.

The model $\mathscr{A}_O$ allows us to give a beautiful "stereometric" interpretation to homogeneous affine coordinates. According to Definition 1, in a projective-affine plane homogeneous affine coordinates are given by some affine coordinate system $O\mathbf{e}_1\mathbf{e}_2$ given in the "affine part" of the plane (in the set of ordinary points). For the projective-affine plane $\mathscr{A}_O$ the "affine part" is identified with the plane π and so for homogeneous affine coordinates to be given in $\mathscr{A}_O$ it is necessary to give in π some affine coordinate system $O'\mathbf{e}_1\mathbf{e}_2$ (the letter O is already employed by us). But instead of giving the point O' it is possible to give its radius vector $\mathbf{e}_3=\overrightarrow{OO'}$ obviously linearly inexpressible in terms of the vectors $\mathbf{e}_1$, $\mathbf{e}_2$ and so constituting together with them a basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ in space.



Conversely, any basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, for which the terminal point O' of the vector $\mathbf{e}_3 = \overrightarrow{OO'}$ is in the plane π and the vectors $\mathbf{e}_1$, $\mathbf{e}_2$ are parallel to the plane, gives an affine coordinate system $O'\mathbf{e}_1\mathbf{e}_2$ in the plane π and hence homogeneous affine coordinates in the plane $\mathscr{A}_O$.

Each point of the plane $\mathscr{A}_O$, i.e. a straight line in space, is uniquely determined by its direction vector. If X, Y, Z are the coordinates of that direction vector (in the basis

$\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$) and $Z \neq 0$, then the straight line intersects the plane π in a point with the coordinates $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$ (relative to the coordinates of the system $O'\mathbf{e}_1\mathbf{e}_2$), and if $Z = 0$, then the straight line is parallel to the straight lines in the plane π having the direction vector with the coordinates X and Y (relative to the basis $\mathbf{e}_1$, $\mathbf{e}_2$). According to Proposition 1 this means that *in the projective-affine plane* $\mathscr{A}_0$ *the homogeneous affine coordinates* determined by the basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ (i.e. by the corresponding affine coordinate system $O'\mathbf{e}_1\mathbf{e}_2$) *are simply the coordinates* X, Y, Z *of the direction vectors of points of that plane considered as straight lines in space.* □

This allows us to write without any additional calculations formulas for transformation from one system of homogeneous affine coordinates to another. Indeed, these must be simply formulas for transformation from the coordinates of vectors in one basis to the coordinates in another. As we know, these formulas are of the form

$$(4)\quad \begin{aligned} \rho X' &= c_{11}X + c_{12}Y + c_{13}Z, \\ \rho Y' &= c_{21}X + c_{22}Y + c_{23}Z, \\ \rho Z' &= c_{31}X + c_{32}Y + c_{33}Z, \end{aligned} \quad \text{where} \quad \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} \neq 0.$$

Here we have introduced in addition an arbitrary factor of proportionality ρ to stress the fact that we are dealing with homogeneous coordinates.

However, it should be taken into account here that we are considering bases related in a definite way to the plane π (the first two vectors of a basis must be parallel to the plane π and the terminal point of the last must be in the plane) rather than arbitrary bases. On taking account of these requirements we immediately see that in formulas (4) the coefficients c_{31} and c_{32} must be zero and the coefficient c_{33} must be equal to unity.

We have thus proved that *two arbitrary systems of homogeneous affine coordinates* $X{:}Y{:}Z$ and $X'{:}Y'{:}Z'$ *in a projective-affine plane are related by formulas of the form*

$$\begin{aligned} \rho X' &= c_{11}X + c_{12}Y + c_{13}Z, \\ \rho Y' &= c_{21}X + c_{22}Y + c_{23}Z, \\ \rho Z' &= Z, \end{aligned}$$

where ρ *is the factor of proportionality and*

$$\begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} \neq 0.$$

This statement holds for any projective-affine plane of course. It could be proved directly by recalling the formulas for the transformation of nonhomogeneous affine coordinates and using Proposition 1. The advantage of the method we have chosen is that it can be applied directly to the bundle $\mathcal{A}_O$ considered as a projective geometry model.

Definition 5. The *projective coordinates* of points of a projective plane $\mathcal{P}_O$ are the coordinates X, Y, Z in an arbitrary basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ of the direction vectors of those points as straight lines in space.

Projective coordinates are certainly homogeneous coordinates. They differ from homogeneous affine coordinates in that the basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ involved in their definition is quite arbitrary and is not related to any plane π.

According to what has been said above *transformation from one system of projective coordinates to another is described by formulas of the form* (4).

Note that Definition 5 gives projective coordinates only, strictly speaking, to the model $\mathcal{P}_O$. In any other model they are most simply defined as coordinates X', Y', Z' related to some projective-affine coordinates by formulas of the form (4).

Great freedom exercised in transformations of projective coordinates allows us to reduce the number of canonical equations of second-degree curves.

For example, on transforming in the equation of a hyperbola $X^2 - Y^2 - Z^2 = 0$ to the new coordinates

$$X' = Y, \ Y' = Z, \ Z' = X$$

and multiplying it by -1 we obtain (after removing the primes) the equation of an ellipse $X^2 + Y^2 - Z^2 = 0$. Similarly, the equation of a parabola $Y^2 = 2XZ$ in the coordinates

$$X' = Y, \ Y' = X - Z, \ Z' = X + Z$$

also coincides with the same equation of an ellipse. This shows that there is only one nonsingular second-degree curve (i.e. such that does not split into straight lines) in a projective real (or complex) plane. If by introducing an ideal straight line we transform our plane into a projective-affine plane, then that curve turns out to be an ellipse if the straight line introduced does not intersect it, a parabola if the straight line is tangent to the curve and, finally, a hyperbola if the straight line intersects the curve in two points.

In a real-complex plane, besides this "real" curve there is only one more curve, an "imaginary" curve $X^2 + Y^2 + Z^2 = 0$ having no real points.

Similarly reduced is the number of the classes of pairs of straight lines.

The resultant theorem has obviously the following form:

Theorem 2 (theorem on the classification of second-degree curves in the projective plane). *In a real-complex projective plane every real second-degree curve is one of the following five types:*

[1] *A nonsingular real curve* ($X^2 + Y^2 - Z^2 = 0$).

[2] *A nonsingular imaginary curve* ($X^2 + Y^2 + Z^2 = 0$).

[3] *A pair of real distinct straight lines* ($X^2 - Y^2 = 0$).

[4] *A pair of imaginary (complex conjugate) straight lines* ($X^2 + Y^2 = 0$).

[5] *A pair of coincident (real) straight lines* ($X^2 = 0$).

In a complex projective plane there are only three types of second-degree curves:

[1] *A nonsingular curve* ($X^2 + Y^2 + Z^2 = 0$).

[2] *A pair of distinct straight lines* ($X^2 + Y^2 = 0$).

[3] *A pair of coincident straight lines* ($X^2 = 0$). □

Written out in parentheses here are canonical equations.

Lecture 26

Coordinate isomorphisms of vector spaces • Coordinate isomorphisms of affine spaces • Projective-affine spaces • Projective spaces • Pencils of planes • Bundles of planes • Extending space with ideal elements • Orthogonal, affine and projective transformations

The method of introducing projective coördinates in terms of projective-affine coordinates is of course rather clumsy and inaesthetic. But in the general case there is no other way out, since the concept of a projective plane was itself defined in terms of projective-affine planes. And it is not at all simple to give this concept an independent definition. We shall do it in a way that is perhaps not the most elegant but seems to be the simplest and rich enough in ideas.

We shall begin by giving a new definition to a vector space.

Note first of all that for any field $\mathbb{K}$ and any $n \geqslant 1$ formulas of the form

$$
\begin{aligned}
y^1 &= c_1^1 x^1 + \ldots + c_n^1 x^n, \\
&\ldots\ldots\ldots\ldots \\
y^n &= c_1^n x^1 + \ldots + c_n^n x^n,
\end{aligned}
\tag{1}
$$

where

$$
\begin{vmatrix} c_1^1 & \ldots & c_n^1 \\ \cdot & \cdot\;\cdot & \cdot \\ c_1^n & \ldots & c_n^n \end{vmatrix} \neq 0,
$$

give bijective mappings of the set $\mathbb{K}^n$ onto itself. These mappings are isomorphisms of the vector space $\mathbb{K}^n$ onto itself (i.e. are its *automorphisms*). They are called *homogeneous* *linear transformations* and obviously constitute a group. This group is designated by the symbol GL $(n; \mathbb{K})$ and is termed a *general linear group*. It is isomorphic to a group of nonsingular matrices of order n.

For any vector space $\mathscr{V}$ of dimension n over a field $\mathbb{K}$ its coordinate isomorphisms (see Lecture 6) are bijective mappings

$$\alpha: \mathscr{V} \to \mathbb{K}^n \tag{2}$$

having the following properties (the notation $\alpha \in \text{Coor}\,(\mathscr{V})$ signifies that α is a coordinate isomorphism of the form (2)):

Property 1. *If* $\alpha \in \text{Coor}\,(\mathscr{V})$ *and* $\varphi \in \text{GL}\,(n; \mathbb{K})$, *then* $\varphi \circ \alpha \in \text{Coor}\,(\mathscr{V})$.

Property 2. *If* $\alpha, \alpha' \in \text{Coor}\,(\mathscr{V})$, *then* $\alpha' \circ \alpha^{-1} \in \text{GL}\,(n; \mathbb{K})$.

Property 2 simply states that the coordinates in two different bases are related by formulas of the form (1) and Property 1 states that if $x^1, \ldots, x^n$ are coordinates, then so are $y^1, \ldots, y^n$ for any transformation (1) (but of course in another basis).

It turns out (and it is a new fact to us) that Properties 1 and 2 fully characterize coordinate isomorphisms:

Proposition 1. *If for some set* $\mathscr{V}$ *a set* Coor $(\mathscr{V})$ *of its bijective mappings* $\mathscr{V} \to \mathbb{K}^n$ *is given having Properties* 1 *and* 2 (*even Property* 2 *alone is sufficient*), *then it is possible to introduce the structure of a vector space into* $\mathscr{V}$ *in a unique way so that mappings of* Coor $(\mathscr{V})$ *should become coordinate isomorphisms* (*all coordinate isomorphisms if Property* 1 *holds*).

Proof. Having chosen an arbitrary mapping $\alpha \in \text{Coor}\,(\mathscr{V})$ transfer with the aid of it linear operations from $\mathbb{K}^n$ into $\mathscr{V}$, i.e. set

$$\mathbf{x}+\mathbf{y}=\alpha^{-1}(\alpha(\mathbf{x})+\alpha(\mathbf{y})),$$
$$k\mathbf{x}=\alpha^{-1}(k\alpha(\mathbf{x}))$$

for any elements $\mathbf{x}, \mathbf{y} \in \mathscr{V}$ and any number $k \in \mathbb{K}$. It is clear that we thus transform $\mathscr{V}$ into a vector space and α into a coordinate isomorphism (corresponding to the basis $\alpha^{-1}(\mathbf{e}_1), \ldots, \alpha^{-1}(\mathbf{e}_n)$, where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is the standard basis of the vector space $\mathbb{K}^n$). So to prove Proposition 1 it suffices to establish that these operations are independent of the choice of α. But this follows directly from Property 2 and the fact that homogeneous linear transformations are automorphisms of the space $\mathbb{K}^n$. Indeed, if $\alpha' = \varphi \circ \alpha$, where $\varphi \in \mathrm{GL}\,(n;\ \mathbb{K})$, then

$$\alpha'^{-1}(\alpha'(\mathbf{x}) + \alpha'(\mathbf{y})) =$$
$$= \alpha^{-1}(\varphi^{-1}(\varphi(\alpha(\mathbf{x})) + \varphi(\alpha(\mathbf{y})))) = \alpha^{-1}(\alpha(\mathbf{x}) + \alpha(\mathbf{y}))$$

and

$$\alpha'^{-1}(k\alpha'(\mathbf{x})) = \alpha^{-1}(\varphi^{-1}(k\varphi(\alpha(\mathbf{x})))) = \alpha^{-1}(k\alpha(\mathbf{x}))$$

for any elements $\mathbf{x}, \mathbf{y} \in \mathscr{V}$ and any number $k \in \mathbb{K}$. The uniqueness of such a vector space structure in $\mathscr{V}$ is obvious. □

We see that Proposition 1 provides us with quite a new axiomatic system of vector spaces. The basic undefined notions of this system are mappings (1), Properties 1 and 2 being its axioms.

Quite similarly one may axiomatically define affine spaces. Here one should begin with the group Aff $(n;\ \mathbb{K})$ of all automorphisms of the set $\mathbb{K}^n$ as an affine space, i.e. of *nonhomogeneous linear transformations* of the form

$$\begin{aligned} y^1 &= c_1^1 x^1 + \ldots + c_n^1 x^n + b^1, \\ &\cdots\cdots\cdots\cdots\cdots \\ y^n &= c_1^n x^1 + \ldots + c_n^n x^n + b^n, \end{aligned}$$

where

$$\begin{vmatrix} c_1^1 & \ldots & c_n^1 \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ c_1^n & \ldots & c_n^n \end{vmatrix} \neq 0.$$

For any affine space $\mathscr{A}$ over a field $\mathbb{K}$ the set Coor $(\mathscr{V})$ of the coordinate isomorphisms

$$\alpha\colon \mathscr{A} \to \mathbb{K}^n$$

has the same Properties 1 and 2 (in which Coor $(\mathscr{V})$ must of course be replaced by Coor $(\mathscr{A})$ and GL $(n;\ \mathbb{K})$ by Aff $(n;\ \mathbb{K})$). We have an analogue of Proposition 1:

Proposition 2. *If for some set $\mathscr{A}$ a set* Coor $(\mathscr{A})$ *of its bijective mappings $\mathscr{A} \to \mathbb{K}^n$ is given having Properties* 1 *and* 2 *(with respect to the group* Aff $(n;\ \mathbb{K})$*), then it is possible to introduce into $\mathscr{A}$ the structure of an affine space over a field $\mathbb{K}$ in a unique way so that mappings of* Coor $(\mathscr{A})$ *should become coordinate isomorphisms.*

Proof. Having chosen in $\mathscr{A}$ some point O, denote by $\operatorname{Coor}^0(\mathscr{A})$ the subset of the set $\operatorname{Coor}(\mathscr{A})$ consisting of all mappings $\alpha \in \operatorname{Coor}(\mathscr{A})$ translating O into a point $(0, \ldots, 0) \in \mathbb{K}^n$. It is easy to see that setting $\mathscr{V} = \mathscr{A}$ and $\operatorname{Coor}(\mathscr{V}) = \operatorname{Coor}^0(\mathscr{A})$ we satisfy all the conditions of Proposition 1. The set $\mathscr{V}$ thus proves to be a vector space. Now determine the mapping $A, B \mapsto \overrightarrow{AB}$ by setting

$$\overrightarrow{AB} = \alpha^{-1}(\alpha(B) - \alpha(A)),$$

where $\alpha \in \operatorname{Coor}^0(\mathscr{A})$. It is easy to check that the vector $\overrightarrow{AB} \in \mathscr{V}$ is independent of the choice of α and that thus the set $\mathscr{A}$ is provided with the structure of an affine space with an associated vector space $\mathscr{V}$. Its uniqueness is obvious. $\square$

For projective-affine spaces to be obtained by the same method (we shall not restrict ourselves to the case $n = 2$ and give a general definition; projective-affine planes we get for $n = 2$), it is necessary to determine beforehand the projective-affine analogue of the space $\mathbb{K}^n$. It is quite clear how it is to be done.

Consider the subset $\mathbb{K}^{n+1} \setminus 0$ of a vector space $\mathbb{K}^{n+1}$ consisting of all nonzero vectors $(x^0, x^1, \ldots, x^n) \neq (0, 0, \ldots, 0)$. The proportional relation

$$(x^0, x^1, \ldots, x^n) \sim (y^0, y^1, \ldots, y^n)$$

if there is a number $k \neq 0$ such that

$$y^0 = kx^0, \; y^1 = kx^1, \; \ldots, \; y^n = kx^n$$

is an equivalence relation on this subset. The equivalence class containing the vector $(x^0, x^1, \ldots, x^n)$ will be designated by the symbol

$$(x^0 : x^1 : \ldots : x^n). \tag{3}$$

(These classes are simply straight lines of the space $\mathbb{K}^{n+1}_{\text{aff}}$ that pass through the point $(0, 0, \ldots, 0)$.)

Definition 1. The set of all equivalence classes (3) is designated by the symbol $\mathbb{K}P^n$ and called an *arithmetical projective space of dimension n over a field* $\mathbb{K}$.

In this space a group Pr-Aff (n; $\mathbb{K}$) functions whose elements are homogeneous linear transformations of the form

$$\begin{aligned}
&\rho y^0 = x^0,\\
&\rho y^1 = c_0^1 x^0 + c_1^1 x^1 + \ldots + c_n^1 x^n,\\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
&\rho y^n = c_0^n x^0 + c_1^n x^1 + \ldots + c_n^n x^n,
\end{aligned}$$

where

$$\begin{vmatrix} c_1^1 & \ldots & c_n^1 \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ c_1^n & \ldots & c_n^n \end{vmatrix} \neq 0.$$

As an abstract group it is isomorphic to the group Aff (n; $\mathbb{K}$).

In particular for $n = 2$ we have a group Pr-Aff (2; $\mathbb{K}$) whose transformations (after the simple redesignation $x^0 = Z$, $x^1 = X$, $x^2 = Y$) can be written

$$\begin{aligned}
&\rho X' = c_{11}X + c_{12}Y + c_{13}Z,\\
&\rho Y' = c_{21}X + c_{22}Y + c_{23}Z,\\
&\rho Z' = Z,
\end{aligned}$$

where

$$\begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} \neq 0.$$

Definition 2. *A projective-affine space of dimension n over a field* $\mathbb{K}$ is an arbitrary set $\mathscr{A}$ for which a set Coor ($\mathscr{A}$) of bijective mappings

$$\alpha\colon \mathscr{A} \to \mathbb{K}\mathrm{P}^n$$

is given such that the following axioms hold:

Axiom 1. If $\alpha \in$ Coor ($\mathscr{A}$) and $\varphi \in$ Pr-Aff (n; $\mathbb{K}$), then $\varphi \circ \alpha \in$ Coor ($\mathscr{A}$).

Axiom 2. If $\alpha, \alpha' \in$ Coor ($\mathscr{A}$), then $\alpha' \circ \alpha^{-1} \in$ Pr-Aff (n; $\mathbb{K}$).

Points $A \in \mathscr{A}$ for which $\alpha(A)$ is of the form $(0 : x^1 : \ldots : x^n)$ at least under one mapping $\alpha \in$ Coor ($\mathscr{A}$) (and hence under all such mappings) are called *ideal points* of a projective-affine space $\mathscr{A}$. The other points of $\mathscr{A}$ are called *ordinary points.*

In the light of what has been said above it is quite clear that projective-affine spaces $\mathscr{A}$ of dimension two coincide in the sense of this definition with projective-affine planes in the sense of Definition 3, Lecture 24. And for any n they coincide with projective-affine spaces which we get as an obvious generalization of Definition 3 of Lecture 24.

It is now absolutely clear how to define projective spaces.

Let Proj $(n;\ \mathbb{K})$ be a group of all transformations of a space $\mathbb{K}\mathrm{P}^n$ acting by formulas of the form

$$\begin{aligned}\rho y^0 &= c_0^0x^0 + c_1^0x^1 + \ldots + c_n^0x^n,\\ \rho y^1 &= c_0^1x^0 + c_1^1x^1 + \ldots + c_n^1x^n,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\\ \rho y^n &= c_0^nx^0 + c_1^nx^1 + \ldots + c_n^nx^n,\end{aligned} \tag{4}$$

where

$$\begin{vmatrix} c_0^0 & c_1^0 & \ldots & c_n^0\\ c_0^1 & c_1^1 & \ldots & c_n^1\\ \ldots & \ldots & \ldots & \ldots\\ c_0^n & c_1^n & \ldots & c_n^n\end{vmatrix} \neq 0.$$

As an abstract group this group is isomorphic to a factor group of the group of all nonsingular matrices of order $n+1$ relative to the normal divisor consisting of all scalar matrices.

Definition 3. *A projective space of dimension n over a field* $\mathbb{K}$ is an arbitrary set $\mathscr{P}$ for which a set Coor $(\mathscr{P})$ of bijective mappings

$$\alpha\colon\ \mathscr{P} \to \mathbb{K}\mathrm{P}^n$$

is given such that the following axioms hold:

Axiom 1. If $\alpha \in$ Coor $(\mathscr{P})$ and $\varphi \in$ Proj $(n;\ \mathbb{K})$, then $\varphi \circ \alpha \in$ Coor $(\mathscr{P})$.

Axiom 2. If $\alpha, \alpha' \in$ Coor $(\mathscr{P})$, then $\alpha' \circ \alpha^{-1} \in$ Proj $(n;\ \mathbb{K})$.

Mappings $\alpha \in$ Coor $(\mathscr{P})$ are called *coordinate isomorphisms*, and for every point $A \in \mathscr{P}$ the numbers $x^0, x^1, \ldots, x^n$ satisfying the relation

$$\alpha(\mathscr{P}) = (x^0 : x^1 : \ldots : x^n)$$

are called the *projective coordinates* of A corresponding to the isomorphism α.

It is clear that for $n = 2$ we obtain projective planes in the sense of Definition 4 of the preceding lecture and projective coordinates as they were defined in the present lecture.

After the excursus into general axiomatic constructions let us return to the more interesting geometrical considerations.

Just as straight lines in the plane it is possible to take planes in space as a basic element of geometry.

Each plane in an affine space is uniquely defined by the coefficients of the equation of the plane

$$Ax + By + Cz + D = 0$$

in the affine coordinates x, y, z; these coordinates may therefore be regarded as its homogeneous coordinates. We shall designate a plane by the symbol $(A:B:C:D)$.

Let

(5) $A_0x + B_0y + C_0z + D_0 = 0$ and $A_1x + B_1y + C_1z + D_1 = 0$

be two distinct planes.

Definition 4. A *pencil of planes* that is determined by the planes (5) is a set of all planes $(A:B:C:D)$ for which there are numbers μ and ν (certainly not both zero) such that

$$A = \mu A_0 + \nu A_1, \quad B = \mu B_0 + \nu B_1,$$
$$C = \mu C_0 + \nu C_1, \quad D = \mu D_0 + \nu D_1$$

(the plane $(A:B:C:D)$ of course depends only on the ratio $\mu:\nu$).

This definition is quite similar to Definition 1 of Lecture 24.

We could further proceed exactly in the way we did in Lecture 24, but we shall instead follow a somewhat different way to avoid dull linear algebra. To do this we shall need the following lemma:

Lemma 1. *For any point of space M_1 (x_1, y_1, z_1) not lying in both planes* (5) *there exists in the pencil* (6) *a unique plane containing the point M_1.*

Proof. If the desired plane does exist, then for the corresponding numbers μ and ν we must have the equation

$$(\mu A_0 + \nu A_1) x_1 + (\mu B_0 + \nu B_1) y_1 + \\ + (\mu C_0 + \nu C_1) z_1 + (\mu D_0 + \nu D_1) = 0,$$

i.e. the equation

$$\mu (A_0 x_1 + B_0 y_1 + C_0 z_1 + D_0) + \\ + \nu (A_1 x_1 + B_1 y_1 + C_1 z_1 + D_1) = 0$$

which uniquely determines the relation

$$(7)\quad \mu : \nu = -(A_1 x_1 + B_1 y_1 + C_1 z_1 + D_1) : \\ : (A_0 x_1 + B_0 y_1 + C_0 z_1 + D_0).$$

This proves the uniqueness of the desired plane.

To prove the existence of the plane it is sufficient to take μ and ν determined by relation (7). □

Definition 5. A pencil (6) is said to be *ordinary* if planes (5) intersect and *ideal* otherwise (if planes (5) are parallel).

Proposition 3. *An ordinary pencil consists of all planes containing the straight line in which planes* (5) *intersect an ideal pencil consisting of all planes parallel to planes* (5).

Proof. Let planes (5) intersect. Then equations (5) taken together are the equations of the straight line of their intersection. If a point M_0 (x_0, y_0, z_0) lies on the straight line (5), then for any μ and ν

$$Ax_0 + By_0 + Cz_0 + D = (\mu A_0 + \nu A_1) x_0 + (\mu B_0 + \nu B_1) y_0 + \\ + (\mu C_0 + \nu C_1) z_0 + (\mu D_0 + \nu D_1) = \\ = \mu (A_0 x_0 + B_0 y_0 + C_0 z_0 + D_0) + \\ + \nu (A_1 x_0 + B_1 y_0 + C_1 z_0 + D_1) = 0,$$

and so the plane (6) passes through that point. Hence the plane (6) contains the straight line (5). Conversely, consider an arbitrary plane containing a straight line (5). Choose in that plane an arbitrary point M_1 (x_1, y_1, z_1) not lying on the straight line (5). Then, according to Lemma 1, there exists in the pencil (6) a unique plane $(A:B:C:D)$ passing through that point. Both planes, the original plane and the constructed one, contain the straight line (5) and the point M_1. So they coincide. Hence the plane in question is in the pencil (6).

Let planes (5) be parallel. Then

$$\frac{A_1}{A_0}=\frac{B_1}{B_0}=\frac{C_1}{C_0},$$

and so for any μ and ν (for which at least one of the numbers $\mu A_0+\nu A_1$, $\mu B_0+\nu B_1$, $\mu C_0+\nu C_1$ is nonzero)

$$\frac{\mu A_0+\nu A_1}{A_0}=\frac{\mu B_0+\nu B_1}{B_0}=\frac{\mu C_0+\nu C_1}{C_0}.$$

This means that any plane of the pencil (6) is parallel to planes (5). Conversely, consider an arbitrary plane parallel to planes (5) and an arbitrary point M_1 (x_1, y_1, z_1) in that plane. According to Lemma 1 there exists in the pencil a unique plane passing through that point. Being parallel to the planes (5) this plane must coincide with the original plane. Therefore the latter plane is in the pencil (6). □

If we now extend the space by adding to each plane an ideal straight line (and hence making it a projective-affine plane) and require that the ideal straight lines of the planes should coincide if and only if the planes are parallel, then each ideal pencil also becomes a pencil of planes passing through a fixed straight line.

Pencils are analogues of straight lines. We now consider analogues of planes.

Let

$$\begin{aligned} &A_0x+B_0y+C_0z+D_0=0,\\ &A_1x+B_1y+C_1z+D_1=0,\\ &A_2x+B_2y+C_2z+D_2=0 \end{aligned} \tag{8}$$

be three planes not in the same pencil (i.e. neither parallel nor passing through the same straight line).

Definition 6. A *bundle of planes* determined by planes (8) is a set of all planes $(A{:}B{:}C{:}D)$ for which there are numbers μ_0, μ_1, μ_2 (certainly not all zero) such that

$$\begin{aligned} A&=\mu_0A_0+\mu_1A_1+\mu_2A_2,\\ B&=\mu_0B_0+\mu_1B_1+\mu_2B_2,\\ C&=\mu_0C_0+\mu_1C_1+\mu_2C_2,\\ D&=\mu_0D_0+\mu_1D_1+\mu_2D_2. \end{aligned} \tag{9}$$

A bundle is said to be *ordinary* if planes (8) have a single point in common (the *centre of the bundle*) and *ideal* if planes (8) do not intersect (i.e. if two of the planes (8) intersect in a straight line parallel to a third plane; this straight line is called the *central straight line* of the ideal bundle).

Denoting the left-hand sides of equations (8) by the symbols f_0, f_1, and f_2 respectively, we may represent the equation of any plane of a bundle as

$$\mu_0 f_0 + \mu_1 f_1 + \mu_2 f_2 = 0. \tag{10}$$

It is often convenient to use such abridged notations.

Proposition 4. *An ordinary bundle consists of all planes passing through its centre, an ideal bundle consisting of all planes parallel to its central straight line.*

Proof. Let a bundle (9) be an ordinary bundle and let $M_0(x_0, y_0, z_0)$ be the centre of the bundle. Then for any μ_0, μ_1, μ_2 the value of the left-hand side of equation (10) at a point M_0 is

$$\mu_0 f_0(x_0, y_0, z_0) + \mu_1 f_1(x_0, y_0, z_0) + \mu_2 f_2(x_0, y_0, z_0)$$

and hence zero. Therefore each plane (10) of the bundle passes through the point M_0. Conversely, consider an arbitrary plane passing through the centre M_0 of a bundle (9) and two distinct points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ in it, not lying on the straight line containing the point M_0. The condition that a plane (10) of the bundle should pass through these points implies that

$$\mu_0 f_0(x_1, y_1, z_1) + \mu_1 f_1(x_1, y_1, z_1) + \mu_2 f_2(x_1, y_1, z_1) = 0,$$
$$\mu_0 f_0(x_2, y_2, z_2) + \mu_1 f_1(x_2, y_2, z_2) + \mu_2 f_2(x_2, y_2, z_2) = 0.$$

These two equations in μ_0, μ_1, μ_2 have a nontrivial solution (unique to within proportionality)

$$\mu_0 : \mu_1 : \mu_2 =$$
$$= \begin{vmatrix} f_1(x_1, y_1, z_1) & f_2(x_1, y_1, z_1) \\ f_1(x_2, y_2, z_2) & f_2(x_2, y_2, z_2) \end{vmatrix} : \begin{vmatrix} f_2(x_1, y_1, z_1) & f_0(x_1, y_1, z_1) \\ f_2(x_2, y_2, z_2) & f_0(x_2, y_2, z_2) \end{vmatrix} :$$
$$: \begin{vmatrix} f_0(x_1, y_1, z_1) & f_1(x_1, y_1, z_1) \\ f_0(x_2, y_2, z_2) & f_1(x_2, y_2, z_2) \end{vmatrix}.$$

The corresponding plane of the bundle passes through three noncollinear points M_0, M_1, M_2 and so coincides with the given plane. Therefore the latter plane is in the bundle.

Let a bundle (9) be an ideal bundle and let **a** (l, m, n) be the direction vector of the central straight line of the bundle. Since the vector **a** is parallel to planes (8) we have

$$A_0l + B_0m + C_0n = 0,$$
$$A_1l + B_1m + C_1n = 0,$$
$$A_2l + B_2m + C_2n = 0.$$

Hence for any μ_0, μ_1, μ_2

$$(\mu_0A_0 + \mu_1A_1 + \mu_2A_2)\, l + (\mu_0B_0 + \mu_1B_1 + \mu_2B_2)\, m + \\ + (\mu_0C_0 + \mu_1C_1 + \mu_2C_2)\, n = 0.$$

This means that any plane of the bundle (9) is parallel to the central straight line of the bundle. Conversely, let $(A:B:C:D)$ be a plane such that

$$Al + Bm + Cn = 0.$$

Then in the matrix

$$\begin{pmatrix} A & B & C & D \\ A_0 & B_0 & C_0 & D_0 \\ A_1 & B_1 & C_1 & D_1 \\ A_2 & B_2 & C_2 & D_2 \end{pmatrix}$$

the columns are linearly dependent (with the coefficients l, m, n, 0) and therefore so are the rows. In addition, under the hypothesis the last three rows are linearly independent (otherwise planes (8) are in the same bundle). Hence the first row is linearly expressible in terms of the others. But this exactly means that the plane $(A:B:C:D)$ is in the bundle (9). $\square$

For ideal bundles to have a centre too it is necessary to add to each straight line an ideal point and to each plane an ideal straight line and assume that the ideal point of a straight line is on the ideal straight line of a plane if and only if the straight line and the plane are parallel.

It is clear that to do this it is sufficient to add to each straight line an ideal point, one and the same for all parallel lines, and declare the set of all ideal points of the straight lines parallel to some plane to be the ideal straight line of that plane. The ideal straight lines of parallel planes will then coincide as required (see above).

It is natural that the set of all ideal points added to a space should be considered to be a plane.

A formal description of the obtained *projective-affine space* can be given according to the already familiar models (see Definition 3 in Lecture 24 or Definition 2 above). Ignoring in this space the difference between ordinary and ideal points yields a projective space (see Definition 4 in Lecture 25 or Definition 3 above).

Unfortunately, there is no possibility to go further into all this here.

We know from the school course what great importance is attached in geometry to the concept of congruent figures and the closely related concept of motion.

Visually, a motion (say of a plane) is a transformation under which "nothing essential changes": lengths and angles remain unchanged, a vector sum goes over into a vector sum, a product of a vector by a number goes over into a product of a vector by a number, etc. From a general point of view this means that a motion is the isomorphism of a Euclidean plane onto itself, i.e. its automorphism. To be precise, since a motion also preserves orientation it is the automorphism of an oriented plane. But we know (see Lecture 14) that the isomorphism $\mathscr{A} \to \mathscr{A}'$ of one Euclidean plane onto another (or onto the same) is established from the equality of coordinates in two systems of rectangular coordinates, i.e. to specify it one needs to choose a rectangular coordinate system $O\mathbf{i}_1 \ldots \mathbf{i}_n$ in the space $\mathscr{A}$, a rectangular coordinate system $O'\mathbf{i}'_1 \ldots \mathbf{i}'_n$ in the space $\mathscr{A}'$ and associate with an arbitrary point $A \in \mathscr{A}$ a point $A' \in \mathscr{A}'$ having in the system $O'\mathbf{i}'_1 \ldots \mathbf{i}'_n$ the same coordinates as those which the point A had in the system $O\mathbf{i}_1 \ldots \mathbf{i}_n$. Since all this is applicable to the case $\mathscr{A}' = \mathscr{A}$ we are concerned with now and since a motion must preserve orientation we thus come to the following formal definition (which we state at once for any n, although we shall only need it for $n = 2$ and $n = 3$):

Definition 7. A *motion* of an n-dimensional Euclidean space $\mathcal{A}$ is an arbitrary transformation of the space established from the equality of coordinates in two rectangular coordinate systems of the same sign. This means that any motion is uniquely determined by two such systems $O\mathbf{i}_1 \ldots \ldots \mathbf{i}_n$ and $O'\mathbf{i}'_1 \ldots \mathbf{i}'_n$ (it being possible to give the first of them arbitrarily) and carries every point A over into a point A' having in the system $O'\mathbf{i}'_1 \ldots \mathbf{i}'_n$ the same coordinates as those which the point A had in the system $O\mathbf{i}_1 \ldots \mathbf{i}_n$. Thus if

$$\overrightarrow{OA} = a_1\mathbf{i}_1 + \ldots + a_n\mathbf{i}_n,$$

then

$$\overrightarrow{O'A'} = a_1\mathbf{i}'_1 + \ldots + a_n\mathbf{i}'_n.$$

It is clear that all motions of the space $\mathcal{A}$ form a group. This group depends to within an isomorphism on the dimension n. It will be denoted by the symbol $\mathrm{Ort}^+(n)$.

If in Definition 7 we remove the requirement that the coordinate systems should be of the same sign, more general transformations are obtained. These are called *orthogonal transformations*. They form a group $\mathrm{Ort}\,(n)$ a subgroup of which is the group of motions $\mathrm{Ort}^+(n)$.

One can similarly introduce the automorphisms of affine spaces. They are called *affine transformations* and represent transformations acting by the equality of coordinates in two affine coordinate systems.

Finally, it is possible to introduce *projective transformations* acting by the equality of coordinates in two systems of projective coordinates.

In Euclidean geometry congruent figures, i.e. figures that can be carried over into each other by some motion (or, more generally, by some orthogonal transformation) are considered to be equal. This definition characterizes Euclidean geometry.

Likewise, regarded as equal in affine geometry are *affinely congruent* figures (sometimes called *affinely equivalent* figures), i.e. figures that can be carried over into each other by an affine transformation, and regarded as equal in projective geometry are *projectively equivalent* figures (that can be carried over into each other by a projective transformation).

That is why projective, affine and orthogonal transformations are very important in geometry.

Unfortunately, we have no time for the study of projective transformations. As to affine and orthogonal transformations we are going to tackle them in our next lecture.

Lecture 27

Expressing an affine transformation in terms of coordinates • Examples of affine transformations • Factorization of affine transformations • Orthogonal transformations • Motions of a plane • Symmetries and glide symmetries • A motion of a plane as a composition of two symmetries • Rotations of a space

Let $x^1, \ldots, x^n$ be fixed affine coordinates in an affine space $\mathscr{A}$. Then any affine transformation Φ of the space $\mathscr{A}$, being a transformation established from the equality of coordinates, determines some other affine coordinates $x^{1'}, \ldots, x^{n'}$ and, conversely, any affine coordinates $x^{1'}, \ldots, x^{n'}$ give some affine transformation Φ.

By definition, for any point $A \in \mathscr{A}$ the "primed" coordinates of a point $A' = \Phi(A)$ coincide with the "unprimed" coordinates of the point A:

$$x^{1'} = x^1, \ \ldots, \ x^{n'} = x^n.$$

But, by the rule of the transformation of affine coordinates, the "unprimed" coordinates of the point A' (denote them by the symbols $y^1, \ldots, y^n$) may be expressed in terms of its "primed" coordinates $x^{1'}, \ldots, x^{n'}$ using formulas of the form

$$\begin{aligned} y^1 &= c_1^1 x^{1'} + \ldots + c_n^1 x^{n'} + b^1, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ y^n &= c_1^n x^{1'} + \ldots + c_n^n x^{n'} + b^n, \end{aligned}$$

where the determinant of the matrix (c_i^j) is nonzero. On replacing here the coordinates $x^{1'}, \ldots, x^{n'}$ of the point A' by the coordinates $x^1, \ldots, x^n$ of the point A, equal to them, we get

$$
(1)\qquad
\begin{aligned}
y^1 &= c_1^1x^1 + \ldots + c_n^1x^n + b^1,\\
&\ldots\ldots\ldots\ldots\ldots\\
y^n &= c_1^nx^1 + \ldots + c_n^nx^n + b^n
\end{aligned}
$$

or, in matrix notation,

$$
(2)\qquad y = Cx + b,
$$

where

$$
x = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}, \quad y = \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix}, \quad b = \begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix}
$$

and

$$
C = \begin{pmatrix} c_1^1 & \ldots & c_n^1 \\ \cdot & \cdot\ \cdot & \cdot \\ c_1^n & \ldots & c_n^n \end{pmatrix}.
$$

Thus we have proved the following theorem:

Theorem 1. *In affine coordinates given in a space $\mathcal{A}$ any affine transformation Φ is uniquely determined by some nonsingular matrix C and a column matrix b. This transformation carries a point A with coordinates $x^1, \ldots, x^n$ over into a point A' with coordinates $y^1, \ldots, y^n$ expressed by formulas* (1) *or* (2). □

The matrix C is usually called a *matrix of transformation* Φ (in a given coordinate system).

Note that formulas (1) are *identical* with the formulas for the transformation of affine coordinates. They are formulas for affine transformation when $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$ are the coordinates of *different* points in the *same* system of coordinates and formulas for transformation of coordinates when $x^1, \ldots, x^n$ and $y', \ldots, y^n$ are the coordinates of the *same* point in *different* systems of coordinates.

Of course, a similar remark holds for any transformations established from the equality of some other coordinates (orthogonal, projective, etc.).

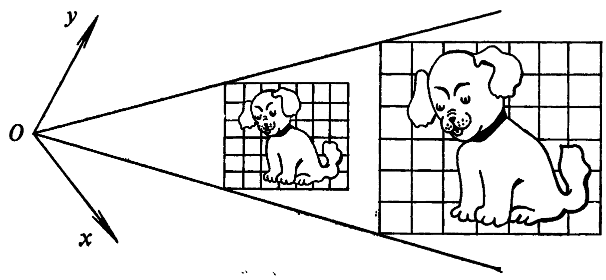
Remark 1. Every affine transformation carries any straight line over into a straight line. It can be proved (we shall not do it) that for $\mathbb{K} = \mathbb{R}$ and $n > 1$ the converse is also true:

every bijective transformation of an affine space preserving the rectilinear arrangement of points is an affine transformation. The restriction $n > 1$ is due to the fact that for $n = 1$ the condition that rectilinearity should be preserved is meaningless. On the other hand, the restriction $\mathbb{K} = \mathbb{R}$ turns out to be essential. For example, in the complex case a transformation carrying every point over into a point with complex conjugate coordinates is not an affine transformation (it is not of the form (1)), although it preserves rectilinear arrangement of points. The fact is that the field $\mathbb{C}$ has a nonidentical automorphism (a complex conjugation) while the field $\mathbb{R}$ has no such automorphisms.

Examples of affine transformations.

It is possible to assume in all the examples for the sake of visualization that the space $\mathscr{A}$ is a Euclidean space and that the coordinates are rectangular coordinates. Of course, any orthogonal transformation is then an affine transformation. We shall give less trivial examples (confining ourselves to the case of a plane).

1. Any homothetic transformation (or homothety) is an affine transformation. Its matrix is (in any system of affine



A homothety

coordinates with the origin at the centre of the homothety) a diagonal matrix with equal diagonal elements, i.e. has the form kE, where E is a unit matrix. (Such matrices are known as *scalar* matrices.)

2. An affine transformation of a plane, given (in coordinates x, y) by the formulas

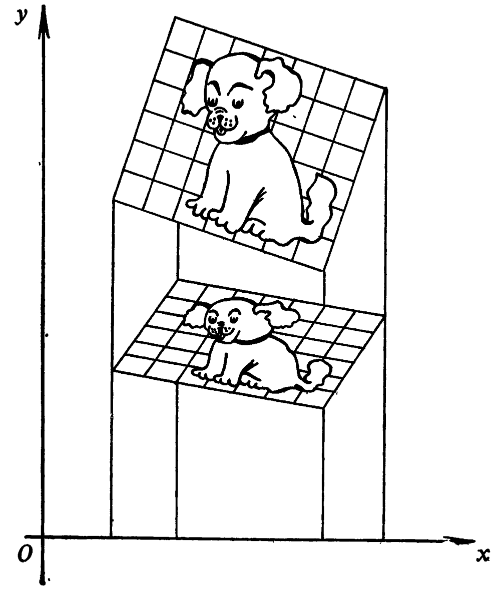
$$x_1 = x, \quad y_1 = ky,$$

where $k > 0$, is called a *compression* toward the axis of abscissae. A combination of two compressions, one toward the axis of abscissae and the other toward the axis of ordinates, is given by the formulas

$$x_1 = kx, \quad y_1 = ly,$$

where $k > 0$, $l > 0$. Its matrix is a diagonal matrix with positive diagonal elements k and l.

It is clear that compressions (and motions) can carry any ellipse over into any other ellipse, for example, into a circle. Thus *all ellipses are affinely congruent*. Similarly, so are



A compression

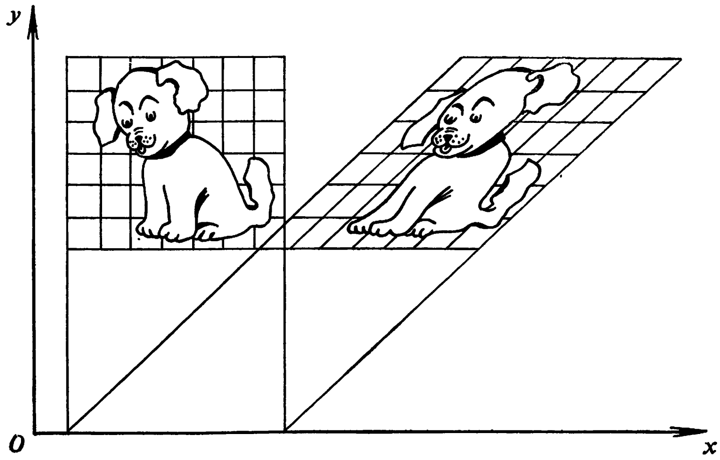
all hyperbolas and all parabolas (the latter are even homothetic, if appropriately located, of course). Analytically this is expressed in there being only one canonical affine equation for ellipses, hyperbolas, and parabolas. See Lecture 22.

3. An affine transformation given by the formulas

$$x_1 = x + py, \quad y_1 = y,$$

where p is an arbitrary number, is called a *shift*. It has the property of preserving the area of plane figures.

It is remarkable that in a Euclidean plane any shift can be represented as a (composition) product of an orthogonal



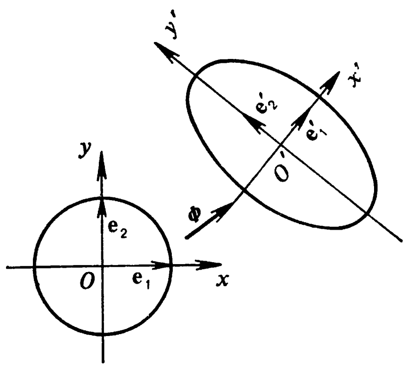
A shift

transformation (motion) and two compressions toward mutually perpendicular axes. Moreover, this holds for any affine transformation.

Proposition 1. *In a Euclidean plane any affine transformation Φ is a composition of an orthogonal transformation and two compressions toward mutually perpendicular axes.*

Proof. Consider in a plane an arbitrary circle of radius 1. An affine transformation Φ carries this circle over into some ellipse, and the centre O of the circle into the centre O' of the ellipse (why?). Let $O'E_1'$ and $O'E_2'$ be straight lines passing through the point O' along the principal directions of the ellipse (the *principal diameters* of the ellipse) and let $\mathbf{e}_1'$ and $\mathbf{e}_2'$ be their direction unit vectors constituting therefore an orthonormal basis. Assuming that E_1' and E_2' are points of intersection of the ellipse with the straight lines $O'E_1'$ and $O'E_2'$, consider their inverse images E_1 and E_2

under the mapping Φ. Then the vectors $\mathbf{e}_1 = \overrightarrow{OE_1}$ and $\mathbf{e}_2 = \overrightarrow{OE_2}$ are unit vectors and the straight lines OE_1 and OE_2 are diameters of the circle. Since an affine transformation is easily seen to carry conjugate diameters over into



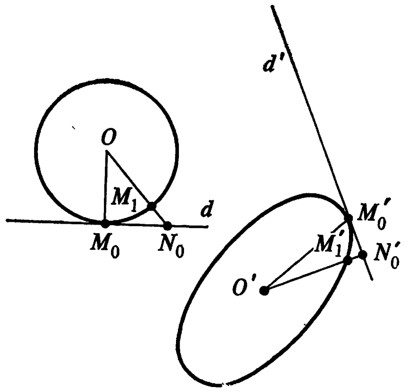
conjugate diameters and the diameters $O'E_1'$ and $O'E_2'$, being principal, are conjugate, so are the circle diameters OE_1 and OE_2. Since conjugate diameters of a circle are perpendicular, this proves that the basis $\mathbf{e}_1$, $\mathbf{e}_2$ is orthonormal.

So we have constructed two rectangular coordinate systems $O\mathbf{e}_1\mathbf{e}_2$ and $O\mathbf{e}_1'\mathbf{e}_2'$ having the property that an affine transformation Φ carries the coordinate axes of the first system over into the coordinate axes of the second system.

To complete the proof it remains to note that the orthogonal transformation established from the equality of these rectangular coordinates obviously differs from the transformation Φ only by two compressions toward the straight lines $O'E_2'$ and OE_1' (with coefficients k and l equal respectively to the lengths of the vectors $\overrightarrow{O'E_1'}$ and $\overrightarrow{O'E_2'}$). □

The key item in this proof is the establishment of the fact that in a plane there are two perpendicular directions which transformation Φ carries over into perpendicular directions. It turns out that this can easily be proved without resorting to the theory of second-degree curves if we use (in conformity with functions on the circle) the *Weierstrass theorem*, known from calculus, which states that in a compact set a continuous function attains its minimum.

Consider to this end the same circle of radius 1 with the centre at the point O. On denoting by $f(M)$ for any point M of the circle the distance of a transformed point $M' = \Phi(M)$ from the point O' we obviously obtain a continuous function $f: M \mapsto f(M)$ on the circle. Let M_0 be the point at which the function attains the minimum. It turns



out that a transformation Φ carries the straight line OM_0 and a straight line d perpendicular to it at the point M_0 (and tangent to the circle) over into perpendicular straight lines $O'M_0'$ and d'. Indeed, if this is not the case and the point M_0' is different from the base N_0' of the perpendicular dropped from the point O' to the straight line d' (and therefore $|O'N_0'| < |O'M_0'| = f(M_0)$), then for a point M_1 in which the straight line ON_0, $N_0 = \Phi^{-1}(N_0')$, intersects our circle we have (in view of the fact that the point M_1 is located on the segment $\overline{ON_0}$ and hence a point $M_1' = f(M_1)$ is located on the segment $\overline{O'N_0'}$) the inequality

$$f(M_1) = |O'M_1'| < |O'N_0'| < f(M_0)$$

contradictory to the choice of the point M_0. □

This argument can be generalized without difficulty to any dimension (the result is of course not two perpendicular directions, but n such directions). Therefore Proposition 1 holds (with appropriate modifications) for affine transformations of a Euclidean space of an arbitrary dimension.

Proposition 1 actually completely reduces the theory of arbitrary affine transformations to that of orthogonal transformations. We shall therefore occupy ourselves with orthogonal transformations (and, in particular, with motions). All coordinates are assumed rectangular.

Being a special case of affine transformations, orthogonal transformations can be written in the same form (2):

(3) $$y = Cx + b,$$

except that for them the matrix C is (in rectangular coordinates) an arbitrary orthogonal (see Lecture 14) matrix rather than an arbitrary nonsingular matrix.

Transformation (3) is a motion if and only if the matrix C is a proper orthogonal matrix (has determinant 1). For this reason motions are also called *proper orthogonal transformations.*

Definition 1. For $C = E$ transformation (3) has the form $y = x + b$, is a motion and is called a (*parallel*) *translation.* It carries any straight line over into a parallel line. Such motions (for $b \neq 0$) leave no point fixed.

Motions leaving some point O fixed are called *rotations with centre at O.* They make up a group designated by the symbol $\mathrm{Rot}_O\,(n)$.

If O is the origin of coordinates, then for a rotation with centre O there is no column matrix b in formula (3) (it is zero). This shows that as an abstract group the group $\mathrm{Rot}_O\,(n)$ is isomorphic to the group of proper orthogonal matrices $\mathrm{SO}\,(n)$.

According to formula (3) any motion is a composition of a rotation $y = Cx$ with centre at the origin and a translation $y = x + b$.

It turns out that *in a plane* (i.e. for $n = 2$) rotations (with an arbitrary centre) and translations exhaust all motions.

Proposition 2. *Any motion of a plane is either a translation or a rotation.*

Proof. It follows from the results of Lecture 14 on the form of orthogonal matrices of the second order that in arbitrary rectangular coordinates x, y every motion of a plane can be written as

$$(4)\qquad \begin{aligned} x' &= x\cos\alpha - y\sin\alpha + b_1,\\ y' &= x\sin\alpha + y\cos\alpha + b_2. \end{aligned}$$

For $\alpha = 0$ it is a translation. We show that for $\alpha \neq 0$ a motion (4) has necessarily a fixed point and therefore is a rotation.

The condition that a point $M_0(x_0, y_0)$ should remain fixed under motion (4) is that

$$x_0 = x_0 \cos\alpha - y_0 \sin\alpha + b_1,$$
$$y_0 = x_0 \sin\alpha + y_0 \cos\alpha + b_2,$$

i.e.

$$(1 - \cos\alpha)\, x_0 + (\sin\alpha)\, y_0 = b_1,$$
$$(-\sin\alpha)\, x_0 + (1 - \cos\alpha)\, y_0 = b_2.$$

Since these equations represent a system of nonhomogeneous linear equations in x_0, y_0 with the determinant

$$\begin{vmatrix} 1-\cos\alpha & \sin\alpha \\ -\sin\alpha & 1-\cos\alpha \end{vmatrix} = (1-\cos\alpha)^2 + \sin^2\alpha = 2(1-\cos\alpha)$$

nonzero for $\alpha \neq 0$, a fixed point M_0 does exist (and is unique). □

Any rotation of a plane with centre at the origin has the form

$$x' = x\cos\alpha - y\sin\alpha,$$
$$y' = x\sin\alpha + y\cos\alpha,$$

where α is some angle uniquely determined always supposing $-\pi < \alpha \leqslant \pi$. This angle is called the *angle of rotation.*

Proposition 2 in particular shows that a motion of a plane cannot have more than one fixed point. This is not so for orthogonal transformations not preserving orientation.

Definition 2. A nonidentity orthogonal transformation of a plane, which has a straight line of fixed points, is called a *symmetry* with respect to (or *reflection* in) that straight line. The straight line consisting of fixed points of symmetry is called the *axis* of symmetry.

Let the orthogonal transformation

$$x' = c_{11}x + c_{12}y + b_1, \quad y' = c_{21}x + c_{22}y + b_2$$

be a symmetry with axis $x = 0$. Then there must hold identically for y the equations

$$0 = c_{12}y + b_1, \quad y = c_{22}y + b_2,$$

whence it follows that $c_{12} = 0$, $c_{22} = 1$ and $b_1 = 0$, $b_2 = 0$. In view of the fact that the matrix C is an orthogonal matrix it is possible (for a nonidentity transformation) only for $c_{21} = 0$ and $c_{11} = -1$. Thus the symmetry under consideration may be expressed by the formulas

$$\begin{aligned} x' &= -x, \\ y' &= y. \end{aligned} \tag{5}$$

In particular we see that symmetry is uniquely determined by its axis. Any symmetry reverses orientation and its square is an identity transformation.

For an arbitrary orientation-reversing orthogonal transformation Φ and an arbitrary symmetry S the composition $S \circ \Phi$ preserves orientation, i.e. is a motion. Since $\Phi = S \circ (S \circ \Phi)$ this proves that *any orientation-reversing orthogonal transformation Φ is a composition of a motion and some symmetry S* (which may be chosen the same for all Φ).

An example of such a transformation is the so-called *glide symmetry* (or glide reflection)

$$x' = -x, \quad y = y + a$$

which is a composition of a symmetry and a translation along the axis of symmetry.

Proposition 3. *Any orientation-reversing orthogonal transformation of a plane is a glide symmetry.*

Proof. It follows immediately from the results obtained above (as well as from what was said in Lecture 14) that any orientation-reversing orthogonal transformation of a plane can be expressed in an arbitrary system of orthogonal coordinates x, y by the formulas

$$\begin{aligned} x' &= x\cos\alpha + y\sin\alpha + b_1, \\ y' &= x\sin\alpha - y\cos\alpha + b_2. \end{aligned} \tag{6}$$

We show that there are numbers l, m such that if Ψ is a translation by a vector $\mathbf{a}\,(l, m)$, then the transformation $\Omega = \Psi^{-1} \circ \Phi$ is a symmetry whose axis is parallel to the vector **a**. It is clear that this will prove the proposition since the transformation $\Phi = \Psi \circ \Omega$ is a glide symmetry.

For any l and m the transformation $\Omega = \Psi^{-1} \circ \Phi$ is obviously given by the formulas

$$x' = x \cos\alpha + y \sin\alpha + b_1 - l,$$
$$y' = x \sin\alpha - y \cos\alpha + b_2 - m.$$

Hence its fixed points are determined from the equations

$$\begin{aligned} x(1-\cos\alpha) - y\sin\alpha &= b_1 - l, \\ -x\sin\alpha + y(1+\cos\alpha) &= b_2 - m. \end{aligned} \tag{7}$$

The problem is to choose l and m in such a way that, first, these equations should determine a straight line (then Ω is a nonidentity—since it reverses the orientation—orthogonal transformation possessing a straight line of fixed points, i.e. is a symmetry). In other words, the numbers l and m must be chosen in such a way that equations (7) should be proportional:

$$\frac{1-\cos\alpha}{-\sin\alpha} = \frac{-\sin\alpha}{1+\cos\alpha} = \frac{b_1 - l}{b_2 - m}.$$

The condition

$$\frac{1-\cos\alpha}{-\sin\alpha} = \frac{-\sin\alpha}{1+\cos\alpha}$$

is satisfied identically, while the condition

$$\frac{-\sin\alpha}{1+\cos\alpha} = \frac{b_1 - l}{b_2 - m}$$

gives for l and m the equation

$$(b_2 - m)\lambda + (b_1 - l) = 0$$

where the notation

$$\lambda = \frac{\sin\alpha}{1+\cos\alpha}$$

is introduced to reduce the formulas. And, secondly, we must require that the resulting straight line should be parallel to the vector **a** (l, m), i.e. that there should hold the equation

$$\frac{l}{m}=\frac{\sin\alpha}{1-\cos\alpha}=\frac{1}{\lambda}.$$

Thus, for l and m we obtain the two equations:

$$(8)\qquad l+m\lambda=b_2\lambda+b_1,\quad \lambda l-m=0.$$

To complete the proof it remains to note that equations (8) have a unique solution

$$l=\frac{b_2\lambda+b_1}{1+\lambda^2},\quad m=\lambda\frac{b_2\lambda+b_1}{1+\lambda^2}.\ \square$$

It is clear that a glide symmetry with $a\neq 0$ leaves no point fixed. Thus, if an orientation-reversing orthogonal transformation has at least one fixed point, then it is a symmetry (and leaves therefore a whole straight line fixed). But for any rotation a composition of it and a symmetry whose axis passes through the centre of rotation is obviously this kind of transformation and so represents a symmetry. Since the square of any symmetry is an identity transformation, this proves that *any rotation of a plane is a composition of two symmetries.*

Similarly a composition of a translation and a symmetry whose axis is perpendicular to the vector of the translation is also an orientation-reversing transformation having a fixed straight line, i.e. a symmetry. For example, if a translation takes place along the axis of abscissae, i.e. has the form

$$x'=x+a,\quad y'=y,$$

then the composition

$$x'=-x+a,\quad y'=y$$

of it and a symmetry with respect to the axis of ordinates has a fixed straight line $x=\frac{a}{2}$. This proves that any translation is also a composition of two symmetries.

Thus *every motion of a plane may be decomposed as a product of two symmetries.* $\square$

We have no time to study orthogonal transformations of a space in as much detail. We shall therefore confine ourselves to a description of its rotations.

By definition every rotation Φ of a space leaves some point of it O fixed. On taking this point as the origin of coordinates we obtain for Φ formulas of the form

$$x' = c_{11}x + c_{12}y + c_{13}z,$$
$$y' = c_{21}x + c_{22}y + c_{23}z,$$
$$z' = c_{31}x + c_{32}y + b_{33}z.$$

Let us try to find all the points $(x, y, z) \neq (0, 0, 0)$ which this rotation leaves on the same straight line as that on which the point $(0, 0, 0)$ lies, i.e. such that

$$x' = \lambda x, \quad y' = \lambda y, \quad z' = \lambda z \tag{9}$$

where λ is some number. Their coordinates x, y, z must satisfy the equations

$$(c_{11} - \lambda)\, x + c_{12}y + c_{13}z = 0,$$
$$c_{21}x + (c_{22} - \lambda)\, y + c_{23}z = 0,$$
$$c_{31}x + c_{32}y + (c_{33} - \lambda)\, z = 0$$

known from algebra to have a nontrivial solution if and only if

$$\begin{vmatrix} c_{11}-\lambda & c_{12} & c_{13} \\ c_{21} & c_{22}-\lambda & c_{23} \\ c_{31} & c_{32} & c_{33}-\lambda \end{vmatrix} = 0. \tag{10}$$

But this last equation is a cubic equation in λ and it is known that any odd-degree equation with real coefficients has necessarily a real root. This proves that there do exist points $(x, y, z) \neq (0, 0, 0)$ having property (9) (λ being the real root of equation (10)). But since the transformation Φ is a rotation, it preserves lengths and so $\lambda = \pm 1$.

This proves that for a rotation Φ there exists a straight line passing through the point O all points of which either remain under the rotation Φ fixed (the case $\lambda = 1$) or go over to positions symmetrical with respect to the point O (the case $\lambda = -1$).

In both cases every point of a plane passing through the point O and perpendicular to the straight line is left by rotation Φ in the same plane. Therefore Φ induces in this plane some orthogonal transformation Φ' with the fixed point O. Since under the hypothesis Φ preserves the orientation of a space and, with $\lambda = -1$, interchanges the sides of a plane, in the case $\lambda = -1$ the transformation Φ' reverses the orientation of the plane. Possessing a fixed point, it is therefore a symmetry of a plane and hence has a straight line of fixed points. All these points are of course fixed points of the rotation Φ as well.

This proves that *any rotation of a space leaves the points of some straight line fixed.*

This straight line is called the *axis of rotation.* In every plane perpendicular to the axis a rotation Φ induces a rotation of the plane through some angle α, the same for all the planes (why?). This angle is called the *angle of a rotation* Φ. It is obvious that the rotation Φ is uniquely determined when we know its axis and angle, so that two rotations coincide if and only if so do their axes and angles. (Recall that the angle α is chosen in the half-interval $-\pi < \alpha \leqslant \pi$; it is clear that rotations through the angles π and $-\pi$ coincide.)

Lecture 28

The Desargues theorem • The Pappus-Pascal theorem • The Fano theorem • The duality principle • Models of the projective plane • Models of the projective straight line and of the projective space • The complex projective straight line

There is a relation between points and straight lines in projective geometry: a point lies on a straight line, and a straight line passes through a point. To emphasize the symmetry of this relation the technical term *incidence* is used and one says that a point is incident with a straight line, and a straight line is incident with a point.

Despite the seeming poorness of this relation, it allows one to state and prove difficult and beautiful theorems. As an example consider the famous Desargues theorem (historically the first theorem of projective geometry):

Theorem 1 (the Desargues theorem). *Let for mutually distinct points* **1**, **2**, **3**, **4**, **5**, **6** *the straight lines* **14**, **25**, and **36** *be incident with the same point other than the points* **1** *to* **6**. *Then the points* **13·46**, **35·62**, **51·24** *are incident with the same straight line.*

Here, say, the symbol **13** designates the straight line passing through points **1** and **3**, and the symbol **13·46** designates the point of intersection of the straight lines **13** and **46**. This theorem does not hold in the affine plane, since straight lines do not always intersect.

The proof of the Desargues theorem, as is the case with other similar theorems, can most easily be carried out in the projective geometry model $\mathscr{P}_O$. Points of this model are the straight lines of a space that pass through a point O. Each of such straight lines is specified by its direction vector $\mathbf{a}$ determined to within proportionality. Therefore, instead of these straight lines it is possible to consider their direction vectors. In other words, we may interpret points of the projective plane $\mathscr{P}_0$ as nonzero vectors in space given to within proportionality.

The points of the straight line passing through the points $\mathbf{a}$ and $\mathbf{b}$ are then of the form $\mu\mathbf{a} + \nu\mathbf{b}$, where $(\mu, \nu) \neq \neq (0, 0)$.

Thus interpreted, the Desargues theorem takes the following "vector" form:

Let for mutually noncollinear vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5, \mathbf{a}_6$ *of a space there exist nonzero numbers* $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$ *such that*

$$(1) \qquad \alpha_1\mathbf{a}_1 + \alpha_4\mathbf{a}_4 = \alpha_2\mathbf{a}_2 + \alpha_5\mathbf{a}_5 = \alpha_3\mathbf{a}_3 + \alpha_6\mathbf{a}_6.$$

Then there exist sets of numbers $(\beta_1, \beta_3, \beta_4, \beta_6)$, $(\gamma_3, \gamma_5, \gamma_6, \gamma_2)$, $(\delta_5, \delta_1, \delta_2, \delta_4)$, *each containing at least one nonzero number, such that we have the equations*

$$\beta_1\mathbf{a}_1 + \beta_3\mathbf{a}_3 = \beta_4\mathbf{a}_4 + \beta_6\mathbf{a}_6,$$
$$\gamma_3\mathbf{a}_3 + \gamma_5\mathbf{a}_5 = \gamma_6\mathbf{a}_6 + \gamma_2\mathbf{a}_2,$$
$$\delta_5\mathbf{a}_5 + \delta_1\mathbf{a}_1 = \delta_2\mathbf{a}_2 + \delta_4\mathbf{a}_4$$

and these three vectors are coplanar.

It is in this form that we shall prove the theorem.

Since the vectors $\mathbf{a}_{1-6}$ are specified only to within proportionality and since none of the numbers α_{1-6} is zero, we can replace each vector $\mathbf{a}_i$ by a vector $\alpha_i\mathbf{a}_i$. Then (1) becomes

$$\mathbf{a}_1 + \mathbf{a}_4 = \mathbf{a}_2 + \mathbf{a}_5 = \mathbf{a}_3 + \mathbf{a}_6$$

whence it immediately follows that

$$\mathbf{a}_1 - \mathbf{a}_3 = -\mathbf{a}_4 + \mathbf{a}_6,$$
$$\mathbf{a}_3 - \mathbf{a}_5 = -\mathbf{a}_6 + \mathbf{a}_2,$$
$$\mathbf{a}_5 - \mathbf{a}_1 = -\mathbf{a}_2 + \mathbf{a}_4.$$

Since

$$(\mathbf{a}_1 - \mathbf{a}_3) + (\mathbf{a}_3 - \mathbf{a}_5) + (\mathbf{a}_5 - \mathbf{a}_1) = 0$$

this proves the Desargues theorem. □

The proof has turned out to be shorter than the statement!

There are four more points in the Desargues theorem besides the points 1 to 6:

14·25 = 14·36, 13·46, 35·62, 51·24.

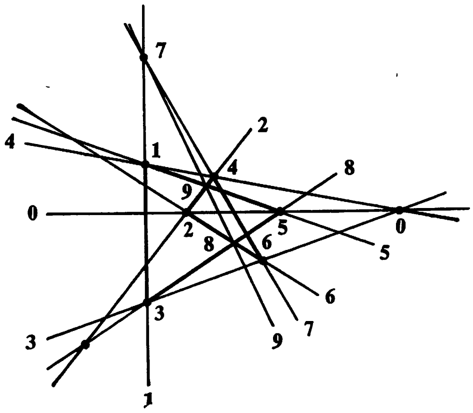
Designate these points by the figures 0, 7, 8, 9 (in the order indicated). Then the ten points 0 to 9 will lie on the ten straight lines:

25, 37, 49, 06, 01, 19, 28, 64, 35, 78.

If we designate these straight lines by the figures 0 to 9 (in the order indicated), then the Desargues theorem can be restated as follows:

(*) *If each of the points* 0 *to* 8 *is incident with the straight line designated by the same figure, then point* 9 *is also incident with straight line* 9.

These ten points and ten straight lines form the so-called *Desargues configuration.* In this configuration any point is



The Desargues configuration

incident with precisely three straight lines and any straight line is incident with precisely three points.

Two triangles are said to be *in perspective* if their vertices are located in pairs on three straight lines intersecting in the same point (the *centre of perspective*). For example, in the Desargues configuration the triangles 135 and 246 are in perspective from the centre of perspective 0. In this terminology the Desargues theorem states that *the points of intersection of corresponding sides of triangles in perspective are on the same straight line* (called the *Desargues straight line*).

It is a remarkable fact (that can be discovered by surveying all the ten points and the ten straight lines) that *in a Desargues configuration all points and straight lines are equivalent*: any of its points (straight lines) is the centre of perspective (the Desargues straight line) of a uniquely defined pair of perspective triangles.

When triples of points 1, 3, 5 and 2, 4, 6 are collinear, the Desargues theorem degenerates into a triviality (all the three points 7, 8, 9 coincide). On the other hand, the following theorem holds for such points:

Theorem 2 (the Pappus-Pascal theorem). *If of the six distinct points* 1, 2, 3, 4, 5, 6 *points* 1, 3, 5 *are incident with the same straight line and points* 2, 4, 6 *are also incident with the same straight line (other than the first) none of these points simultaneously lying on both straight lines, then the three points* 14·23, 45·36, 52·61 *are incident with the same straight line.*

Proof. In a "vector" interpretation the theorem states that *if*

(i) *the vectors* $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5, \mathbf{a}_6$ *are mutually noncollinear,*

(ii) *the vectors* $\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5$ *as well as the vectors* $\mathbf{a}_2, \mathbf{a}_4, \mathbf{a}_6$ *are coplanar,*

(iii) *none of the vectors* $\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5$ *is coplanar with the vectors* $\mathbf{a}_2, \mathbf{a}_4, \mathbf{a}_6$ *and none of the vectors* $\mathbf{a}_2, \mathbf{a}_4, \mathbf{a}_6$ *is coplanar with the vectors* $\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5$,

then the vectors $\mathbf{a}_7, \mathbf{a}_8, \mathbf{a}_9$ *expressed by the formulas*

$$\mathbf{a}_7 = \mu_1\mathbf{a}_1 + \mu_4\mathbf{a}_4 = \nu_2\mathbf{a}_2 + \nu_3\mathbf{a}_3,$$

$$\mathbf{a}_8 = \mu_2\mathbf{a}_2 + \mu_5\mathbf{a}_5 = \nu_1\mathbf{a}_1 + \nu_6\mathbf{a}_6,$$

$$\mathbf{a}_9 = \mu_3\mathbf{a}_3 + \mu_6\mathbf{a}_6 = \nu_4\mathbf{a}_4 + \nu_5\mathbf{a}_5,$$

are coplanar.

Let $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2$ be vectors such that

$$\mathbf{a}_1 \wedge \mathbf{a}_3 = \mathbf{e}_0 \wedge \mathbf{e}_1, \quad \mathbf{a}_2 \wedge \mathbf{a}_4 = \mathbf{e}_0 \wedge \mathbf{e}_2$$

(see Propositions 2 and 3 of Lecture 7). It is clear that the vectors $\mathbf{e}_0$, $\mathbf{e}_1$, $\mathbf{e}_2$ are linearly independent and the vectors $\mathbf{a}_{1-6}$ are expressible in terms of them. In the expression of the vectors $\mathbf{a}_1$, $\mathbf{a}_3$, $\mathbf{a}_5$ (the vectors $\mathbf{a}_2$, $\mathbf{a}_4$, $\mathbf{a}_6$) in terms of the vectors $\mathbf{e}_0$, $\mathbf{e}_1$, $\mathbf{e}_2$ the coefficient of the vector $\mathbf{e}_2$ (the vector $\mathbf{e}_1$) is equal to zero and that of the vector $\mathbf{e}_1$ (the vector $\mathbf{e}_2$) is nonzero. Since the vectors $\mathbf{a}_{1-6}$ are specified only to within collinearity, we may assume without loss of generality that

$$\mathbf{a}_1 = k_1\mathbf{e}_0 + \mathbf{e}_1, \quad \mathbf{a}_2 = k_2\mathbf{e}_0 + \mathbf{e}_2,$$
$$\mathbf{a}_3 = k_3\mathbf{e}_0 + \mathbf{e}_1, \quad \mathbf{a}_4 = k_4\mathbf{e}_0 + \mathbf{e}_2,$$
$$\mathbf{a}_5 = k_5\mathbf{e}_0 + \mathbf{e}_1, \quad \mathbf{a}_6 = k_6\mathbf{e}_0 + \mathbf{e}_2,$$

where k_{1-6} are some numbers. On substituting these expressions in the formula for the vector $\mathbf{a}_7$ we at once see that

$$\mathbf{a}_7 = (\mu_1 k_1 + \mu_4 k_4)\,\mathbf{e}_0 + \mu_1\mathbf{e}_1 + \mu_4\mathbf{e}_2 = \\ = (\nu_2 k_2 + \nu_3 k_3)\,\mathbf{e}_0 + \nu_3\mathbf{e}_1 + \nu_2\mathbf{e}_2.$$

Since the vectors $\mathbf{e}_0$, $\mathbf{e}_1$, $\mathbf{e}_2$ are linearly independent, this equation may hold if and only if $\mu_1 = \nu_3$, $\mu_4 = \nu_2$ and $\mu_1 k_1 + \mu_4 k_4 = \nu_2 k_2 + \nu_3 k_3$. Consequently,

$$\mu_1 (k_1 - k_3) = \mu_4 (k_2 - k_4).$$

Since the vector $\mathbf{a}_7$ is also specified only to within collinearity, we may assume that $\mu_1 = k_2 - k_4$ (it is obvious that $k_2 \neq k_4$) and hence that $\mu_4 = k_1 - k_3$. Thus

$$\mathbf{a}_7 = [(k_2 - k_4)\,k_1 + (k_1 - k_3)\,k_4]\,\mathbf{e}_0 + (k_2 - k_4)\,\mathbf{e}_1 + \\ + (k_1 - k_3)\,\mathbf{e}_2 = (k_1k_2 - k_3k_4)\,\mathbf{e}_0 + \\ + (k_2 - k_4)\,\mathbf{e}_1 + (k_1 - k_3)\,\mathbf{e}_2.$$

It can be shown in a similar way that

$$\mathbf{a}_8 = (k_1k_2 - k_5k_6)\,\mathbf{e}_0 + (k_2 - k_6)\,\mathbf{e}_1 + (k_1 - k_5)\,\mathbf{e}_2,$$
$$\mathbf{a}_9 = (k_3k_4 - k_5k_6)\,\mathbf{e}_0 + (k_4 - k_6)\,\mathbf{e}_1 + (k_3 - k_5)\,\mathbf{e}_2.$$

But then $\mathbf{a}_7 + \mathbf{a}_9 = \mathbf{a}_8$, and therefore the vectors $\mathbf{a}_7$, $\mathbf{a}_8$, $\mathbf{a}_9$ are coplanar. □

In the Pappus-Pascal theorem, besides points 1 to 6 there are other points, $7 = 14 \cdot 23$, $8 = 45 \cdot 36$, and $9 =$

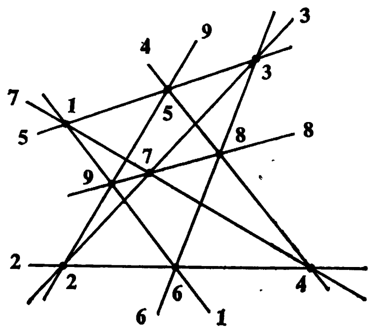
= 52·61, as well as nine straight lines

69, 46, 27, 58, 13, 38, 14, 78, 25.

Designating these straight lines by the figures 1 to 9 (in the order indicated) we may restate the Pappus-Pascal theorem as follows:

(*) *If each of the points* 1 *to* 8 *is incident with the straight line designated by the same figure, then point* 9 *is also incident with straight line* 9.

Thus the Pappus-Pascal theorem allows the same wording of the statement as that of the Desargues theorem! The only difference is what straight lines are in what way designated. It turns out that a statement of the type (*) (but



The Pappus-Pascal configuration

in general with a different number of straight lines and points) suits a whole group of theorems. These are called *configuration theorems.* The Desargues and Pappus-Pascal theorems are the simplest and at the same time most important among them.

The configuration of nine points and nine straight lines that is obtained in the Pappus-Pascal theorem is called a *Pappus configuration.* Just as in the Desargues configuration, in a Pappus configuration three straight lines pass through any point and three points lie on any straight line. In addition, all points and straight lines are also equivalent in Pappus configurations.

An example of a nonconfiguration theorem of projective geometry is the following:

Theorem 3 (the Fano theorem). *If no three of four points* 1, 2, 3, 4 *are collinear, then neither are the points*

$$12 \cdot 34, \quad 13 \cdot 24, \quad 14 \cdot 23.$$

Proof. We must show that if of four vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, $\mathbf{a}_4$ no three are coplanar, then neither are vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ of the form

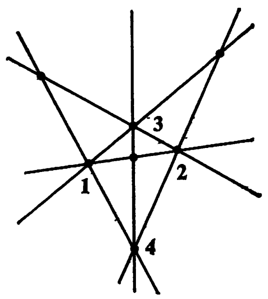
$$\begin{aligned} \mathbf{a} &= \alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 = \alpha_3\mathbf{a}_3 + \alpha_4\mathbf{a}_4, \\ \mathbf{b} &= \beta_1\mathbf{a}_1 + \beta_3\mathbf{a}_3 = \beta_2\mathbf{a}_2 + \beta_4\mathbf{a}_4, \\ \mathbf{c} &= \gamma_2\mathbf{a}_2 + \gamma_3\mathbf{a}_3 = \gamma_1\mathbf{a}_1 + \gamma_4\mathbf{a}_4. \end{aligned} \tag{2}$$

But indeed, according to these equations,

$$\begin{aligned} \alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 - \alpha_3\mathbf{a}_3 - \alpha_4\mathbf{a}_4 &= 0, \\ \beta_1\mathbf{a}_1 - \beta_2\mathbf{a}_2 + \beta_3\mathbf{a}_3 - \beta_4\mathbf{a}_4 &= 0, \\ \gamma_1\mathbf{a}_1 - \gamma_2\mathbf{a}_2 - \gamma_3\mathbf{a}_3 + \gamma_4\mathbf{a}_4 &= 0. \end{aligned}$$

Since under the hypothesis the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are non-coplanar (i.e. linearly independent), these equations are possible if and only if their coefficients are proportional. Thus there are numbers $k \neq 0$ and $l \neq 0$ such that

$$\begin{aligned} \beta_1 &= k\alpha_1, & \beta_2 &= -k\alpha_2, \\ \beta_3 &= -k\alpha_3, & \beta_4 &= k\alpha_4, \\ \gamma_1 &= l\alpha_1, & \gamma_2 &= -l\alpha_2, \\ \gamma_3 &= l\alpha_3, & \gamma_4 &= -l\alpha_4. \end{aligned}$$



The Fano theorem

On the other hand, the first of the equations (2) implies that the vectors $\mathbf{a}$, $\mathbf{d}$, $\mathbf{c}$ have in the basis $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ the coordinates

$$(\alpha_1, \alpha_2, 0), \ (\beta_1, 0, \beta_3), \ (0, \gamma_2, \gamma_3).$$

Since

$$\begin{vmatrix} \alpha_1 & \alpha_2 & 0 \\ \beta_1 & 0 & \beta_3 \\ 0 & \gamma_2 & \gamma_3 \end{vmatrix} = -\alpha_2\beta_1\gamma_3 - \alpha_1\beta_3\gamma_2 = 2kl\,(\alpha_1\alpha_2\alpha_3) \neq 0$$

these vectors are not coplanar. □

The appearance in the last formula of two means that the Fano theorem holds if and only if the characteristic of the ground field $\mathbb{K}$ is *other than two*. At the same time the Desargues and Pappus-Pascal theorems hold for *any* ground field $\mathbb{K}$.

In the proofs of the preceding theorems we used vectors only for the sake of simplifying and reducing the computations. It could be possible, if desired, to develop essentially the same arguments in an arbitrary system of projective coordinates.

As we know, in plane projective geometry points and straight lines play perfectly symmetrical roles. Analytically this finds expression in the incidence condition

$$AX + BY + CZ = 0 \tag{3}$$

of a point $(X:Y:Z)$ and a straight line $(A:B:C)$ being absolutely symmetrical with respect to the coordinates of the point and the straight line. Therefore, if we associate with every point (straight line) a straight line (point) with the same coordinates, then the incidence between points and straight lines will not fail, i.e., for example, points incident with some straight line will be carried over into straight lines incident with the point corresponding to this straight line. This proves the following general principle.

The duality principle for the projective plane. *If in some true sentence about points and straight lines and about incidences between them we interchange the words "point" and "straight line", then we again obtain a true sentence.*

This new sentence is called the *dual* of the original sentence. Note that this principle *is not* a theorem of projective geometry since it tells us not of points and straight lines but of theorems.

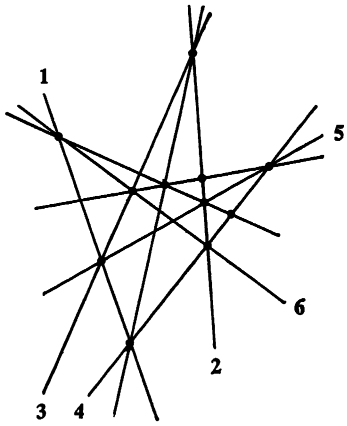
The assertion that "any two points are incident with the same straight line" is the dual of the assertion that "any two straight lines are incident with the same point". Note that in affine geometry the first assertion is true, while the second is not.

A more interesting example is obtained if we consider the dual of the Pappus-Pascal theorem. It is called the Pappus-Brianchon theorem.

Theorem 4 (the Pappus-Brianchon theorem). *If of six distinct straight lines* **1, 2, 3, 4, 5, 6** *straight lines* **1, 3, 5** *are incident with the same point and straight lines* **2, 4, 6** *are also incident with the same point (other than the first), none of these straight lines simultaneously containing both points, then the three straight lines*

14·23, 45·36, 52·61

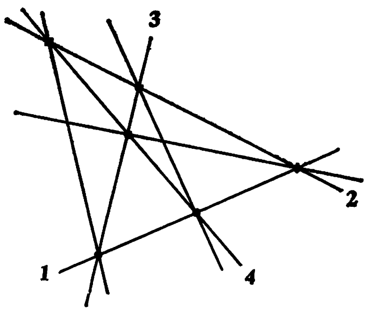
are also incident with the same point. □



The Pappus-Brianchon theorem

Here, say, **14** denotes the point of intersection of straight lines **1** and **4** while **14·23** denotes the straight line passing through the points **14** and **23**.

Just as the Pappus-Pascal theorem, the Pappus-Brianchon theorem also asserts the existence of some configuration. It is remarkable that *we obtain here the same Pappus configuration.*



The dual of the Fano theorem

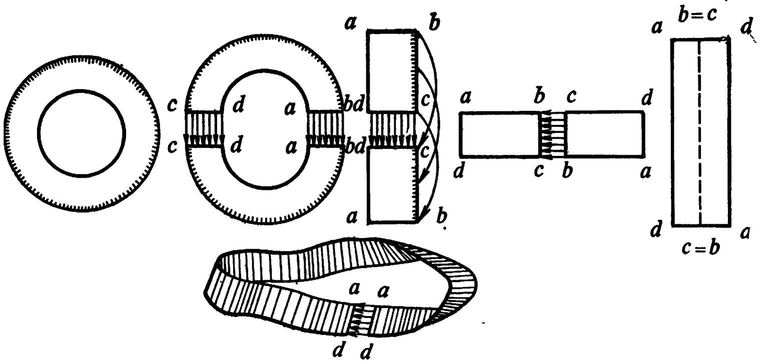
Indeed, the dual of the statement (*) obviously coincides with itself (this statement is said to be *self-dual*). The Pappus-Pascal theorem guarantees the construction of this configuration beginning with two triples of points, while the Pappus-Brianchon theorem does so beginning with two triples of straight lines. □

For the same reason the dual of the Desargues theorem (which certainly does not coincide with this theorem, and what is more, as can be easily seen, is its converse) guarantees the construction of the same Desargues configuration from ten points and ten straight lines.

The dual of the Fano theorem states that for no four straight lines **1**, **2**, **3**, **4** of which no three pass through the same point the straight lines **12·34**, **13·24**, **14·23** (the "generalized diagonals of a quadrangle") pass through the same point either.

In conclusion we consider the question of visualizing the models of the real projective plane.

The "circle" model is a circle diametrically opposite points of which are identified ("glued"). Unfortunately, this gluing is not possible in three-dimensional space, even if we imagine a circle that is made of thin rubber and allow

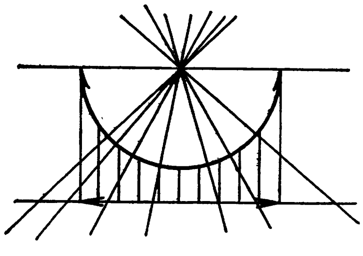


Consecutive steps in the gluing of the projective plane "with a hole" into a Möbius strip

it to be arbitrarily bent and stretched. The situation will change if we cut a smaller circle in it leaving an annulus with the diametrically opposite points of the external circumference identified.

On cutting the annulus along two radii *ab* and *cd* we obtain two rectangles which can now be glued as required without difficulty. After gluing we obtain a rectangle two opposite sides of which consist of the edges of the radial cuts. On gluing these sides, for which purpose the rectangle has to be twisted (and prestretched), we obtain in space a surface called a *Möbius strip*.

A Möbius strip has an edge that is a circumference (it is the edge of the circle cut out of the projective plane at the beginning). Therefore, to obtain again a projective plane it is sufficient to glue the Möbius strip over with the circle along its boundary circumference. However, it cannot be done without self-intersections in a three-dimensional space. One can do it only in a four-dimensional space, by choosing an arbitrary point (outside the three-dimensional space containing the Möbius strip) and constructing over the Möbius strip a cone with vertex at that point.



A model of the projective straight line

Anyone knows the remarkable properties of the Möbius strip (for example, its onesidedness) from books on popular mathematics. Therefore we shall not describe them anew here.

A similar representation of the projective straight line (of a one-dimensional real projective space) presents no difficulty. The model of the projective straight line, similar to the "bundle" model of the projective plane, is a pencil of straight lines in a plane with centre at some point O. On considering a semicircle with centre at O and "straightening" it into a segment we obtain as a model of the projective straight line (similar to the "circle" model) a segment (i.e. a "one-dimensional circle") whose end points are identified. On making this identification we obtain a circle from the segment. Thus *the model of the projective straight line is a circle.*

For a (three-dimensional) projective space the analogue of the "circle" model is the "ball" model. In this model points of a projective space are represented by inner points of a ball and pairs of diametrically opposite points of its boundary sphere.

Take, for example, a ball of radius π. Let O be its centre. Any point $M \neq 0$ uniquely determines a straight line OM and a number $\alpha = |OM|$, $0 < \alpha \leqslant \pi$. Associate with this point a counterclockwise (if viewed along the axis in the direction of the vector $\overrightarrow{OM}$) rotation about the axis OM through an angle α. Associate with the point O an identity transformation (rotation through an angle 0). It is clear that by an appropriate choice of the point M in the ball we may thus obtain any rotation of the space. To two different points of the ball will then correspond different rotations, provided these points are not diametrically opposite points in the boundary of the ball, in which case the same rotation is obtained.

This shows that a model of a three-dimensional projective space is the group Rot_O (3) of rotations of the space. Since this group is isomorphic to the group of proper orthogonal matrices SO (3) we finally find that *the set* SO (3) *of proper orthogonal matrices of the third order is a model of the three-dimensional real projective space.*

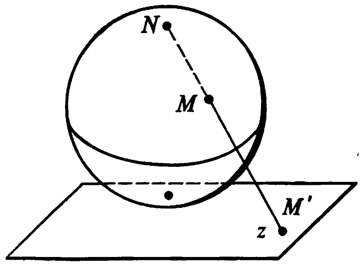
We see in particular that matrices of SO(3) can be parametrized in quadruples $(a^0 : a^1 : a^2 : a^3)$ of numbers specified to within proportionality. It is possible to write beautiful explicit formulas for this parametrization, but we haven't the time.

A similar visualization of complex projective spaces $\mathbb{C}P^n$ is possible only for $n = 1$ (since even the complex projective plane $\mathbb{C}P^2$ is of real dimension 4 and therefore cannot be visualized). As to the complex projective straight line $\mathbb{C}P^1$ it is by definition a set of proportionality classes $(z_0 : z_1)$ of pairs of complex numbers (z_0, z_1). For $z_0 \neq 0$ every such class is uniquely characterized by a complex number $z = z_1/z_0$, while for $z_0 = 0$ there is only one class, $(0 : 1)$.

Designating the last class by the symbol ∞ we thus find that *a model of the complex projective straight line is the set of complex numbers* $\mathbb{C}$ *augmented with the symbol* ∞.

Every complex number $z = x + iy$ can be represented in a plane (with a given system of rectangular coordinates) by a point (x, y). Therefore the set $\mathbb{C}$ is often referred to as the *plane of complex numbers* (the sometimes used term "complex plane" is very unhappy, since it is employed as the name of the plane over the field $\mathbb{C}$, a plane that, from a "real" viewpoint, is of dimension 4), while the set $\mathbb{C}$ with the added symbol ∞ is called the *augmented plane of complex numbers* $\mathbb{C}^+$.

Consider on a sphere S in a three-dimensional space an arbitrary point N and a plane π tangent to S at the diametrically opposite point. For any point M of the sphere S distinct from the point N a straight line NM intersecting the plane π is then determined. On associating with the point M the point of intersection M' of the straight line NM and the plane π we obtain a bijective mapping $M \mapsto M'$ (called the *stereographic projection*) of the sphere S, without the point N, onto the plane π.



The Riemann sphere

Taking as π the plane of complex numbers we are in particular enabled to represent complex numbers as points of a sphere "punctured" at one point N. On assigning to this point the symbol ∞ we thus obtain a bijective correspondence between the points of the augmented plane $\mathbb{C}^+$ and those of the sphere S. On these grounds the augmented plane $\mathbb{C}^+$ is often referred to as the *sphere of complex numbers* or *Riemann sphere*.

We thus see that *the sphere is a model of the complex projective straight line.*

Lecture 29

Linear fractional transformations • Linear transformations • Inversion • Inversions and linear fractional transformations • Two properties of linear fractional transformations • Fixed points of linear fractional transformations • Parabolic, elliptical, hyperbolic and loxodromic linear fractional transformations • The three-point theorem • The multiplier of linear fractional nonparabolic transformation • Classification of linear fractional transformations • Stereographic projection formulas • Rotations of a sphere as linear fractional transformations of a plane • Isometries of a cube

---

For any complex numbers a, b, c, d satisfying the relation

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \neq 0$$

the formula

$$w = \frac{az+b}{cz+d} \tag{1}$$

defines the transformation $z \mapsto w$ of the augmented plane $\mathbb{C}^+$ of complex numbers into itself. (It is assumed by definition that $w = \frac{a}{c}$ for $z = \infty$ and $w = \infty$ for $z = -\frac{d}{c}$; in particular, if $c = 0$, then $w = \infty$ for $z = \infty$.)

Definition 1. Transformations $\mathbb{C}^+ \to \mathbb{C}^+$ of the form (1) are called *linear fractional transformations.*

These transformations are closely connected with transformations of the complex projective straight line $\mathbb{C}P^1$. Indeed, as a transformation established from the equality of projective coordinates, every projective transformation $\mathbb{C}P^n \to \mathbb{C}P^n$ must be written in homogeneous projective coordinates using the same formulas as those used to write a transformation of projective coordinates (see Lecture 27), i.e. formulas (4) of Lecture 26. For $n = 1$ these formulas have, to within the notation, the form

$$\begin{aligned}\rho w_0 &= cz_0 + dz_1 \\ \rho w_1 &= az_0 + bz_1\end{aligned} \quad \text{where} \quad \begin{vmatrix} c & d \\ a & b \end{vmatrix} \neq 0$$

and therefore on transforming from them to the nonhomogeneous coordinates $z = \frac{z_1}{z_0}$ and $w = \frac{w_1}{w_0}$ we obtain precisely transformation (1). This proves that *linear fractional transformations* (1) *are simply projective transformations of the complex projective straight line in the model* $\mathbb{C}^+$. □

It is clear that *all linear fractional transformations form a group*, i.e. a composition (product) of two linear fractional transformations and an inverse of a linear fractional transformation are linear fractional transformations. An inverse of linear fractional transformation (1) is given by the formula

$$z = \frac{-dw + b}{cw - a}.$$

According to what has been said above, the group of linear fractional transformations is isomorphic to the group Proj (1, $\mathbb{C}$).

Transformation (1) remains unchanged when the coefficients a, b, c, d are all simultaneously multiplied by the same number $\rho \neq 0$. Therefore any linear fractional transformation can be written in the form (1) with the additional condition

$$ad - bc = 1.$$

Such a notation of linear fractional transformations will be called a *normed* notation. Every linear fractional transformation allows two normalizations going over into each other when all the coefficients are multiplied by -1.

For $a = d = 1$ and $c = 0$ we have the transformation

$$w = z + b. \tag{2}$$

Interpreting the complex number $b = b_1 + ib_2$ as a vector (with coordinates b_1 and b_2) we see that this transformation is a translation by a vector b.

Let $d = 1$, $b = c = 0$ and hence

(3) $$w = az.$$

For $|a| = 1$, i.e. when

$$a = \cos\alpha + i \sin\alpha$$

(or, otherwise, $a = e^{i\alpha}$), transformation (3) has the form

(4) $$w = e^{i\alpha}z$$

and is a rotation through an angle α (to discover this it is sufficient to write the transformation (4) in rectangular coordinates $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$). When $a = r > 0$ is real and positive transformation (3) is a homothety with coefficient r. In the general case $a = re^{i\alpha}$ and transformation (3) is a composition of a rotation through an angle $\alpha = \arg a$ and a homothety with coefficient $r = |a|$.

Transformations (2) and (3) are both special cases of linear fractional transformations of the form

(5) $$w = az + b$$

obtained for $c = 0$ and $d = 1$. Such transformations are called *linear* transformations.

Every linear transformation is a composition of a translation (2) and a transformation (3). In particular, for $|a| = 1$ a linear transformation is a motion (and conversely, any motion of a plane is a linear transformation (5) with $|a| = 1$). It follows that *linear transformations are exactly similarity transformations.* □

Note that notation (5) of linear transformations is not a normed notation (for $a \neq 1$).

Let T be an arbitrary circle of radius R and let M_0 be its centre. Points M and N (distinct from the point M_0) are said to be *symmetrical with respect to the circle T* if

(a) these points lie on the same ray emanating from the point M_0; and

(b) we have

$$|M_0M| \cdot |M_0N| = R^2.$$

The point M_0 is defined to be symmetrical to the point ∞.

If the points M_0, M, and N are represented by complex numbers z_0, z, and w, then conditions (a) and (b) may be written as:

(a) the arguments of the numbers $z - z_0$ and $w - z_0$ are equal:

$$\arg (z - z_0) = \arg (w - z_0),$$

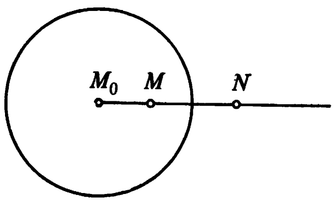
(b) for their moduli we have

$$| z - z_0 | \cdot | w - z_0 | = R^2.$$

According to condition (b) the complex numbers $w - z_0$ and $\frac{R^2}{\overline{z - z_0}}$ have equal moduli and according to condition (a) they have equal arguments as well. Therefore these numbers are equal:

$$w - z_0 = \frac{R^2}{\overline{z - z_0}}. \tag{6}$$

Thus points representing the complex numbers z and w (commonly called simply points z and w) are symmetrical with respect to a circle of radius R with centre at a point z_0 if and only if relation (6) holds.



Points symmetrical with respect to a circle

Formula (6) may be considered to give some transformation $z \mapsto w$ of the augmented plane $\mathbb{C}^+$ of complex numbers.

Definition 2. Transformation (6) of the plane $\mathbb{C}^+$ carrying every point z into a point w symmetrical with respect to a circle T is called an *inversion* with respect to the circle T.

It carries the centre of this circle, M_0, over into the point ∞, and conversely, the point ∞ into the centre M_0.

In the special case where T is a *unit circle* (of radius 1 with centre at a point $z_0 = 0$) formula (6) becomes

$$w = \frac{1}{\bar{z}}. \tag{7}$$

It follows from this formula that by composing this transformation with a transformation $z \mapsto \bar{z}$ we obtain a linear fractional transformation

$$w = -\frac{1}{z}. \tag{8}$$

The transformation $z \mapsto \bar{z}$ is a symmetry with respect to the abscissa axis $y = 0$. We thus see that linear fractional transformation (8) is a composition of a symmetry with respect to a straight line and an inversion with respect to a circle.

In the theory of linear fractional transformations it is appropriate to rank all straight lines with circles. The formal analytical basis for this is provided by the fact that the equation

$$E(x^2 + y^2) + Ax + By + C = 0 \tag{9}$$

gives a circle when $E \neq 0$ (and when $A^2 + B^2 > 4EC$) and a straight line when $E = 0$. Accordingly, it is appropriate to say that an *inversion with respect to a straight line* is a symmetry with respect to it. In this terminology we may thus state that *linear fractional transformation* (8) *is a composition of two inversions.* □

An inversion with respect to a circle of radius R and with centre at a point $z_0 = 0$ may be expressed by the formula

$$w = \frac{R^2}{\bar{z}}.$$

Its composition with an inversion (7) is therefore of the form

$$w = R^2 z,$$

i.e. is a homothety with coefficient R^2. Since any positive real number may be represented by R^2, this in particular proves that *any homothety is a composition of two inversions.* □

As we know (see Lecture 27), any motion of a plane is also a composition of two inversions (symmetries with respect to straight lines). Therefore *any linear transformation* (*similarity transformation*) *is a composition of four inversions.* □

The formula

$$\frac{az+b}{cz+d} = \frac{a}{c} - \frac{1}{c(cz+d)}$$

(we assume that $ad - bc = 1$) shows that for $c \neq 0$ linear fractional transformation (1) is a composition of a linear transformation $z \mapsto cz + d$, a transformation $z \mapsto \frac{1}{z}$ and a linear transformation $z \mapsto -\frac{1}{c} z + \frac{a}{c}$. This, together with what has been said above, proves the following proposition.

Proposition 1. *Any linear fractional transformation is a composition of an even number of inversions.* □

Since

$$z\bar{z} = x^2 + y^2, \quad x = \frac{z+\bar{z}}{2}, \quad y = \frac{z-\bar{z}}{2i},$$

we can write equation (9) of a circle (or straight line) as

$$Ez\bar{z} + A\left(\frac{z+\bar{z}}{2}\right) + B\left(\frac{z-\bar{z}}{2i}\right) + C = 0,$$

i.e. as

$$Ez\bar{z} + \bar{P}z + P\bar{z} + C = 0 \tag{10}$$

where E and C are real numbers and $P = \frac{A}{2} - i\frac{B}{2} i$ is some complex number.

Transformation (8) carries this circle over into a curve with the equation

$$E + \bar{P}\bar{z} + Pz + Cz\bar{z} = 0,$$

i.e. again into a circle (that is a straight line when $C = 0$). Since, as was noted above, any linear fractional transformation is either a linear transformation, i.e. a similarity transformation (trivially carrying any circle over into a circle) or a composition of transformation (8) and two linear transformations we thus see that any *linear fractional transformation carries a circle or a straight line over into a circle or a straight line.* □

This property of linear fractional transformations is usually called their *circle property*.

Since any inversion (6) differs from a linear fractional transformation by some symmetry, the *circle property holds for every inversion as well.* □

Lemma 1. *Two circles (or straight lines)*

(11) $$E(x^2+y^2)+Ax+By+C=0$$

and

(12) $$E_1(x^2+y^2)+A_1x+B_1y+C_1=0$$

are orthogonal (intersect at right angles) if and only if

(13) $$AA_1+BB_1=2(EC_1+E_1C).$$

Proof. When $E \neq 0$ the centre of circle (11) has coordinates $\left(-\frac{A}{2E}, -\frac{B}{2E}\right)$ and the square of radius R^2 of the circle may be expressed by the formula

(14) $$R^2=\frac{A^2+B^2-4EC}{4E^2}.$$

Similarly, when $E_1 \neq 0$ the centre of a circle (12) has coordinates $\left(-\frac{A_1}{2E_1}, -\frac{B_1}{2E_1}\right)$ and the square of the radius R_1^2 of the circle can be expressed by the formula

(15) $$R_1^2=\frac{A_1^2+B_1^2-4E_1C_1}{4E_1^2}.$$

On the other hand, a trivial elementary-geometry argument shows that circles (11) and (12) are orthogonal if and only if the sum $R^2+R_1^2$ of the squares of their radii is equal to the square of the distance between their centres, i.e. if

$$R^2+R_1^2=\left(\frac{A}{2E}-\frac{A_1}{2E_1}\right)^2+\left(\frac{B}{2E}-\frac{B_1}{2E_1}\right)^2.$$

On substituting in this relation expressions (14) and (15) we obtain, after simplifications, condition (13).

When $E_1=0$, i.e. when circle (12) is in fact a straight line $A_1x+B_1y+C_1=0$, the statement that the straight line and the circle (11) are orthogonal implies that the straight line passes through the centre of the circle (11), i.e. that

$$A_1\left(-\frac{A}{2E}\right)+B_1\left(-\frac{B}{2E}\right)+C_1=0.$$

After simplifications we again obtain condition (13) (with $E_1=0$).

1/2 21-01283

Finally, for $E = 0$ and $E_1 = 0$ condition (13) becomes $AA_1 + BB_1 = 0$, the perpendicularity condition of two straight lines we know. □

When equations (11) and (12) are written in the form (10), i.e. as

$$Ez\bar{z} + \bar{P}z + P\bar{z} + C + 0$$

and

$$E_1 z\bar{z} + \bar{P}z + P_1\bar{z}_1 + C_1 = 0,$$

condition (13) becomes

$$P\bar{P}_1 + \bar{P}P_1 = EC_1 + E_1C. \tag{16}$$

As shown above, transformation (8) carries a circle (11) over into a circle for which the role of the coefficient E is played by the coefficient C, and that of the coefficient C is played by the coefficient E, and the coefficient P is replaced by a complex conjugate number $\bar{P}$.

In a short but clear notation this yields

$$E \Rightarrow C, \quad C \Rightarrow E, \quad P \Rightarrow \bar{P}. \tag{17}$$

Similarly for the circle (12)

$$E_1 \Rightarrow C_1, \quad C_1 \Rightarrow E_1, \quad P_1 \Rightarrow \bar{P}_1. \tag{18}$$

But it is clear that replacements (17) and (18) leave relation (16) invariant (carry it into itself). This proves that a linear fractional transformation (8) preserves the orthogonality of circles, i.e. carries orthogonal circles over into orthogonal circles.

Since any linear fractional transformation is a composition of linear transformations and a transformation (8) it follows that *any linear fractional transformation also preserves orthogonality of circles* (for linear transformations, being similarity transformations, do possess this property). □

Remark. As a matter of fact linear fractional transformations preserve any (and not only right) angles (are said to be *conformal transformations*), but proving this fact falls outside the limits of this exposition.

From the elementary-geometry theorem on the square of the length of a tangent to a circle it immediately follows that points M and N are symmetrical with respect to a circle T if and only if any circle passing through these points is orthogonal to the circle T. It is clear that this statement remains valid also when T is a straight line. Since every linear fractional transformation Φ carries orthogonal circles (straight lines) over into orthogonal circles (straight lines) it follows that *points M and N symmetrical with respect to an arbitrary circle T are carried by every linear fractional transformation Φ over into points $M' = \Phi(M)$ and $N' = \Phi(N)$ symmetrical with respect to a circle $T' = \Phi(T)$.* □

A point which a linear fractional transformation (1) leaves fixed is called a *fixed point* of the transformation. Fixed points are determined from the equation

$$z = \frac{az+b}{cz+d},$$

i.e. from

$$cz^2 + (d - a)z - b = 0. \tag{19}$$

Since a quadratic equation has not more than two roots, it follows that *a nonidentity linear fractional transformation has not more than two fixed points.* □

Since by virtue of the normalization condition ($ad - bc = 1$) we have

$$(d - a)^2 + 4bc = (a + d)^2 - 4,$$

the fixed points of a normed linear fractional transformation (1) *can be expressed by the formula*

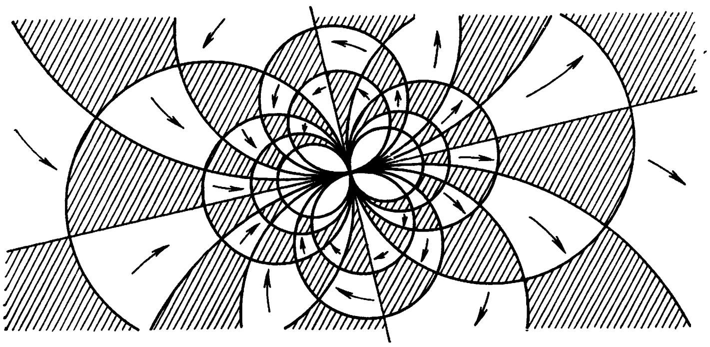
$$z_{1,2} = \frac{a-d \pm \sqrt{(a+d)^2 - 4}}{2c}. \tag{20}$$

When $c = 0$ (for a linear transformation) one of the fixed points is the point ∞ and the other is a point $\frac{b}{d-a}$ (which, with $d = a$, i.e. for a translation, is also the point ∞)

A linear fractional transformation is said to be *parabolic* if it has only one fixed point. From formula (20) it follows that *a normed transformation* (1) *is a parabolic transformation if and only if the number $a + d$ is real and equal to ± 2.* □

Parabolic transformations with fixed point ∞ are translations and nothing but translations.

Let a linear fractional transformation (1) have two fixed points z_1 and z_2. Consider a family $\mathscr{K}$ of circles passing through these points and a family $\mathscr{H}$ of circles orthogonal to all circles of the family $\mathscr{K}$. It immediately follows from



A parabolic transformation with a finite fixed point. Each shaded region goes over into the next one in the direction indicated by the arrow

the circle property of linear fractional transformations that transformation (1) carries each circle of the family $\mathscr{K}$ over into a circle of the same family and since, in addition, it preserves the orthogonality of circles it also carries each circle of the family $\mathscr{H}$ over into a circle of the same family.

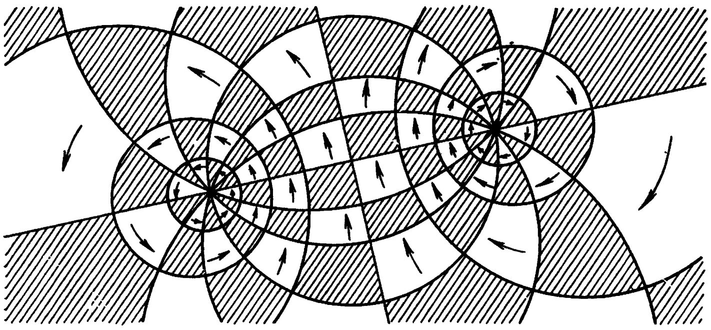
It may happen that under transformation (1) each circle of the family $\mathscr{H}$ is carried into itself, i.e. any of its points is carried over into a (generally speaking, different) point of the same circle. Such a linear fractional transformation is called an *elliptical* transformation. Under an elliptical transformation each point moves round that circle of the family $\mathscr{H}$ on which it lies.

One example of an elliptical transformation is a rotation with fixed points 0 and ∞. The family $\mathscr{H}$ consists of concentric circles with centres at the point 0 and the family $\mathscr{K}$ of straight lines passing through this point.

Transformation (1) is said to be *hyperbolic* if, on the contrary, it carries into itself each circle of the family $\mathscr{K}$

and, in addition, each of the arcs into which the circles of $\mathscr{K}$ are divided by the fixed points.

An example of a hyperbolic transformation is a homothety $w = az$, $a > 0$, with fixed points 0 and ∞. The families $\mathscr{H}$ and $\mathscr{K}$ for this homothety are the same as those for a rotation $w = e^{i\alpha}z$.



An elliptical transformation with finite fixed points

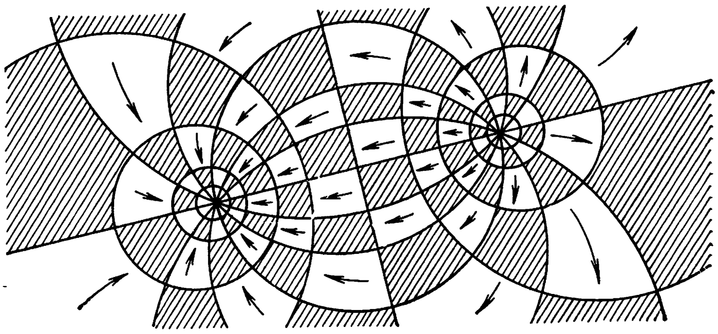
Generally speaking, a linear fractional transformation with two fixed points is neither hyperbolic nor elliptical. Such a transformation is called a *loxodromic* transformation. One example is a composition of a hyperbolic and an elliptical transformation with the same fixed points (say, a composition of a homothety and a rotation).

Proposition 2. *For any two triples of distinct points of the plane* $\mathbb{C}^+$, (z_1, z_2, z_3) and (w_1, w_2, w_3), *there exists a unique linear fractional transformation* Φ *carrying the first triple over into the second.*

Proof. The uniqueness of the transformation Φ is obvious. Indeed, if there is another linea.· fractional transformation, Ψ, carrying the triple (z_1, z_2, z_3) over into the triple (w_1, w_2, w_3), then the transformation $\Psi^{-1} \circ \Phi$ will leave the three points z_1, z_2, z_3 fixed, which, as we know, is impossible for a nonidentity linear fractional transformation.

As far as existence is concerned, it is sufficient to prove it just for the special case where $(w_1, w_2, w_3) = (0, \infty, 1)$, since if Φ carries (z_1, z, z_3) over into $(0, \infty, 1)$ and Ψ carries (w_1, w_2, w_3) over into $(0, \infty, 1)$, then $\Psi^{-1} \circ \Phi$ will carry (z_1, z_2, z_3) over into (w_1, w_2, w_3).

On the other hand, it is easy to directly select a linear fractional transformation carrying (z_1, z_2, z_3) over into $(0, \infty, 1)$. Indeed, for such a transformation the numerator



A hyperbolic transformation with finite fixed points

must vanish when $z = z_1$, and the denominator must vanish when $z = z_2$, while for $z = z_3$ the numerator and denominator must take equal values. It is clear that this condition is satisfied by the transformation

$$w = \frac{z - z_1}{z - z_2} \cdot \frac{z_3 - z_2}{z_3 - z_1}. \quad \square \tag{21}$$

Remark 1. In formula (21) all points z_1, z_2, z_3 are assumed to be finite (distinct from the point ∞). For $z_1 = \infty$ we should set

$$w = \frac{z_3 - z_2}{z - z_2},$$

and, respectively,

$$w = \frac{z - z_1}{z_3 - z_1} \quad \text{when} \quad z_2 = \infty,$$

$$w = \frac{z - z_1}{z - z_2} \quad \text{when} \quad z_3 = \infty.$$

Each of these formulas is obtained from the general formula (21) by cancelling the two multipliers containing ∞.

Remark 2. The presented proof of Proposition 2 allows us to write at once for a linear fractional transformation carrying a triple (z_1, z_2, z_3) over into a triple (w_1, w_2, w_3) a formula relating z and w. It is indeed clear that such a formula has the form

$$\frac{w-w_1}{w-w_2}\cdot\frac{w_3-w_2}{w_3-w_1}=\frac{z-z_1}{z-z_2}\cdot\frac{z_3-z_2}{z_3-z_1}. \tag{22}$$

To obtain an explicit expression (1) of the point w in terms of the point z, it is necessary to solve this equation for w.

Of course, if there is a point ∞ among the points z_1, z_2, z_3 or the points w_1, w_2, w_3, formula (22) should be modified in accordance with remark 1.

Formula (22) may be regarded as a relation to be satisfied by two quadruples (z_1, z_2, z_3, z_4) and (w_1, w_2, w_3, w_4) of points of the plane $\mathbb{C}^+$ for the second quadruple to be obtained from the first by some (according to Proposition 2, unique) linear fractional transformation. Denoting for uniformity z by z_4 and w by w_4 and calling the number

$$\frac{z_4-z_1}{z_4-z_2}\cdot\frac{z_3-z_2}{z_3-z_1}=\frac{z_4-z_1}{z_4-z_2}:\frac{z_3-z_1}{z_3-z_2}$$

a *cross ratio* of the points z_1, z_2, z_3, z_4 we may state this fact as the following proposition.

Proposition 3. *Four points z_1, z_2, z_3, z_4 of the plane $\mathbb{C}^+$ can be carried over into points w_1, w_2, w_3, w_4 by some linear fractional transformation if and only if the cross ratios of these points are equal.* □

Now we can easily write a general form of a linear fractional transformation Φ with given fixed points z_1, z_2. Indeed, let z_3 be any other point and let $w_3 = \Phi(z_3)$. Then Φ can be described as a linear fractional transformation carrying a triple (z_1, z_2, z_3) over into a triple (z_1, z_2, w_3) and hence given by the formula

$$\frac{w-z_1}{w-z_2}\cdot\frac{w_3-z_2}{w_3-z_1}=\frac{z-z_1}{z-z_2}\cdot\frac{z_3-z_2}{z_3-z_1}.$$

Setting

$$K=\frac{w_3-z_1}{w_3-z_2}\cdot\frac{z_3-z_2}{z_3-z_1}, \tag{23}$$

we can write this formula as

$$\frac{w-z_1}{w-z_2}=K\frac{z-z_1}{z-z_2}. \tag{24}$$

This is the general formula giving a linear fractional transformation Φ with fixed points z_1, z_2 (distinct from the point ∞; if say $z_2 = \infty$, then the formula takes the form $w - z_1 = K(z - z_1)$).

The number K in formula (24) is called the *multiplier* of a linear fractional transformation Φ (with two fixed points, i.e. not a parabolic one). It may be any (nonzero) complex number.

To clarify the meaning of formula (24) we introduce into consideration a linear fractional transformation

$$\Omega\colon z \mapsto \frac{z-z_1}{z-z_2}$$

and a linear fractional transformation

$$\mathrm{K}\colon z \mapsto Kz. \tag{25}$$

Then formula (24) is equivalent to the equation

$$\Omega \circ \Phi = \mathrm{K} \circ \Omega,$$

i.e. to the equation

$$\Phi = \Omega^{-1} \circ \mathrm{K} \circ \Omega,$$

signifying in the language of group theory that transformation Φ is *conjugate* in the group of all linear fractional transformations to transformation K with respect to a transformation Ω.

Transformation Ω carries points z_1 and z_2 over into the points 0 and ∞ respectively, and hence the circles of the family $\mathscr{K}$ into straight lines passing through the point 0 and the circles of the family $\mathscr{H}$ into concentric circles with centre at the point 0. When $K > 0$ transformation K moves the points along these straight lines, interchanging the circles, and conversely when $|K| = 1$ it moves every point round the circles, interchanging the straight lines. After that a transformation Ω^{-1} again carries straight lines over into the circles of the family $\mathscr{K}$ and the concentric circles into the circles of the family $\mathscr{H}$. Therefore, when $K > 0$

transformation Φ moves the points of a plane round the circles of the family $\mathscr{K}$, interchanging the circles of the family $\mathscr{H}$, i.e. is a hyperbolic transformation, and conversely when $|K| = 1$ it moves the points of a plane round the circles of the family $\mathscr{H}$, interchanging the circles of the family $\mathscr{K}$, i.e. is an elliptical transformation. With K nonreal and $|K| \neq 1$ transformation Φ is loxodromic.

This reasoning may be presented in a more compact form taking into account that in fact it proves that *for any linear fractional transformations* Ω *and* K *the transformation* $\Omega^{-1} \circ \mathrm{K} \circ \Omega$ *is of the same type as the transformation* K. After that all that is necessary is to recall that transformation (25) is hyperbolic for $K > 0$, elliptical for $|K| = 1$ and loxodromic for all other K.

What remains now is to learn how to compute the multiplier K directly from the notation (1) (assumed to be normed) of a linear fractional transformation Φ. To do this we shall use formula (23) applying it to the case where $z_3 = \infty$ and hence $w_3 = a/c$ (it is clear that the left hand side of the formula does not depend on the choice of z_3). According to this formula

$$K = \frac{a - cz_1}{a - cz_2},$$

where z_1, z_2 are the roots (20) of the quadratic equation (19). Since according to the Viète formulas

$$z_1 + z_2 = -\frac{d-a}{c}, \quad z_1 z_2 = -\frac{b}{c}$$

and so

$$z_1^2 + z_2^2 = (z_1 + z_2)^2 - 2z_1 z_2 = \frac{(d-a)^2 + 2bc}{c^2},$$

we have

$$K + \frac{1}{K} = \frac{a - cz_1}{a - cz_2} + \frac{a - cz_2}{a - cz_1} = \frac{(a - cz_1)^2 + (a - cz_2)^2}{(a - cz_1)(a - cz_2)} =$$

$$= \frac{2a^2 - 2ac(z_1 + z_2) + c^2(z_1^2 + z_2^2)}{a^2 - ac(z_1 + z_2) + c^2 z_1 z_2} = \frac{2a^2 + 2a(d-a) + (d-a)^2 + 2bc}{a^2 + a(d-a) - bc} =$$

$$= \frac{(a+d)^2 - 2(ad - bc)}{ad - bc} = (a + d)^2 - 2.$$

This proves that *the multiplier K is a root of the quadratic equation*

$$K+\frac{1}{K}=(a+d)^2-2. \quad \square \tag{26}$$

(We remark that the multiplier K of a linear fractional transformation with two fixed points depends on the way the points are numbered; with a different numbering it is replaced by its inverse. That is why equation (26) has two roots K and K^{-1}.)

From equation (26) it immediately follows (it is sufficient to write the explicit formula for its roots) that K is real and positive (and hence transformation (1) is hyperbolic) if and only if the number $a + d$ is real and $|a + d| > 2$. Similarly $|K| = 1$ (and hence transformation (1) is elliptical) if and only if $a + d$ is real and $|a + d| < 2$. Therefore transformation (1) is loxodromic if and only if the number $a + d$ is nonreal (recall that when $|a + d| = 2$ we go outside the class of linear fractional transformations with two fixed points to obtain parabolic transformations, i.e. transformations having only one fixed point).

This proves the following final theorem.

Theorem 5 (classification of linear fractional transformations). *A linear fractional transformation in the normed notation*

$$w=\frac{az+d}{cz+d}$$

is nonloxodromic if and only if the number $a + d$ is real. When this condition holds the transformation is

$$\left.\begin{array}{r}\textit{hyperbolic}\\ \textit{parabolic}\\ \textit{elliptical}\end{array}\right\}\ \textit{if and only if}\ \left\{\begin{array}{l}|a+d|<2,\\ |a+d|=2,\\ |a+d|<2.\quad\square\end{array}\right.$$

In conclusion we shall describe one interesting class of linear fractional transformations. As a preliminary we shall find explicit formulas for a stereographic projection of a sphere S in the plane $\mathbb{C}^+$.

Let S be a *unit sphere* given in rectangular coordinates ξ, η, ζ by the equation

$$\xi^2+\eta^2+\zeta^2=1$$

(we now cannot use the conventional symbols x, y, z for coordinates, since we employ the symbol z to denote complex numbers). As the projection plane identifiable with the plane $\mathbb{C}^+$ of complex numbers $z = x + iy$ we now take the equatorial plane $\zeta = 0$ of the sphere S (which is equivalent to an additional homothety with the coefficient 1/2), taking as the pole N of the stereographic projection the point (0, 0, 1). The straight line NM passing through the point N and an arbitrary point $M(\xi, \eta, \zeta)$ of the sphere S has canonical equations of the form

$$\frac{X}{\xi}=\frac{Y}{\eta}=\frac{Z-1}{\zeta-1} \tag{27}$$

(we use X, Y, Z here to denote the "moving" coordinates of the points of a straight line) and therefore intersects the projection plane in a point $\left(\frac{-\xi}{\zeta-1}, \frac{-\eta}{\zeta-1}, 0\right)$. This means that a stereographic projection carries the point $M(\xi, \eta, \zeta)$ over into a point

$$z=\frac{\xi}{1-\zeta}+i\frac{\eta}{1-\zeta}=\frac{\xi+i\eta}{1-\zeta} \tag{28}$$

of the plane $\mathbb{C}^+$ (for $\xi = 0$, $\eta = 0$, $\zeta = 1$ we naturally put $z = \infty$).

Let $M^*(-\xi, -\eta, -\zeta)$ be the point diametrically opposite to the point $M(\xi, \eta, \zeta)$ and let z^* be the corresponding point of the plane $\mathbb{C}^+$. Then

$$z^*=-\frac{\xi+i\eta}{1+\zeta} \tag{29}$$

and hence

$$z^*\bar{z}=-\frac{\xi^2+\eta^2}{1-\zeta^2}=-1.$$

Conversely, if $z^*\bar{z} = -1$ and z is given by formula (28), then for z^* we have formula (29). This proves that *points z and z^* of the plane $\mathbb{C}^+$ are images under a stereographic projection of diametrically opposite points of a sphere if and only if*

$$z^*=-\frac{1}{\bar{z}}. \quad \square \tag{30}$$

Consider now an arbitrary rotation Σ of a sphere S. Going over with the aid of a stereographic projection $\Pi: S \to \mathbb{C}^+$ from S to $\mathbb{C}^+$ we obtain a linear fractional transformation $\Phi = \Pi \circ \Sigma \circ \Pi^{-1}$ of the plane $\mathbb{C}^+$ which is said to *represent in* $\mathbb{C}^+$ the rotation Σ. Since the rotation Σ carries diametrically opposite points over into diametrically opposite points and the transformation Φ carries points connected by relation (30) over into points also connected by relation (30). This means that if Φ is given by the (normed) formula (1), then for any points z and z^* connected by the relation $z^*\bar{z} = -1$ there must hold the relation

$$\left(\frac{az^*+b}{cz^*+d}\right)\left(\frac{\bar{a}\bar{z}+\bar{b}}{\bar{c}\bar{z}+\bar{d}}\right) = -1,$$

i.e. we must have identically for z the equation

$$\left(\frac{a-b\bar{z}}{c-d\bar{z}}\right)\left(\frac{\bar{a}\bar{z}+\bar{b}}{\bar{c}\bar{z}+\bar{d}}\right) = -1,$$

equivalent to

$$(a-b\bar{z})(\bar{a}\bar{z}+\bar{b}) + (c-d\bar{z})(\bar{c}\bar{z}+\bar{d}) = 0.$$

This is possible if and only if all the coefficients are zero, i.e. if

$$a\bar{b}+c\bar{d} = 0, \quad a\bar{a}-b\bar{b}+c\bar{c}-d\bar{d} = 0, \quad b\bar{a}+d\bar{c} = 0.$$

The first (and third) relations imply that there exists $\lambda \neq 0$ such that $d = \lambda\bar{a}$ and $b = -\lambda\bar{c}$. Therefore the second relation takes the form $(1 - \lambda\bar{\lambda})(a\bar{a} + c\bar{c}) = 0$, which (since $a\bar{a} + c\bar{c} > 0$) is possible only for $1 - \lambda\bar{\lambda} = 0$, i.e. for $|\lambda| = 1$.

On the other hand, the normalization condition $ad - bc = 1$ is now equivalent to the equation $\lambda(a\bar{a} + c\bar{c}) = 1$ from which it follows in particular that the number λ is real and positive. Therefore $\lambda = 1$ and hence $d = \bar{a}$, $b = -\bar{c}$. This proves that *a linear fractional transformation of the plane* $\mathbb{C}^+$ *that is a rotation of a sphere is given by a formula of the form*

$$w = \frac{az-\bar{c}}{cz+\bar{a}} \quad \textit{where} \quad a\bar{a}+c\bar{c} = 1. \quad \square \tag{31}$$

For transformation (31) the number $a + d = 2 \operatorname{Re} a$ is real. In addition, if $c \neq 0$, then $|a|^2 = a\bar{a} = 1 - c\bar{c} < 1$ and so $|\operatorname{Re} a| \leqslant |a| < 1$ and if $\operatorname{Im} a \neq 0$, then $|\operatorname{Re} a| < |a| \leqslant 1$, so that in both cases $|\operatorname{Re} a| < 1$ and hence $|a + d| < 2$. Since for $c = 0$ and $\operatorname{Im} a = 0$ transformation (31) is obviously an identity transformation this proves that *every nonidentity transformation* (31) *is elliptical.* $\square$

Its multiplier K is therefore of the form $e^{i\theta}$. It is possible to show (do it yourself!) that the angle θ coincides with the angle of the initial rotation of the sphere.

Remark. The presented derivation of formula (31) contains a significant gap. Namely, we have tacitly assumed and taken for granted that transformation $\Phi = \Pi \circ \Sigma \circ \Pi^{-1}$ of the plane $\mathbb{C}^+$ corresponding to rotation Σ is a linear fractional transformation. This fact is sufficiently hard to prove and we restrict ourselves to outlining the main steps in its proof.

It is easy to see that a stereographic projection Π maps every circle lying on a sphere into a circle (or a straight line) in a plane. Since rotation Σ obviously carries circles over into circles it follows that transformation Φ has the circle property. Let us try to prove therefore that any transformation (bijective mapping) Φ of the plane $\mathbb{C}^+$ onto itself having the circle property is linear fractional. (We warn the reader at once that this statement is incorrect and so we shall fail to obtain a proof.)

Let transformation Φ carry the point ∞ over into a point z_0. Consider an arbitrary linear fractional transformation Ω carrying the point z_0 back into the point ∞. Then a transformation $\Psi = \Omega \circ \Phi$ having as before the circle property will leave the point ∞ fixed. It is therefore sufficient to consider the transformation $\Psi\colon \mathbb{C}^+ \to \mathbb{C}^+$ leaving the point ∞ fixed and having the circle property. But since $\Psi(\infty) = \infty$ every such transformation can be assumed to be a transformation of an ordinary plane (unaugmented with the point ∞). On the other hand, straight lines in the plane $\mathbb{C}^+$ are exactly the circles passing through the point ∞ and therefore under the transformation Ψ these straight lines are again carried over into straight lines. In an unaugmented plane we thus have a transformation carrying straight lines over into straight lines. As we know (see Remark 1 in Lecture 27), such a transformation is necessarily an affine transformation. So in an unaugmented plane, Ψ is an affine transformation having the circle property.

It is easy to comprehend that such a transformation is necessarily orthogonal, i.e. is either a motion (case I) or a composition of a motion and a symmetry (case II). In case I everything is proved, since every motion is known to be a linear fractional (even a linear) transformation. However, case II shows us that the statement we are trying to prove is false, for the symmetry $z \mapsto \bar{z}$ is not a linear fractional transformation.

To mend the matters it is necessary to learn how to distinguish between linear fractional transformations and transformations of the symmetry type. It turns out that it is possible to correctly distinguish in the class of transformations having the circle property a subclass of transformations preserving orientation (which at present we can do only for affine transformations), and our restricting ourselves to this subclass will automatically exclude case II.

Thus *any transformation of the plane* $\mathbb{C}^+$ *having the circle property and preserving orientation is a linear fractional transformation.* □

After that, to fill in the gap in the proof of formula (31) it only remains to show that the transformation $\Phi = \Pi \circ \Sigma \circ \Pi^{-1}$ is orientation-preserving, which reduces to an automatic check.

For every rotation of a sphere the points in which the axis of rotation intersects the sphere are the fixed points of rotation. We shall call them the *poles of rotation.*

Since the poles of rotation are diametrically opposite points of the sphere, their images under a stereographic projection are connected by relation (30). Since these images are fixed points of a linear fractional transformation representing a given rotation it follows that linear fractional transformations representing rotations of a sphere are uniquely characterized as elliptical transformations whose fixed points have the form z_1 and $-\bar{z}_1^{-1}$. Therefore these transformations may be expressed by the formula

$$\frac{w-z_1}{w+\bar{z}_1^{-1}}=K\frac{z-z_1}{z+\bar{z}_1^{-1}},$$

where $K = e^{i\theta}$, i.e. by the formula

$$\frac{w-z_1}{w\bar{z}_1+1}=K\frac{z-z_1}{z\bar{z}_1+1}.$$

On solving this equation for w we have the formula

$$(32)\qquad w=\frac{(K+z_1\bar{z}_1)\,z+(1-K)\,z_1}{(1-K)\,\bar{z}_1\cdot z+(1+Kz_1\bar{z}_1)},\quad K=e^{i\theta}.$$

Just as formula (31) this one gives a general form of linear fractional transformations of an augmented plane that are rotations of a sphere. The argument θ of the multiplier K is the angle of the corresponding rotation of the sphere and the number z_1 is the image of one of the poles of this rotation. If we choose the other (diametrically opposite) pole, then the angle θ will change the sign.

The notation (32) of linear fractional transformations representing rotations is not normed. Since

$$(K+z_1\bar{z}_1)(1+Kz_1\bar{z}_1)-(1-K)^2 z_1\bar{z}_1=$$
$$=(1+z_1\bar{z}_1)^2K=(1+z_1\bar{z}_1)^2e^{i\theta},$$

to obtain a normed notation we should multiply all the coefficients of formula (32) by $\lambda = (1 + z_1\bar{z}_1)^{-1}e^{-i\frac{\theta}{2}}$.

Therefore on setting

$$a=\lambda(K+z_1\bar{z}_1),\quad c=\lambda(1-K)\bar{z}_1$$

we obtain from formula (32) formula (31). But formula (32) is more convenient to use in calculations.

Consider, for example, a rotation about the axis of abscissae. Its poles are points $(\pm 1, 0, 0)$ projectible into points ± 1 of an augmented plane. We can therefore apply formula (32) putting $z_1 = 1$. Consequently, a linear fractional transformation corresponding to a rotation of a sphere through an angle θ about the axis of abscissae is given by the formula

$$w=\frac{(1+K)\,z+(1-K)}{(1-K)\,z+(1+K)}\,,\quad K=e^{i\theta}.$$

By multiplying the numerator and denominator by $\frac{1}{2}e^{-i\frac{\theta}{2}}$, setting $\alpha=\frac{\theta}{2}$ and using the Euler formulas

$$\cos\alpha=\frac{e^{i\alpha}+e^{-i\alpha}}{2}\,,\quad \sin\alpha=\frac{e^{i\alpha}-e^{-i\alpha}}{2i}$$

we obtain the normed form

$$w=\frac{\cos\alpha\cdot z-i\sin\alpha}{-i\sin\alpha\cdot z+\cos\alpha}$$

of this transformation. When $\alpha=\pi/4,\ \pi/2,\ 3\pi/4$, i.e. when $\theta=\pi/2,\ \pi,\ 3\pi/2$, we have the transformations

(33) $$w=\frac{z-i}{-iz+1},\quad \frac{1}{z},\quad \frac{z+i}{iz+1}.$$

On taking the ordinate axis as the axis of rotation we must set in formula (32) $z_1=i$ (or $z_1=-i$). Consequently, the corresponding linear fractional transformation can be written as

$$w=\frac{(1+K)\,z+(1-K)\,i}{-(1-K)\,iz+(1+K)},\quad K=e^{i\theta},$$

and hence $\left(\text{after multiplying by } \frac{1}{2}e^{-i\frac{\theta}{2}}\right)$ as

$$w=\frac{\cos\alpha\cdot z+\sin\alpha}{-\sin\alpha\cdot z+\cos\alpha}\,,\quad 0\leqslant\alpha<\pi,$$

where $\alpha=\theta/2$. When $\alpha=\pi/4,\ \pi/2,\ 3\pi/4$, i.e. when $\theta=\pi/2,\ \pi,\ 3\pi/2$, we have the transformations

(34) $$w=\frac{z+1}{z-1},\quad -\frac{1}{z},\quad \frac{z-1}{z+1}.$$

If the axis of rotation is vertical (is the applicate axis), then the fixed points of the corresponding linear fractional transformation are the points 0 and ∞ and hence it is a rotation

$$w=e^{i\theta}z$$

of the augmented plane. When $\theta = \frac{\pi}{2}$, π, $\frac{3\pi}{2}$, we have the transformations

(35) $$w = iz, \quad -z, \quad -iz.$$

Let us now take as the axis of rotation the bisectrix of the first coordinate octant intersecting the sphere in the point $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$. For this rotation $z_1 = \frac{1+i}{\sqrt{3}-1}$. For an arbitrary K formula (32) yields an involved and inaesthetic expression, but for $\theta = 2\pi/3$, i.e. for $K = \frac{-1+i\sqrt{3}}{2}$, we obtain, after cancelling by $\frac{(2+\sqrt{3})\sqrt{3}+i\sqrt{3}}{2}$, the formula $w = \frac{z+1}{-iz+i} = i\frac{z+1}{z-1}$ and for $\theta = 4\pi/3$, i.e. for $K = \frac{-1-i\sqrt{3}}{2}$, we have, after cancelling by $\frac{(2+\sqrt{3})\sqrt{3}-i\sqrt{3}}{2}$, the formula $w = \frac{z+i}{z-i}$. Thus in this case

(36) $$w = i\frac{z+1}{z-1}, \quad \frac{z+i}{z-i}.$$

Similarly if the axis of rotation intersects the sphere in the point $\left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$, then $z_1 = \frac{-1+i}{\sqrt{3}-1}$ and for $K = \frac{-1 \pm i\sqrt{3}}{2}$ we have the transformation

(37) $$w = -\frac{z+i}{z-i}, \quad i\frac{z-1}{z+1},$$

and if the axis of rotation intersects the sphere in the point $\left(\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$ or in the point $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right)$, then for the same K we have the transformations

(38) $$w = \frac{z-i}{z+i}, \quad -i\frac{z+1}{z-1}$$

and

(39) $$w = -\frac{z-i}{z+i}, \quad -i\frac{z-1}{z+1}.$$

But if the axis of rotation is the bisectrix of the first coordinate quadrant in a plane $\xi = 0$ intersecting the sphere in a point $\left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, then $z_1 = \frac{i}{\sqrt{2}-1}$ and for $\theta = \pi$, i.e. for $K = -1$, we have the transformation

(40) $$w = i\,\frac{z+i}{z-i}.$$

Similarly a rotation through the angle π about the bisectrix of the second coordinate quadrant in the plane $\xi = 0$ intersecting the sphere in the point $\left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ may be represented by the transformation

(41) $$w = -i\,\frac{z-i}{z+i}$$

and rotations through the angle π about the bisectrices of the coordinate quadrants in the planes $\eta = 0$ and $\zeta = 0$ can be represented by the transformations

(42) $$w = \frac{z+1}{z-1}, \quad -\frac{z-1}{z+1}, \quad \frac{i}{z}, \quad -\frac{i}{z}.$$

Rotations represented by linear fractional transformations (33) to (42) are characterized by the fact that they carry into itself a cube inscribed in a sphere $\left(\text{with vertices at points } \left(\pm\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}}\right)\right)$ i.e. what are said to be *isometries* of that cube.

Transformations (33), (34) and (35) represent rotations of a cube through angles $\pi/2$, π, and $3\pi/2$ about each of the three axes joining the centres of opposite faces of the cube, transformations (36), (37), (38) and (39) represent rotations of a cube through angles $2\pi/3$ and $4\pi/3$ about each of the four axes joining diametrically opposite vertices of the cube, and finally, transformations (40), (41) and (42) represent rotations of a cube through the angle π about each of the six axes joining the midpoints of diametrically opposite edges of the cube. To these transformations yet another one should be added, the identity transformation leaving all points of a cube fixed. Thus we have 24 =

$= 3 \times 3 + 4 \times 2 + 6 + 1$ isometries of a cube in all. It can easily be proved (do it!) that a cube has no other isometries, i.e. the 24 rotations enumerated exhaust all isometries of a cube.

Transformations (35) plus the identity transformation can be written as a single formula

(43) $$w = i^k z, \quad k = 0,\ 1,\ 2,\ 3.$$

Similarly the second of the transformations (33), the third of (42), the second of (34) and the fourth of (42) can also be written as a single formula

(44) $$w = \frac{i^k}{z}, \quad k = 0,\ 1,\ 2,\ 3.$$

In exactly the same way the formula

(45) $$w = i^k \frac{z+1}{z-1}, \quad k = 0,\ 1,\ 2,\ 3$$

gives the first of the transformations (42), the first of (36), the first of (34) and the second of (38); the formula

(46) $$w = i^k \frac{z-1}{z+1}, \quad k = 0,\ 1,\ 2,\ 3$$

gives the third of the transformations (34), the second of (37), the second of (42) and the second of (39); the formula

(47) $$w = i^k \frac{z+i}{z-i}, \quad k = 0,\ 1,\ 2,\ 3$$

gives the second of the transformations (36), the transformation (40), the first of the transformations (37) and the third of (33); and finally, the formula

(48) $$w = i^k \frac{z-i}{z+i}, \quad k = 0,\ 1,\ 2,\ 3$$

gives the first of the transformations (38), the first of (33), the first of (39) and the transformation (41).

Summing up we obtain the following proposition.

Proposition 4. *There are exactly* 24 *isometries of a cube. In the augmented plane of complex numbers they can be represented by linear fractional transformations* (43) to (48). □

A similar description can be made of the isometries of other regular polyhedra. A tetrahedron has 12 isometries which can be represented by transformations (43), (44), (47) and (48) for $k = 0, 2$ and transformations (45) and (46) for $k = 1, 3$. The isometries of an octahedron coincide with those of a cube and an icosahedron (or, equivalently, a dodecahedron) allows 60 isometries which can be represented by the linear fractional transformations

$$w = \zeta^k z, \quad \frac{-\zeta^k}{z}, \quad \zeta^k \frac{(1+\zeta)\zeta^l \cdot z + 1}{\zeta^{l+1} \cdot z - (1+\zeta)}, \quad -\zeta^k \frac{\zeta^l z - (1+\zeta)}{(1+\zeta)\zeta^{l-1} \cdot z + 1},$$

where $\zeta = e^{\frac{2\pi i}{5}}$, and $k, l = 0, 1, 2, 3, 4$.

SUBJECT INDEX

УДК 516=20

Михаил Михайлович Постников

ЛЕКЦИИ ПО ГЕОМЕТРИИ

Семестр I

АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ

Контрольный редактор С. М. Головина

Редактор Н. М. Привато

Художник Н. В. Зотова

Технический редактор М. А. Страшнова

Корректоры О. Д. Эшлиман, С. И. Каштанова

ИБ № 2966

Сдано в набор 24.02.81.
Подписано к печати 22.10.81.
Формат $84 \times 108^1/_{32}$.
Бумага типографская № 1.
Гарнитура обыкновенная. Печать высокая.
Усл. печ. л. 18,06. Усл. кр.-отт. 18,43.
Уч.-изд. л. 15,8. Изд. № 17/1243.
Тираж 7450 экз. Зак. 01283. Цена 1 р. 61 к.

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2.

Ордена Трудового Красного Знамени
Московская типография № 7 «Искра революции»
Союзполиграфпрома
Государственного Комитета СССР
по делам издательств, полиграфии
и книжной торговли.
Москва 103001, Трехпрудный пер., 9.

1702040000

П $\frac{20203-18}{041(01)-82}$ 81—82, ч. 2

Printed in the Union of Soviet Socialist Republics

M. Postnikov received his doctor's degree (Phys.-Math.) in 1954, from 1955 to 1960 was Professor of the Higher Algebra Department (Moscow State University) and since 1965 he has been Professor of the University's Department of Higher Geometry and Topology.
He is a senior research worker at the Steklov Institute of Mathematics. In 1961 Prof. Postnikov was awarded Lenin Prize for a series of works on the Homotopy Theory of Continuous Maps. He is the author of several monographs and textbooks some of which have been translated into foreign languages.