Chapter 7
Euclidean Spaces

The notions entering into the definition of a vector space do not provide a way of formulating multidimensional analogues of the length of a vector, the angle between vectors, and volumes. Yet such concepts appear in many branches of mathematics and physics, and we shall study such concepts in this chapter. All the vector spaces that we shall consider here will be real (with the exception of certain special cases in which complex vector spaces will be considered as a means of studying real spaces).

7.1 The Definition of a Euclidean Space

Definition 7.1 A *Euclidean space* is a real vector space on which is defined a fixed symmetric bilinear form whose associated quadratic form is positive definite.

The vector space itself will be denoted as a rule by L, and the fixed symmetric bilinear form will be denoted by $(\boldsymbol{x}, \boldsymbol{y})$. Such an expression is also called the *inner product* of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Let us now reformulate the definition of a Euclidean space using this terminology.

A *Euclidean space* is a real vector space L in which to every pair of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ there corresponds a real number $(\boldsymbol{x}, \boldsymbol{y})$ such that the following conditions are satisfied:

(1) $(\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}) = (\boldsymbol{x}_1, \boldsymbol{y}) + (\boldsymbol{x}_2, \boldsymbol{y})$ for all vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y} \in \mathsf{L}$.
(2) $(\alpha \boldsymbol{x}, \boldsymbol{y}) = \alpha(\boldsymbol{x}, \boldsymbol{y})$ for all vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{L}$ and real number α.
(3) $(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y}, \boldsymbol{x})$ for all vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{L}$.
(4) $(\boldsymbol{x}, \boldsymbol{x}) > 0$ for $\boldsymbol{x} \neq \boldsymbol{0}$.

Properties (1)–(3) show that the function $(\boldsymbol{x}, \boldsymbol{y})$ is a symmetric bilinear form on L, and in particular, that $(\boldsymbol{0}, \boldsymbol{y}) = 0$ for every vector $\boldsymbol{y} \in \mathsf{L}$. It is only property (4) that expresses the specific character of a Euclidean space.

The expression $(\boldsymbol{x}, \boldsymbol{x})$ is frequently denoted by $(\boldsymbol{x}^2)$; it is called the *scalar square* of the vector $\boldsymbol{x}$. Thus property (4) implies that the quadratic form corresponding to the bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ is positive definite.

I.R. Shafarevich, A.O. Remizov, *Linear Algebra and Geometry*,
DOI 10.1007/978-3-642-30994-6_7, © Springer-Verlag Berlin Heidelberg 2013

Let us point out some obvious consequences of these definitions. For a fixed vector $\boldsymbol{y} \in \mathsf{L}$, where L is a Euclidean space, conditions (1) and (2) in the definition can be formulated in such a way that the function $\boldsymbol{f}_{\boldsymbol{y}}(\boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{y})$ with argument $\boldsymbol{x}$ is linear. Thus we have a mapping $\boldsymbol{y} \mapsto \boldsymbol{f}_{\boldsymbol{y}}$ of the vector space L to L^*. Condition (4) in the definition of Euclidean space shows that the kernel of this mapping is equal to $(\mathbf{0})$. Indeed, $\boldsymbol{f}_{\boldsymbol{y}} \neq \mathbf{0}$ for every $\boldsymbol{y} \neq \mathbf{0}$, since $\boldsymbol{f}_{\boldsymbol{y}}(\boldsymbol{y}) = (\boldsymbol{y}^2) > 0$. If the dimension of the space L is finite, then by Theorems 3.68 and 3.78, this mapping is an isomorphism. Moreover, we should note that in contrast to the construction used for proving Theorem 3.78, we have now constructed an isomorphism $\mathsf{L} \xrightarrow{\sim} \mathsf{L}^*$ without using the specific choice of a basis in L. Thus we have a certain natural isomorphism $\mathsf{L} \xrightarrow{\sim} \mathsf{L}^*$ defined only by the imposition of an inner product on L. In view of this, in the case of a finite-dimensional Euclidean space L, we shall in what follows sometimes identify L and L^*. In other words, as for any bilinear form, for the inner product $(\boldsymbol{x}, \boldsymbol{y})$ there exists a unique linear transformation $\mathcal{A} : \mathsf{L} \to \mathsf{L}^*$ such that $(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{A}(\boldsymbol{y})(\boldsymbol{x})$. The previous reasoning shows that in the case of a Euclidean space, the transformation $\mathcal{A}$ is an isomorphism, and in particular, the bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ is nonsingular. Let us give some examples of Euclidean spaces.

Example 7.2 The plane, in which for $(\boldsymbol{x}, \boldsymbol{y})$ is taken the well-known inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$ as studied in analytic geometry, that is, the product of the vectors' lengths and the cosine of the angle between them, is a Euclidean space.

Example 7.3 The space $\mathbb{R}^n$ consisting of rows (or columns) of length n, in which the inner product of rows $\boldsymbol{x} = (\alpha_1, \dots, \alpha_n)$ and $\boldsymbol{y} = (\beta_1, \dots, \beta_n)$ is defined by the relation

$$(\boldsymbol{x}, \boldsymbol{y}) = \alpha_1\beta_1 + \alpha_2\beta_2 + \cdots + \alpha_n\beta_n, \tag{7.1}$$

is a Euclidean space.

Example 7.4 The vector space L consisting of polynomials of degree at most n with real coefficients, defined on some interval $[a, b]$, is a Euclidean space. For two polynomials $f(t)$ and $g(t)$, their inner product is defined by the relation

$$(f, g) = \int_a^b f(t)g(t)\,dt. \tag{7.2}$$

Example 7.5 The vector space L consisting of all real-valued continuous functions on the interval $[a, b]$ is a Euclidean space. For two such functions $f(t)$ and $g(t)$, we shall define their inner product by equality (7.2).

Example 7.5 shows that a Euclidean space, like a vector space, does not have to be finite-dimensional.[1] In the sequel, we shall be concerned exclusively with finite-dimensional Euclidean spaces, on which the inner product is sometimes called the

[1] Infinite-dimensional Euclidean spaces are usually called *pre-Hilbert spaces*. An especially important role in a number of branches of mathematics and physics is played by the so-called *Hilbert*

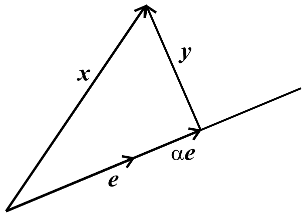


Fig. 7.1 Orthogonal projection

scalar product (because the inner product of two vectors is a scalar) or *dot product* (because the notation $\boldsymbol{x} \cdot \boldsymbol{y}$ is frequently used instead of $(\boldsymbol{x}, \boldsymbol{y})$).

Example 7.6 Every subspace L' of a Euclidean space L is itself a Euclidean space if we define on it the form $(\boldsymbol{x}, \boldsymbol{y})$ exactly as on the space L.

In analogy with Example 7.2, we make the following definition.

Definition 7.7 The *length* of a vector $\boldsymbol{x}$ in a Euclidean space is the nonnegative value $\sqrt{(\boldsymbol{x}^2)}$. The length of a vector $\boldsymbol{x}$ is denoted by $|\boldsymbol{x}|$.

We note that we have here made essential use of property (4), by which the length of a nonnull vector is a positive number.

Following the same analogy, it is natural to define the *angle* φ between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ by the condition

$$\cos\varphi = \frac{(\boldsymbol{x}, \boldsymbol{y})}{|\boldsymbol{x}| \cdot |\boldsymbol{y}|}, \quad 0 \le \varphi \le \pi. \tag{7.3}$$

However, such a number φ exists only if the expression on the right-hand side of equality (7.3) does not exceed 1 in absolute value. Such is indeed the case, and the proof of this fact will be our immediate objective.

Lemma 7.8 *Given a vector $\boldsymbol{e} \neq \boldsymbol{0}$, every vector $\boldsymbol{x} \in \mathsf{L}$ can be expressed in the form*

$$\boldsymbol{x} = \alpha\boldsymbol{e} + \boldsymbol{y}, \quad (\boldsymbol{e}, \boldsymbol{y}) = 0, \tag{7.4}$$

for some scalar α and vector $\boldsymbol{y} \in \mathsf{L}$; see Fig. 7.1.

Proof Setting $\boldsymbol{y} = \boldsymbol{x} - \alpha\boldsymbol{e}$, we obtain α from the condition $(\boldsymbol{e}, \boldsymbol{y}) = 0$. This is equivalent to $(\boldsymbol{x}, \boldsymbol{e}) = \alpha(\boldsymbol{e}, \boldsymbol{e})$, which implies that $\alpha = (\boldsymbol{x}, \boldsymbol{e})/|\boldsymbol{e}|^2$. We remark that $|\boldsymbol{e}| \neq 0$, since by assumption, $\boldsymbol{e} \neq \boldsymbol{0}$. □

---

spaces, which are pre-Hilbert spaces that have the additional property of *completeness*, just for the case of infinite dimension. (Sometimes, in the definition of pre-Hilbert space, the condition $(\boldsymbol{x}, \boldsymbol{x}) > 0$ is replaced by the weaker condition $(\boldsymbol{x}, \boldsymbol{x}) \ge 0$.)

Definition 7.9 The vector αe from relation (7.4) is called the *orthogonal projection* of the vector x onto the line $\langle e \rangle$.

Theorem 7.10 *The length of the orthogonal projection of a vector x is at most its length $|x|$.*

Proof Indeed, since by definition, $x = \alpha e + y$ and $(e, y) = 0$, it follows that

$$|x|^2 = \big(x^2\big) = (\alpha e + y, \alpha e + y) = |\alpha e|^2 + |y|^2 \geq |\alpha e|^2,$$

and this implies that

$$|x| \geq |\alpha e|. \tag{7.5}$$

□

This leads directly to the following necessary theorem.

Theorem 7.11 *For arbitrary vectors x and y in a Euclidean space, the following inequality holds*:

$$\big|(x, y)\big| \leq |x| \cdot |y|. \tag{7.6}$$

Proof If one of the vectors x, y is equal to zero, then the inequality (7.6) is obvious, and is reduced to the equality $0 = 0$. Now suppose that neither vector is the null vector. In this case, let us denote by αy the orthogonal projection of the vector x onto the line $\langle y \rangle$. Then by (7.4), we have the relationship $x = \alpha y + z$, where $(y, z) = 0$. From this we obtain the equality

$$(x, y) = (\alpha y + z, y) = (\alpha y, y) = \alpha |y|^2.$$

This means that $|(x, y)| = |\alpha| \cdot |y|^2 = |\alpha y| \cdot |y|$. But by Theorem 7.10, we have the inequality $|\alpha y| \leq |x|$, and consequently, $|(x, y)| \leq |x| \cdot |y|$. □

Inequality (7.6) goes by a number of names, but it is generally known as the Cauchy–Schwarz inequality. From it we can derive the well-known *triangle inequality* from elementary geometry. Indeed, suppose that the vectors $x = \overrightarrow{AB}$, $y = \overrightarrow{BC}$, $z = \overrightarrow{CA}$ correspond to the sides of a triangle ABC. Then we have the relationship $x + y + z = \mathbf{0}$, from which with the help of (7.6) we obtain the inequality

$$\begin{aligned} |z|^2 = (x + y, x + y) = |x|^2 + 2(x, y) + |y|^2 &\leq |x|^2 + 2\big|(x, y)\big| + |y|^2 \\ &\leq |x|^2 + 2|x| \cdot |y| + |y|^2 = \big(|x| + |y|\big)^2, \end{aligned}$$

from which clearly follows the triangle inequality $|z| \leq |x| + |y|$.

Thus from Theorem 7.11 it follows that there exists a number φ that satisfies the equality (7.3). This number is what is called the *angle* between the vectors x and y. Condition (7.3) determines the angle uniquely if we assume that $0 \leq \varphi \leq \pi$.

Definition 7.12 Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be *orthogonal* if their inner product is equal to zero: $(\boldsymbol{x}, \boldsymbol{y}) = 0$.

Let us note that this repeats the definition given in Sect. 6.2 for a bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{y})$. By the definition given above in (7.3), the angle between orthogonal vectors is equal to $\frac{\pi}{2}$.

For a Euclidean space, there is a useful criterion for the linear independence of vectors. Let $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ be m vectors in the Euclidean space L.

Definition 7.13 The *Gram determinant*, or *Gramian*, of a system of vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ is the determinant

$$G(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m) = \begin{vmatrix} (\boldsymbol{a}_1, \boldsymbol{a}_1) & (\boldsymbol{a}_1, \boldsymbol{a}_2) & \cdots & (\boldsymbol{a}_1, \boldsymbol{a}_m) \\ (\boldsymbol{a}_2, \boldsymbol{a}_1) & (\boldsymbol{a}_2, \boldsymbol{a}_2) & \cdots & (\boldsymbol{a}_2, \boldsymbol{a}_m) \\ \vdots & \vdots & \ddots & \vdots \\ (\boldsymbol{a}_m, \boldsymbol{a}_1) & (\boldsymbol{a}_m, \boldsymbol{a}_2) & \cdots & (\boldsymbol{a}_m, \boldsymbol{a}_m) \end{vmatrix}. \tag{7.7}$$

Theorem 7.14 *If the vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ are linearly dependent, then the Gram determinant $G(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m)$ is equal to zero, while if they are linearly independent, then $G(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m) > 0$.*

Proof If the vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ are linearly dependent, then as was shown in Sect. 3.2, one of the vectors can be expressed as a linear combination of the others. Let it be the vector $\boldsymbol{a}_m$, that is, $\boldsymbol{a}_m = \alpha_1 \boldsymbol{a}_1 + \cdots + \alpha_{m-1} \boldsymbol{a}_{m-1}$. Then from the properties of the inner product, it follows that for every $i = 1, \dots, m$, we have the equality

$$(\boldsymbol{a}_m, \boldsymbol{a}_i) = \alpha_1 (\boldsymbol{a}_1, \boldsymbol{a}_i) + \alpha_2 (\boldsymbol{a}_2, \boldsymbol{a}_i) + \cdots + \alpha_{m-1} (\boldsymbol{a}_{m-1}, \boldsymbol{a}_i).$$

From this it is clear that if we subtract from the last row of the determinant (7.7), all the previous rows multiplied by coefficients $\alpha_1, \dots, \alpha_{m-1}$, then we obtain a determinant with a row consisting entirely of zeros. Therefore, $G(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m) = 0$.

Now suppose that vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ are linearly independent. Let us consider in the subspace $\mathsf{L}' = \langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_m \rangle$, the quadratic form $(\boldsymbol{x}^2)$. Setting $\boldsymbol{x} = \alpha_1 \boldsymbol{a}_1 + \cdots + \alpha_m \boldsymbol{a}_m$, we may write it in the form

$$\big((\alpha_1 \boldsymbol{a}_1 + \cdots + \alpha_m \boldsymbol{a}_m)^2\big) = \sum_{i,j=1}^{m} \alpha_i \alpha_j (\boldsymbol{a}_i, \boldsymbol{a}_j).$$

It is easily seen that this quadratic form is positive definite, and its determinant coincides with the Gram determinant $G(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m)$. By Theorem 6.19, it now follows that $G(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m) > 0$. □

Theorem 7.14 is a broad generalization of the Cauchy–Schwarz inequality. Indeed, for $m = 2$, inequality (7.6) is obvious (it becomes an equality) if vectors $\boldsymbol{x}$

and $\boldsymbol{y}$ are linearly dependent. However, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly independent, then their Gram determinant is equal to

$$G(\boldsymbol{x},\boldsymbol{y}) = \begin{vmatrix} (\boldsymbol{x},\boldsymbol{x}) & (\boldsymbol{x},\boldsymbol{y}) \\ (\boldsymbol{x},\boldsymbol{y}) & (\boldsymbol{y},\boldsymbol{y}) \end{vmatrix}.$$

The inequality $G(\boldsymbol{x},\boldsymbol{y}) > 0$ established in Theorem 7.14 gives us (7.6). In particular, we see that inequality (7.6) becomes an equality *only* if the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are proportional. We remark that this is easy to derive if we examine the proof of Theorem 7.11.

Definition 7.15 Vectors $\boldsymbol{e}_1,\dots,\boldsymbol{e}_m$ in a Euclidean space form an *orthonormal* system if

$$(\boldsymbol{e}_i,\boldsymbol{e}_j) = 0 \quad \text{for } i \neq j, \qquad (\boldsymbol{e}_i,\boldsymbol{e}_i) = 1, \tag{7.8}$$

that is, if these vectors are mutually orthogonal and the length of each of them is equal to 1. If $m = n$ and the vectors $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n$ form a basis of the space, then such a basis is called an *orthonormal basis*.

It is obvious that the Gram determinant of an orthonormal basis is equal to 1.

We shall now use the fact that a quadratic form $(\boldsymbol{x}^2)$ is positive definite and apply to it formula (6.28), in which by the definition of positive definiteness, $s = n$. This result can now be reformulated as an assertion about the existence of a basis $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n$ of the space L in which the scalar square of a vector $\boldsymbol{x} = \alpha_1\boldsymbol{e}_1 + \cdots + \alpha_n\boldsymbol{e}_n$ is equal to the sum of the squares of its coordinates, that is, $(\boldsymbol{x}^2) = \alpha_1^2 + \cdots + \alpha_n^2$. In other words, we have the following result.

Theorem 7.16 *Every Euclidean space has an orthonormal basis.*

Remark 7.17 In an orthonormal basis, the inner product of $\boldsymbol{x} = (\alpha_1,\dots,\alpha_n)$ and $\boldsymbol{y} = (\beta_1,\dots,\beta_n)$ has a particularly simple form, given by formula (7.1). Accordingly, in an orthonormal basis, the scalar square of an arbitrary vector is equal to the sum of the squares of its coordinates, while its length is equal to the square root of the sum of the squares.

The lemma establishing the decomposition (7.4) has an important and far-reaching generalization. To formulate it, we recall that in Sect. 3.7, for every subspace $\mathsf{L}' \subset \mathsf{L}$ we defined its annihilator $(\mathsf{L}')^a \subset \mathsf{L}^*$, while earlier in this section, we showed that an arbitrary Euclidean space L of finite dimension can be identified with its dual space L^*. As a result, we can view $(\mathsf{L}')^a$ as a subspace of the original space L. In this light, we shall call it the *orthogonal complement* of the subspace L' and denote it by $(\mathsf{L}')^\perp$. If we recall the relevant definitions, we obtain that the orthogonal complement $(\mathsf{L}')^\perp$ of the subspace $\mathsf{L}' \subset \mathsf{L}$ consists of all vectors $\boldsymbol{y} \in \mathsf{L}$ for which the following condition holds:

$$(\boldsymbol{x},\boldsymbol{y}) = 0 \quad \text{for all } \boldsymbol{x} \in \mathsf{L}'. \tag{7.9}$$

On the other hand, $(\mathsf{L}')^\perp$ is the subspace $(\mathsf{L}')^\perp_\varphi$, defined for the case that the bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y})$ is given by $\varphi(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{y})$; see p. 198.

A basic property of the orthogonal complement in a finite-dimensional Euclidean space is contained in the following theorem.

Theorem 7.18 *For an arbitrary subspace* L_1 *of a finite-dimensional Euclidean space* L, *the following holds*:

$$\mathsf{L} = \mathsf{L}_1 \oplus \mathsf{L}_1^\perp. \tag{7.10}$$

In the case $\mathsf{L}_1 = \langle \boldsymbol{e} \rangle$, Theorem 7.18 follows from Lemma 7.8.

Proof of Theorem 7.18 In the previous chapter, we saw that every quadratic form $\psi(\boldsymbol{x})$ in some basis of a vector space L can be reduced to the canonical form (6.22), and in the case of a real vector space, to the form (6.28) for some scalars $0 \le s \le r$, where s is the index of inertia and r is the rank of the quadratic form $\psi(\boldsymbol{x})$, or equivalently, the rank of the symmetric bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y})$ associated with $\psi(\boldsymbol{x})$ by the relationship (6.11). We recall that a bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y})$ is nonsingular if $r = n$, where $n = \dim \mathsf{L}$.

The condition of positive definiteness for the form $\psi(\boldsymbol{x})$ is equivalent to the condition that all scalars $\lambda_1, \dots, \lambda_n$ in (6.22) be positive, or equivalently, that the equality $s = r = n$ hold in formula (6.28). From this it follows that a symmetric bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y})$ associated with a positive definite quadratic form $\psi(\boldsymbol{x})$ is nonsingular on the space L as well as on every subspace $\mathsf{L}' \subset \mathsf{L}$. To complete the proof, it suffices to recall that by definition, the quadratic form $(\boldsymbol{x}^2)$ associated with the inner product $(\boldsymbol{x}, \boldsymbol{y})$ is positive definite and to use Theorem 6.9 for the bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{y})$. □

From relationship (3.54) for the annihilator (see Sect. 3.7) or from Theorem 7.18, it follows that

$$\dim(\mathsf{L}_1)^\perp = \dim \mathsf{L} - \dim \mathsf{L}_1.$$

The map that is the projection of the space L onto the subspace L_1 parallel to $\mathsf{L}_1^\perp$ (see the definition on p. 103) is called the *orthogonal projection* of L onto L_1. Then the projection of the vector $\boldsymbol{x} \in \mathsf{L}$ onto the subspace L_1 is called its *orthogonal projection* onto L_1. This is a natural generalization of the notion introduced above of orthogonal projection of a vector onto a line. Similarly, for an arbitrary subset $X \subset \mathsf{L}$, we can define its orthogonal projection onto L_1.

The Gram determinant is connected to the notion of *volume* in a Euclidean space, generalizing the notion of the length of a vector.

Definition 7.19 The *parallelepiped spanned by vectors* $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ is the collection of all vectors $\alpha_1 \boldsymbol{a}_1 + \cdots + \alpha_m \boldsymbol{a}_m$ for all $0 \le \alpha_i \le 1$. It is denoted by $\varPi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m)$. A *base* of the parallelepiped $\varPi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m)$ is a parallelepiped spanned by any $m-1$ vectors among $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$, for example, $\varPi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_{m-1})$.

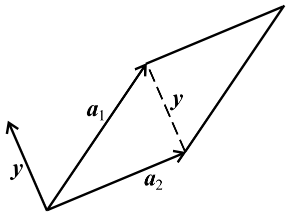


Fig. 7.2 Altitude of a parallelepiped

In the case of the plane (see Example 7.2), we have parallelepipeds $\Pi(\boldsymbol{a}_1)$ and $\Pi(\boldsymbol{a}_1, \boldsymbol{a}_2)$. By definition, $\Pi(\boldsymbol{a}_1)$ is the segment whose beginning and end coincide with the beginning and end of the vector $\boldsymbol{a}_1$, while $\Pi(\boldsymbol{a}_1, \boldsymbol{a}_2)$ is the parallelogram constructed from the vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$.

We return now to the consideration of an arbitrary parallelepiped

$$\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m),$$

and we define the subspace $\mathsf{L}_1 = \langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_{m-1} \rangle$. To this case we may apply the notion introduced above of orthogonal projection of the space L. By the decomposition (7.10), the vector $\boldsymbol{a}_m$ can be uniquely represented in the form $\boldsymbol{a}_m = \boldsymbol{x} + \boldsymbol{y}$, where $\boldsymbol{x} \in \mathsf{L}_1$ and $\boldsymbol{y} \in \mathsf{L}_1^{\perp}$. The vector $\boldsymbol{y}$ is called the *altitude* of the parallelepiped $\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m)$ *dropped* to the base $\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_{m-1})$. The construction we have described is depicted in Fig. 7.2 for the case of the plane.

Now we can introduce the concept of *volume* of a parallelepiped

$$\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m),$$

or more precisely, its *unoriented volume*. This is by definition a nonnegative number, denoted by $V(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m)$ and defined by induction on m. In the case $m = 1$, it is equal to $V(\boldsymbol{a}_1) = |\boldsymbol{a}_1|$, and in the general case, $V(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m)$ is the product of $V(\boldsymbol{a}_1, \dots, \boldsymbol{a}_{m-1})$ and the length of the altitude of the parallelepiped $\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m)$ dropped to the base $\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_{m-1})$.

The following is a numerical expression for the unoriented volume:

$$V^2(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m) = G(\boldsymbol{a}_1, \dots, \boldsymbol{a}_m). \tag{7.11}$$

This relationship shows the geometric meaning of the Gram determinant.

Formula (7.11) is obvious for $m = 1$, and in the general case, it is proved by induction on m. According to (7.10), we may represent the vector $\boldsymbol{a}_m$ in the form $\boldsymbol{a}_m = \boldsymbol{x} + \boldsymbol{y}$, where $\boldsymbol{x} \in \mathsf{L}_1 = \langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_{m-1} \rangle$ and $\boldsymbol{y} \in \mathsf{L}_1^{\perp}$. Then $\boldsymbol{a}_m = \alpha_1 \boldsymbol{a}_1 + \cdots + \alpha_{m-1} \boldsymbol{a}_{m-1} + \boldsymbol{y}$. We note that $\boldsymbol{y}$ is the altitude of our parallelepiped dropped to the base $\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_{m-1})$. Let us recall formula (7.7) for the Gram determinant and subtract from its last column, each of the other columns multiplied by $\alpha_1, \dots, \alpha_{m-1}$.

As a result, we obtain

$$G(\boldsymbol{a}_1,\dots,\boldsymbol{a}_m)=\begin{vmatrix}(\boldsymbol{a}_1,\boldsymbol{a}_1) & (\boldsymbol{a}_1,\boldsymbol{a}_2) & \cdots & 0\\ (\boldsymbol{a}_2,\boldsymbol{a}_1) & (\boldsymbol{a}_2,\boldsymbol{a}_2) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ (\boldsymbol{a}_{m-1},\boldsymbol{a}_1) & (\boldsymbol{a}_{m-1},\boldsymbol{a}_2) & \cdots & 0\\ (\boldsymbol{a}_m,\boldsymbol{a}_1) & (\boldsymbol{a}_m,\boldsymbol{a}_2) & \cdots & (\boldsymbol{y},\boldsymbol{a}_m)\end{vmatrix}, \tag{7.12}$$

and moreover, $(\boldsymbol{y},\boldsymbol{a}_m)=(\boldsymbol{y},\boldsymbol{y})=|\boldsymbol{y}|^2$, since $\boldsymbol{y}\in\mathsf{L}_1^{\perp}$.

Expanding the determinant (7.12) along its last column, we obtain the equality

$$G(\boldsymbol{a}_1,\dots,\boldsymbol{a}_m)=G(\boldsymbol{a}_1,\dots,\boldsymbol{a}_{m-1})|\boldsymbol{y}|^2.$$

Let us recall that by construction, $\boldsymbol{y}$ is the altitude of the parallelepiped $\Pi(\boldsymbol{a}_1,\dots,\boldsymbol{a}_m)$ dropped to the base $\Pi(\boldsymbol{a}_1,\dots,\boldsymbol{a}_{m-1})$. By the induction hypothesis, we have $G(\boldsymbol{a}_1,\dots,\boldsymbol{a}_{m-1})=V^2(\boldsymbol{a}_1,\dots,\boldsymbol{a}_{m-1})$, and this implies

$$G(\boldsymbol{a}_1,\dots,\boldsymbol{a}_m)=V^2(\boldsymbol{a}_1,\dots,\boldsymbol{a}_{m-1})|\boldsymbol{y}|^2=V^2(\boldsymbol{a}_1,\dots,\boldsymbol{a}_m).$$

Thus the concept of unoriented volume that we have introduced differs from the volume and area about which we spoke in Sects. 2.1 and 2.6, since the unoriented volume cannot assume negative values. This explains the term "unoriented." We shall now formulate a second way of looking at the volume of a parallelepiped, one that generalizes the notions of volume and area about which we spoke earlier and differs from unoriented volume by the sign ±1. By Theorem 7.14, of interest is only the case in which the vectors $\boldsymbol{a}_1,\dots,\boldsymbol{a}_m$ are linearly independent. Then we may consider the space $\mathsf{L}=\langle\boldsymbol{a}_1,\dots,\boldsymbol{a}_m\rangle$ with basis $\boldsymbol{a}_1,\dots,\boldsymbol{a}_m$.

Thus we are given n vectors $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$, where $n=\dim\mathsf{L}$. We consider the matrix A, whose jth column consists of the coordinates of the vector $\boldsymbol{a}_j$ relative to some orthonormal basis $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n$:

$$A=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix}.$$

An easy verification shows that in the matrix A^*A, the intersection of the ith row and jth column contains the element $(\boldsymbol{a}_i,\boldsymbol{a}_j)$. This implies that the determinant of the matrix A^*A is equal to $G(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)$, and in view of the equalities $|A^*A|=|A^*|\cdot|A|=|A|^2$, we obtain $|A|^2=G(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)$. On the other hand, from formula (7.11), it follows that $G(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)=V^2(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)$, and this implies that

$$|A|=\pm V(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n).$$

The determinant of the matrix A is called the *oriented volume* of the n-dimensional parallelepiped $\Pi(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)$. It is denoted by $v(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)$. Thus the oriented and unoriented volumes are related by the equality

$$V(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n) = \bigl|v(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)\bigr|.$$

Since the determinant of a matrix does not change under the transpose operation, it follows that $v(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n) = |A^*|$. In other words, for computing the oriented volume, one may write the coordinates of the generators of the parallelepiped $\boldsymbol{a}_i$ not in the columns of the matrix, but in the rows, which is sometimes more convenient.

It is obvious that the sign of the oriented volume *depends* on the choice of orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$. This dependence is suggested by the term "oriented." We shall have more to say about this in Sect. 7.3.

The volume possesses some important properties.

Theorem 7.20 *Let $\mathcal{C} : \mathsf{L} \to \mathsf{L}$ be a linear transformation of the Euclidean space L of dimension n. Then for any n vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ in this space, one has the relationship*

$$v\bigl(\mathcal{C}(\boldsymbol{a}_1), \dots, \mathcal{C}(\boldsymbol{a}_n)\bigr) = |\mathcal{C}| v(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n). \tag{7.13}$$

Proof We shall choose an orthonormal basis of the space L. Suppose that the transformation $\mathcal{C}$ has matrix C in this basis and that the coordinates $\alpha_1, \dots, \alpha_n$ of an arbitrary vector $\boldsymbol{a}$ are related to the coordinates $\beta_1, \dots, \beta_n$ of its image $\mathcal{C}(\boldsymbol{a})$ by the relationship (3.25), or in matrix notation, (3.27). Let A be the matrix whose columns consist of the coordinates of the vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$, and let A' be the matrix whose columns consist of the coordinates of the vectors $\mathcal{C}(\boldsymbol{a}_1), \dots, \mathcal{C}(\boldsymbol{a}_n)$. Then it is obvious that we have the relationship $A' = CA$, from which it follows that $|A'| = |C| \cdot |A|$.

To complete the proof, it remains to note that $|\mathcal{C}| = |C|$, and by the definition of oriented volume, we have the equalities $v(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n) = |A|$ and $v(\mathcal{C}(\boldsymbol{a}_1), \dots, \mathcal{C}(\boldsymbol{a}_n)) = |A'|$. □

It follows from this theorem, of course, that

$$V\bigl(\mathcal{C}(\boldsymbol{a}_1), \dots, \mathcal{C}(\boldsymbol{a}_n)\bigr) = \bigl||A|\bigr| V(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n), \tag{7.14}$$

where $||A||$ denotes the absolute value of the determinant of the matrix A.

Using the concepts introduced thus far, we may define an analogue of the volume $V(M)$ for a very broad class of sets M containing all the sets actually encountered in mathematics and physics. This is the subject of what is called *measure theory*, but since it is a topic that is rather far removed from linear algebra, it will not concern us here. Let us note only that the important relationship (7.14) remains valid here:

$$V\bigl(\mathcal{C}(M)\bigr) = \bigl||A|\bigr| V(M). \tag{7.15}$$

An interesting example of a set in an n-dimensional Euclidean space is the *ball* $B(r)$ of radius r, namely the set of all vectors $\boldsymbol{x} \in \mathsf{L}$ such that $|\boldsymbol{x}| \le r$. The set of vectors $\boldsymbol{x} \in \mathsf{L}$ for which $|\boldsymbol{x}| = r$ is called the *sphere* $S(r)$ of radius r. From the relationship (7.15) it follows that $V(B(r)) = V_n r^n$, where $V_n = V(B(1))$. The calculation of the interesting geometric constant V_n is a question from analysis, related to the theory of the *gamma function* Γ. Here we shall simply quote the result:

$$V_n = \frac{\pi^{n/2}}{\Gamma(n/2+1)}.$$

It follows from the theory of the gamma function that if n is an even number ($n = 2m$), then $V_n = \pi^m/m!$, and if n is odd ($n = 2m+1$), then $V_n = 2^{m+1}\pi^m/(1 \cdot 3 \cdots (2m+1))$.

7.2 Orthogonal Transformations

Let L_1 and L_2 be Euclidean spaces of the same dimension with inner products $(\boldsymbol{x}, \boldsymbol{y})_1$ and $(\boldsymbol{x}, \boldsymbol{y})_2$ defined on them. We shall denote the length of a vector $\boldsymbol{x}$ in the spaces L_1 and L_2 by $|\boldsymbol{x}|_1$ and $|\boldsymbol{x}|_2$, respectively.

Definition 7.21 An *isomorphism* of Euclidean spaces L_1 and L_2 is an isomorphism $\mathcal{A} : \mathsf{L}_1 \to \mathsf{L}_2$ of the underlying vector spaces that preserves the inner product, that is, for arbitrary vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{L}_1$, the following relationship holds:

$$(\boldsymbol{x}, \boldsymbol{y})_1 = \big(\mathcal{A}(\boldsymbol{x}), \mathcal{A}(\boldsymbol{y})\big)_2. \tag{7.16}$$

If we substitute the vector $\boldsymbol{y} = \boldsymbol{x}$ into equality (7.16), we obtain that $|\boldsymbol{x}|_1^2 = |\mathcal{A}(\boldsymbol{x})|_2^2$, and this implies that $|\boldsymbol{x}|_1 = |\mathcal{A}(\boldsymbol{x})|_2$, that is, the isomorphism $\mathcal{A}$ preserves the lengths of vectors.

Conversely, if $\mathcal{A} : \mathsf{L}_1 \to \mathsf{L}_2$ is an isomorphism of vector spaces that preserves the lengths of vectors, then $|\mathcal{A}(\boldsymbol{x}+\boldsymbol{y})|_2^2 = |\boldsymbol{x}+\boldsymbol{y}|_1^2$, and therefore,

$$\big|\mathcal{A}(\boldsymbol{x})\big|_2^2 + 2\big(\mathcal{A}(\boldsymbol{x}), \mathcal{A}(\boldsymbol{y})\big)_2 + \big|\mathcal{A}(\boldsymbol{y})\big|_2^2 = |\boldsymbol{x}|_1^2 + 2(\boldsymbol{x}, \boldsymbol{y})_1 + |\boldsymbol{y}|_1^2.$$

But by assumption, we also have the equalities $|\mathcal{A}(\boldsymbol{x})|_2 = |\boldsymbol{x}|_1$ and $|\mathcal{A}(\boldsymbol{y})|_2 = |\boldsymbol{y}|_1$, which implies that $(\boldsymbol{x}, \boldsymbol{y})_1 = (\mathcal{A}(\boldsymbol{x}), \mathcal{A}(\boldsymbol{y}))_2$. This, strictly speaking, is a consequence of the fact (Theorem 6.6) that a symmetric bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ is determined by the quadratic form $(\boldsymbol{x}, \boldsymbol{x})$, and here we have simply repeated the proof given in Sect. 4.1.

If the spaces L_1 and L_2 have the same dimension, then from the fact that the linear transformation $\mathcal{A} : \mathsf{L}_1 \to \mathsf{L}_2$ preserves the lengths of vectors, it already follows that it is an isomorphism. Indeed, as we saw in Sect. 3.5, it suffices to verify that the kernel of the transformation $\mathcal{A}$ is equal to $(\boldsymbol{0})$. But if $\mathcal{A}(\boldsymbol{x}) = \boldsymbol{0}$, then $|\mathcal{A}(\boldsymbol{x})|_2 = 0$, which implies that $|\boldsymbol{x}|_1 = 0$, that is, $\boldsymbol{x} = \boldsymbol{0}$.

Theorem 7.22 *All Euclidean spaces of a given finite dimension are isomorphic to each other.*

Proof From the existence of an orthonormal basis, it follows at once that every n-dimensional Euclidean space is isomorphic to the Euclidean space in Example 7.3. Indeed, let $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ be an orthonormal basis of a Euclidean space L. Assigning to each vector $\boldsymbol{x} \in \mathsf{L}$ the row of its coordinates in the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, we obtain an isomorphism of the space L and the space $\mathbb{R}^n$ of rows of length n with inner product (7.1) (see the remarks on p. 218). It is easily seen that isomorphism is an equivalence relation (p. xii) on the set of Euclidean spaces, and by transitivity, it follows that all Euclidean spaces of dimension n are isomorphic to each other. □

Theorem 7.22 is analogous to Theorem 3.64 for vector spaces, and its general meaning is the same (this is elucidated in detail in Sect. 3.5). For example, using Theorem 7.22, we could have proved the inequality (7.6) differently from how it was done in the preceding section. Indeed, it is completely obvious (the inequality is reduced to an equality) if the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly dependent. If, on the other hand, they are linearly independent, then we can consider the subspace $\mathsf{L}' = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. By Theorem 7.22, it is isomorphic to the plane (Example 7.2 in the previous section), where this inequality is well known. Therefore, it must also be correct for arbitrary vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

Definition 7.23 A linear transformation $\mathcal{U}$ of a Euclidean space L into itself that preserves the inner product, that is, satisfies the condition that for all vectors $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$(\boldsymbol{x}, \boldsymbol{y}) = \big(\mathcal{U}(\boldsymbol{x}), \mathcal{U}(\boldsymbol{y})\big), \tag{7.17}$$

is said to be *orthogonal*.

This is clearly a special case of an isomorphism of Euclidean spaces L_1 and L_2 that coincide.

It is also easily seen that an orthogonal transformation $\mathcal{U}$ takes an orthonormal basis to another orthonormal basis, since from the conditions (7.8) and (7.17), it follows that $\mathcal{U}(\boldsymbol{e}_1), \dots, \mathcal{U}(\boldsymbol{e}_n)$ is an orthonormal basis if $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ is. Conversely, if a linear transformation $\mathcal{U}$ takes *some* orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ to another orthonormal basis, then for vectors $\boldsymbol{x} = \alpha_1 \boldsymbol{e}_1 + \cdots + \alpha_n \boldsymbol{e}_n$ and $\boldsymbol{y} = \beta_1 \boldsymbol{e}_1 + \cdots + \beta_n \boldsymbol{e}_n$, we have

$$\mathcal{U}(\boldsymbol{x}) = \alpha_1 \mathcal{U}(\boldsymbol{e}_1) + \cdots + \alpha_n \mathcal{U}(\boldsymbol{e}_n), \qquad \mathcal{U}(\boldsymbol{y}) = \beta_1 \mathcal{U}(\boldsymbol{e}_1) + \cdots + \beta_n \mathcal{U}(\boldsymbol{e}_n).$$

Since both $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and $\mathcal{U}(\boldsymbol{e}_1), \dots, \mathcal{U}(\boldsymbol{e}_n)$ are orthonormal bases, it follows by (7.1) that both the left- and right-hand sides of relationship (7.17) are equal to the expression $\alpha_1\beta_1 + \cdots + \alpha_n\beta_n$, that is, relationship (7.17) is satisfied, and this implies that $\mathcal{U}$ is an orthogonal transformation.

We note the following important reformulation of this fact: for any two orthonormal bases of a Euclidean space, there exists a unique orthogonal transformation that takes the first basis into the second.

Let $U = (u_{ij})$ be the matrix of a linear transformation $\mathcal{U}$ in some orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$. It follows from what has gone before that the transformation $\mathcal{U}$ is orthogonal if and only if the vectors $\mathcal{U}(\boldsymbol{e}_1), \dots, \mathcal{U}(\boldsymbol{e}_n)$ form an orthonormal basis. But by the definition of the matrix U, the vector $\mathcal{U}(\boldsymbol{e}_i)$ is equal to $\sum_{k=1}^{n} u_{ki}\boldsymbol{e}_k$, and since $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ is an orthonormal basis, we have

$$\big(\mathcal{U}(\boldsymbol{e}_i), \mathcal{U}(\boldsymbol{e}_j)\big) = u_{1i}u_{1j} + u_{2i}u_{2j} + \cdots + u_{ni}u_{nj}.$$

The expression on the right-hand side is equal to the element c_{ij}, where the matrix (c_{ij}) is equal to U^*U. This implies that the condition of orthogonality of the transformation $\mathcal{U}$ can be written in the form

$$U^*U = E, \tag{7.18}$$

or equivalently, $U^* = U^{-1}$. This equality is equivalent to

$$UU^* = E, \tag{7.19}$$

and can be expressed as relationships among the elements of the matrix U:

$$u_{i1}u_{j1} + \cdots + u_{in}u_{jn} = 0 \quad \text{for } i \neq j, \qquad u_{i1}^2 + \cdots + u_{in}^2 = 1. \tag{7.20}$$

The matrix U satisfying the relationship (7.18) or the equivalent relationship (7.19) is said to be *orthogonal*.

The concept of an orthonormal basis of a Euclidean space can be interpreted more graphically using the notion of flag (see the definition on p. 101). Namely, we associate with an orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ the flag

$$(\mathbf{0}) \subset \mathsf{L}_1 \subset \mathsf{L}_2 \subset \cdots \subset \mathsf{L}_n = \mathsf{L}, \tag{7.21}$$

in which the subspace L_i is equal to $\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_i \rangle$, and the pair $(\mathsf{L}_{i-1}, \mathsf{L}_i)$ is directed in the sense that L_i^+ is the half-space of L_i containing the vector $\boldsymbol{e}_i$. In the case of a Euclidean space, the essential fact is that we obtain a *bijection* between orthonormal bases and flags.

For the proof of this, we have only to verify that the orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ is uniquely determined by its associated flag. Let this basis be associated with the flag (7.21). If we have already constructed an orthonormal system of vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{i-1}$ such that $\mathsf{L}_{i-1} = \langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_{i-1} \rangle$, then we should consider the orthogonal complement $\mathsf{L}_{i-1}^\perp$ of the subspace L_{i-1} in L_i. Then $\dim \mathsf{L}_{i-1}^\perp = 1$ and $\mathsf{L}_{i-1}^\perp = \langle \boldsymbol{e}_i \rangle$, where the vector $\boldsymbol{e}_i$ is uniquely defined up to the factor ± 1. This factor can be selected unambiguously based on the condition $\boldsymbol{e}_i \in \mathsf{L}_i^+$.

An observation made earlier can now be interpreted as follows: For any two flags Φ_1 and Φ_2 of a Euclidean space L, there exists a unique orthogonal transformation that maps Φ_1 to Φ_2.

Our next goal will be the construction of an orthonormal basis in which a given orthogonal transformation $\mathcal{U}$ has the simplest matrix possible. By Theorem 4.22, the transformation $\mathcal{U}$ has a one- or two-dimensional invariant subspace L'. It is clear that the restriction of $\mathcal{U}$ to the subspace L' is again an orthogonal transformation.

Let us determine first the sort of transformation that this can be, that is, what sorts of orthogonal transformations of one- and two-dimensional spaces exist.

If $\dim \mathsf{L}' = 1$, then $\mathsf{L}' = \langle \boldsymbol{e} \rangle$ for some nonnull vector $\boldsymbol{e}$. Then $\mathcal{U}(\boldsymbol{e}) = \alpha \boldsymbol{e}$, where α is some scalar. From the orthogonality of the transformation $\mathcal{U}$, we obtain that

$$(\boldsymbol{e}, \boldsymbol{e}) = (\alpha \boldsymbol{e}, \alpha \boldsymbol{e}) = \alpha^2 (\boldsymbol{e}, \boldsymbol{e}),$$

from which it follows that $\alpha^2 = 1$, and this implies that $\alpha = \pm 1$. Consequently, in a one-dimensional space L', there exist two orthogonal transformations: the identity $\mathcal{E}$, for which $\mathcal{E}(\boldsymbol{x}) = \boldsymbol{x}$ for all vectors $\boldsymbol{x}$, and the transformation $\mathcal{U}$ such that $\mathcal{U}(\boldsymbol{x}) = -\boldsymbol{x}$. It is obvious that $\mathcal{U} = -\mathcal{E}$.

Now let $\dim \mathsf{L}' = 2$, in which case L' is isomorphic to the plane with inner product (7.1). It is well known from analytic geometry that an orthogonal transformation of the plane is either a rotation through some angle φ about the origin or a reflection with respect to some line l. In the first case, the orthogonal transformation $\mathcal{U}$ in an arbitrary orthonormal basis of the plane has matrix

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}. \tag{7.22}$$

In the second case, the plane can be represented in the form of the direct sum $\mathsf{L}' = l \oplus l^{\perp}$, where l and $l^{\perp}$ are lines, and for a vector $\boldsymbol{x}$ we have the decomposition $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$, where $\boldsymbol{y} \in l$ and $\boldsymbol{z} \in l^{\perp}$, while the vector $\mathcal{U}(\boldsymbol{x})$ is equal to $\boldsymbol{y} - \boldsymbol{z}$. If we choose an orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$ in such a way that the vector $\boldsymbol{e}_1$ lies on the line l, then the transformation $\mathcal{U}$ will have matrix

$$U = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{7.23}$$

But we shall not presuppose this fact from analytic geometry, and instead show that it derives from simple considerations in linear algebra. Let $\mathcal{U}$ have, in some orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$, the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{7.24}$$

that is, it maps the vector $x\boldsymbol{e}_1 + y\boldsymbol{e}_2$ to $(ax + by)\boldsymbol{e}_1 + (cx + dy)\boldsymbol{e}_2$. The fact that $\mathcal{U}$ preserves the length of a vector gives the relationship

$$(ax + by)^2 + (cx + dy)^2 = x^2 + y^2$$

for all x and y. Substituting in turn $(1, 0)$, $(0, 1)$, and $(1, 1)$ for (x, y), we obtain

$$a^2 + c^2 = 1, \qquad b^2 + d^2 = 1, \qquad ab + cd = 0. \tag{7.25}$$

From the relationship (7.19), it follows that $|UU^*| = 1$, and since $|U^*| = |U|$, it follows that $|U|^2 = 1$, and this implies that $|U| = \pm 1$. We need to consider separately the cases of different signs.

If $|U| = -1$, then the characteristic polynomial $|U - tE|$ of the matrix (7.24) is equal to $t^2 - (a+d)t - 1$ and has positive discriminant. Therefore, the matrix (7.24) has two real eigenvalues λ_1 and λ_2 of opposite signs (since by Viète's theorem, $\lambda_1\lambda_2 = -1$) and two associated eigenvectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. Examining the restriction of $\mathcal{U}$ to the one-dimensional invariant subspaces $\langle \boldsymbol{e}_1 \rangle$ and $\langle \boldsymbol{e}_2 \rangle$, we arrive at the one-dimensional case considered above, from which, in particular, it follows that the values λ_1 and λ_2 are equal to ± 1. Let us show that the vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are orthogonal. By the definition of eigenvectors, we have the equalities $\mathcal{U}(\boldsymbol{e}_i) = \lambda_i \boldsymbol{e}_i$, from which we have

$$\big(\mathcal{U}(\boldsymbol{e}_1), \mathcal{U}(\boldsymbol{e}_2)\big) = (\lambda_1 \boldsymbol{e}_1, \lambda_2 \boldsymbol{e}_2) = \lambda_1\lambda_2(\boldsymbol{e}_1, \boldsymbol{e}_2). \tag{7.26}$$

But since the transformation $\mathcal{U}$ is orthogonal, it follows that $(\mathcal{U}(\boldsymbol{e}_1), \mathcal{U}(\boldsymbol{e}_2)) = (\boldsymbol{e}_1, \boldsymbol{e}_2)$, and from (7.26), we obtain the equality $(\boldsymbol{e}_1, \boldsymbol{e}_2) = \lambda_1\lambda_2(\boldsymbol{e}_1, \boldsymbol{e}_2)$. Since λ_1 and λ_2 have opposite signs, it follows that $(\boldsymbol{e}_1, \boldsymbol{e}_2) = 0$. Choosing eigenvectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ of unit length and such that $\lambda_1 = 1$ and $\lambda_2 = -1$, we obtain the orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$ in which the transformation $\mathcal{U}$ has matrix (7.23). We then have the decomposition $\mathsf{L} = l \oplus l^\perp$, where $l = \langle \boldsymbol{e}_1 \rangle$ and $l^\perp = \langle \boldsymbol{e}_2 \rangle$, and the transformation $\mathcal{U}$ is a reflection in the line l.

But if $|U| = 1$, then by relationship (7.25) for a, b, c, d, it is easy to derive, keeping in mind that $ad - bc = 1$, that there exists an angle φ such that $a = d = \cos\varphi$ and $c = -b = \sin\varphi$, that is, the matrix (7.24) has the form (7.22).

As a basis for examining the general case, we have the following theorem.

Theorem 7.24 *If a subspace L' is invariant with respect to an orthogonal transformation $\mathcal{U}$, then its orthogonal complement $(\mathsf{L}')^\perp$ is also invariant with respect to $\mathcal{U}$.*

Proof We must show that for every vector $\boldsymbol{y} \in (\mathsf{L}')^\perp$, we have $\mathcal{U}(\boldsymbol{y}) \in (\mathsf{L}')^\perp$. If $\boldsymbol{y} \in (\mathsf{L}')^\perp$, then $(\boldsymbol{x}, \boldsymbol{y}) = 0$ for all $\boldsymbol{x} \in \mathsf{L}'$. From the orthogonality of the transformation $\mathcal{U}$, we obtain that $(\mathcal{U}(\boldsymbol{x}), \mathcal{U}(\boldsymbol{y})) = (\boldsymbol{x}, \boldsymbol{y}) = 0$. Since $\mathcal{U}$ is a bijective mapping from L to L, its restriction to the invariant subspace L' is a bijection from L' to L'. In other words, every vector $\boldsymbol{x}' \in \mathsf{L}'$ can be represented in the form $\boldsymbol{x}' = \mathcal{U}(\boldsymbol{x})$, where $\boldsymbol{x}$ is some other vector in L'. Consequently, $(\boldsymbol{x}', \mathcal{U}(\boldsymbol{y})) = 0$ for every vector $\boldsymbol{x}' \in \mathsf{L}'$, and this implies that $\mathcal{U}(\boldsymbol{y}) \in (\mathsf{L}')^\perp$. □

Remark 7.25 In the proof of Theorem 7.24, we nowhere used the positive definiteness of the quadratic form $(\boldsymbol{x}, \boldsymbol{x})$ associated with the inner product $(\boldsymbol{x}, \boldsymbol{y})$. Indeed, this theorem holds as well for an arbitrary nonsingular bilinear form $(\boldsymbol{x}, \boldsymbol{y})$. The condition of nonsingularity is required in order that the restriction of the transformation $\mathcal{U}$ to an invariant subspace be a bijection, without which the theorem would not be true.

Definition 7.26 Subspaces L_1 and L_2 of a Euclidean space are said to be mutually *orthogonal* if $(\boldsymbol{x}, \boldsymbol{y}) = 0$ for all vectors $\boldsymbol{x} \in \mathsf{L}_1$ and $\boldsymbol{y} \in \mathsf{L}_2$. In such a case, we write

$\mathsf{L}_1 \perp \mathsf{L}_2$. The decomposition of a Euclidean space as a direct sum of orthogonal subspaces is called an *orthogonal decomposition*.

If $\dim \mathsf{L} > 2$, then by Theorem 4.22, the transformation $\mathcal{U}$ has a one- or two-dimensional invariant subspace. Thus using Theorem 7.24 as many times as necessary (depending on $\dim \mathsf{L}$), we obtain the orthogonal decomposition

$$\mathsf{L} = \mathsf{L}_1 \oplus \mathsf{L}_2 \oplus \cdots \oplus \mathsf{L}_k, \quad \text{where } \mathsf{L}_i \perp \mathsf{L}_j \text{ for all } i \neq j, \tag{7.27}$$

with all subspaces L_i invariant with respect to the transformation $\mathcal{U}$ and of dimension 1 or 2.

Combining the orthonormal bases of the subspaces $\mathsf{L}_1, \dots, \mathsf{L}_k$ and choosing a convenient ordering, we obtain the following result.

Theorem 7.27 *For every orthogonal transformation there exists an orthonormal basis in which the matrix of the transformation has the block-diagonal form*

$$\begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & 0 & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & A_{\varphi_1} & & \\ & & 0 & & & & & \ddots & \\ & & & & & & & & A_{\varphi_r} \end{pmatrix}, \tag{7.28}$$

where

$$A_{\varphi_i} = \begin{pmatrix} \cos\varphi_i & -\sin\varphi_i \\ \sin\varphi_i & \cos\varphi_i \end{pmatrix}, \tag{7.29}$$

$\varphi_i \neq \pi k$, $k \in \mathbb{Z}$.

Let us note that the determinants of all the matrices (7.29) are equal to 1, and therefore, for a proper orthogonal transformation (see the definition on p. 135), the number of -1's on the main diagonal in (7.28) is even, and for an improper orthogonal transformation, that number is odd.

Let us now look at what the theorems we have proved give us in the cases $n = 1, 2, 3$ familiar from analytic geometry.

For $n = 1$, there exist, as we have already seen, altogether two orthogonal transformations, namely $\mathcal{E}$ and $-\mathcal{E}$, the first of which is proper, and the second, improper.

For $n = 2$, a proper orthogonal transformation is a rotation of the plane through some angle φ. In an arbitrary orthonormal basis, its matrix has the form A_φ from (7.29), with no restriction on the angle φ. For the improper transformation appearing

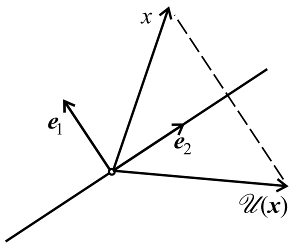


Fig. 7.3 Reflection of the plane with respect to a line

in (7.28), the number -1 must be encountered an odd number of times, that is, once. This implies that in some orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$, its matrix has the form

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This transformation is a reflection of the plane with respect to the line $\langle \boldsymbol{e}_2 \rangle$ (Fig. 7.3).

Let us now consider the case $n = 3$. Since the characteristic polynomial of the transformation $\mathcal{U}$ has odd degree 3, it must have at least one real root. This implies that in the representation (7.28), the number $+1$ or -1 must appear on the main diagonal of the matrix.

Let us consider proper transformations first. In this case, for the matrix (7.28), we have only one possibility:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}.$$

If the matrix is written in the basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$, then the transformation $\mathcal{U}$ does not change the points of the line $l = \langle \boldsymbol{e}_1 \rangle$ and represents a rotation through the angle φ in the plane $\langle \boldsymbol{e}_2, \boldsymbol{e}_3 \rangle$. In this case, we say that the transformation $\mathcal{U}$ is a *a rotation of the plane through the angle φ about the axis l*. That every proper orthogonal transformation of a three-dimensional Euclidean space possesses a "rotational axis" is a result first proved by Euler. We shall discuss the mechanical significance of this assertion later, in connection with motions of affine spaces.

Finally, if an orthogonal transformation is improper, then in expression (7.28), we have only the possibility

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}.$$

In this case, the orthogonal transformation $\mathcal{U}$ reduces to a rotation about the l-axis with a simultaneous *reflection* with respect to the plane $l^\perp$.

7.3 Orientation of a Euclidean Space*

In a Euclidean space, as in any real vector space, there are defined the notions of equal and opposite *orientations* of two bases and *orientation* of the space (see Sect. 4.4). But in Euclidean spaces, these notions possess certain specific features.

Let $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and $\boldsymbol{e}'_1, \dots, \boldsymbol{e}'_n$ be two *orthonormal* bases of a Euclidean space L. By general definition, they have *equal orientations* if the transformation from one basis to the other is proper. This implies that for a transformation $\mathcal{U}$ such that

$$\mathcal{U}(\boldsymbol{e}_1) = \boldsymbol{e}'_1, \qquad \dots, \qquad \mathcal{U}(\boldsymbol{e}_n) = \boldsymbol{e}'_n,$$

the determinant of its matrix is positive. But in the case that both bases under consideration are orthonormal, the mapping $\mathcal{U}$, as we know, is orthogonal, and its matrix U satisfies the relationship $|U| = \pm 1$. This implies that $\mathcal{U}$ is a proper transformation if and only if $|U| = 1$, and it is improper if and only if $|U| = -1$. We have the following analogue to Theorems 4.38–4.40 of Sect. 4.4.

Theorem 7.28 *Two orthogonal transformations of a real Euclidean space can be continuously deformed into each other if and only if the signs of their determinants coincide.*

The definition of a continuous deformation repeats here the definition given in Sect. 4.4 for the set $\mathfrak{A}$, but now consisting only of *orthogonal* matrices (or transformations). Since the product of any two orthogonal transformations is again orthogonal, Lemma 4.37 (p. 159) is also valid in this case, and we shall make use of it.

Proof of Theorem 7.28 Let us show that an arbitrary proper orthogonal transformation $\mathcal{U}$ can be continuously deformed into the identity. Since the condition of continuous deformability defines an equivalence relation on the set of orthogonal transformations, then by transitivity, the assertion of the theorem will follow for all proper transformations.

Thus we must prove that there exists a family of orthogonal transformations $\mathcal{U}_t$ depending continuously on the parameter $t \in [0, 1]$ for which $\mathcal{U}_0 = \mathcal{E}$ and $\mathcal{U}_1 = \mathcal{U}$. The continuous dependence of $\mathcal{U}_t$ implies that when it is represented in an arbitrary basis, all the elements of the matrices of the transformations $\mathcal{U}_t$ are continuous functions of t. We note that this is a not at all obvious corollary to Theorem 4.38. Indeed, it did not guarantee us that all the intermediate transformations $\mathcal{U}_t$ for $0 < t < 1$ are orthogonal. A possible "bad" deformation $\mathcal{A}_t$ taking us out of the domain of orthogonal transformations is depicted as the dotted line in Fig. 7.4.

We shall use Theorem 7.27 and examine the orthonormal basis in which the matrix of the transformation $\mathcal{U}$ has the form (7.28). The transformation $\mathcal{U}$ is proper if and only if the number of instances of -1 on the main diagonal of (7.28) is odd. We observe that the second-order matrix

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

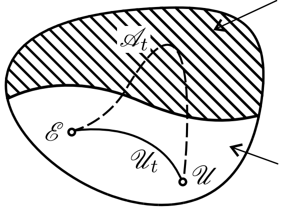


Fig. 7.4 Deformation taking us outside the domain of orthogonal transformations

can also be written in the form (7.29) for $\varphi_i = \pi$. Thus a proper orthogonal transformation can be written in a suitable orthonormal basis in block-diagonal form

$$\begin{pmatrix} E & & & \\ & A_{\varphi_1} & & \\ & & \ddots & \\ & & & A_{\varphi_k} \end{pmatrix}, \tag{7.30}$$

where the arguments φ_i can now be taken to be any values. Formula (7.30) in fact gives a continuous deformation of the transformation $\mathcal{U}$ into $\mathcal{E}$. To maintain agreement with our notation, let us examine the transformations $\mathcal{U}_t$ having in this same basis the matrix

$$\begin{pmatrix} E & & & \\ & A_{t\varphi_1} & & \\ & & \ddots & \\ & & & A_{t\varphi_k} \end{pmatrix}. \tag{7.31}$$

Then it is clear first of all that the transformation $\mathcal{U}_t$ is orthogonal for every t, and secondly, that $\mathcal{U}_0 = \mathcal{E}$ and $\mathcal{U}_1 = \mathcal{U}$. This gives us a proof of the theorem in the case of a proper transformation.

Let us now consider improper orthogonal transformations and show that any such transformation $\mathcal{V}$ can be continuously deformed into a reflection with respect to a hyperplane, that is, into a transformation $\mathcal{F}$ having in some orthonormal basis the matrix

$$F = \begin{pmatrix} -1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}. \tag{7.32}$$

Let us choose an arbitrary orthonormal basis of the vector space and suppose that in this basis, the improper orthogonal transformation $\mathcal{V}$ has matrix V. Then it is obvious that the transformation $\mathcal{U}$ with matrix $U = VF$ in this same basis is a proper orthogonal transformation. Taking into account the obvious relationship $F^{-1} = F$, we have $V = UF$, that is, $\mathcal{V} = \mathcal{U}\mathcal{F}$. We shall use the family $\mathcal{U}_t$ effecting a continuous deformation of the proper transformation $\mathcal{U}$ into $\mathcal{E}$. From the preceding

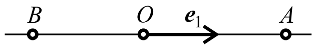


Fig. 7.5 Oriented length

equality, with the help of Lemma 4.37, we obtain the continuous family $\mathcal{V}_t = \mathcal{U}_t\mathcal{F}$, where $\mathcal{V}_0 = \mathcal{E}\mathcal{F} = \mathcal{F}$ and $\mathcal{V}_1 = \mathcal{U}\mathcal{F} = \mathcal{V}$. Thus the family $\mathcal{V}_t = \mathcal{U}_t\mathcal{F}$ effects the deformation of the improper transformation $\mathcal{V}$ into $\mathcal{F}$. □

In analogy to what we did in Sect. 4.4, Theorem 7.28 gives us the following topological result: the set of orthogonal transformations consists of two path-connected components: the proper and improper orthogonal transformations.

Exactly as in Sect. 4.4, from what we have proved, it also follows that two equally oriented orthogonal bases can be continuously deformed into each other. That is, if $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and $\boldsymbol{e}'_1, \dots, \boldsymbol{e}'_n$ are orthogonal bases with the same orientation, then there exists a family of orthonormal bases $\boldsymbol{e}_1(t), \dots, \boldsymbol{e}_n(t)$ depending continuously on the parameter $t \in [0, 1]$ such that $\boldsymbol{e}_i(0) = \boldsymbol{e}_i$ and $\boldsymbol{e}_i(1) = \boldsymbol{e}'_i$. In other words, the concept of orientation of a space is the same whether we define it in terms of an arbitrary basis or an orthonormal one. We shall further examine oriented Euclidean spaces, choosing an orientation arbitrarily. This choice makes it possible to speak of positively and negatively oriented orthonormal bases.

Now we can compare the concepts of oriented and unoriented volume. These two numbers differ by the factor ± 1 (unoriented volumes are nonnegative by definition). When the oriented volume of a parallelepiped $\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$ in a space L of dimension n was introduced, we noted that its definition depends on the choice of some orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$. Since we are assuming that the space L is oriented, we can include in the definition of oriented volume of a parallelepiped $\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$ the condition that the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ used in the definition of $v(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$ be positively oriented. Then the number $v(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$ does not depend on the choice of basis (that is, it remains unchanged if instead of $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, we take any other orthonormal positively oriented basis $\boldsymbol{e}'_1, \dots, \boldsymbol{e}'_n$). This follows immediately from formula (7.13) for the transformation $\mathcal{C} = \mathcal{U}$ and from the fact that the transformation $\mathcal{U}$ taking one basis to the other is orthogonal and proper, that is, $|\mathcal{U}| = 1$.

We can now say that the oriented volume $v(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$ is positive (and consequently equal to the unoriented volume) if the bases $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ are equally oriented, and is negative (that is, it differs from the unoriented volume by a sign) if these bases have opposite orientations. For example, on the line (Fig. 7.5), the length of the segment OA is equal to 2, while the length of the segment OB is equal to -2.

Thus, we may say that for the parallelepiped $\Pi(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$, its oriented volume is its "volume with orientation."

If we choose a coordinate origin on the real line, then a basis of it consists of a single vector, and vectors $\boldsymbol{e}_1$ and $\alpha\boldsymbol{e}_1$ are equally oriented if they lie to one side of the origin, that is, $\alpha > 0$. The choice of orientation on the line, one might say, corresponds to the choice of "right" and "left."

In the real plane, the orientation given by the basis $\boldsymbol{e}_1, \boldsymbol{e}_2$ is determined by the "direction of rotation" from $\boldsymbol{e}_1$ to $\boldsymbol{e}_2$: clockwise or counterclockwise. Equally oriented bases $\boldsymbol{e}_1, \boldsymbol{e}_2$ and $\boldsymbol{e}'_1, \boldsymbol{e}'_2$ (Fig. 7.6(a) and (b)) can be continuously transformed

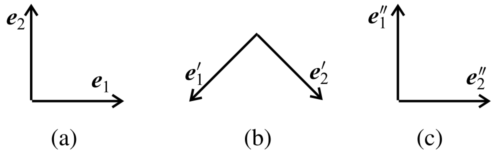


Fig. 7.6 Oriented bases of the plane

one into the other, while oppositely oriented bases cannot even if they form equal figures (Fig. 7.6(a) and (c)), since what is required for this is a reflection, that is, an improper transformation.

In real three-dimensional space, the orientation is defined by a basis of three orthonormal vectors. We again meet with two opposite orientations, which are represented by our right and left hands (see Fig. 7.7(a)). Another method of providing an orientation in three-dimensional space is defined by a helix (Fig. 7.7(b)). In this case, the orientation is defined by the direction in which the helix turns as it rises—clockwise or counterclockwise.[2]

7.4 Examples*

Example 7.29 By the term "figure" in a Euclidean space L we shall understand an arbitrary subset $S \subset \mathsf{L}$. Two figures S and S' contained in a Euclidean space M of dimension n are said to be *congruent*, or *geometrically identical*, if there exists an orthogonal transformation $\mathcal{U}$ of the space M taking S to S'. We shall be interested in the following question: When are figures S and S' congruent, that is, when do we have $\mathcal{U}(S) = S'$?

Let us first deal with the case in which the figures S and S' consist of collections of m vectors: $S = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_m)$ and $S' = (\boldsymbol{a}'_1, \dots, \boldsymbol{a}'_m)$ with $m \le n$. For S and S' to be congruent is equivalent to the existence of an orthogonal transformation $\mathcal{U}$ such that $\mathcal{U}(\boldsymbol{a}_i) = \boldsymbol{a}'_i$ for all $i = 1, \dots, m$. For this, of course, it is necessary that the

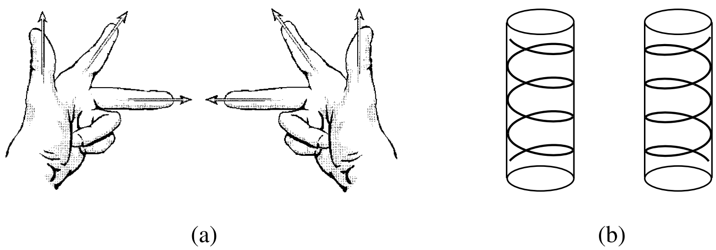


Fig. 7.7 Different orientations of three-dimensional space

[2]The molecules of amino acids likewise determine a certain orientation of space. In biology, the two possible orientations are designated by D (right = *dexter* in Latin) and L (left = *laevus*). For some unknown reason, they all determine the same orientation, namely the counterclockwise one.

following equality holds:

$$(\boldsymbol{a}_i, \boldsymbol{a}_j) = \big(\boldsymbol{a}'_i, \boldsymbol{a}'_j\big), \quad i, j = 1, \ldots, m. \tag{7.33}$$

Let us assume that vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ are linearly independent, and we shall then prove that the condition (7.33) is sufficient. By Theorem 7.14, in this case we have $G(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m) > 0$, and by assumption, $G(\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_m) = G(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)$. From this same theorem, it follows that the vectors $\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_m$ will also be linearly independent.

Let us set

$$\mathsf{L} = \langle \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \rangle, \qquad \mathsf{L}' = \big\langle \boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_m \big\rangle, \tag{7.34}$$

and consider first the case $m = n$. Let $\mathsf{M} = \langle \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \rangle$. We shall consider the transformation $\mathcal{U} : \mathsf{M} \to \mathsf{M}$ given by the conditions $\mathcal{U}(\boldsymbol{a}_i) = \boldsymbol{a}'_i$ for all $i = 1, \ldots, m$. Obviously, such a transformation is uniquely determined, and by the relationship

$$\left(\mathcal{U}\left(\sum_{i=1}^m \alpha_i \boldsymbol{a}_i\right), \mathcal{U}\left(\sum_{j=1}^m \beta_j \boldsymbol{a}_j\right)\right) = \left(\sum_{i=1}^m \alpha_i \boldsymbol{a}'_i, \sum_{j=1}^m \beta_j \boldsymbol{a}'_j\right) = \sum_{i,j=1}^m \alpha_i \beta_j \big(\boldsymbol{a}'_i, \boldsymbol{a}'_j\big)$$

and equality (7.33), it is orthogonal.

Let $m < n$. Then we have the decomposition $\mathsf{M} = \mathsf{L} \oplus \mathsf{L}^\perp = \mathsf{L}' \oplus (\mathsf{L}')^\perp$, where the subspaces L and L' of the space M are defined by formula (7.34). By what has gone before, there exists an isomorphism $\mathcal{V} : \mathsf{L} \to \mathsf{L}'$ such that $\mathcal{V}(\boldsymbol{a}_i) = \boldsymbol{a}'_i$ for all $i = 1, \ldots, m$. The orthogonal complements $\mathsf{L}^\perp$ and $(\mathsf{L}')^\perp$ of these subspaces have dimension $n - m$, and consequently, are also isomorphic (Theorem 7.22). Let us choose an arbitrary isomorphism $\mathcal{W} : \mathsf{L}^\perp \to (\mathsf{L}')^\perp$. As a result of the decomposition $\mathsf{M} = \mathsf{L} \oplus \mathsf{L}^\perp$, an arbitrary vector $\boldsymbol{x} \in \mathsf{M}$ can be uniquely represented in the form $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$, where $\boldsymbol{y} \in \mathsf{L}$ and $\boldsymbol{z} \in \mathsf{L}^\perp$. Let us define the linear transformation $\mathcal{U} : \mathsf{M} \to \mathsf{M}$ by the formula $\mathcal{U}(\boldsymbol{x}) = \mathcal{V}(\boldsymbol{y}) + \mathcal{W}(\boldsymbol{z})$. By construction, $\mathcal{U}(\boldsymbol{a}_i) = \boldsymbol{a}'_i$ for all $i = 1, \ldots, m$, and a trivial verification shows that the transformation $\mathcal{U}$ is orthogonal.

Let us now consider the case that $S = l$ and $S' = l'$ are lines, and consequently, consist of an infinite number of vectors. It suffices to set $l = \langle \boldsymbol{e} \rangle$ and $l' = \langle \boldsymbol{e}' \rangle$, where $|\boldsymbol{e}| = |\boldsymbol{e}'| = 1$, and to use the fact that there exists an orthogonal transformation $\mathcal{U}$ of the space M taking $\boldsymbol{e}$ to $\boldsymbol{e}'$. Thus any two lines are congruent.

The next case in order of increasing complexity is that in which figures S and S' each consist of two lines: $S = l_1 \cup l_2$ and $S' = l'_1 \cup l'_2$. Let us set $l_i = \langle \boldsymbol{e}_i \rangle$ and $l'_i = \langle \boldsymbol{e}'_i \rangle$, where $|\boldsymbol{e}_i| = |\boldsymbol{e}'_i| = 1$ for $i = 1$ and 2. Now, however, vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are no longer defined uniquely, but can be replaced by $-\boldsymbol{e}_1$ or $-\boldsymbol{e}_2$. In this case, their lengths do not change, but the inner product $(\boldsymbol{e}_1, \boldsymbol{e}_2)$ can change their sign, that is, what remains unchanged is only their absolute value $|(\boldsymbol{e}_1, \boldsymbol{e}_2)|$. Based on previous considerations, we may say that figures S and S' are congruent if and only if $|(\boldsymbol{e}_1, \boldsymbol{e}_2)| = |(\boldsymbol{e}'_1, \boldsymbol{e}'_2)|$. If φ is the angle between the vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, then we see that the lines l_1 and l_2 determine $|\cos\varphi|$, or equivalently the angle φ, for which $0 \le \varphi \le \frac{\pi}{2}$. In textbooks on geometry, one often reads about two angles between straight lines, the "acute" and "obtuse" angles, but we shall choose only the one that is acute or a right angle. This angle φ is called the *angle between the lines* l_1 *and* l_2. The previous exposition shows that two pairs of lines l_1, l_2 and l_1', l_2' are congruent if and only if the angles between them thus defined coincide.

The case in which a figure S consists of a line l and a plane L ($\dim l = 1$, $\dim \mathsf{L} = 2$) is also related, strictly speaking, to elementary geometry, since $\dim(l + \mathsf{L}) \le 3$, and the figure $S = l \cup \mathsf{L}$ can be embedded in three-dimensional space. But we shall consider it from a more abstract point of view, using the language of Euclidean spaces. Let $l = \langle \boldsymbol{e} \rangle$ and let $\boldsymbol{f}$ be the orthogonal projection of $\boldsymbol{e}$ onto L. The angle φ between the lines l and $l' = \langle \boldsymbol{f} \rangle$ is called the *angle between* l *and* L (as already mentioned above, it is acute or right). The cosine of this angle can be calculated according to the following formula:

$$\cos\varphi = \frac{|(\boldsymbol{e}, \boldsymbol{f})|}{|\boldsymbol{e}| \cdot |\boldsymbol{f}|}. \tag{7.35}$$

Let us show that if the angle between the line l and the plane L is equal to the angle between the line l' and the plane L', then the figures $S = l \cup \mathsf{L}$ and $S' = l' \cup \mathsf{L}'$ are congruent. First of all, it is obvious that there exists an orthogonal transformation taking L to L', so that we may consider that $\mathsf{L} = \mathsf{L}'$. Let $l = \langle \boldsymbol{e} \rangle$, $|\boldsymbol{e}| = 1$ and $l' = \langle \boldsymbol{e}' \rangle$, $|\boldsymbol{e}'| = 1$, and let us denote by $\boldsymbol{f}$ and $\boldsymbol{f}'$ the orthogonal projections $\boldsymbol{e}$ and $\boldsymbol{e}'$ onto L. By assumption,

$$\frac{|(\boldsymbol{e}, \boldsymbol{f})|}{|\boldsymbol{e}| \cdot |\boldsymbol{f}|} = \frac{|(\boldsymbol{e}', \boldsymbol{f}')|}{|\boldsymbol{e}'| \cdot |\boldsymbol{f}'|}. \tag{7.36}$$

Since $\boldsymbol{e}$ and $\boldsymbol{e}'$ can be represented in the form $\boldsymbol{e} = \boldsymbol{f} + \boldsymbol{x}$ and $\boldsymbol{e}' = \boldsymbol{f}' + \boldsymbol{y}$, where $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{L}^\perp$, it follows that $|(\boldsymbol{e}, \boldsymbol{f})| = |\boldsymbol{f}|^2$, $|(\boldsymbol{e}', \boldsymbol{f}')| = |\boldsymbol{f}'|^2$. Moreover, $|\boldsymbol{e}| = |\boldsymbol{e}'| = 1$, and the relationship (7.36) shows that $|\boldsymbol{f}| = |\boldsymbol{f}'|$.

Since $\boldsymbol{e} = \boldsymbol{x} + \boldsymbol{f}$, we have $|\boldsymbol{e}|^2 = |\boldsymbol{x}|^2 + 2(\boldsymbol{x}, \boldsymbol{f}) + |\boldsymbol{f}|^2$, from which, if we take into account the equalities $|\boldsymbol{e}|^2 = 1$ and $(\boldsymbol{x}, \boldsymbol{f}) = 0$, we obtain $|\boldsymbol{x}|^2 = 1 - |\boldsymbol{f}|^2$ and analogously, $|\boldsymbol{y}|^2 = 1 - |\boldsymbol{f}'|^2$. From this follows the equality $|\boldsymbol{x}| = |\boldsymbol{y}|$. Let us define the orthogonal transformation $\mathcal{U}$ of the space $\mathsf{M} = \mathsf{L} \oplus \mathsf{L}^\perp$ whose restriction to the plane L carries the vector $\boldsymbol{f}$ to $\boldsymbol{f}'$ (this is possible because $|\boldsymbol{f}| = |\boldsymbol{f}'|$), while the restriction to its orthogonal complement $\mathsf{L}^\perp$ takes the vector $\boldsymbol{x}$ to $\boldsymbol{y}$ (which is possible on account of the equality $|\boldsymbol{x}| = |\boldsymbol{y}|$). Clearly, $\mathcal{U}$ takes $\boldsymbol{e}$ to $\boldsymbol{e}'$ and hence l to l', and by construction, the plane L in both figures is one and the same, and the transformation $\mathcal{U}$ takes it into itself.

We encounter a new and more interesting situation when we consider the case in which a figure S consists of a pair of planes L_1 and L_2 ($\dim \mathsf{L}_1 = \dim \mathsf{L}_2 = 2$). If $\mathsf{L}_1 \cap \mathsf{L}_2 \ne (\mathbf{0})$, then $\dim(\mathsf{L}_1 + \mathsf{L}_2) \le 3$, and we are dealing with a question from elementary geometry (which, however, can be considered simply in the language of Euclidean spaces). Therefore, we shall assume that $\mathsf{L}_1 \cap \mathsf{L}_2 = (\mathbf{0})$ and similarly, that $\mathsf{L}_1' \cap \mathsf{L}_2' = (\mathbf{0})$. When are figures $S = \mathsf{L}_1 \cup \mathsf{L}_2$ and $S' = \mathsf{L}_1' \cup \mathsf{L}_2'$ congruent? It turns out that for this to occur, it is necessary that there be agreement of not one (as in the examples considered above) but two parameters, which can be interpreted as *two angles* between the planes L_1 and L_2.

We shall consider all possible straight lines lying in the plane L_1 and the angles that they form with the plane L_2. To this end, we recall the geometric interpretation of the angle between a line l and a plane L. If $l = \langle \boldsymbol{e} \rangle$, where $|\boldsymbol{e}| = 1$, then the angle φ between l and L is determined by formula (7.35) with the condition $0 \le \varphi \le \frac{\pi}{2}$, where $\boldsymbol{f}$ is the orthogonal projection of the vector $\boldsymbol{e}$ onto L. From this, it follows that $\boldsymbol{e} = \boldsymbol{f} + \boldsymbol{x}$, where $\boldsymbol{x} \in \mathsf{L}^\perp$, and this implies that $(\boldsymbol{e}, \boldsymbol{f}) = (\boldsymbol{f}, \boldsymbol{f}) + (\boldsymbol{x}, \boldsymbol{f}) = |\boldsymbol{f}|^2$, whence the relationship (7.35) gives $|\cos\varphi| = |\boldsymbol{f}|$. In other words, to consider all the angles between lines lying in the plane L_1 and the plane L_2, we must consider the circle in L_1 consisting of all vectors of length 1 and the lengths of the orthogonal projections of these vectors onto the plane L_2. In order to write down these angles in a formula, we shall consider the orthogonal projection $\mathsf{M} \to \mathsf{L}_2$ of the space M onto the plane L_2. Let us denote by $\mathscr{P}$ the restriction of this linear transformation to the plane L_1. Then the angles of interest to us are given by the formula $|\cos\varphi| = |\mathscr{P}(\boldsymbol{e})|$, where $\boldsymbol{e}$ are all possible vectors in the plane L_1 of unit length. We restrict our attention to the case in which the linear transformation $\mathscr{P}$ is an isomorphism. The case in which this does not occur, that is, when the kernel of the transformation $\mathscr{P}$ is not equal to $(\mathbf{0})$ and the image is not equal to L_2, is dealt with similarly.

Since $\mathscr{P}$ is an isomorphism, there is an inverse transformation $\mathscr{P}^{-1} : \mathsf{L}_2 \to \mathsf{L}_1$. Let us choose in the planes L_1 and L_2 orthonormal bases $\boldsymbol{e}_1, \boldsymbol{e}_2$ and $\boldsymbol{g}_1, \boldsymbol{g}_2$. Let the vector $\boldsymbol{e} \in \mathsf{L}_1$ have unit length. We set $\boldsymbol{f} = \mathscr{P}(\boldsymbol{e})$, and assuming that $\boldsymbol{f} = x_1\boldsymbol{g}_1 + x_2\boldsymbol{g}_2$, we shall obtain equations for the coordinates x_1 and x_2. Let us set

$$\mathscr{P}^{-1}(\boldsymbol{g}_1) = \alpha\boldsymbol{e}_1 + \beta\boldsymbol{e}_2, \qquad \mathscr{P}^{-1}(\boldsymbol{g}_2) = \gamma\boldsymbol{e}_1 + \delta\boldsymbol{e}_2.$$

Since $\boldsymbol{f} = \mathscr{P}(\boldsymbol{e})$, it follows that

$$\boldsymbol{e} = \mathscr{P}^{-1}(\boldsymbol{f}) = x_1\mathscr{P}^{-1}(\boldsymbol{g}_1) + x_2\mathscr{P}^{-1}(\boldsymbol{g}_2) = (\alpha x_1 + \gamma x_2)\boldsymbol{e}_1 + (\beta x_1 + \delta x_2)\boldsymbol{e}_2,$$

and the condition $|\mathscr{P}^{-1}(\boldsymbol{f})| = 1$, which we shall write in the form $|\mathscr{P}^{-1}(\boldsymbol{f})|^2 = 1$, reduces to the equality $(\alpha x_1 + \gamma x_2)^2 + (\beta x_1 + \delta x_2)^2 = 1$, that is,

$$\bigl(\alpha^2 + \beta^2\bigr)x_1^2 + 2(\alpha\gamma + \beta\delta)x_1x_2 + \bigl(\gamma^2 + \delta^2\bigr)x_2^2 = 1. \tag{7.37}$$

Equation (7.37) with variables x_1, x_2 defines a second-degree curve in the rectangular coordinate system determined by the vectors $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$. This curve is bounded, since $|\boldsymbol{f}| \le |\boldsymbol{e}|$ ($\boldsymbol{f}$ is the orthogonal projection of the vector $\boldsymbol{e}$), and this implies that $(\boldsymbol{f}^2) \le 1$, that is, $x_1^2 + x_2^2 \le 1$. As one learns in a course on analytic geometry, such a curve is an ellipse. In our case, it has its center of symmetry at the origin O, that is, it is unchanged by a change of variables $x_1 \to -x_1$, $x_2 \to -x_2$ (see Fig. 7.8).

It is known from analytic geometry that an ellipse has two distinguished points A and A', symmetric with respect to the origin, such that the length $|OA| = |OA'|$ is greater than $|OC|$ for all other points C of the ellipse. The segment $|OA| = |OA'|$ is called the *semimajor axis* of the ellipse. Similarly, there exist points B and B' symmetric with respect to the origin such that the segment $|OB| = |OB'|$ is shorter than every other segment $|OC|$. The segment $|OB| = |OB'|$ is called the *semiminor axis* of the ellipse.

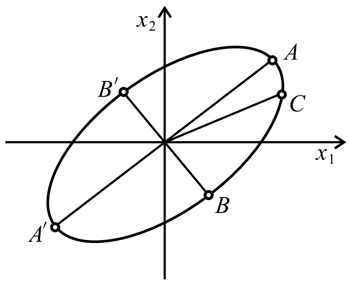


Fig. 7.8 *Ellipse* described by equation (7.37)

Let us recall that the length of an arbitrary line segment $|OC|$, where C is any point on the ellipse, gives us the value $\cos\varphi$, where φ is the angle between a certain line contained in L_1 and the plane L_2. From this it follows that $\cos\varphi$ attains its maximum for one value of φ, while for some other value of φ it attains its minimum. Let us denote these angles by φ_1 and φ_2 respectively. By definition, $0 \le \varphi_1 \le \varphi_2 \le \frac{\pi}{2}$. It is these *two* angles that are called the *angles between the planes* L_1 *and* L_2.

The case that we have omitted, in which the transformation $\mathcal{P}$ has a nonnull kernel, reduces to the case in which the ellipse depicted in Fig. 7.8 shrinks to a line segment.

It now remains for us to check that if both angles between the planes $(\mathsf{L}_1, \mathsf{L}_2)$ are equal to the corresponding angles between the planes $(\mathsf{L}_1', \mathsf{L}_2')$, then the figures $S = \mathsf{L}_1 \cup \mathsf{L}_2$ and $S' = \mathsf{L}_1' \cup \mathsf{L}_2'$ will be congruent, that is, there exists an orthogonal transformation $\mathcal{U}$ taking the plane L_i into L_i', $i = 1, 2$.

Let φ_1 and φ_2 be the angles between L_1 and L_2, equal, by hypothesis, to the angles between L_1' and L_2'. Reasoning as previously (in the case of the angle between a line and a plane), we can find an orthogonal transformation that takes L_2 to L_2'. This implies that we may assume that $\mathsf{L}_2 = \mathsf{L}_2'$. Let us denote this plane by L. Here, of course, the angles φ_1 and φ_2 remain unchanged. Let $\cos\varphi_1 \le \cos\varphi_2$ for the pair of planes L_1 and L. This implies that $\cos\varphi_1$ and $\cos\varphi_2$ are the lengths of the semiminor and semimajor axes of the ellipse that we considered above. This is also the case for the pair of planes L_1' and L. By construction, this means that $\cos\varphi_1 = |\boldsymbol{f}_1| = |\boldsymbol{f}_1'|$ and $\cos\varphi_2 = |\boldsymbol{f}_2| = |\boldsymbol{f}_2'|$, where the vectors $\boldsymbol{f}_i \in \mathsf{L}$ are orthogonal projections of the vectors $\boldsymbol{e}_i \in \mathsf{L}_1$ of length 1. Reasoning similarly, we obtain the vectors $\boldsymbol{f}_i' \in \mathsf{L}$ and $\boldsymbol{e}_i' \in \mathsf{L}_1'$, $i = 1, 2$.

Since $|\boldsymbol{f}_1| = |\boldsymbol{f}_1'|$, $|\boldsymbol{f}_2| = |\boldsymbol{f}_2'|$, and since by well-known properties of the ellipse, its semimajor and semiminor axes are orthogonal, we can find an orthogonal transformation of the space M that takes $\boldsymbol{f}_1$ to $\boldsymbol{f}_1'$ and $\boldsymbol{f}_2$ to $\boldsymbol{f}_2'$, and having done so, assume that $\boldsymbol{f}_1 = \boldsymbol{f}_1'$ and $\boldsymbol{f}_2 = \boldsymbol{f}_2'$. But since an ellipse is defined by its semiaxes, it follows that the ellipses C_1 and C_1' that are obtained in the plane L from the planes L_1 and L_1' simply coincide. Let us consider the orthogonal projections of the space M to the plane L. Let us denote by $\mathcal{P}$ its restriction to the plane L_1, and by $\mathcal{P}'$ its restriction to the plane L_1'.

We shall assume, as we did previously, that the transformations $\mathcal{P} : \mathsf{L}_1 \to \mathsf{L}$ and $\mathcal{P}' : \mathsf{L}_1' \to \mathsf{L}$ are isomorphisms of the corresponding linear spaces, but it is not at all necessary that they be isomorphisms of Euclidean spaces. Let us represent this with arrows in a commutative diagram

$$\begin{tikzcd} \mathsf{L}_1 \arrow[dd, "\mathcal{V}"'] \arrow[dr, "\mathcal{P}"] & \\ & \mathsf{L} \\ \mathsf{L}_1' \arrow[ur, "\mathcal{P}'"'] & \end{tikzcd} \tag{7.38}$$

and let us show that the transformations $\mathcal{P}$ and $\mathcal{P}'$ differ from each other by an isomorphism of Euclidean spaces L_1 and L_1'. In other words, we claim that the transformation $\mathcal{V} = (\mathcal{P}')^{-1}\mathcal{P}$ is an isomorphism of the Euclidean spaces L_1 and L_1'.

As the product of isomorphisms of linear spaces, the transformation $\mathcal{V}$ is also an isomorphism, that is, a bijective linear transformation. It remains for us to verify that $\mathcal{V}$ preserves the inner product. As noted above, to do this, it suffices to verify that $\mathcal{V}$ preserves the lengths of vectors. Let $\boldsymbol{x}$ be a vector in L. If $\boldsymbol{x} = \mathbf{0}$, then the vector $\mathcal{V}(\boldsymbol{x})$ is equal to $\mathbf{0}$ by the linearity of $\mathcal{V}$, and the assertion is obvious. If $\boldsymbol{x} \neq \mathbf{0}$, then we set $\boldsymbol{e} = \alpha^{-1}\boldsymbol{x}$, where $\alpha = |\boldsymbol{x}|$, and then $|\boldsymbol{e}| = 1$. The vector $\mathcal{P}(\boldsymbol{e})$ is contained in the ellipse C in the plane L. Since $C = C'$, it follows that $\mathcal{P}(\boldsymbol{e}) = \mathcal{P}'(\boldsymbol{e}')$, where $\boldsymbol{e}'$ is some vector in the plane L_1' and $|\boldsymbol{e}'| = 1$. From this we obtain the equality $(\mathcal{P}')^{-1}\mathcal{P}(\boldsymbol{e}) = \boldsymbol{e}'$, that is, $\mathcal{V}(\boldsymbol{e}) = \boldsymbol{e}'$ and $|\boldsymbol{e}'| = 1$, which implies that $|\mathcal{V}(\boldsymbol{x})| = \alpha = |\boldsymbol{x}|$, which is what we had to prove.

We shall now consider a basis of the plane L consisting of vectors $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ lying on the semimajor and semiminor axes of the ellipse $C = C'$, and augment it with vectors $\boldsymbol{e}_1, \boldsymbol{e}_2$, where $\mathcal{P}(\boldsymbol{e}_i) = \boldsymbol{f}_i$. We thereby obtain four vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{f}_1, \boldsymbol{f}_2$ in the space $\mathsf{L}_1 + \mathsf{L}$ (it is easily verified that they are linearly independent). Similarly, in the space $\mathsf{L}_1' + \mathsf{L}$, we shall construct four vectors $\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{f}_1, \boldsymbol{f}_2$. We shall show that there exists an orthogonal transformation of the space M taking the first set of four vectors into the second. To do so, it suffices to prove that the inner products of the associated vectors (in the order in which we have written them) coincide. Here what is least trivial is the relationship $(\boldsymbol{e}_1', \boldsymbol{e}_2') = (\boldsymbol{e}_1, \boldsymbol{e}_2)$, but it follows from the fact that $\boldsymbol{e}_i' = \mathcal{V}(\boldsymbol{e}_i)$, where $\mathcal{V}$ is an isomorphism of the Euclidean spaces L_1 and L_1'. The relationship $(\boldsymbol{e}_1', \boldsymbol{f}_1) = (\boldsymbol{e}_1, \boldsymbol{f}_1)$ is a consequence of the fact that $\boldsymbol{f}_1$ is an orthogonal projection, $(\boldsymbol{e}_1, \boldsymbol{f}_1) = |\boldsymbol{f}_1|^2$, and similarly, $(\boldsymbol{e}_1', \boldsymbol{f}_1) = |\boldsymbol{f}_1|^2$. The remaining relationships are even more obvious.

Thus the figures $S = \mathsf{L}_1 \cup \mathsf{L}_2$ and $S' = \mathsf{L}_1' \cup \mathsf{L}_2'$ are congruent if and only if *both* angles between the planes $\mathsf{L}_1, \mathsf{L}_2$ and $\mathsf{L}_1', \mathsf{L}_2'$ coincide. With the help of theorems to be proved in Sect. 7.5, it will be easy for the reader to investigate the case of a pair of subspaces $\mathsf{L}_1, \mathsf{L}_2 \subset \mathsf{M}$ of arbitrary dimension. In this case, the answer to the question whether two pairs of subspaces $S = \mathsf{L}_1 \cup \mathsf{L}_2$ and $S' = \mathsf{L}_1' \cup \mathsf{L}_2'$ are congruent is determined by the agreement of two finite sets of numbers that can be interpreted as "angles" between the subspaces $\mathsf{L}_1, \mathsf{L}_2$ and $\mathsf{L}_1', \mathsf{L}_2'$.

Example 7.30 When the senior of the two authors of this textbook gave the course on which it is based (this was probably in 1952 or 1953) at Moscow State University, he told his students about a question that had arisen in the work of A.N. Kolmogorov, A.A. Petrov, and N.V. Smirnov, the answer to which in one particular case had been obtained by A.I. Maltsev. This question was presented by the professor as an example of an unsolved problem that had been worked on by noted mathematicians yet could be formulated entirely in the language of linear algebra. At the next lecture, that is, a week later, one of the students in the class came up to him and said that he had found a solution to the problem.[3]

The question posed by A.N. Kolmogorov et al. was this: In a Euclidean space L of dimension n, we are given n nonnull mutually orthogonal vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$, that is, $(\boldsymbol{x}_i, \boldsymbol{x}_j) = 0$ for all $i \neq j$, $i, j = 1, \dots, n$. For what values $m < n$ does there exist an m-dimensional subspace $\mathsf{M} \subset \mathsf{L}$ such that the orthogonal projections of the vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ to it all have the same length? A.I. Maltsev showed that if all the vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ have the same length, then there exists such a subspace M of each dimension $m < n$.

The general case is approached as follows. Let us set $|\boldsymbol{x}_i| = \alpha_i$ and assume that there exists an m-dimensional subspace M such that the orthogonal projections of all vectors $\boldsymbol{x}_i$ to it have the same length α. Let us denote by $\mathcal{P}$ the orthogonal mapping to the subspace M, so that $|\mathcal{P}(\boldsymbol{x}_i)| = \alpha$. Let us set $\boldsymbol{f}_i = \alpha_i^{-1}\boldsymbol{x}_i$. Then the vectors $\boldsymbol{f}_1, \dots, \boldsymbol{f}_n$ form an orthonormal basis of the space L. Conversely, let us select in L an orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ such that the vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_m$ form a basis in M, that is, for the decomposition

$$\mathsf{L} = \mathsf{M} \oplus \mathsf{M}^{\perp}, \tag{7.39}$$

we join the orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_m$ of the subspace M to the orthonormal basis $\boldsymbol{e}_{m+1}, \dots, \boldsymbol{e}_n$ of the subspace $\mathsf{M}^{\perp}$.

Let $\boldsymbol{f}_i = \sum_{k=1}^{n} u_{ki}\boldsymbol{e}_k$. Then we can interpret the matrix $U = (u_{ki})$ as the matrix of the linear transformation $\mathcal{U}$, written in terms of the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, taking vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ to vectors $\boldsymbol{f}_1, \dots, \boldsymbol{f}_n$. Since both sets of vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and $\boldsymbol{f}_1, \dots, \boldsymbol{f}_n$ are orthonormal bases, it follows that $\mathcal{U}$ is an orthogonal transformation, in particular, by formula (7.18), satisfying the relationship

$$UU^* = E. \tag{7.40}$$

From the decomposition (7.39) we see that every vector $\boldsymbol{f}_i$ can be uniquely represented in the form of a sum $\boldsymbol{f}_i = \boldsymbol{u}_i + \boldsymbol{v}_i$, where $\boldsymbol{u}_i \in \mathsf{M}$ and $\boldsymbol{v}_i \in \mathsf{M}^{\perp}$. By definition, the orthogonal projection of the vector $\boldsymbol{f}_i$ onto the subspace M is equal to $\mathcal{P}(\boldsymbol{f}_i) = \boldsymbol{u}_i$. By construction of the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, it follows that

$$\mathcal{P}(\boldsymbol{f}_i) = \sum_{k=1}^{m} u_{ki}\boldsymbol{e}_k.$$

[3]It was published as L.B. Nisnevich, V.I. Bryzgalov, "On a problem of n-dimensional geometry," *Uspekhi Mat. Nauk* 8:4(56) (1953), 169–172.

By assumption, we have the equalities $|\mathcal{P}(\boldsymbol{f}_i)|^2 = |\mathcal{P}(\alpha_i^{-1}\boldsymbol{x}_i)|^2 = \alpha^2\alpha_i^{-2}$, which in coordinates assume the form

$$\sum_{k=1}^{m} u_{ki}^2 = \alpha^2\alpha_i^{-2}, \quad i = 1, \dots, n.$$

If we sum these relationships for all $i = 1, \dots, n$ and change the order of summation in the double sum, then taking into account the relationship (7.40) for the orthogonal matrix U, we obtain the equality

$$\alpha^2 \sum_{i=1}^{n} \alpha_i^{-2} = \sum_{i=1}^{n}\sum_{k=1}^{m} u_{ki}^2 = \sum_{k=1}^{m}\sum_{i=1}^{n} u_{ki}^2 = m, \tag{7.41}$$

from which it follows that α can be expressed in terms of $\alpha_1, \dots, \alpha_n$, and m by the formula

$$\alpha^2 = m\left(\sum_{i=1}^{n} \alpha_i^{-2}\right)^{-1}. \tag{7.42}$$

From this, in view of the equalities $|\mathcal{P}(\boldsymbol{f}_i)|^2 = |\mathcal{P}(\alpha_i^{-1}\boldsymbol{x}_i)|^2 = \alpha^2\alpha_i^{-2}$, we obtain the expressions

$$\left|\mathcal{P}(\boldsymbol{f}_i)\right|^2 = m\left(\alpha_i^2 \sum_{i=1}^{n} \alpha_i^{-2}\right)^{-1}, \quad i = 1, \dots, n.$$

By Theorem 7.10, we have $|\mathcal{P}(\boldsymbol{f}_i)| \le |\boldsymbol{f}_i|$, and since by construction, $|\boldsymbol{f}_i| = 1$, we obtain the inequalities

$$m\left(\alpha_i^2 \sum_{i=1}^{n} \alpha_i^{-2}\right)^{-1} \le 1, \quad i = 1, \dots, n,$$

from which it follows that

$$\alpha_i^2 \sum_{i=1}^{n} \alpha_i^{-2} \ge m, \quad i = 1, \dots, n. \tag{7.43}$$

Thus the inequalities (7.43) are *necessary* for the solvability of the problem. Let us show that they are also sufficient.

Let us consider first the case $m = 1$. We observe that in this situation, the inequalities (7.43) are automatically satisfied for an arbitrary collection of positive numbers $\alpha_1, \dots, \alpha_n$. Therefore, for an arbitrary system of mutually orthogonal vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ in L, we must produce a line $\mathsf{M} \subset \mathsf{L}$ such that the orthogonal projections of all these vectors onto it have the same length. For this, we shall take as such a line $\mathsf{M} = \langle \boldsymbol{y} \rangle$ with the vectors

$$\boldsymbol{y} = \sum_{i=1}^{n} \frac{(\alpha_1 \cdots \alpha_n)^2}{\alpha_i^2} \boldsymbol{x}_i,$$

where as before, $\alpha_i^2 = (\boldsymbol{x}_i, \boldsymbol{x}_i)$. Since $\frac{(\boldsymbol{x}_i, \boldsymbol{y})}{|\boldsymbol{y}|^2} \boldsymbol{y} \in \mathsf{M}$ and $(\boldsymbol{x}_i - \frac{(\boldsymbol{x}_i, \boldsymbol{y})}{|\boldsymbol{y}|^2} \boldsymbol{y}, \boldsymbol{y}) = 0$, it follows that the orthogonal projection of the vector $\boldsymbol{x}_i$ onto the line M is equal to

$$\mathcal{P}(\boldsymbol{x}_i) = \frac{(\boldsymbol{x}_i, \boldsymbol{y})}{|\boldsymbol{y}|^2} \boldsymbol{y}.$$

Clearly, the length of each such projection

$$\left|\mathcal{P}(\boldsymbol{x}_i)\right| = \frac{|(\boldsymbol{x}_i, \boldsymbol{y})|}{|\boldsymbol{y}|} = \frac{(\alpha_1 \cdots \alpha_n)^2}{|\boldsymbol{y}|}$$

does not depend on the index of the vector $\boldsymbol{x}_i$. Thus we have proved that for an arbitrary system of n nonnull mutually orthogonal vectors in an n-dimensional Euclidean space, there exists a line such that the orthogonal projections of all vectors onto it have the same length.

To facilitate understanding in what follows, we shall use the symbol $P(m,n)$ to denote the following assertion: If the lengths $\alpha_1, \dots, \alpha_n$ of a system of mutually orthogonal vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ in an n-dimensional Euclidean space L satisfy condition (7.43), then there exists an m-dimensional subspace $\mathsf{M} \subset \mathsf{L}$ such that the orthogonal projections $\mathcal{P}(\boldsymbol{x}_1), \dots, \mathcal{P}(\boldsymbol{x}_n)$ of the vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ onto it have the same length α, expressed by the formula (7.42). Using this convention, we may say that we have proved the assertion $P(1,n)$ for all $n > 1$.

Before passing to the case of arbitrary m, let us recast the problem in a more convenient form. Let $\beta_1, \dots, \beta_n$ be arbitrary numbers satisfying the following condition:

$$\beta_1 + \cdots + \beta_n = m, \quad 0 < \beta_i \le 1, i = 1, \dots, n. \tag{7.44}$$

Let us denote by $P'(m,n)$ the following assertion: In the Euclidean space L there exist an orthonormal basis $\boldsymbol{g}_1, \dots, \boldsymbol{g}_n$ and an m-dimensional subspace $\mathsf{L}' \subset \mathsf{L}$ such that the orthogonal projections $\mathcal{P}'(\boldsymbol{g}_i)$ of the basis vectors onto L' have length $\sqrt{\beta_i}$, that is,

$$\left|\mathcal{P}'(\boldsymbol{g}_i)\right|^2 = \beta_i, \quad i = 1, \dots, n.$$

Lemma 7.31 *The assertions $P(m,n)$ and $P'(m,n)$ with a suitable choice of numbers $\alpha_1, \dots, \alpha_n$ and $\beta_1, \dots, \beta_n$ are equivalent.*

Proof Let us first prove that the assertion $P'(m,n)$ follows from the assertion $P(m,n)$. Here we are given a collection of numbers $\beta_1, \dots, \beta_n$ satisfying the condition (7.44), and it is known that the assertion $P(m,n)$ holds for arbitrary positive numbers $\alpha_1, \dots, \alpha_n$ satisfying condition (7.43). For the numbers $\beta_1, \dots, \beta_n$ and arbitrary orthonormal basis $\boldsymbol{g}_1, \dots, \boldsymbol{g}_n$ we define vectors $\boldsymbol{x}_i = \beta_i^{-1/2}\boldsymbol{g}_i$, $i = 1, \dots, n$. It is clear that these vectors are mutually orthogonal, and furthermore, $|\boldsymbol{x}_i| = \beta_i^{-1/2}$. Let us prove that the numbers $\alpha_i = \beta_i^{-1/2}$ satisfy the inequalities (7.43). Indeed, if we take into account the condition (7.44), we have

$$\alpha_i^2 \sum_{i=1}^{n} \alpha_i^{-2} = \beta_i^{-1} \sum_{i=1}^{n} \beta_i = \beta_i^{-1} m \geq m.$$

The assertion $P(m,n)$ says that in the space L there exists an m-dimensional subspace M such that the lengths of the orthogonal projections of the vectors $\boldsymbol{x}_i$ onto it are equal to

$$\left|\mathcal{P}(\boldsymbol{x}_i)\right| = \alpha = \sqrt{m\left(\sum_{i=1}^{n} \alpha_i^{-2}\right)^{-1}} = \sqrt{m\left(\sum_{i=1}^{n} \beta_i\right)^{-1}} = 1.$$

But then the lengths of the orthogonal projections of the vectors $\boldsymbol{g}_i$ onto the same subspace M are equal to $|\mathcal{P}(\boldsymbol{g}_i)| = |\mathcal{P}(\sqrt{\beta_i}\boldsymbol{x}_i)| = \sqrt{\beta_i}$.

Now let us prove that the assertion $P'(m,n)$ yields $P(m,n)$. Here we are given a collection of nonnull mutually orthogonal vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ of length $|\boldsymbol{x}_i| = \alpha_i$, and moreover, the numbers α_i satisfy the inequalities (7.43). Let us set

$$\beta_i = \alpha_i^{-2} m \left(\sum_{i=1}^{n} \alpha_i^{-2}\right)^{-1}$$

and verify that β_i satisfies conditions (7.44). The equality $\beta_1 + \cdots + \beta_n = m$ clearly follows from the definition of the numbers β_i. From the inequalities (7.43) it follows that

$$\alpha_i^2 \geq \left(m \sum_{i=1}^{n} \alpha_i^{-2}\right)^{-1},$$

and this implies that

$$\beta_i = \alpha_i^{-2} m \left(\sum_{i=1}^{n} \alpha_i^{-2}\right)^{-1} \leq 1.$$

The assertion $P'(m,n)$ says that there exist an orthonormal basis $\boldsymbol{g}_1, \dots, \boldsymbol{g}_n$ of the space L and an m-dimensional subspace $\mathsf{L}' \subset \mathsf{L}$ such that the lengths of the orthogonal projections of the vectors $\boldsymbol{g}_i$ onto it are equal to $|\mathcal{P}'(\boldsymbol{g}_i)| = \sqrt{\beta_i}$. But then the orthogonal projections of the mutually orthogonal vectors $\beta_i^{-1/2}\boldsymbol{g}_i$ onto the same subspace L' will have the same length, namely 1.

To prove the assertion $P(m,n)$ for given vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$, it now suffices to consider the linear transformation $\mathcal{U}$ of the space L mapping the vectors $\boldsymbol{g}_i$ to

$\mathcal{U}(\boldsymbol{g}_i) = \boldsymbol{f}_i$, where $\boldsymbol{f}_i = \alpha_i^{-1}\boldsymbol{x}_i$. Since the bases $\boldsymbol{g}_1, \dots, \boldsymbol{g}_n$ and $\boldsymbol{f}_1, \dots, \boldsymbol{f}_n$ are orthonormal, it follows that $\mathcal{U}$ is an orthogonal transformation, and therefore, the orthogonal projections of the $\boldsymbol{x}_i$ onto the m-dimensional subspace $\mathsf{M} = \mathcal{U}(\mathsf{L}')$ have the same length. Moreover, by what we have proved above, this length is equal to the number α determined by formula (7.42). This completes the proof of the lemma. □

Thanks to the lemma, we may prove the assertion $P'(m,n)$ instead of the assertion $P(m,n)$. We shall do so by induction on m and n. We have already proved the base case of the induction ($m = 1$, $n > 1$). The inductive step will be divided into three parts:

(1) From assertion $P'(m,n)$ for $2m \le n+1$ we shall derive $P'(m,n+1)$.
(2) We shall prove that the assertion $P'(m,n)$ implies $P'(n,m-n)$.
(3) We shall prove that the assertion $P'(m+1,n)$ for all $n > m+1$ is a consequence of the assertion $P'(m',n)$ for all $m' \le m$ and $n > m'$.

Part 1: From assertion $P'(m,n)$ for $2m \le n+1$, we derive $P'(m,n+1)$. We shall consider the collection of positive numbers $\beta_1, \dots, \beta_n, \beta_{n+1}$ satisfying conditions (7.44) with n replaced by $n+1$, with $2m \le (n+1)$. Without loss of generality, we may assume that $\beta_1 \ge \beta_2 \ge \cdots \ge \beta_{n+1}$. Since $\beta_1 + \cdots + \beta_{n+1} = m$ and $n+1 \ge 2m$, it follows that $\beta_n + \beta_{n+1} \le 1$. Indeed, for example for odd n, the contrary assumption would give the inequality

$$\underbrace{\beta_1 + \beta_2 \ge \cdots \ge \beta_n + \beta_{n+1}}_{(n+1)/2 \text{ sums}} > 1,$$

from which clearly follows $\beta_1 + \cdots + \beta_{n+1} > (n+1)/2 \ge m$, which contradicts the assumption that has been made.

Let us consider the $(n+1)$-dimensional Euclidean space L and decompose it as a direct sum $\mathsf{L} = \langle \boldsymbol{e} \rangle \oplus \langle \boldsymbol{e} \rangle^\perp$, where $\boldsymbol{e} \in \mathsf{L}$ is an arbitrary vector of length 1. By the induction hypothesis, the assertion $P'(m,n)$ holds for numbers $\beta_1, \dots, \beta_{n-1}$ and $\beta = \beta_n + \beta_{n+1}$ and the n-dimensional Euclidean space $\langle \boldsymbol{e} \rangle^\perp$. This implies that in the space $\langle \boldsymbol{e} \rangle^\perp$, there exist an orthonormal basis $\boldsymbol{g}_1, \dots, \boldsymbol{g}_n$ and an m-dimensional subspace L' such that the squares of the lengths of the orthogonal projections of the vectors $\boldsymbol{g}_i$ onto L' are equal to

$$\left|\mathcal{P}'(\boldsymbol{g}_i)\right|^2 = \beta_i, \quad i = 1, \dots, n-1, \qquad \left|\mathcal{P}'(\boldsymbol{g}_n)\right|^2 = \beta_n + \beta_{n+1}.$$

We shall denote by $\bar{\mathcal{P}} : \mathsf{L} \to \mathsf{L}'$ the orthogonal projection of the space L onto L' (in this case, of course, $\bar{\mathcal{P}}(\boldsymbol{e}) = \boldsymbol{0}$), and we construct in L an orthonormal basis $\bar{\boldsymbol{g}}_1, \dots, \bar{\boldsymbol{g}}_{n+1}$ for which $|\bar{\mathcal{P}}(\bar{\boldsymbol{g}}_i)|^2 = \beta_i$ for all $i = 1, \dots, n+1$.

Let us set $\bar{\boldsymbol{g}}_i = \boldsymbol{g}_i$ for $i = 1, \dots, n-2$ and $\bar{\boldsymbol{g}}_n = a\boldsymbol{g}_n + b\boldsymbol{e}$, $\bar{\boldsymbol{g}}_{n+1} = c\boldsymbol{g}_n + d\boldsymbol{e}$, where the numbers a, b, c, d are chosen in such a way that the following conditions are satisfied:

$$\begin{gathered} |\bar{\boldsymbol{g}}_n| = |\bar{\boldsymbol{g}}_{n+1}| = 1, \qquad (\bar{\boldsymbol{g}}_n, \bar{\boldsymbol{g}}_{n+1}) = 0, \\ \left|\bar{\mathcal{P}}(\bar{\boldsymbol{g}}_n)\right|^2 = \beta_n, \qquad \left|\bar{\mathcal{P}}(\bar{\boldsymbol{g}}_{n+1})\right|^2 = \beta_{n+1}. \end{gathered} \tag{7.45}$$

Then the system of vectors $\bar{\boldsymbol{g}}_1, \dots, \bar{\boldsymbol{g}}_{n+1}$ proves the assertion $P'(m, n+1)$.

The relationships (7.45) can be rewritten in the form

$$a^2 + b^2 = c^2 + d^2 = 1, \qquad ac + bd = 0,$$
$$a^2(\beta_n + \beta_{n+1}) = \beta_n, \qquad c^2(\beta_n + \beta_{n+1}) = \beta_{n+1}.$$

It is easily verified that these relationships will be satisfied if we set

$$b = \pm c, \qquad d = \mp a, \qquad a = \sqrt{\frac{\beta_n}{\beta_n + \beta_{n+1}}}, \qquad c = \sqrt{\frac{\beta_{n+1}}{\beta_n + \beta_{n+1}}}.$$

Before proceeding to part 2, let us make the following observation.

Proposition 7.32 *To prove the assertion $P'(m, n)$, we may assume that $\beta_i < 1$ for all $i = 1, \dots, n$.*

Proof Let $1 = \beta_1 = \cdots = \beta_k > \beta_{k+1} \geq \cdots \geq \beta_n > 0$. We choose in the n-dimensional vector space L an arbitrary subspace L_k of dimension k and consider the orthogonal decomposition $\mathsf{L} = \mathsf{L}_k \oplus \mathsf{L}_k^\perp$. We note that

$$1 > \beta_{k+1} \geq \cdots \geq \beta_n > 0 \quad \text{and} \quad \beta_{k+1} + \cdots + \beta_n = m - k.$$

Therefore, if the assertion $P'(m-k, n-k)$ holds for the numbers $\beta_{k+1}, \dots, \beta_n$, then in $\mathsf{L}_k^\perp$, there exist a subspace L'_k of dimension $m-k$ and an orthonormal basis $\boldsymbol{g}_{k+1}, \dots, \boldsymbol{g}_n$ such that $|\mathcal{P}(\boldsymbol{g}_i)|^2 = \beta_i$ for $i = k+1, \dots, n$, where $\mathcal{P} : \mathsf{L}_k^\perp \to \mathsf{L}'_k$ is an orthogonal projection.

We now set $\mathsf{L}' = \mathsf{L}_k \oplus \mathsf{L}'_k$ and choose in L_k an arbitrary orthonormal basis $\boldsymbol{g}_1, \dots, \boldsymbol{g}_k$. Then if $\mathcal{P}' : \mathsf{L} \to \mathsf{L}'$ is the orthogonal projection, we have that $|\mathcal{P}'(\boldsymbol{g}_i)|^2 = 1$ for $i = 1, \dots, k$ and $|\mathcal{P}'(\boldsymbol{g}_i)|^2 = \beta_i$ for $i = k+1, \dots, n$. □

Part 2: Assertion $P'(m, n)$ implies assertion $P'(n, m-n)$. Let us consider n numbers $\beta_1 \geq \cdots \geq \beta_n$ satisfying condition (7.44) in which the number m is replaced by $n - m$. We must construct an orthogonal projection $\mathcal{P}' : \mathsf{L} \to \mathsf{L}'$ of the n-dimensional Euclidean space L onto the $(m-n)$-dimensional subspace L' and an orthonormal basis $\boldsymbol{g}_1, \dots, \boldsymbol{g}_n$ in L for which the conditions $|\mathcal{P}'(\boldsymbol{g}_i)|^2 = \beta_i$, $i = 1, \dots, n$, are satisfied. By a previous observation, we may assume that all β_i are less than 1. Then the numbers $\beta'_i = 1 - \beta_i$ satisfy conditions (7.44), and by assertion $P'(m, n)$, there exist an orthonormal projection $\bar{\mathcal{P}} : \mathsf{L} \to \bar{\mathsf{L}}$ of the space L onto the m-dimensional subspace $\bar{\mathsf{L}}$ and an orthonormal basis $\boldsymbol{g}_1, \dots, \boldsymbol{g}_n$ for which the conditions $|\bar{\mathcal{P}}(\boldsymbol{g}_i)|^2 = \beta'_i$ are satisfied. For the desired $(m-n)$-dimensional subspace we shall take $\mathsf{L}' = \bar{\mathsf{L}}^\perp$ and denote by $\mathcal{P}'$ the orthogonal projection onto L'. Then for each $i = 1, \dots, n$, the equalities

$$\boldsymbol{g}_i = \bar{\mathcal{P}}(\boldsymbol{g}_i) + \mathcal{P}'(\boldsymbol{g}_i), \qquad 1 = |\boldsymbol{g}_i|^2 = \left|\bar{\mathcal{P}}(\boldsymbol{g}_i)\right|^2 + \left|\mathcal{P}'(\boldsymbol{g}_i)\right|^2 = \beta'_i + \left|\mathcal{P}'(\boldsymbol{g}_i)\right|^2$$

are satisfied, from which it follows that $|\mathcal{P}'(\boldsymbol{g}_i)|^2 = 1 - \beta_i' = \beta_i$.

Part 3: Assertion $P'(m+1, n)$ for all $n > m+1$ is a consequence of $P'(m', n)$ for all $m' \le m$ and $n > m'$. By our assumption, the assertion $P'(m, n)$ holds in particular for $n = 2m+1$. By part 2, we may assert that $P'(m+1, 2m+1)$ holds, and since $2(m+1) \le (2m+1)+1$, then by virtue of part 1, we may conclude that $P'(m+1, n)$ holds for all $n \ge 2m+1$. It remains to prove the assertions $P'(m+1, n)$ for $m+2 \le n \le 2m$. But these assertions follow from $P'(n-(m+1), n)$ by part 2. It is necessary only to verify that the inequalities $1 \le n-(m+1) \le m$ are satisfied, which follows directly from the assumption that $m+2 \le n \le 2m$.

7.5 Symmetric Transformations

As we observed at the beginning of Sect. 7.1, for a Euclidean space L, there exists a natural isomorphism $\mathsf{L} \xrightarrow{\sim} \mathsf{L}^*$ that allows us to identify in this case the space L^* with L. In particular, using the definition given in Sect. 3.7, we may define for an arbitrary basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ of the space L the *dual basis* $\boldsymbol{f}_1, \dots, \boldsymbol{f}_n$ of the space L by the condition $(\boldsymbol{f}_i, \boldsymbol{e}_i) = 1$, $(\boldsymbol{f}_i, \boldsymbol{e}_j) = 0$ for $i \ne j$. Thus an orthonormal basis is one that is its own dual.

In the same way, we can assume that for an arbitrary linear transformation $\mathcal{A} : \mathsf{L} \to \mathsf{L}$, the dual transformation $\mathcal{A}^* : \mathsf{L}^* \to \mathsf{L}^*$ defined in Sect. 3.7 is a linear transformation of the Euclidean space L into itself and is determined by the condition

$$\bigl(\mathcal{A}^*(\boldsymbol{x}), \boldsymbol{y}\bigr) = \bigl(\boldsymbol{x}, \mathcal{A}(\boldsymbol{y})\bigr) \tag{7.46}$$

for all vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{L}$. By Theorem 3.81, the matrix of the linear transformation $\mathcal{A}$ in an arbitrary basis of the space L and the matrix of the dual transformation $\mathcal{A}^*$ in the dual basis are transposes of each other. In particular, the matrices of the transformations $\mathcal{A}$ and $\mathcal{A}^*$ in an arbitrary orthonormal basis are transposes of each other. This is in accord with the notation A^* that we have chosen for the transpose matrix. It is easily verified also that conversely, if the matrices of transformations $\mathcal{A}$ and $\mathcal{B}$ in some orthonormal basis are transposes of each other, then the transformations $\mathcal{A}$ and $\mathcal{B}$ are dual.

As an example, let us consider the orthogonal transformation $\mathcal{U}$, for which by definition, the condition $(\mathcal{U}(\boldsymbol{x}), \mathcal{U}(\boldsymbol{y})) = (\boldsymbol{x}, \boldsymbol{y})$ is satisfied. By formula (7.46), we have the equality $(\mathcal{U}(\boldsymbol{x}), \mathcal{U}(\boldsymbol{y})) = (\boldsymbol{x}, \mathcal{U}^*\mathcal{U}(\boldsymbol{y}))$, from which follows $(\boldsymbol{x}, \mathcal{U}^*\mathcal{U}(\boldsymbol{y})) = (\boldsymbol{x}, \boldsymbol{y})$. This implies that $(\boldsymbol{x}, \mathcal{U}^*\mathcal{U}(\boldsymbol{y}) - \boldsymbol{y}) = 0$ for all vectors $\boldsymbol{x}$, from which follows the equality $\mathcal{U}^*\mathcal{U}(\boldsymbol{y}) = \boldsymbol{y}$ for all vectors $\boldsymbol{y} \in \mathsf{L}$. In other words, the fact that $\mathcal{U}^*\mathcal{U}$ is equal to $\mathcal{E}$, the identity transformation, is equivalent to the property of orthogonality of the transformation $\mathcal{U}$. In matrix form, this is the relationship (7.18).

Definition 7.33 A linear transformation $\mathcal{A}$ of a Euclidean space is called *symmetric* or *self-dual* if $\mathcal{A}^* = \mathcal{A}$.

In other words, for a symmetric transformation $\mathcal{A}$ and arbitrary vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, the following condition must be satisfied:

$$\big(\mathcal{A}(\boldsymbol{x}), \boldsymbol{y}\big) = \big(\boldsymbol{x}, \mathcal{A}(\boldsymbol{y})\big), \tag{7.47}$$

that is, the bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y}) = (\mathcal{A}(\boldsymbol{x}), \boldsymbol{y})$ is symmetric. As we have seen, from this it follows that in an arbitrary orthonormal basis, the matrix of the transformation $\mathcal{A}$ is symmetric.

Symmetric linear transformations play a very large role in mathematics and its applications. Their most essential applications relate to quantum mechanics, where symmetric transformations of infinite-dimensional Hilbert space (see the note on p. 214) correspond to what are called *observed* physical quantities. We shall, however, restrict our attention to finite-dimensional spaces. As we shall see in the sequel, even with this restriction, the theory of symmetric linear transformations has a great number of applications.

The following theorem gives a basic property of symmetric linear transformations of finite-dimensional Euclidean spaces.

Theorem 7.34 *Every symmetric linear transformation of a real vector space has an eigenvector.*

In view of the very large number of applications of this theorem, we shall present three proofs, based on different principles.

Proof of Theorem 7.34 **First proof**. Let $\mathcal{A}$ be a symmetric linear transformation of a Euclidean space L. If $\dim \mathsf{L} > 2$, then by Theorem 4.22, it has a one- or two-dimensional invariant subspace L'. It is obvious that the restriction of the transformation $\mathcal{A}$ to the invariant subspace L' is also a symmetric transformation. If $\dim \mathsf{L}' = 1$, then we have $\mathsf{L}' = \langle \boldsymbol{e} \rangle$, where $\boldsymbol{e} \neq \boldsymbol{0}$, and this implies that $\boldsymbol{e}$ is an eigenvector. Consequently, to prove the theorem, it suffices to show that a symmetric linear transformation in the two-dimensional subspace L' has an eigenvector. Choosing in L' an orthonormal basis, we obtain for $\mathcal{A}$ a symmetric matrix in this basis:

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

In order to find an eigenvector of the transformation $\mathcal{A}$, we must find a *real* root of the polynomial $|A - tE|$. This polynomial has the form

$$(a - t)(c - t) - b^2 = t^2 - (a + c)t + ac - b^2$$

and has a real root if and only if its discriminant in nonnegative. But the discriminant of this quadratic trinomial is equal to

$$(a + c)^2 - 4\big(ac - b^2\big) = (a - c)^2 + 4b^2 \geq 0,$$

and the proof is complete.

Second proof. The second proof is based on the complexification $\mathsf{L}^{\mathbb{C}}$ of the real vector space L. Following the construction presented in Sect. 4.3, we may extend the transformation $\mathcal{A}$ to the vectors of the space $\mathsf{L}^{\mathbb{C}}$. By Theorem 4.18, the obtained transformation $\mathcal{A}^{\mathbb{C}} : \mathsf{L}^{\mathbb{C}} \to \mathsf{L}^{\mathbb{C}}$ will already have an eigenvector $\boldsymbol{e} \in \mathsf{L}^{\mathbb{C}}$ and eigenvalue $\lambda \in \mathbb{C}$, so that $\mathcal{A}^{\mathbb{C}}(\boldsymbol{e}) = \lambda \boldsymbol{e}$.

We shall extend the inner product $(\boldsymbol{x}, \boldsymbol{y})$ from the space L to $\mathsf{L}^{\mathbb{C}}$ so that it determines there a Hermitian form (see the definition on p. 210). It is clear that this can be accomplished in only one way: defining two vectors $\boldsymbol{a}_1 = \boldsymbol{x}_1 + i\boldsymbol{y}_1$ and $\boldsymbol{a}_2 = \boldsymbol{x}_2 + i\boldsymbol{y}_2$ of the space $\mathsf{L}^{\mathbb{C}}$, we obtain the inner product according to the formula

$$(\boldsymbol{a}_1, \boldsymbol{a}_2) = (\boldsymbol{x}_1, \boldsymbol{x}_2) + (\boldsymbol{y}_1, \boldsymbol{y}_2) + i\big((\boldsymbol{y}_1, \boldsymbol{x}_2) - (\boldsymbol{x}_1, \boldsymbol{y}_2)\big). \tag{7.48}$$

The verification of the fact that the inner product $(\boldsymbol{a}_1, \boldsymbol{a}_2)$ thus defined actually determines in $\mathsf{L}^{\mathbb{C}}$ a Hermitian form is reduced to the verification of sesquilinearity (in this case, it suffices to consider separately the product of a vector $\boldsymbol{a}_1$ and a vector $\boldsymbol{a}_2$ by a real number and by i) and the property of being Hermitian. Here all calculations are completely trivial, and we shall omit them.

An important new property of the inner product $(\boldsymbol{a}_1, \boldsymbol{a}_2)$ that we have obtained is its positive definiteness, that is, like the scalar product $(\boldsymbol{a}, \boldsymbol{a})$, it is real (this follows from the Hermitian property) and $(\boldsymbol{a}, \boldsymbol{a}) > 0$, $\boldsymbol{a} \neq \boldsymbol{0}$ (this is a direct consequence of formula (7.48), for $\boldsymbol{x}_1 = \boldsymbol{x}_2$, $\boldsymbol{y}_1 = \boldsymbol{y}_2$). It is obvious that for the new inner product we also have an analogue of the relationship (7.47), that is,

$$\big(\mathcal{A}^{\mathbb{C}}(\boldsymbol{a}_1), \boldsymbol{a}_2\big) = \overline{\big(\boldsymbol{a}_1, \mathcal{A}^{\mathbb{C}}(\boldsymbol{a}_2)\big)}; \tag{7.49}$$

in other words, the form $\varphi(\boldsymbol{a}_1, \boldsymbol{a}_2) = (\mathcal{A}^{\mathbb{C}}(\boldsymbol{a}_1), \boldsymbol{a}_2)$ is Hermitian. Let us apply (7.49) to the vectors $\boldsymbol{a}_1 = \boldsymbol{a}_2 = \boldsymbol{e}$. Then we obtain $(\lambda\boldsymbol{e}, \boldsymbol{e}) = (\boldsymbol{e}, \lambda\boldsymbol{e})$. Taking into account the Hermitian property, we have the equalities $(\lambda\boldsymbol{e}, \boldsymbol{e}) = \lambda(\boldsymbol{e}, \boldsymbol{e})$ and $(\boldsymbol{e}, \lambda\boldsymbol{e}) = \overline{\lambda}(\boldsymbol{e}, \boldsymbol{e})$, from which it follows that $\lambda(\boldsymbol{e}, \boldsymbol{e}) = \overline{\lambda}(\boldsymbol{e}, \boldsymbol{e})$. Since $(\boldsymbol{e}, \boldsymbol{e}) > 0$, we derive from this that $\lambda = \overline{\lambda}$, that is, the number λ is real. Thus the characteristic polynomial $|\mathcal{A}^{\mathbb{C}} - t\mathcal{E}|$ of the transformation $\mathcal{A}^{\mathbb{C}}$ has a real root λ. But a basis of the space L as a space over $\mathbb{R}$ is a basis of the space $\mathsf{L}^{\mathbb{C}}$ over $\mathbb{C}$, and the matrix of the transformation $\mathcal{A}^{\mathbb{C}}$ in this basis coincides with the matrix of the transformation $\mathcal{A}$. In other words, $|\mathcal{A}^{\mathbb{C}} - t\mathcal{E}| = |\mathcal{A} - t\mathcal{E}|$, which implies that the characteristic polynomial $|\mathcal{A} - t\mathcal{E}|$ of the transformation $\mathcal{A}$ has a real root λ, and this implies that the transformation $\mathcal{A} : \mathsf{L} \to \mathsf{L}$ has an eigenvector in the space L.

Third proof. The third proof rests on certain facts from analysis, which we now introduce. We first observe that a Euclidean space can be naturally converted into a metric space by defining the distance $r(\boldsymbol{x}, \boldsymbol{y})$ between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ by the relationship $r(\boldsymbol{x}, \boldsymbol{y}) = |\boldsymbol{x} - \boldsymbol{y}|$. Thus in the Euclidean space L we have the notions of convergence, limit, continuous functions, and closed and bounded sets; see p. xvii.

The *Bolzano–Weierstrass theorem* asserts that for an arbitrary closed and bounded set X in a finite-dimensional Euclidean space L and arbitrary continuous function $\varphi(\boldsymbol{x})$ on X there exists a vector $\boldsymbol{x}_0 \in X$ at which $\varphi(\boldsymbol{x})$ assumes its maximum value: that is, $\varphi(\boldsymbol{x}_0) \geq \varphi(\boldsymbol{x})$ for all $\boldsymbol{x} \in X$. This theorem is well known from real analysis in the case that the set X is an interval of the real line. Its proof in the general case is exactly the same and is usually presented somewhat later. Here we shall use the theorem without offering a proof.

Let us apply the Bolzano–Weierstrass theorem to the set X consisting of all vectors $\boldsymbol{x}$ of the space L such that $|\boldsymbol{x}| = 1$, that is, to the sphere of radius 1, and to the function $\varphi(\boldsymbol{x}) = (\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$. This function is continuous not only on X, but also on the entire space L. Indeed, it suffices to choose in the space L an arbitrary basis and to write down in it the inner product $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$ as a quadratic form in the coordinates of the vector $\boldsymbol{x}$. Of importance to us is solely the fact that as a result, we obtain a *polynomial* in the coordinates. After this, it suffices to use the well-known theorem that states that the sum and product of continuous functions are continuous. Then the question is reduced to a verification of the fact that an arbitrary coordinate of the vector $\boldsymbol{x}$ is a continuous function of $\boldsymbol{x}$, but this is completely obvious.

Thus the function $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$ assumes its maximum over the set X at some $\boldsymbol{x}_0 = \boldsymbol{e}$. Let us denote this value by λ. Consequently, $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x})) \leq \lambda$ for every $\boldsymbol{x}$ for which $|\boldsymbol{x}| = 1$. For every nonnull vector $\boldsymbol{y}$, we set $\boldsymbol{x} = \boldsymbol{y}/|\boldsymbol{y}|$. Then $|\boldsymbol{x}| = 1$, and applying to this vector the inequality above, we see that $(\boldsymbol{y}, \mathcal{A}(\boldsymbol{y})) \leq \lambda(\boldsymbol{y}, \boldsymbol{y})$ for all $\boldsymbol{y}$ (this obviously holds as well for $\boldsymbol{y} = \boldsymbol{0}$).

Let us prove that the number λ is an eigenvalue of the transformation $\mathcal{A}$. To this end, let us write the condition that defines λ in the form

$$\bigl(\boldsymbol{y}, \mathcal{A}(\boldsymbol{y})\bigr) \leq \lambda(\boldsymbol{y}, \boldsymbol{y}), \quad \lambda = \bigl(\boldsymbol{e}, \mathcal{A}(\boldsymbol{e})\bigr), |\boldsymbol{e}| = 1, \tag{7.50}$$

for an arbitrary vector $\boldsymbol{y} \in \mathsf{L}$.

Let us apply (7.50) to the vector $\boldsymbol{y} = \boldsymbol{e} + \varepsilon \boldsymbol{z}$, where both the scalar ε and vector $\boldsymbol{z} \in \mathsf{L}$ are thus far arbitrary. Expanding the expressions $(\boldsymbol{y}, \mathcal{A}(\boldsymbol{y})) = (\boldsymbol{e} + \varepsilon \boldsymbol{z}, \mathcal{A}(\boldsymbol{e}) + \varepsilon \mathcal{A}(\boldsymbol{z}))$ and $(\boldsymbol{y}, \boldsymbol{y}) = (\boldsymbol{e} + \varepsilon \boldsymbol{z}, \boldsymbol{e} + \varepsilon \boldsymbol{z})$, we obtain the inequality

$$\begin{aligned}
&\bigl(\boldsymbol{e}, \mathcal{A}(\boldsymbol{e})\bigr) + \varepsilon\bigl(\boldsymbol{e}, \mathcal{A}(\boldsymbol{z})\bigr) + \varepsilon\bigl(\boldsymbol{z}, \mathcal{A}(\boldsymbol{e})\bigr) + \varepsilon^2\bigl(\mathcal{A}(\boldsymbol{z}), \mathcal{A}(\boldsymbol{z})\bigr) \\
&\quad \leq \lambda\bigl((\boldsymbol{e}, \boldsymbol{e}) + \varepsilon(\boldsymbol{e}, \boldsymbol{z}) + \varepsilon(\boldsymbol{z}, \boldsymbol{e}) + \varepsilon^2(\boldsymbol{z}, \boldsymbol{z})\bigr).
\end{aligned}$$

In view of the symmetry of the transformation $\mathcal{A}$, on the basis of the properties of Euclidean spaces and recalling that $(\boldsymbol{e}, \boldsymbol{e}) = 1$, $(\boldsymbol{e}, \mathcal{A}(\boldsymbol{e})) = \lambda$, after canceling the common term $(\boldsymbol{e}, \mathcal{A}(\boldsymbol{e})) = \lambda(\boldsymbol{e}, \boldsymbol{e})$ on both sides of the above inequality, we obtain

$$2\varepsilon\bigl(\boldsymbol{e}, \mathcal{A}(\boldsymbol{z}) - \lambda \boldsymbol{z}\bigr) + \varepsilon^2\bigl(\bigl(\mathcal{A}(\boldsymbol{z}), \mathcal{A}(\boldsymbol{z})\bigr) - \lambda(\boldsymbol{z}, \boldsymbol{z})\bigr) \leq 0. \tag{7.51}$$

Let us now note that every expression $a\varepsilon + b\varepsilon^2$ in the case $a \neq 0$ assumes a positive value for some ε. For this it is necessary to choose a value $|\varepsilon|$ sufficiently small that $a + b\varepsilon$ has the same sign as a, and then to choose the appropriate sign for ε. Thus the inequality (7.51) always leads to a contradiction except in the case $(\boldsymbol{e}, \mathcal{A}(\boldsymbol{z}) - \lambda \boldsymbol{z}) = 0$.

If for some vector $\boldsymbol{z} \neq \boldsymbol{0}$, we have $\mathcal{A}(\boldsymbol{z}) = \lambda \boldsymbol{z}$, then $\boldsymbol{z}$ is an eigenvector of the transformation $\mathcal{A}$ with eigenvalue λ, which is what we wished to prove. But if

$\mathcal{A}(z) - \lambda z \neq \mathbf{0}$ for all $z \neq \mathbf{0}$, then the kernel of the transformation $\mathcal{A} - \lambda\mathcal{E}$ is equal to $(\mathbf{0})$. From Theorem 3.68 it follows that then the transformation $\mathcal{A} - \lambda\mathcal{E}$ is an isomorphism, and its image is equal to all of the space L. This implies that for arbitrary $\boldsymbol{u} \in \mathsf{L}$, it is possible choose a vector $z \in \mathsf{L}$ such that $\boldsymbol{u} = \mathcal{A}(z) - \lambda z$. Then taking into account relationship $(\boldsymbol{e}, \mathcal{A}(z) - \lambda z) = 0$, we obtain that an arbitrary vector $\boldsymbol{u} \in \mathsf{L}$ satisfies the equality $(\boldsymbol{e}, \boldsymbol{u}) = 0$. But this is impossible at least for $\boldsymbol{u} = \boldsymbol{e}$, since $|\boldsymbol{e}| = 1$. □

The further theory of symmetric transformations is constructed on the basis of some very simple considerations.

Theorem 7.35 *If a subspace L' of a Euclidean space L is invariant with respect to the symmetric transformation $\mathcal{A}$, then its orthogonal complement $(\mathsf{L}')^{\perp}$ is also invariant.*

Proof The result is a direct consequence of the definitions. Let $\boldsymbol{y}$ be a vector in $(\mathsf{L}')^{\perp}$. Then $(\boldsymbol{x}, \boldsymbol{y}) = 0$ for all $\boldsymbol{x} \in \mathsf{L}'$. In view of the symmetry of the transformation $\mathcal{A}$, we have the relationship

$$\big(\boldsymbol{x}, \mathcal{A}(\boldsymbol{y})\big) = \big(\mathcal{A}(\boldsymbol{x}), \boldsymbol{y}\big),$$

while taking into account the invariance of L' yields that $\mathcal{A}(\boldsymbol{x}) \in \mathsf{L}'$. This implies that $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{y})) = 0$ for all vectors $\boldsymbol{x} \in \mathsf{L}'$, that is, $\mathcal{A}(\boldsymbol{y}) \in (\mathsf{L}')^{\perp}$, and this completes the proof of the theorem. □

Combining Theorems 7.34 and 7.35 yields a fundamental result in the theory of symmetric transformations.

Theorem 7.36 *For every symmetric transformation $\mathcal{A}$ of a Euclidean space L of finite dimension, there exists an orthonormal basis of this space consisting of eigenvectors of the transformation $\mathcal{A}$.*

Proof The proof is by induction on the dimension of the space L. Indeed, by Theorem 7.34, the transformation $\mathcal{A}$ has at least one eigenvector $\boldsymbol{e}$. Let us set

$$\mathsf{L} = \langle \boldsymbol{e} \rangle \oplus \langle \boldsymbol{e} \rangle^{\perp},$$

where $\langle \boldsymbol{e} \rangle^{\perp}$ has dimension $n - 1$, and by Theorem 7.35, is invariant with respect to $\mathcal{A}$. By the induction hypothesis, in the space $\langle \boldsymbol{e} \rangle^{\perp}$ there exists a required basis. If we add the vector $\boldsymbol{e}$ to this basis, we obtain the desired basis in L. □

Let us discuss this result. For a symmetric transformation $\mathcal{A}$, we have an orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ consisting of eigenvectors. But to what extent is such a basis uniquely determined? Suppose the vector $\boldsymbol{e}_i$ has the associated eigenvalue λ_i.

Then in our basis, the transformation $\mathcal{A}$ has matrix

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \tag{7.52}$$

But as we saw in Sect. 4.1, the eigenvalues of a linear transformation $\mathcal{A}$ coincide with the roots of the characteristic polynomial

$$|\mathcal{A} - t\mathcal{E}| = |A - tE| = \prod_{i=1}^{n} (\lambda_i - t).$$

Thus the eigenvalues $\lambda_1, \dots, \lambda_n$ of the transformation $\mathcal{A}$ are uniquely determined. Suppose that the distinct values among them are $\lambda_1, \dots, \lambda_k$. If we assemble all the vectors of the constructed orthonormal basis that correspond to one and the same eigenvalue λ_i (from the set $\lambda_1, \dots, \lambda_k$ of distinct eigenvalues) and consider the subspace spanned by them, then we obviously obtain the eigensubspace L_{λ_i} (see the definition on p. 138). We then have the orthogonal decomposition

$$\mathsf{L} = \mathsf{L}_{\lambda_1} \oplus \cdots \oplus \mathsf{L}_{\lambda_k}, \quad \text{where } \mathsf{L}_{\lambda_i} \perp \mathsf{L}_{\lambda_j} \text{ for all } i \neq j. \tag{7.53}$$

The restriction of $\mathcal{A}$ to the eigensubspace L_{λ_i} gives a transformation $\lambda_i \mathcal{E}$, and in this subspace, *every* orthonormal basis consists of eigenvectors (with eigenvalue λ_i).

Thus we see that a symmetric transformation $\mathcal{A}$ uniquely defines only the eigensubspace L_{λ_i}, while in each of them, one can choose an orthonormal basis as one likes. On combining these bases, we obtain an arbitrary basis of the space L satisfying the conditions of Theorem 7.36.

Let us note that every eigenvector of the transformation $\mathcal{A}$ lies in one of the subspaces L_{λ_i}. If two eigenvectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are associated with different eigenvalues $\lambda_i \neq \lambda_j$, then they lie in different subspaces L_{λ_i} and L_{λ_j}, and in view of the orthogonality of the decomposition (7.53), they must be orthogonal. We thus obtain the following result.

Theorem 7.37 *The eigenvectors of a symmetric transformation corresponding to different eigenvalues are orthogonal.*

We note that this theorem can also be easily proved by direct calculation.

Proof of Theorem 7.37 Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be eigenvectors of a symmetric transformation $\mathcal{A}$ corresponding to distinct eigenvalues λ_i and λ_j. Let us substitute the expressions $\mathcal{A}(\boldsymbol{x}) = \lambda_i \boldsymbol{x}$ and $\mathcal{A}(\boldsymbol{y}) = \lambda_j \boldsymbol{y}$ into the equality $(\mathcal{A}(\boldsymbol{x}), \boldsymbol{y}) = (\boldsymbol{x}, \mathcal{A}(\boldsymbol{y}))$. From this we obtain $(\lambda_i - \lambda_j)(\boldsymbol{x}, \boldsymbol{y}) = 0$, and since $\lambda_i \neq \lambda_j$, we have $(\boldsymbol{x}, \boldsymbol{y}) = 0$. □

Theorem 7.36 is often formulated conveniently as a theorem about quadratic forms using Theorem 6.3 from Sect. 6.1 and the possibility of identifying the space

L^* with L if the space L is equipped with an inner product. Indeed, Theorem 6.3 shows that every bilinear form φ on a Euclidean space L can be represented in the form

$$\varphi(\boldsymbol{x}, \boldsymbol{y}) = \big(\boldsymbol{x}, \mathcal{A}(\boldsymbol{y})\big), \tag{7.54}$$

where $\mathcal{A}$ is the linear transformation of the space L to L^* uniquely defined by the bilinear form φ; that is, if we make the identification of L^* with L, it is a transformation of the space L into itself.

It is obvious that the symmetry of the transformation $\mathcal{A}$ coincides with the symmetry of the bilinear form φ. Therefore, the bijection between symmetric bilinear forms and linear transformations established above yields the same correspondence between quadratic forms and symmetric linear transformations of a Euclidean space L. Moreover, in view of relationship (7.54), to the symmetric transformation $\mathcal{A}$ there corresponds the quadratic form

$$\psi(\boldsymbol{x}) = \big(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x})\big),$$

and every quadratic form $\psi(\boldsymbol{x})$ has a unique representation in this form.

If in some basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, the transformation $\mathcal{A}$ has a diagonal matrix (7.52), then for the vector $\boldsymbol{x} = x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n$, the quadratic form $\psi(\boldsymbol{x})$ has in this basis the canonical form

$$\psi(\boldsymbol{x}) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2. \tag{7.55}$$

Thus Theorem 7.36 is equivalent to the following.

Theorem 7.38 *For any quadratic form in a finite-dimensional Euclidean space, there exists an orthonormal basis in which it has the canonical form* (7.55).

Theorem 7.38 is sometimes conveniently formulated as a theorem about arbitrary vector spaces.

Theorem 7.39 *For two quadratic forms in a finite-dimensional vector space, one of which is positive definite, there exists a basis (not necessarily orthonormal) in which they both have canonical form* (7.55).

In this case, we say that in a suitable basis, these quadratic forms *are reduced to a sum of squares* (even if there are negative coefficients λ_i in formula (7.55)).

Proof of Theorem 7.39 Let $\psi_1(\boldsymbol{x})$ and $\psi_2(\boldsymbol{x})$ be two such quadratic forms, one of which, let it be $\psi_1(\boldsymbol{x})$, is positive definite. By Theorem 6.10, there exists, in the vector space L in question, a basis in which the form $\psi_1(\boldsymbol{x})$ has the canonical form (7.55). Since by assumption, the quadratic form $\psi_1(\boldsymbol{x})$ is positive definite, it follows that in formula (7.55), all the numbers λ_i are positive, and therefore, there exists a basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ of the space L in which $\psi_1(\boldsymbol{x})$ is brought into the form

$$\psi(\boldsymbol{x}) = x_1^2 + \cdots + x_n^2. \tag{7.56}$$

Let us consider as the scalar product $(\boldsymbol{x}, \boldsymbol{y})$ in the space L the symmetric bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y})$, associated by Theorem 6.6 with the quadratic form $\psi_1(\boldsymbol{x})$. We thereby convert L into a Euclidean space.

As can be seen from formulas (6.14) and (7.56), the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ for this inner product is orthonormal. Then by Theorem 7.38, there exists an orthonormal basis $\boldsymbol{e}'_1, \dots, \boldsymbol{e}'_n$ of the space L in which the form $\psi_2(\boldsymbol{x})$ has canonical form (7.55). But since the basis $\boldsymbol{e}'_1, \dots, \boldsymbol{e}'_n$ is orthonormal with respect to the inner product that we defined with the help of the quadratic form $\psi_1(\boldsymbol{x})$, then in this basis, $\psi_1(\boldsymbol{x})$ as before takes the form (7.56), and that completes the proof of the theorem. □

Remark 7.40 It is obvious that Theorem 7.39 remains true if in its formulation we replace the condition of *positive* definiteness of one of the forms by the condition of *negative* definiteness. Indeed, if $\psi(\boldsymbol{x})$ is a negative definite quadratic form, then the form $-\psi(\boldsymbol{x})$ is positive definite, and both of these assume canonical form in one and the same basis.

Without the assumption of positive (or negative) definiteness of one of the quadratic forms, Theorem 7.39 is no longer true. To prove this, let us derive one *necessary* (but not sufficient) condition for two quadratic forms $\psi_1(\boldsymbol{x})$ and $\psi_2(\boldsymbol{x})$ to be simultaneously reduced to a sum of squares. Let A_1 and A_2 be their matrices in some basis. If the quadratic forms $\psi_1(\boldsymbol{x})$ and $\psi_2(\boldsymbol{x})$ are simultaneously reducible to sums of squares, then in some other basis, their matrices A'_1 and A'_2 will be diagonal, that is,

$$A'_1 = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{pmatrix}, \qquad A'_2 = \begin{pmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_n \end{pmatrix}.$$

Then the polynomial $|A'_1 t + A'_2|$ is equal to $\prod_{i=1}^n (\alpha_i t + \beta_i)$, that is, it can be factored as a product of linear factors $\alpha_i t + \beta_i$. But by formula (6.10) for replacing the matrix of a bilinear form through a change of basis, the matrices A_1, A'_1 and A_2, A'_2 are related by

$$A'_1 = C^* A_1 C, \qquad A'_2 = C^* A_2 C,$$

where C is some nonsingular matrix, that is, $|C| \neq 0$. Therefore,

$$\left|A'_1 t + A'_2\right| = \left|C^*(A_1 t + A_2)C\right| = \left|C^*\right||A_1 t + A_2||C|,$$

from which taking into account the equality $|C^*| = |C|$, we obtain the relationship

$$|A_1 t + A_2| = |C|^{-2}\left|A'_1 t + A'_2\right|,$$

from which it follows that the polynomial $|A_1 t + A_2|$ can also be factored into linear factors. Thus for two quadratic forms $\psi_1(\boldsymbol{x})$ and $\psi_2(\boldsymbol{x})$ with matrices A_1 and A_2 to be simultaneously reduced each to a sum of squares, it is *necessary* that the polynomial $|A_1 t + A_2|$ be factorable into real linear factors.

Now for $n = 2$ we set $\psi_1(\boldsymbol{x}) = x_1^2 - x_2^2$ and $\psi_2(\boldsymbol{x}) = x_1 x_2$. These quadratic forms are neither positive definite nor negative definite. Their matrices have the form

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and it is obvious that the polynomial $|A_1 t + A_2| = -(t^2 + 1)$ cannot be factored into real linear factors. This implies that the quadratic forms $\psi_1(\boldsymbol{x})$ and $\psi_2(\boldsymbol{x})$ cannot simultaneously be reduced to sums of squares.

The question of reducing pairs of quadratic forms with complex coefficients to sums of squares (with the help of a complex linear transformation) is examined in detail, for instance, in the book *The Theory of Matrices*, by F.R. Gantmacher. See the references section.

Remark 7.41 The last proof of Theorem 7.34 that we gave makes it possible to interpret the largest eigenvalue λ of a symmetric transformation $\mathcal{A}$ as the maximum of the quadratic form $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$ on the sphere $|\boldsymbol{x}| = 1$. Let λ_i be the other eigenvalues, so that $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x})) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$. Then λ is the greatest among the λ_i. Indeed, let us assume that the eigenvalues are numbered in descending order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then

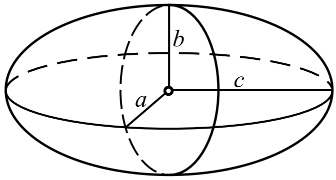
$$\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2 \leq \lambda_1 \bigl(x_1^2 + \cdots + x_n^2\bigr),$$

and the maximum value of the form $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$ on the sphere $|\boldsymbol{x}| = 1$ is equal to λ_1 (it is attained at the vector with coordinates $x_1 = 1, x_2 = \cdots = x_n = 0$). This implies that $\lambda_1 = \lambda$.

There is an analogous characteristic for the other eigenvalues λ_i as well, namely the *Courant–Fischer theorem*, which we shall present without proof. Let us consider all possible vector subspaces $\mathsf{L}' \subset \mathsf{L}$ of dimension k. We restrict the quadratic form $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$ to the subspace L' and examine its values at the intersection of L' with the unit sphere, that is, the set of all vectors $\boldsymbol{x} \in \mathsf{L}'$ that satisfy $|\boldsymbol{x}| = 1$. By the Bolzano–Weierstrass theorem, the restriction of the form $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$ to L' assumes a maximum value λ' at some point of the sphere, which, of course depends on the subspace L'. The Courant–Fischer theorem asserts that the smallest number thus obtained (as the subspace L' ranges over all subspaces of dimension k) is equal to the eigenvalue λ_{n-k+1}.

Remark 7.42 Eigenvectors are connected with the question of finding maxima and minima. Let $f(x_1, \ldots, x_n)$ be a real-valued differentiable function of n real variables. A point at which all the derivatives of the function f with respect to the variables $(x_1, \ldots, x_n)$, that is, the derivatives in all directions from this point, are equal to zero is called a *critical point* of the function. It is proved in real analysis that with some natural constraints, this condition is necessary (but not sufficient) for the function f to assume a maximum or minimum value at the point in question. Let us consider a quadratic form $f(\boldsymbol{x}) = (\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$ on the unit sphere $|\boldsymbol{x}| = 1$. It is not difficult to show that for an arbitrary point on this sphere, all points sufficiently

Fig. 7.9 An ellipsoid



close to it can be written in some system of coordinates such that our function f can be viewed as a function of these coordinates. Then the critical points of the function $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$ are exactly those points of the sphere that are eigenvectors of the symmetric transformation $\mathcal{A}$.

Example 7.43 Let an ellipsoid be given in three-dimensional space with coordinates x, y, z by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{7.57}$$

The expression on the left-hand side of (7.57) can be written in the form $\psi(\boldsymbol{x}) = (\boldsymbol{x}, \mathcal{A}(\boldsymbol{x}))$, where

$$\boldsymbol{x} = (x, y, z), \qquad \mathcal{A}(\boldsymbol{x}) = \left(\frac{x}{a^2}, \frac{y}{b^2}, \frac{z}{c^2}\right).$$

Let us assume that $0 < a < b < c$. Then the maximum value that the quadratic form $\psi(\boldsymbol{x})$ takes on the sphere $|\boldsymbol{x}| = 1$ is $\lambda = 1/a^2$. It is attained on the vectors $(\pm 1, 0, 0)$. If $|\psi(\boldsymbol{x})| \leq \lambda$ for $|\boldsymbol{x}| = 1$, then for an arbitrary vector $\boldsymbol{y} \neq \boldsymbol{0}$, setting $\boldsymbol{x} = \boldsymbol{y}/|\boldsymbol{y}|$, we obtain $|\psi(\boldsymbol{y})| \leq \lambda |\boldsymbol{y}|^2$. For the vector $\boldsymbol{y} = \boldsymbol{0}$, this inequality is obvious. Therefore, it holds in general for all $\boldsymbol{y}$. For $|\psi(\boldsymbol{y})| = 1$, it then follows that $|\boldsymbol{y}|^2 \geq 1/\lambda$. This implies that the shortest vector $\boldsymbol{y}$ satisfying equation (7.57) is the vector $(\pm a, 0, 0)$. The line segments beginning at the point $(0, 0, 0)$ and ending at the points $(\pm a, 0, 0)$ are called the *semiminor axes* of the ellipsoid (sometimes, this same term denotes their length). Similarly, the smallest value that the quadratic form $\psi(\boldsymbol{x})$ attains on the sphere $|\boldsymbol{x}| = 1$ is equal to $1/c^2$. It attains this value at vectors $(0, 0, \pm 1)$ on the unit sphere. Line segments corresponding to vectors $(0, 0, \pm c)$ are called *semimajor axes* of the ellipsoid. A vector $(0, \pm b, 0)$ corresponds to a critical point of the quadratic form $\psi(\boldsymbol{x})$ that is neither a maximum nor a minimum. Such a point is called a *minimax*, that is, as it moves from this point in one direction, the function $\psi(\boldsymbol{x})$ will increase, while in moving in another direction it will decrease (see Fig. 7.9). The line segments corresponding to the vectors $(0, \pm b, 0)$ are called the *median semiaxes* of the ellipsoid.

Everything presented thus far in this chapter (with the exception of Sect. 7.3 on the orientation of a real Euclidean space) can be transferred verbatim to complex Euclidean spaces if the inner product is defined using the positive definite Hermitian form $\varphi(\boldsymbol{x}, \boldsymbol{y})$. The condition of positive definiteness means that for the associated quadratic Hermitian form $\psi(\boldsymbol{x}) = \varphi(\boldsymbol{x}, \boldsymbol{x})$, the inequality $\psi(\boldsymbol{x}) > 0$ is satisfied for all $\boldsymbol{x} \neq \mathbf{0}$. If we denote, as before, the inner product by $(\boldsymbol{x}, \boldsymbol{y})$, the last condition can be written in the form $(\boldsymbol{x}, \boldsymbol{x}) > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$.

The dual transformation $\mathcal{A}^*$, as previously, is defined by condition (7.46). But now, the matrix of the transformation $\mathcal{A}^*$ in an orthonormal basis is obtained from the matrix of the transformation $\mathcal{A}$ not simply by taking the transpose, but by taking the complex conjugate of the transpose. The analogue of a symmetric transformation is defined as a transformation $\mathcal{A}$ whose associated bilinear form $(\boldsymbol{x}, \mathcal{A}(\boldsymbol{y}))$ is Hermitian.

It is a fundamental fact that in quantum mechanics, one deals with *complex* space. We can formulate what was stated earlier in the following form: *observed* physical quantities correspond to Hermitian forms in infinite-dimensional complex Hilbert space.

The theory of Hermitian transformations in the finite-dimensional case is constructed even more simply than the theory of symmetric transformations in real spaces, since there is no need to prove analogues of Theorem 7.34: we know already that an arbitrary linear transformation of a complex vector space has an eigenvector. From the definition of being Hermitian, it follows that the eigenvalues of a Hermitian transformation are real. The theorems proved in this section are valid for Hermitian forms (with the same proofs).

In the complex case, a transformation $\mathcal{U}$ preserving the inner product is called *unitary*. The reasoning carried out in Sect. 7.2 shows that for a unitary transformation $\mathcal{U}$, there exists an orthonormal basis consisting of eigenvectors, and all eigenvalues of the transformation $\mathcal{U}$ are complex numbers of modulus 1.

7.6 Applications to Mechanics and Geometry*

We shall present two examples from two different areas—mechanics and geometry—in which the theorems of the previous section play a key role. Since these questions will be taken up in other courses, we shall allow ourselves to be brief in both the definitions and the proofs.

Example 7.44 Let us consider the motion of a mechanical system in a small neighborhood of its equilibrium position. One says that such a system possesses *n degrees of freedom* if in some region, its state is determined by n so-called *generalized coordinates* $q_1, \dots, q_n$, which we shall consider the coordinates of a vector $\boldsymbol{q}$ in some coordinate system, and where we will take the origin $\mathbf{0}$ to be the equilibrium position of our system. The motion of the system determines the dependence of a vector $\boldsymbol{q}$ on time t. We shall assume that the equilibrium position under investigation is determined by a strict local minimum of its *potential energy* Π. If this value is equal to c, and the potential energy is a function $\Pi(q_1, \dots, q_n)$ in the generalized coordinates (it is assumed that it does not depend on time), then this implies that $\Pi(0, \dots, 0) = c$ and $\Pi(q_1, \dots, q_n) > c$ for all remaining values $q_1, \dots, q_n$ close to zero. From the fact that a critical point of the function Π corresponds to the minimum value, we may conclude that at the point $\mathbf{0}$, all partial derivatives $\partial\Pi/\partial q_i$

become zero. Therefore, for an expansion of the function $\Pi(q_1,\dots,q_n)$ as a series in powers of the variables $q_1,\dots,q_n$ at the point $\mathbf{0}$, the linear terms will be equal to zero, and we obtain the expression $\Pi(q_1,\dots,q_n)=c+\sum_{i,j=1}^n b_{ij}q_iq_j+\cdots$, where b_{ij} are certain constants, and the ellipsis indicates terms of degree greater than 2. Since we are considering motions not far from the point $\mathbf{0}$, we can disregard those values. It is in this approximation that we shall consider this problem. That is, we set

$$\Pi(q_1,\dots,q_n)=c+\sum_{i,j=1}^n b_{ij}q_iq_j.$$

Since $\Pi(q_1,\dots,q_n)>c$ for all values $q_1,\dots,q_n$ not equal to zero, the quadratic form $\sum_{i,j=1}^n b_{ij}q_iq_j$ will be positive definite.

Kinetic energy T is a quadratic form in so-called *generalized velocities dq_1/dt, $\dots, dq_n/dt$*, which are also denoted by $\dot q_1,\dots,\dot q_n$, that is,

$$T=\sum_{i,j=1}^n a_{ij}\dot q_i\dot q_j, \tag{7.58}$$

where $a_{ij}=a_{ji}$ are functions of $\boldsymbol{q}$ (we assume that they do not depend on time t). Considering as we did for potential energy only those values q_i close to zero, we may replace all the functions a_{ij} in (7.58) by constants $a_{ij}(\mathbf{0})$, which is what we shall now assume. Kinetic energy is always positive except in the case that all $\dot q_i$ are equal to 0, and therefore, the quadratic form (7.58) is positive definite.

Motion in a broad class of mechanical systems (so-called *natural systems*) is described by a rather complex system of differential equations—*second-order Lagrange equations*:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot q_i}\right)-\frac{\partial T}{\partial q_i}=-\frac{\partial \Pi}{\partial q_i},\quad i=1,\dots,n. \tag{7.59}$$

Application of Theorem 7.39 makes it possible to reduce these equations in the given situation to much simpler ones. To this end, let us find a coordinate system in which the quadratic form $\sum_{i,j=1}^n a_{ij}x_ix_j$ can be brought into the form $\sum_{i=1}^n x_i^2$, and the quadratic form $\sum_{i,j=1}^n b_{ij}x_ix_j$ into the form $\sum_{i=1}^n \lambda_i x_i^2$. Then in this case, the form $\sum_{i,j=1}^n b_{ij}x_ix_j$ is positive definite, which implies that all λ_i are positive. In this system of coordinates (we shall again denote them by $q_1,\dots,q_n$), the system of equations (7.59) is decomposed into the independent equations

$$\frac{d^2q_i}{dt^2}=-\lambda_iq_i,\quad i=1,\dots,n, \tag{7.60}$$

which have the solutions $q_i=c_i\cos\sqrt{\lambda_i}t+d_i\sin\sqrt{\lambda_i}t$, where c_i and d_i are arbitrary constants. This shows that "small oscillations" are periodic in each coordinate q_i. Since they are bounded, it follows that our equilibrium position $\mathbf{0}$ is *stable*. If we were to examine the state of equilibrium at a point that was a critical point of potential energy $\varPi$ but not a strict minimum, then in the equations (7.60) we would not be able to guarantee that all the λ_i were positive. Then for those i for which $\lambda_i < 0$, we would obtain the solutions $q_i = c_i \cosh\sqrt{-\lambda_i}t + d_i \sinh\sqrt{-\lambda_i}t$, which can grow without bound with the growth of t. Just as for $\lambda_i = 0$, we would obtain an unbounded solution $q_i = c_i + d_i t$.

Strictly speaking, we have done only the following altogether: we have replaced the given conditions of our problem with conditions close to them, with the result that the problem became much simpler. Such a procedure is usual in the theory of differential equations, where it is proved that solutions to a simplified system of equations are in a certain sense similar to the solutions of the initial system. And moreover, the degree of this deviation can be estimated as a function of the values of the terms that we have ignored. This estimation takes place in a finite interval of time whose length also depends on the value of the ignored terms. This justifies the simplifications that we have made.

A beautiful example, which played an important role historically, is given by lateral oscillations of a string of beads.[4]

Suppose we have a weightless and ideally flexible thread fixed at the ends. On it are securely fastened n beads with masses $m_1, \dots, m_n$, and suppose they divide the thread into segments of lengths $l_0, l_1, \dots, l_n$. We shall assume that in its initial state, the thread lies along the x-axis, and we shall denote by $y_1, \dots, y_n$ the displacements of the beads along the y-axis. Then the kinetic energy of this system has the form

$$T = \frac{1}{2}\sum_{i=1}^{n} m_i \dot{y}_i^2.$$

Assuming the tension of the thread to be constant (as we may because the displacements are small) and equal to σ, we obtain for the potential energy the expression $\varPi = \sigma \Delta l$, where $\Delta l = \sum_{i=0}^{n} \Delta l_i$ is the change in length of the entire thread, and Δl_i is the change in length of the portion of the thread corresponding to l_i. Then we know the Δl_i in terms of the l_i:

$$\Delta l_i = \sqrt{l_i^2 + (y_{i+1} - y_i)^2} - l_i, \quad i = 0, \dots, n,$$

where $y_0 = y_{n+1} = 0$. Expanding this expression as a sum in $y_{i+1} - y_i$, we obtain quadratic terms $\sum_{i=0}^{n} \frac{1}{2l_i}(y_{i+1} - y_i)^2$, and we may set

$$\varPi = \frac{\sigma}{2}\sum_{i=0}^{n} \frac{1}{l_i}(y_{i+1} - y_i)^2, \quad y_0 = y_{n+1} = 0.$$

[4] This example is taken from Gantmacher and Krein's book *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*, Moscow 1950, English translation, AMS Chelsea Publishing, 2002.

Thus in this case, the problem is reduced to simultaneously expressing two quadratic forms in the variables $y_1, \dots, y_n$ as sums of squares:

$$T = \frac{1}{2} \sum_{i=0}^{n} m_i \dot{y}_i^2, \qquad \varPi = \frac{\sigma}{2} \sum_{i=0}^{n} \frac{1}{l_i} (y_{i+1} - y_i)^2, \quad y_0 = y_{n+1} = 0.$$

But if the masses of all the beads are equal and they divide the thread into equal segments, that is, $m_i = m$ and $l_i = l/(n+1)$, $i = 1, \dots, n$, then all the formulas can be written in a more explicit form. In this case, we are speaking about the simultaneous representation as the sum of squares of two forms:

$$T = \frac{m}{2} \sum_{i=1}^{n} \dot{y}_i^2, \qquad \varPi = \frac{\sigma(n+1)}{l} \left(\sum_{i=1}^{n} y_i^2 - \sum_{i=0}^{n} y_i y_{i+1} \right), \quad y_0 = y_{n+1} = 0.$$

Therefore, we must use an orthogonal transformation (preserving the form $\sum_{i=1}^{n} y_i^2$) to express as a sum of squares the form $\sum_{i=0}^{n} y_i y_{i+1}$ with matrix

$$A = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \ddots & 0 & 0 \\ 0 & 1 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

It would have been possible to take the standard route: find the eigenvalues $\lambda_1, \dots, \lambda_n$ as roots of the determinant $|A - tE|$ and eigenvectors $\boldsymbol{y}$ from the system of equations

$$A\boldsymbol{y} = \lambda \boldsymbol{y}, \tag{7.61}$$

where $\lambda = \lambda_i$ and $\boldsymbol{y}$ is the column of unknowns $y_1, \dots, y_n$. But it is simpler to use equations (7.61) directly. They give a system of n equations in the unknowns $y_1, \dots, y_n$:

$$y_2 = 2\lambda y_1, \qquad y_1 + y_3 = 2\lambda y_2, \qquad \dots,$$
$$y_{n-2} + y_n = 2\lambda y_{n-1}, \qquad y_{n-1} = 2\lambda y_n,$$

which can be written in the form

$$y_{k-1} + y_{k+1} = 2\lambda y_k, \quad k = 1, \dots, n, \tag{7.62}$$

where we set $y_0 = y_{n+1} = 0$. The system of equations (7.62) is called a *recurrence relation*, whereby each value y_{k+1} is expressed in terms of the two preceding values: y_k and y_{k-1}. Thus if we know two adjacent values, then we can use relationship

(7.62) to construct all the y_k. The condition $y_0 = y_{n+1} = 0$ is called a *boundary* condition.

Let us note that for $\lambda = \pm 1$, the equation (7.62) with boundary condition $y_0 = y_{n+1} = 0$ has only the null solution: $y_0 = \cdots = y_{n+1} = 0$. Indeed, for $\lambda = 1$, we obtain

$$y_2 = 2y_1, \qquad y_3 = 3y_1, \qquad \ldots, \qquad y_n = ny_1, \qquad y_{n+1} = (n+1)y_1,$$

from which by $y_{n+1} = 0$ it follows that $y_1 = 0$, and all y_k are equal to 0. Similarly, for $\lambda = -1$, we obtain

$$y_2 = -2y_1, \qquad y_3 = 3y_1, \qquad y_4 = -4y_1, \qquad \ldots,$$
$$y_n = (-1)^{n-1} n y_1, \qquad y_{n+1} = (-1)^n (n+1) y_1,$$

from which by $y_{n+1} = 0$ it follows as well that $y_1 = 0$, and again all the y_k are equal to zero. Thus for $\lambda = \pm 1$, the system of equations (7.61) has as its only solution the vector $\boldsymbol{y} = \boldsymbol{0}$, which by definition, cannot be an eigenvector. In other words, this implies that the numbers ± 1 are not eigenvalues of the matrix A.

There is a lovely formula for solving equation (7.62) with boundary condition $y_0 = y_{n+1} = 0$. Let us denote by α and β the roots of the quadratic equation $z^2 - 2\lambda z + 1 = 0$. By the above reasoning, $\lambda \neq \pm 1$, and therefore, the numbers α and β are distinct and cannot equal ± 1. Direct substitution shows that then for arbitrary A and B, the sequence $y_k = A\alpha^k + B\beta^k$ satisfies the relationship (7.62). The coefficients A and B taken to satisfy $y_0 = 0$, y_1 are given. The following y_k, as we have seen, are determined by the relationship (7.62), and this implies that again they are given by our formula. The conditions $y_0 = 0$, y_1 fixed give $B = -A$ and $A(\alpha - \beta) = y_1$, whence $A = y_1/(\alpha - \beta)$. Thus we obtain the expression

$$y_k = \frac{y_1}{\alpha - \beta}\big(\alpha^k - \beta^k\big). \tag{7.63}$$

We now use the condition $y_{n+1} = 0$, which gives $\alpha^{n+1} = \beta^{n+1}$. Moreover, since α and β are roots of the polynomial $z^2 - 2\lambda z + 1$, we have $\alpha\beta = 1$, whence $\beta = \alpha^{-1}$, which implies that $\alpha^{2(n+1)} = 1$. From this (taking into account that $\alpha \neq \pm 1$), we obtain

$$\alpha = \cos\left(\frac{\pi j}{n+1}\right) + i \sin\left(\frac{\pi j}{n+1}\right),$$

where i is the imaginary unit, and the number j assumes the values $1, \ldots, n$. Again using the equation $z^2 - 2\lambda z + 1 = 0$, whose roots are α and β, we obtain n distinct values for λ:

$$\lambda_j = \cos\left(\frac{\pi j}{n+1}\right), \quad j = 1, \ldots, n,$$

since $j = n+2, \ldots, 2n+1$ give the same values λ_j. These are precisely the eigenvalues of the matrix A. For the eigenvector $\boldsymbol{y}_j$ of the associated eigenvalue λ_j, we obtain by formula (7.63) its coordinates $y_{1j}, \dots, y_{nj}$ in the form

$$y_{kj} = \sin\left(\frac{\pi kj}{n+1}\right), \quad k = 1, \dots, n.$$

These formulas were derived by d'Alembert and Daniel Bernoulli. Passing to the limit as $n \to \infty$, Lagrange derived from these the law of vibrations of a uniform string.

Example 7.45 Let us consider in an n-dimensional real Euclidean space L the subset X given by the equation

$$F(x_1, \dots, x_n) = 0 \tag{7.64}$$

in some coordinate system. Such a subset X is called a *hypersurface* and consists of all vectors $\boldsymbol{x} = (x_1, \dots, x_n)$ of the Euclidean space L whose coordinates satisfy the equation[5] (7.64). Using the change-of-coordinates formula (3.36), we see that the property of the subset $X \subset \mathsf{L}$ being a hypersurface does not depend on the choice of coordinates, that is, on the choice of the basis of L. Then if we assume that the beginning of every vector is located at a single fixed point, then every vector $\boldsymbol{x} = (x_1, \dots, x_n)$ can be identified with its endpoint, a point of the given space. In order to conform to more customary terminology, as we continue with this example, we shall call the vectors $\boldsymbol{x}$ of which the hypersurface X consists its *points*.

We shall assume that $F(\mathbf{0}) = 0$ and that the function $F(x_1, \dots, x_n)$ is differentiable in each of its arguments as many times as necessary. It is easily verified that this condition also does not depend on the choice of basis. Let us assume in addition that $\mathbf{0}$ is not a critical point of the hypersurface X, that is, that not all partial derivatives $\partial F(\mathbf{0})/\partial x_i$ are equal to zero. In other words, if we introduce the vector $\operatorname{grad} F = (\partial F/\partial x_1, \dots, \partial F/\partial x_n)$, called the *gradient* of the function F, then this implies that $\operatorname{grad} F(\mathbf{0}) \neq \mathbf{0}$.

We shall be interested in *local* properties of the hypersurface X, that is, properties associated with points close to $\mathbf{0}$. With the assumptions that we have made, the *implicit function theorem*, known from analysis, shows that near $\mathbf{0}$, the coordinates $x_1, \dots, x_n$ of each point of the hypersurface X can be represented as a function of $n-1$ arguments $u_1, \dots, u_{n-1}$, and furthermore, for each point, the values $u_1, \dots, u_{n-1}$ are uniquely determined. It is possible to choose as $u_1, \dots, u_{n-1}$ some $n-1$ of the coordinates $x_1, \dots, x_n$, after determining the remaining coordinate x_k from equation (7.64), for which must be satisfied only the condition $\frac{\partial F}{\partial x_k}(\mathbf{0}) \neq 0$ for the given k, which holds because of the assumption $\operatorname{grad} F(\mathbf{0}) \neq \mathbf{0}$. The functions that determine the dependence of the coordinates $x_1, \dots, x_n$ of a point of the hyperplane X on the arguments $u_1, \dots, u_{n-1}$ are differentiable at all arguments as many times as the original function $F(x_1, \dots, x_n)$.

[5]The more customary point of view, when the hypersurface (for example, a curve or surface) consists of *points*, requires the consideration of an n-dimensional space consisting of *points* (otherwise *affine* space), which will be introduced in the following chapter.

The *hyperplane* defined by the equation

$$\sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(\mathbf{0}) x_i = 0$$

is called the *tangent space* or *tangent hyperplane* to the hypersurface X at the point $\mathbf{0}$ and is denoted by $T_{\mathbf{0}}X$. In the case that the basis of the Euclidean space L is orthonormal, this equation can also be written in the form $(\operatorname{grad} F(\mathbf{0}), \boldsymbol{x}) = 0$. As a subspace of the Euclidean space L, the tangent space $T_{\mathbf{0}}X$ is also a Euclidean space.

The set of vectors depending on the parameter t taking values on some interval of the real line, that is, $\boldsymbol{x}(t) = (x_1(t), \ldots, x_n(t))$, is called a *smooth curve* if all functions $x_i(t)$ are differentiable a sufficient number of times and if for every value of the parameter t, not all the derivatives dx_i/dt are equal to zero. In analogy to what was said above about hypersurfaces, we may visualize the curve as consisting of *points* $A(t)$, where each $A(t)$ is the endpoint of some vector $\boldsymbol{x}(t)$, while all the vectors $\boldsymbol{x}(t)$ begin at a certain fixed point O. In what follows, we shall refer to the vectors $\boldsymbol{x}$ that constitute the curve as its *points*.

We say that a curve γ *passes through* the point $\boldsymbol{x}_0$ if $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ for some value of the parameter t_0. It is clear that here we may always assume that $t_0 = 0$. Indeed, let us consider a different curve $\widetilde{\boldsymbol{x}}(t) = (\widetilde{x}_1(t), \ldots, \widetilde{x}_n(t))$, where the functions $\widetilde{x}_i(t)$ are equal to $x_i(t + t_0)$. This can also be written in the form $\widetilde{\boldsymbol{x}}(\tau) = \boldsymbol{x}(t)$, where we have introduced a new parameter τ related to the old one by $\tau = t - t_0$.

Generally speaking, for a curve we may make an arbitrary *change of parameter* by the formula $t = \psi(\tau)$, where the function ψ defines a continuously differentiable bijective mapping of one interval to another. Under such a change, a curve, considered as a set of points (or vectors), will remain the same. From this it follows that one and the same curve can be written in a variety of ways using various parameters.[6]

We now introduce the vector $\frac{d\boldsymbol{x}}{dt} = (\frac{dx_1}{dt}, \ldots, \frac{dx_n}{dt})$. Suppose the curve γ passes through the point $\mathbf{0}$ for $t = 0$. Then the vector $\boldsymbol{p} = \frac{d\boldsymbol{x}}{dt}(0)$ is called a *tangent vector* to the curve γ at the point $\mathbf{0}$. It depends, of course, on the choice of parameter t defining the curve. Under a change of parameter $t = \psi(\tau)$, we have

$$\frac{d\boldsymbol{x}}{d\tau} = \frac{d\boldsymbol{x}}{dt} \cdot \frac{dt}{d\tau} = \frac{d\boldsymbol{x}}{dt} \cdot \psi'(\tau), \tag{7.65}$$

and the tangent vector $\boldsymbol{p}$ is multiplied by a constant equal to the value of the derivative $\psi'(0)$. Using this fact, it is possible to arrange things so that $|\frac{d\boldsymbol{x}}{dt}(t)| = 1$ for all t close to 0. Such a parameter is said to be *natural*. The condition that the curve $\boldsymbol{x}(t)$ belong to the hyperplane (7.64) gives the equality $F(\boldsymbol{x}(t)) = 0$, which is satisfied for all t. Differentiating this relationship with respect to t, we obtain that the vector $\boldsymbol{p}$ lies in the space $T_{\mathbf{0}}X$. And conversely, an arbitrary vector contained in $T_{\mathbf{0}}X$ can

[6]For example, the circle of radius 1 with center at the origin with Cartesian coordinates x, y can be defined not only by the formula $x = \cos t$, $y = \sin t$, but also by the formula $x = \cos\tau$, $y = -\sin\tau$ (with the replacement $t = -\tau$), or by the formula $x = \sin\tau$, $y = \cos\tau$ (replacement $t = \frac{\pi}{2} - \tau$).

be represented in the form $\frac{d\boldsymbol{x}}{dt}(0)$ for some curve $\boldsymbol{x}(t)$. This curve, of course, is not uniquely determined. Curves whose tangent vectors $\boldsymbol{p}$ are proportional are said to be *tangent* at the point $\mathbf{0}$.

Let us denote by $\boldsymbol{n}$ a unit vector orthogonal to the tangent space $T_{\mathbf{0}}X$. There are two such vectors, $\boldsymbol{n}$ and $-\boldsymbol{n}$, and we shall choose one of them. For example, we may set

$$\boldsymbol{n} = \frac{\operatorname{grad} F}{|\operatorname{grad} F|}(\mathbf{0}). \tag{7.66}$$

We define the vector $\frac{d^2\boldsymbol{x}}{dt^2}$ as $\frac{d}{dt}(\frac{d\boldsymbol{x}}{dt})$ and set

$$Q = \left(\frac{d^2\boldsymbol{x}}{dt^2}(0), \boldsymbol{n}\right). \tag{7.67}$$

Proposition 7.46 *The value Q depends only on the vector $\boldsymbol{p}$; namely, it is a quadratic form in its coordinates.*

Proof It suffices to verify this assertion by substituting in (7.67) for the vector $\boldsymbol{n}$, any vector proportional to it, for example, $\operatorname{grad} F(\mathbf{0})$. Since by assumption, the curve $\boldsymbol{x}(t)$ is contained in the hyperplane (7.64), it follows that $F(x_1(t), \dots, x_n(t)) = 0$. Differentiating this equality twice with respect to t, we obtain

$$\sum_{i=1}^{n} \frac{\partial F}{\partial x_i}\frac{dx_i}{dt} = 0, \qquad \sum_{i,j=1}^{n} \frac{\partial^2 F}{\partial x_i\, \partial x_j}\frac{dx_i}{dt}\frac{dx_j}{dt} + \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}\frac{d^2 x_i}{dt^2} = 0.$$

Setting here $t = 0$, we see that

$$\left(\frac{d^2\boldsymbol{x}}{dt^2}(0), \operatorname{grad} F(\mathbf{0})\right) = -\sum_{i,j=1}^{n} \frac{\partial^2 F}{\partial x_i\, \partial x_j}(\mathbf{0}) p_i p_j,$$

where $\boldsymbol{p} = (p_1, \dots, p_n)$. This proves the assertion. □

The form $Q(\boldsymbol{p})$ is called the *second quadratic form* of the hypersurface. The form $(\boldsymbol{p}^2)$ is called the *first quadratic form* when $T_{\mathbf{0}}X$ is taken as a subspace of a Euclidean space L. We observe that the second quadratic form requires the selection of one of two unit vectors ($\boldsymbol{n}$ or $-\boldsymbol{n}$) orthogonal to $T_{\mathbf{0}}X$. This is frequently interpreted as the selection of *one side* of the hypersurface in a neighborhood of the point $\mathbf{0}$.

The first and second quadratic forms give us the possibility to obtain an expression for the curvature of certain curves $\boldsymbol{x}(t)$ lying in the hypersurface X. Let us suppose that a curve is the intersection of a plane L' containing the point $\mathbf{0}$ and the hypersurface X (even if only in an arbitrarily small neighborhood of the point $\mathbf{0}$). Such a curve is called a *plane section* of the hypersurface. If we define the curve $\boldsymbol{x}(t)$ in such a way that t is a natural parameter, then its *curvature* at the point $\mathbf{0}$ is the number

$$k = \left| \frac{d^2\boldsymbol{x}}{dt^2}(0) \right|.$$

We assume that $k \neq 0$ and set

$$\boldsymbol{m} = \frac{1}{k} \cdot \frac{d^2\boldsymbol{x}}{dt^2}(0).$$

The vector $\boldsymbol{m}$ has length 1 by definition. It is said to be *normal* to the curve $\boldsymbol{x}(t)$ at the point $\mathbf{0}$. If the curve $\boldsymbol{x}(t)$ is a plane section of the hypersurface, then $\boldsymbol{x}(t)$ lies in the plane L' (for all sufficiently small t), and consequently, the vector

$$\frac{d\boldsymbol{x}}{dt} = \lim_{h \to 0} \frac{\boldsymbol{x}(t+h) - \boldsymbol{x}(t)}{h}$$

also lies in the plane L'. Therefore, this holds as well for the vector $d^2\boldsymbol{x}/dt^2$, which implies that it holds as well for the normal $\boldsymbol{m}$. If the curve γ is defined in terms of the natural parameter t, then

$$\left| \frac{d\boldsymbol{x}}{dt} \right|^2 = \left(\frac{d\boldsymbol{x}}{dt}, \frac{d\boldsymbol{x}}{dt} \right) = 1.$$

Differentiating this equality with respect to t, we obtain that the vectors $d^2\boldsymbol{x}/dt^2$ and $d\boldsymbol{x}/dt$ are orthogonal. Hence the normal $\boldsymbol{m}$ to the curve γ is orthogonal to an arbitrary tangent vector (for arbitrary definition of the curve γ in the form $\boldsymbol{x}(t)$ with natural parameter t), and the vector $\boldsymbol{m}$ is defined uniquely up to sign. It is obvious that $\mathsf{L}' = \langle \boldsymbol{m}, \boldsymbol{p} \rangle$, where $\boldsymbol{p}$ is an arbitrary tangent vector.

By definition (7.67) of the second quadratic form Q and taking into account the equality $|\boldsymbol{m}| = |\boldsymbol{n}| = 1$, we obtain the expression

$$Q(\boldsymbol{p}) = (k\boldsymbol{m}, \boldsymbol{n}) = k(\boldsymbol{m}, \boldsymbol{n}) = k \cos\varphi, \tag{7.68}$$

where φ is the angle between the vectors $\boldsymbol{m}$ and $\boldsymbol{n}$. The expression $k \cos\varphi$ is denoted by $\widetilde{k}$ and is called the *normal curvature* of the hypersurface X *in the direction* $\boldsymbol{p}$. We recall that here $\boldsymbol{n}$ denotes the chosen unit vector orthogonal to the tangent space $T_{\mathbf{0}}X$, and $\boldsymbol{m}$ is the normal to the curve to which the vector $\boldsymbol{p}$ is tangent. An analogous formula for an arbitrary parametric definition of the curve $\boldsymbol{x}(t)$ (where t is not necessarily a natural parameter) also uses the first quadratic form. Namely, if τ is another parameter, while t is a natural parameter, then by formula (7.65), now instead of the vector $\boldsymbol{p}$, we obtain $\boldsymbol{p}' = \boldsymbol{p}\psi'(0)$. Since Q is a quadratic form, it follows that $Q(\boldsymbol{p}\psi'(0)) = \psi'(0)^2 Q(\boldsymbol{p})$, and instead of formula (7.68), we now obtain

$$\frac{Q(\boldsymbol{p})}{(\boldsymbol{p}^2)} = k \cos\varphi. \tag{7.69}$$

Here the first quadratic form $(\boldsymbol{p}^2)$ is already involved as well as the second quadratic form $Q(\boldsymbol{p})$, but now (7.69), in contrast to (7.68), holds for an arbitrary choice of parameter t on the curve γ.

The point of the term *normal curvature* given above is the following. The section of the hypersurface X by the plane L' is said to be *normal* if $\boldsymbol{n} \in \mathsf{L}'$. The vector $\boldsymbol{n}$ defined by formula (7.66) is orthogonal to the tangent plane $T_{\mathbf{0}}X$. But in the plane L' there is also the vector $\boldsymbol{p}$ tangent to the curve γ, and the normal vector $\boldsymbol{m}$ orthogonal to it. Thus in the case of a normal section $\boldsymbol{n} = \pm\boldsymbol{m}$, this means that in formula (7.68), the angle φ is equal to 0 or π. Conversely, from the equality $|\cos\varphi| = 1$, it follows that $\boldsymbol{n} \in \mathsf{L}'$. Thus in the case of a normal section, the normal curvature $\widetilde{k}$ differs from k only by the factor ± 1 and is defined by the relationship

$$\widetilde{k} = \frac{Q(\boldsymbol{p})}{|\boldsymbol{p}|^2}.$$

Since $\mathsf{L}' = \langle \boldsymbol{m}, \boldsymbol{p} \rangle$, it follows that all normal sections correspond to straight lines in the plane L'. For each line, there exists a unique normal section containing this line. In other words, we "rotate" the plane L' about the vector $\boldsymbol{m}$, considering all obtained planes $\langle \boldsymbol{m}, \boldsymbol{p} \rangle$, where $\boldsymbol{p}$ is a vector in the tangent hyperplane $T_{\mathbf{0}}X$. Thus all normal sections of the hypersurface X are obtained.

We shall now employ Theorem 7.38. In our case, it gives an orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{n-1}$ in the tangent hyperplane $T_{\mathbf{0}}X$ (viewed as a subspace of the Euclidean space L) in which the quadratic form $Q(\boldsymbol{p})$ is brought into canonical form. In other words, for the vector $\boldsymbol{p} = u_1\boldsymbol{e}_1 + \cdots + u_{n-1}\boldsymbol{e}_{n-1}$, the second quadratic form takes the form $Q(\boldsymbol{p}) = \lambda_1 u_1^2 + \cdots + \lambda_{n-1}u_{n-1}^2$. Since the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{n-1}$ is orthonormal, we have in this case

$$\frac{u_i}{|p_i|} = \frac{(\boldsymbol{p}, \boldsymbol{e}_i)}{|p_i|} = \cos\varphi_i, \tag{7.70}$$

where φ_i is the angle between the vectors $\boldsymbol{p}$ and $\boldsymbol{e}_i$. From this we obtain for the normal curvature $\widetilde{k}$ of the normal section γ, the formula

$$\widetilde{k} = \frac{Q(\boldsymbol{p})}{|\boldsymbol{p}|^2} = \sum_{i=1}^{n-1} \lambda_i \left(\frac{u_i}{|\boldsymbol{p}|} \right)^2 = \sum_{i=1}^{n-1} \lambda_i \cos^2\varphi_i, \tag{7.71}$$

where $\boldsymbol{p}$ is an arbitrary tangent vector to the curve γ at the point $\mathbf{0}$. Relationships (7.70) and (7.71) are called *Euler's formula*. The numbers λ_i are called *principal curvatures* of the hypersurface X at the point $\mathbf{0}$.

In the case $n = 3$, the hypersurface (7.64) is an ordinary surface and has two principal curvatures λ_1 and λ_2. Taking into account the fact that $\cos^2\varphi_1 + \cos^2\varphi_2 = 1$, Euler's formula takes the form

$$\widetilde{k} = \lambda_1 \cos^2\varphi_1 + \lambda_2 \cos^2\varphi_2 = (\lambda_1 - \lambda_2)\cos^2\varphi_1 + \lambda_2. \tag{7.72}$$

Suppose $\lambda_1 \geq \lambda_2$. Then from (7.72), it is clear that the normal curvature $\widetilde{k}$ assumes a maximum (equal to λ_1) for $\cos^2\varphi_1 = 1$ and a minimum (equal to λ_2) for

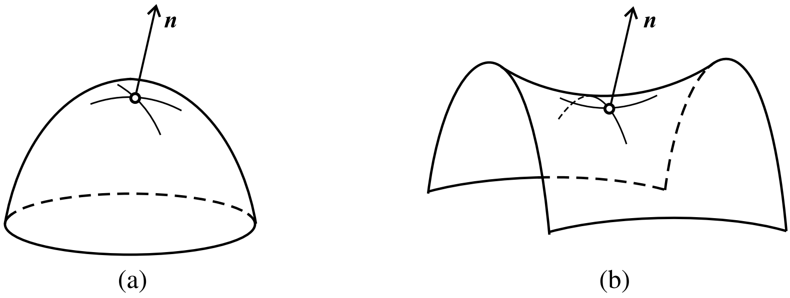


Fig. 7.10 Elliptic (**a**) and hyperbolic (**b**) points

$\cos^2 \varphi_1 = 0$. This assertion is called the *extremal property* of the principal curvatures of the surface. If λ_1 and λ_2 have the same sign ($\lambda_1\lambda_2 > 0$), then as can be seen from (7.72), an arbitrary normal section of a surface at a given point **0** has its curvature of the same sign, and therefore, all normal sections have convexity in the same direction, and near the point **0**, the surface lies *on one side* of its tangent plane; see Fig. 7.10(a). Such points are called *elliptic*. If λ_1 and λ_2 have different signs ($\lambda_1\lambda_2 < 0$), then as can be seen from formula (7.72), there exist normal sections with opposite directions of convexity, and at points near **0**, the surface is located *on different sides* of its tangent plane; see Fig. 7.10(b). Such points are called *hyperbolic*.[7]

From all this discussion, it is evident that the product of principal curvatures $\kappa = \lambda_1\lambda_2$ characterizes some important properties of a surface (called "internal geometric properties" of the surface). This product is called the *Gaussian* or *total curvature* of the surface.

7.7 Pseudo-Euclidean Spaces

Many of the theorems proved in the previous sections of this chapter remain valid if in the definition of Euclidean space we forgo the requirement of positive definiteness of the quadratic form $(\boldsymbol{x}^2)$ or replace it with something weaker. Without this condition, the inner product $(\boldsymbol{x}, \boldsymbol{y})$ does not differ at all from an arbitrary symmetric bilinear form. As Theorem 6.6 shows, it is uniquely defined by the quadratic form $(\boldsymbol{x}^2)$.

We thus obtain a theory that fully coincides with the theory of quadratic forms that we presented in Chap. 6. The fundamental theorem (on bringing a quadratic form into canonical form) consists in the existence of an orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, that is, a basis for which $(\boldsymbol{e}_i, \boldsymbol{e}_j) = 0$ for all $i \neq j$. Then for the vector $x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n$, the quadratic form $(\boldsymbol{x}^2)$ is equal to $\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$.

[7]Examples of surfaces consisting entirely of elliptic points are ellipsoids, hyperboloids of two sheets, and elliptic paraboloids, while surfaces consisting entirely of hyperbolic points include hyperboloids of one sheet and hyperbolic paraboloids.

Moreover, this is true for vector spaces and bilinear forms over an arbitrary field $\mathbb{K}$ of characteristic different from 2. The concept of an isomorphism of spaces makes sense also in this case; as previously, it is necessary to require that the scalar product $(\boldsymbol{x}, \boldsymbol{y})$ be preserved.

The theory of such spaces (defined up to isomorphism) with a bilinear or quadratic form is of great interest (for example, in the case $\mathbb{K} = \mathbb{Q}$, the field of rational numbers). But here we are interested in real spaces. In this case, formula (6.28) and Theorem 6.17 (law of inertia) show that up to isomorphism, a space is uniquely defined by its rank and the index of inertia of the associated quadratic form.

We shall further restrict attention to an examination of real vector spaces with a nonsingular symmetric bilinear form $(\boldsymbol{x}, \boldsymbol{y})$. Let us recall that the nonsingularity of a bilinear form implies that its rank (that is, the rank of its matrix in an arbitrary basis of the space) is equal to $\dim \mathsf{L}$. In other words, this means that its radical is equal to $(\mathbf{0})$; that is, if the vector $\boldsymbol{x}$ is such that $(\boldsymbol{x}, \boldsymbol{y}) = 0$ for all vectors $\boldsymbol{y} \in \mathsf{L}$, then $\boldsymbol{x} = \mathbf{0}$ (see Sect. 6.2). For a Euclidean space, this condition follows automatically from property (4) of the definition (it suffices to set there $\boldsymbol{y} = \boldsymbol{x}$).

Formula (6.28) shows that with these conditions, there exists a basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ of the space L for which

$$(\boldsymbol{e}_i, \boldsymbol{e}_j) = 0 \quad \text{for } i \neq j, \qquad \left(\boldsymbol{e}_i^2\right) = \pm 1.$$

Such a basis is called, as it was previously, *orthonormal*. In it, the form $(\boldsymbol{x}^2)$ can be written in the form

$$\left(\boldsymbol{x}^2\right) = x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_n^2,$$

and the number s is called the *index of inertia* of both the quadratic form $(\boldsymbol{x}^2)$ and the pseudo-Euclidean space L.

A new difficulty appears that was not present for Euclidean spaces if the quadratic form $(\boldsymbol{x}^2)$ is neither positive nor negative definite, that is, if its index of inertia s is positive but less than n. In this case, the restriction of the bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ to the subspace $\mathsf{L}' \subset \mathsf{L}$ can turn out to be singular, even if the original bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ in L was nonsingular. For example, it is clear that in L, there exists a vector $\boldsymbol{x} \neq \mathbf{0}$ for which $(\boldsymbol{x}^2) = 0$, and then the restriction of $(\boldsymbol{x}, \boldsymbol{y})$ to a one-dimensional subspace $\langle \boldsymbol{x} \rangle$ is singular (identically equal to zero).

Thus let us consider a vector space L with a nonsingular symmetric bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ defined on it. In this case, we shall use many concepts and much of the notation used for Euclidean spaces earlier. Hence, vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are called *orthogonal* if $(\boldsymbol{x}, \boldsymbol{y}) = 0$. Subspaces L_1 and L_2 are called *orthogonal* if $(\boldsymbol{x}, \boldsymbol{y}) = 0$ for all vectors $\boldsymbol{x} \in \mathsf{L}_1$ and $\boldsymbol{y} \in \mathsf{L}_2$, and we express this by writing $\mathsf{L}_1 \perp \mathsf{L}_2$. The orthogonal complement of the subspace $\mathsf{L}' \subset \mathsf{L}$ with respect to the bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ is denoted by $(\mathsf{L}')^{\perp}$. However, there is an important difference from the case of Euclidean spaces, in connection with which it will be useful to give the following definition.

Definition 7.47 A subspace $\mathsf{L}' \subset \mathsf{L}$ is said to be *nondegenerate* if the bilinear form obtained by restricting the form $(\boldsymbol{x}, \boldsymbol{y})$ to L' is nonsingular. In the contrary case, L' is said to be *degenerate*.

By Theorem 6.9, in the case of a nondegenerate subspace L' we have the orthogonal decomposition

$$\mathsf{L} = \mathsf{L}' \oplus \left(\mathsf{L}'\right)^{\perp}. \tag{7.73}$$

In the case of a Euclidean space, as we have seen, every subspace L' is nondegenerate, and the decomposition (7.73) holds without any additional conditions. As the following example will show, in a pseudo-Euclidean space, the condition of nondegeneracy of a subspace L' for the decomposition (7.73) is in fact essential.

Example 7.48 Let us consider a three-dimensional space L with a symmetric bilinear form defined in some chosen basis by the formula

$$(\boldsymbol{x}, \boldsymbol{y}) = x_1 y_1 + x_2 y_2 - x_3 y_3,$$

where the x_i are the coordinates of the vector $\boldsymbol{x}$, and the y_i are the coordinates of the vector $\boldsymbol{y}$. Let $\mathsf{L}' = \langle \boldsymbol{e} \rangle$, where the vector $\boldsymbol{e}$ has coordinates $(0, 1, 1)$. Then as is easily verified, $(\boldsymbol{e}, \boldsymbol{e}) = 0$, and therefore, the restriction of the form $(\boldsymbol{x}, \boldsymbol{y})$ to L' is identically equal to zero. This implies that the subspace L' is degenerate. Its orthogonal complement $(\mathsf{L}')^{\perp}$ is two-dimensional and consists of all vectors $\boldsymbol{z} \in \mathsf{L}$ with coordinates (z_1, z_2, z_3) for which $z_2 = z_3$. Consequently, $\mathsf{L}' \subset (\mathsf{L}')^{\perp}$, and the intersection $\mathsf{L}' \cap (\mathsf{L}')^{\perp} = \mathsf{L}'$ contains nonnull vectors. This implies that the sum $\mathsf{L}' + (\mathsf{L}')^{\perp}$ is not a direct sum. Furthermore, it is obvious that $\mathsf{L}' + (\mathsf{L}')^{\perp} \neq \mathsf{L}$.

It follows from the nonsingularity of a bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ that the determinant of its matrix (in an arbitrary basis) is different from zero. If this matrix is written in the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, then its determinant is equal to

$$\begin{vmatrix} (\boldsymbol{e}_1, \boldsymbol{e}_1) & (\boldsymbol{e}_1, \boldsymbol{e}_2) & \cdots & (\boldsymbol{e}_1, \boldsymbol{e}_n) \\ (\boldsymbol{e}_2, \boldsymbol{e}_1) & (\boldsymbol{e}_2, \boldsymbol{e}_2) & \cdots & (\boldsymbol{e}_2, \boldsymbol{e}_n) \\ \vdots & \vdots & \ddots & \vdots \\ (\boldsymbol{e}_n, \boldsymbol{e}_1) & (\boldsymbol{e}_n, \boldsymbol{e}_2) & \cdots & (\boldsymbol{e}_n, \boldsymbol{e}_n) \end{vmatrix}, \tag{7.74}$$

and just as in the case of a Euclidean space, we shall call this its *Gram determinant* of the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$. Of course, this determinant depends on the choice of basis, but its *sign* does not depend on the basis. Indeed, if A and A' are matrices of our bilinear form in two different bases, then they are related by the equality $A' = C^* A C$, where C is a nonsingular transition matrix, from which it follows that $|A'| = |A| \cdot |C|^2$. Thus the sign of the Gram determinant is the same for all bases.

As noted above, for a nondegenerate subspace $\mathsf{L}' \subset \mathsf{L}$, we have the decomposition (7.73), which yields the equality

$$\dim \mathsf{L} = \dim \mathsf{L}' + \dim\left(\mathsf{L}'\right)^{\perp}. \tag{7.75}$$

But equality (7.75) holds as well for every subspace $\mathsf{L}' \subset \mathsf{L}$, although as we saw in Example 7.48, the decomposition (7.73) may already not hold in the general case.

Indeed, by Theorem 6.3, we can write an arbitrary bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ in the space L in the form $(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \mathcal{A}(\boldsymbol{y}))$, where $\mathcal{A} : \mathsf{L} \to \mathsf{L}^*$ is some linear transformation. From the nonsingularity of the bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ follows the nonsingularity of the transformation $\mathcal{A}$. In other words, the transformation $\mathcal{A}$ is an isomorphism, that is, its kernel is equal to $(\mathbf{0})$, and in particular, for an arbitrary subspace $\mathsf{L}' \subset \mathsf{L}$, we have the equality $\dim \mathcal{A}(\mathsf{L}') = \dim \mathsf{L}'$. On the other hand, we can write the orthogonal complement $(\mathsf{L}')^{\perp}$ in the form $(\mathcal{A}(\mathsf{L}'))^a$, using the notion of the annihilator introduced in Sect. 3.7. On the basis of what we have said above and formula (3.54) for the annihilator, we have the relationship

$$\dim\bigl(\mathcal{A}\bigl(\mathsf{L}'\bigr)\bigr)^a = \dim \mathsf{L} - \dim \mathcal{A}\bigl(\mathsf{L}'\bigr) = \dim \mathsf{L} - \dim \mathsf{L}',$$

that is, $\dim(\mathsf{L}')^{\perp} = \dim \mathsf{L} - \dim \mathsf{L}'$. We note that this argument holds for vector spaces L defined not only over the real numbers, but over any field.

The spaces that we have examined are defined (up to isomorphism) by the index of inertia s, which can take values from 0 to n. By what we have said above, the sign of the Gram determinant of an arbitrary basis is equal to $(-1)^{n-s}$. It is obvious that if we replace the inner product $(\boldsymbol{x}, \boldsymbol{y})$ in the space L by $-(\boldsymbol{x}, \boldsymbol{y})$, we shall preserve all of its essential properties, but the index of inertia s will be replaced by $n - s$, whence in what follows, we shall assume that $n/2 \le s \le n$. The case $s = n$ corresponds to a Euclidean space. There exists, however, a phenomenon whose explanation is at present not completely clear; the most interesting questions in mathematics and physics were until now connected with two types of spaces: those in which the index of inertia s is equal to n and those for which $s = n - 1$. The theory of Euclidean spaces $(s = n)$ has been up till now the topic of this chapter. In the remaining part, we shall consider the other case: $s = n - 1$. In the sequel, we shall call such spaces *pseudo-Euclidean spaces* (although sometimes, this term is used when $(\boldsymbol{x}, \boldsymbol{y})$ is an arbitrary nonsingular symmetric bilinear form neither positive nor negative definite, that is, with index of inertia $s \ne 0, n$).

Thus a pseudo-Euclidean space of dimension n is a vector space L equipped with a symmetric bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ such that in some basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, the quadratic form $(\boldsymbol{x}^2)$ takes the form

$$x_1^2 + \cdots + x_{n-1}^2 - x_n^2. \tag{7.76}$$

As in the case of a Euclidean space, we shall, as we did previously, call such bases *orthonormal*.

The best-known application of pseudo-Euclidean spaces is related to the *special theory of relativity*. According to an idea put forward by Minkowski, in this theory, one considers a four-dimensional space whose vectors are called *space–time events* (we mentioned this earlier, on p. 86). They have coordinates (x, y, z, t), and the space is equipped with a quadratic form $x^2 + y^2 + z^2 - t^2$ (here the speed of light is assumed to be 1). The pseudo-Euclidean space thus obtained is called *Minkowski space*. By analogy with the physical sense of these concepts in Minkowski space, in an arbitrary pseudo-Euclidean space, a vector $\boldsymbol{x}$ is said to be *spacelike* if $(\boldsymbol{x}^2) > 0$,

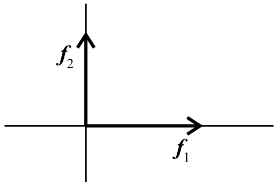


Fig. 7.11 A pseudo-Euclidean plane

while such a vector is said to be *timelike* if $(\boldsymbol{x}^2) < 0$, and *lightlike*, or *isotropic*, if $(\boldsymbol{x}^2) = 0$.[8]

Example 7.49 Let us consider the simplest case of a pseudo-Euclidean space L with $\dim \mathsf{L} = 2$ and index of inertia $s = 1$. By the general theory, in this space there exists an orthonormal basis, in this case the basis $\boldsymbol{e}_1, \boldsymbol{e}_2$, for which

$$\left(\boldsymbol{e}_1^2\right) = 1, \qquad \left(\boldsymbol{e}_2^2\right) = -1, \qquad (\boldsymbol{e}_1, \boldsymbol{e}_2) = 0, \tag{7.77}$$

and the scalar square of the vector $\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2$ is equal to $(\boldsymbol{x}^2) = x_1^2 - x_2^2$. However, it is easier to write the formulas connected with the space L in the basis consisting of lightlike vectors $\boldsymbol{f}_1, \boldsymbol{f}_2$, after setting

$$\boldsymbol{f}_1 = \frac{\boldsymbol{e}_1 + \boldsymbol{e}_2}{2}, \qquad \boldsymbol{f}_2 = \frac{\boldsymbol{e}_1 - \boldsymbol{e}_2}{2}. \tag{7.78}$$

Then $(\boldsymbol{f}_1^2) = (\boldsymbol{f}_2^2) = 0$, $(\boldsymbol{f}_1, \boldsymbol{f}_2) = \frac{1}{2}$, and the scalar square of the vector $\boldsymbol{x} = x_1\boldsymbol{f}_1 + x_2\boldsymbol{f}_2$ is equal to $(\boldsymbol{x}^2) = x_1x_2$. The lightlike vectors are located on the coordinate axes; see Fig. 7.11. The timelike vectors comprise the second and fourth quadrants, and the spacelike vectors make up the first and third quadrants.

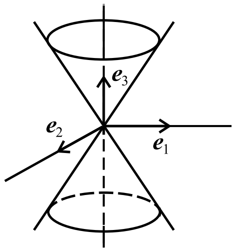
Definition 7.50 The set $V \subset \mathsf{L}$ consisting of all lightlike vectors of a pseudo-Euclidean space is called the *light cone* (or *isotropic cone*).

That we call the set V a *cone* suggests that if it contains some vector $\boldsymbol{e}$, then it contains the entire straight line $\langle \boldsymbol{e} \rangle$, which follows at once from the definition. The set of timelike vectors is called the *interior* of the cone V, while the set of spacelike vectors makes up its *exterior*. In the space from Example 7.49, the light cone V is the union of two straight lines $\langle \boldsymbol{f}_1 \rangle$ and $\langle \boldsymbol{f}_2 \rangle$. A more visual representation of the light cone is given by the following example.

---

[8] We remark that this terminology differs from what is generally used: Our "spacelike" vectors are usually called "timelike," and conversely. The difference is explained by the condition $s = n - 1$ that we have assumed. In the conventional definition of Minkowski space, one usually considers the quadratic form $-x^2 - y^2 - z^2 + t^2$, with index of inertia $s = 1$, and we need to multiply it by -1 in order that the condition $s \geq n/2$ be satisfied.

Fig. 7.12 The light cone



Example 7.51 We consider the pseudo-Euclidean space L with $\dim \mathsf{L} = 3$ and index of inertia $s = 2$. With the selection of an orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ such that

$$\big(\boldsymbol{e}_1^2\big) = \big(\boldsymbol{e}_2^2\big) = 1, \qquad \big(\boldsymbol{e}_3^2\big) = -1, \qquad (\boldsymbol{e}_i, \boldsymbol{e}_j) = 0 \quad \text{for all } i \neq j,$$

the light cone V is defined by the equation $x_1^2 + x_2^2 - x_3^2 = 0$. This is an ordinary right circular cone in three-dimensional space, familiar from a course in analytic geometry; see Fig. 7.12.

We now return to the general case of a pseudo-Euclidean space L of dimension n and consider the light cone V in L in greater detail. First of all, let us verify that it is "completely circular." By this we mean the following.

Lemma 7.52 *Although the cone V contains along with every vector $\boldsymbol{x}$ the entire line $\langle \boldsymbol{x} \rangle$, it contains no two-dimensional subspace.*

Proof Let us assume that V contains a two-dimensional subspace $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$. We choose a vector $\boldsymbol{e} \in \mathsf{L}$ such that $(\boldsymbol{e}^2) = -1$. Then the line $\langle \boldsymbol{e} \rangle$ is a nondegenerate subspace of L, and we can use the decomposition (7.73):

$$\mathsf{L} = \langle \boldsymbol{e} \rangle \oplus \langle \boldsymbol{e} \rangle^{\perp}. \tag{7.79}$$

From the law of inertia it follows that $\langle \boldsymbol{e} \rangle^{\perp}$ is a Euclidean space. Let us apply the decomposition (7.79) to our vectors $\boldsymbol{x}, \boldsymbol{y} \in V$. We obtain

$$\boldsymbol{x} = \alpha \boldsymbol{e} + \boldsymbol{u}, \qquad \boldsymbol{y} = \beta \boldsymbol{e} + \boldsymbol{v}, \tag{7.80}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in the Euclidean space $\langle \boldsymbol{e} \rangle^{\perp}$, while α and β are some scalars.

The conditions $(\boldsymbol{x}^2) = 0$ and $(\boldsymbol{y}^2) = 0$ can be written as $\alpha^2 = (\boldsymbol{u}^2)$ and $\beta^2 = (\boldsymbol{v}^2)$. Using the same reasoning for the vector $\boldsymbol{x} + \boldsymbol{y} = (\alpha + \beta)\boldsymbol{e} + \boldsymbol{u} + \boldsymbol{v}$, which by the assumption $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \subset V$ is also contained in V, we obtain the equality

$$(\alpha + \beta)^2 = (\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{v}) = \big(\boldsymbol{u}^2\big) + 2(\boldsymbol{u}, \boldsymbol{v}) + \big(\boldsymbol{v}^2\big) = \alpha^2 + 2(\boldsymbol{u}, \boldsymbol{v}) + \beta^2.$$

Canceling the terms α^2 and β^2 on the left- and right-hand sides of the equality, we obtain that $\alpha\beta = (\boldsymbol{u}, \boldsymbol{v})$, that is, $(\boldsymbol{u}, \boldsymbol{v})^2 = \alpha^2\beta^2 = (\boldsymbol{u}^2) \cdot (\boldsymbol{v}^2)$. Thus for the vectors

$\boldsymbol{u}$ and $\boldsymbol{v}$ in the Euclidean space $\langle \boldsymbol{e} \rangle^{\perp}$, the Cauchy–Schwarz inequality reduces to an equality, from which it follows that $\boldsymbol{u}$ and $\boldsymbol{v}$ are proportional (see p. 218). Let $\boldsymbol{v} = \lambda \boldsymbol{u}$. Then the vector $\boldsymbol{y} - \lambda \boldsymbol{x} = (\beta - \lambda\alpha)\boldsymbol{e}$ is also lightlike. Since $(\boldsymbol{e}^2) = -1$, it follows that $\beta = \lambda\alpha$. But then from the relationship (7.80), it follows that $\boldsymbol{y} = \lambda \boldsymbol{x}$, and this contradicts the assumption $\dim\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 2$. □

Let us select an arbitrary timelike vector $\boldsymbol{e} \in \mathsf{L}$. Then in the orthogonal complement $\langle \boldsymbol{e} \rangle^{\perp}$ of the line $\langle \boldsymbol{e} \rangle$, the bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ determines a positive definite quadratic form. This implies that $\langle \boldsymbol{e} \rangle^{\perp} \cap V = (\boldsymbol{0})$, and the hyperplane $\langle \boldsymbol{e} \rangle^{\perp}$ divides the set $V \setminus \boldsymbol{0}$ into two parts, V_+ and V_-, consisting of vectors $\boldsymbol{x} \in V$ such that in each part, the condition $(\boldsymbol{e}, \boldsymbol{x}) > 0$ or $(\boldsymbol{e}, \boldsymbol{x}) < 0$ is respectively satisfied. We shall call these sets V_+ and V_- *poles* of the light cone V. In Fig. 7.12, the plane $\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle$ divides V into "upper" and "lower" poles V_+ and V_- for the vector $\boldsymbol{e} = \boldsymbol{e}_3$.

The partition $V \setminus \boldsymbol{0} = V_+ \cup V_-$ that we have constructed rested on the choice of some timelike vector $\boldsymbol{e}$, and ostensibly, it must depend on it (for example, a change in the vector $\boldsymbol{e}$ to $-\boldsymbol{e}$ interchanges the poles V_+ and V_-). We shall now show that the decomposition $V \setminus \boldsymbol{0} = V_+ \cup V_-$, without taking into account how we designate each pole, does not depend on the choice of vector $\boldsymbol{e}$, that is, it is a property of the pseudo-Euclidean space itself. To do so, we shall require the following, almost obvious, assertion.

Lemma 7.53 *Let* L' *be a subspace of the pseudo-Euclidean space* L *of dimension* $\dim \mathsf{L}' \geq 2$. *Then the following statements are equivalent*:

(1) L' *is a pseudo-Euclidean space.*
(2) L' *contains a timelike vector.*
(3) L' *contains two linearly independent lightlike vectors.*

Proof If L' is a pseudo-Euclidean space, then statements (2) and (3) obviously follow from the definition of a pseudo-Euclidean space.

Let us show that statement (2) implies statement (1). Suppose L' contains a timelike vector $\boldsymbol{e}$. That is, $(\boldsymbol{e}^2) < 0$, whence the subspace $\langle \boldsymbol{e} \rangle$ is nondegenerate, and therefore, we have the decomposition (7.79), and moreover, as follows from the law of inertia, the subspace $\langle \boldsymbol{e} \rangle^{\perp}$ is a Euclidean space. If the subspace L' were degenerate, then there would exist a nonnull vector $\boldsymbol{u} \in \mathsf{L}'$ such that $(\boldsymbol{u}, \boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathsf{L}'$, and in particular, for vectors $\boldsymbol{e}$ and $\boldsymbol{u}$. The condition $(\boldsymbol{u}, \boldsymbol{e}) = 0$ implies that the vector $\boldsymbol{u}$ is contained in $\langle \boldsymbol{e} \rangle^{\perp}$, while the condition $(\boldsymbol{u}, \boldsymbol{u}) = 0$ implies that the vector $\boldsymbol{u}$ is lightlike. But this is impossible, since the subspace $\langle \boldsymbol{e} \rangle^{\perp}$ is a Euclidean space and cannot contain lightlike vectors. This contradiction shows that the subspace L' is nondegenerate, and therefore, it exhibits the decomposition (7.73). Taking into account the law of inertia, it follows from this that the subspace L' is a pseudo-Euclidean space.

Let us show that statement (3) implies statement (1). Suppose the subspace L' contains linearly independent lightlike vectors $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$. We shall show that the plane $\Pi = \langle \boldsymbol{f}_1, \boldsymbol{f}_2 \rangle$ contains a timelike vector $\boldsymbol{e}$. Then obviously, $\boldsymbol{e}$ is contained

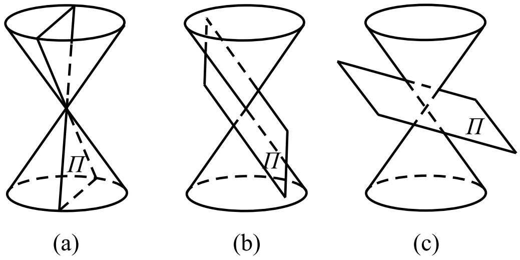


Fig. 7.13 The *plane* Π in a three-dimensional pseudo-Euclidean space

in L', and by what was proved above, the subspace L' is a pseudo-Euclidean space. Every vector $\boldsymbol{e} \in \Pi$ can be represented in the form $\boldsymbol{e} = \alpha \boldsymbol{f}_1 + \beta \boldsymbol{f}_2$. From this, we obtain $(\boldsymbol{e}^2) = 2\alpha\beta(\boldsymbol{f}_1, \boldsymbol{f}_2)$. We note that $(\boldsymbol{f}_1, \boldsymbol{f}_2) \neq 0$, since in the contrary case, for each vector $\boldsymbol{e} \in \Pi$, the equality $(\boldsymbol{e}^2) = 0$ would be satisfied, implying that the plane Π lies completely in the light cone V, which contradicts Lemma 7.52. Thus $(\boldsymbol{f}_1, \boldsymbol{f}_2) \neq 0$, and choosing coordinates α and β such that the sign of their product is opposite to the sign of $(\boldsymbol{f}_1, \boldsymbol{f}_2)$, we obtain the vector $\boldsymbol{e}$, for which $(\boldsymbol{e}^2) < 0$. □

Example 7.54 Let us consider the three-dimensional pseudo-Euclidean space L from Example 7.51 and a plane Π in L. The property of a plane Π being a Euclidean space, a pseudo-Euclidean space, or degenerate is clearly illustrated in Fig. 7.13.

In Fig. 7.13(a), the plane Π intersects the light cone V in two lines, corresponding to two linearly independent lightlike vectors. Clearly, this is equivalent to the condition that Π also intersects the interior of the light cone, which consists of timelike vectors, and therefore is a pseudo-Euclidean plane. In Fig. 7.13(c), it is shown that the plane Π intersects V only in its vertex, that is, $\Pi \cap V = (\mathbf{0})$. This implies that the plane Π is a Euclidean space, since every nonnull vector $\boldsymbol{e} \in \Pi$ lies outside the cone V, that is, $(\boldsymbol{e}^2) > 0$.

Finally, in Fig. 7.13(b) is shown the intermediate variant: the plane Π intersects the cone V in a single line, that is, it is tangent to it. Since the plane Π contains lightlike vectors (lying on this line), it follows that it cannot be a Euclidean space, and since it does not contain timelike vectors, it follows by Lemma 7.53 that it cannot be a pseudo-Euclidean space. This implies that Π is degenerate.

This is not difficult to verify in another way if we write down the matrix of the restriction of the inner product to the plane Π. Suppose that in the orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ from Example 7.49, this plane is defined by the equation $x_3 = \alpha x_1 + \beta x_2$. Then the vectors $\boldsymbol{g}_1 = \boldsymbol{e}_1 + \alpha \boldsymbol{e}_3$ and $\boldsymbol{g}_2 = \boldsymbol{e}_2 + \beta \boldsymbol{e}_3$ form a basis of Π in which the restriction of the inner product has matrix $\begin{pmatrix} 1-\alpha^2 & -\alpha\beta \\ -\alpha\beta & 1-\beta^2 \end{pmatrix}$ with determinant $\Delta = (1-\alpha^2)(1-\beta^2) - (\alpha\beta)^2$. On the other hand, the assumption of tangency of the plane Π and cone V amounts to the discriminant of the quadratic form $x_1^2 + x_2^2 - (\alpha x_1 + \beta x_2)^2$ in the variables x_1 and x_2 being equal to zero. It is easily verified that this discriminant is equal to $-\Delta$, and this implies that it is zero precisely when the determinant of this matrix is zero.

Theorem 7.55 *The partition of the light cone V into two poles V_+ and V_- does not depend on the choice of timelike vector $\boldsymbol{e}$. In particular, the linearly independent lightlike vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ lie in one pole if and only if $(\boldsymbol{x}, \boldsymbol{y}) < 0$.*

Proof Let us assume that for some choice of timelike vector $\boldsymbol{e}$, the lightlike vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ lie in one pole of the light cone V, and let us show that then, for any choice $\boldsymbol{e}$, they will always belong to the same pole. The case that the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are proportional, that is, $\boldsymbol{y} = \lambda \boldsymbol{x}$, is obvious. Indeed, since $\langle \boldsymbol{e} \rangle^\perp \cap V = (\boldsymbol{0})$, it follows that $(\boldsymbol{e}, \boldsymbol{x}) \neq 0$, and this implies that the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to one pole if and only if $\lambda > 0$, independent of the choice of the vector $\boldsymbol{e}$.

Now let us consider the case that $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly independent. Then $(\boldsymbol{x}, \boldsymbol{y}) \neq 0$, since otherwise, the entire plane $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ would be contained in the light cone V, which by Lemma 7.52, is impossible. Let us prove that regardless of what timelike vector $\boldsymbol{e}$ we have chosen for the partition $V \setminus \boldsymbol{0} = V_+ \cup V_-$, the vectors $\boldsymbol{x}, \boldsymbol{y} \in V \setminus \boldsymbol{0}$ belong to one pole if and only if $(\boldsymbol{x}, \boldsymbol{y}) < 0$. Let us note that this question, strictly speaking, relates not to the entire space L, but only to the subspace $\langle \boldsymbol{e}, \boldsymbol{x}, \boldsymbol{y} \rangle$, whose dimension, by the assumptions we have made, is equal to 2 or 3, depending on whether the vector $\boldsymbol{e}$ does or does not lie in the plane $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$.

Let us consider first the case $\dim \langle \boldsymbol{e}, \boldsymbol{x}, \boldsymbol{y} \rangle = 2$, that is, $\boldsymbol{e} \in \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. Then let us set $\boldsymbol{e} = \alpha \boldsymbol{x} + \beta \boldsymbol{y}$. Consequently, $(\boldsymbol{e}, \boldsymbol{x}) = \beta (\boldsymbol{x}, \boldsymbol{y})$ and $(\boldsymbol{e}, \boldsymbol{y}) = \alpha (\boldsymbol{x}, \boldsymbol{y})$, since $\boldsymbol{x}, \boldsymbol{y} \in V$. By definition, vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are in the same pole if and only if $(\boldsymbol{e}, \boldsymbol{x})(\boldsymbol{e}, \boldsymbol{y}) > 0$. But since $(\boldsymbol{e}, \boldsymbol{x})(\boldsymbol{e}, \boldsymbol{y}) = \alpha\beta (\boldsymbol{x}, \boldsymbol{y})^2$, this condition is equivalent to the inequality $\alpha\beta > 0$. The vector $\boldsymbol{e}$ is timelike, and therefore, $(\boldsymbol{e}^2) < 0$, and in view of the equality $(\boldsymbol{e}^2) = 2\alpha\beta(\boldsymbol{x}, \boldsymbol{y})$, we obtain that the condition $\alpha\beta > 0$ is equivalent to $(\boldsymbol{x}, \boldsymbol{y}) < 0$.

Let us now consider the case that $\dim \langle \boldsymbol{e}, \boldsymbol{x}, \boldsymbol{y} \rangle = 3$. The space $\langle \boldsymbol{e}, \boldsymbol{x}, \boldsymbol{y} \rangle$ contains the timelike vector $\boldsymbol{e}$. Consequently, by Lemma 7.53, it is a pseudo-Euclidean space, and its subspace $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is nondegenerate, since $(\boldsymbol{x}, \boldsymbol{y}) \neq 0$ and $(\boldsymbol{x}^2) = (\boldsymbol{y}^2) = 0$. Thus here the decomposition (7.73) takes the form

$$\langle \boldsymbol{e}, \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle \oplus \langle \boldsymbol{h} \rangle, \tag{7.81}$$

where the space $\langle \boldsymbol{h} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle^\perp$ is one-dimensional. On the left-hand side of the decomposition (7.81) stands a three-dimensional pseudo-Euclidean space, and the space $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is a two-dimensional pseudo-Euclidean space; therefore, by the law of inertia, the space $\langle \boldsymbol{h} \rangle$ is a Euclidean space. Thus for the vector $\boldsymbol{e}$, we have the representation

$$\boldsymbol{e} = \alpha \boldsymbol{x} + \beta \boldsymbol{y} + \gamma \boldsymbol{h}, \qquad (\boldsymbol{h}, \boldsymbol{x}) = 0, \qquad (\boldsymbol{h}, \boldsymbol{y}) = 0.$$

From this follows the equality

$$(\boldsymbol{e}, \boldsymbol{x}) = \beta (\boldsymbol{x}, \boldsymbol{y}), \qquad (\boldsymbol{e}, \boldsymbol{y}) = \alpha (\boldsymbol{x}, \boldsymbol{y}), \qquad \left(\boldsymbol{e}^2\right) = 2\alpha\beta(\boldsymbol{x}, \boldsymbol{y}) + \gamma^2 \left(\boldsymbol{h}^2\right).$$

Taking into account the fact that $(\boldsymbol{e}^2) < 0$ and $(\boldsymbol{h}^2) > 0$, from the last of these relationships, we obtain that $\alpha\beta(\boldsymbol{x}, \boldsymbol{y}) < 0$. The condition that the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ lie in one pole can be expressed as the inequality $(\boldsymbol{e}, \boldsymbol{x})(\boldsymbol{e}, \boldsymbol{y}) > 0$, that is, $\alpha\beta > 0$.

Since $\alpha\beta(\boldsymbol{x}, \boldsymbol{y}) < 0$, it follows as in the previous case that this is equivalent to the condition $(\boldsymbol{x}, \boldsymbol{y}) < 0$. $\square$

Remark 7.56 As we did in Sect. 3.2 in connection with the partition of a vector space L by a hyperplane L', it is possible to ascertain that the partition of the set $V \setminus \mathbf{0}$ coincides with its partition into two path-connected components V_+ and V_-. From this we can obtain another proof of Theorem 7.55 without using any formulas.

A pseudo-Euclidean space emerges in the following remarkable relationship.

Theorem 7.57 *For every pair of timelike vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, the reverse of the Cauchy–Schwarz inequality is satisfied*:

$$(\boldsymbol{x}, \boldsymbol{y})^2 \geq \left(\boldsymbol{x}^2\right) \cdot \left(\boldsymbol{y}^2\right), \tag{7.82}$$

which reduces to an equality if and only if $\boldsymbol{x}$ and $\boldsymbol{y}$ are proportional.

Proof Let us consider the subspace $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$, in which are contained all the vectors of interest to us. If the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are proportional, that is, $\boldsymbol{y} = \lambda \boldsymbol{x}$, where λ is some scalar, then the inequality (7.82) obviously reduces to a tautological equality. Thus we may assume that $\dim\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 2$, that is, we may suppose ourselves to be in the situation considered in Example 7.49.

As we have seen, in the space $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$, there exists a basis $\boldsymbol{f}_1, \boldsymbol{f}_2$ for which the relationship $(\boldsymbol{f}_1^2) = (\boldsymbol{f}_2^2) = 0$, $(\boldsymbol{f}_1, \boldsymbol{f}_2) = \frac{1}{2}$ holds. Writing the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in this basis, we obtain the expressions

$$\boldsymbol{x} = x_1 \boldsymbol{f}_1 + x_2 \boldsymbol{f}_2, \qquad \boldsymbol{y} = y_1 \boldsymbol{f}_1 + y_2 \boldsymbol{f}_2,$$

from which it follows that

$$\left(\boldsymbol{x}^2\right) = x_1 x_2, \qquad \left(\boldsymbol{y}^2\right) = y_1 y_2, \qquad (\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}(x_1 y_2 + x_2 y_1).$$

Substituting these expressions into (7.82), we see that we have to verify the inequality $(x_1 y_2 + x_2 y_1)^2 \geq 4 x_1 x_2 y_1 y_2$. Having carried out in the last inequality the obvious transformations, we see that this is equivalent to the inequality

$$(x_1 y_2 - x_2 y_1)^2 \geq 0, \tag{7.83}$$

which holds for all real values of the variables. Moreover, it is obvious that the inequality (7.83) reduces to an equality if and only if $x_1 y_2 - x_2 y_1 = 0$, that is, if and only if the determinant $\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}$ equals 0, and this implies that the vectors $\boldsymbol{x} = (x_1, x_2)$ and $\boldsymbol{y} = (y_1, y_2)$ are proportional. $\square$

From Theorem 7.57 we obtain the following useful corollary.

Corollary 7.58 *Two timelike vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot be orthogonal.*

Proof Indeed, if $(\boldsymbol{x}, \boldsymbol{y}) = 0$, then from the inequality (7.82), it follows that $(\boldsymbol{x}^2) \cdot (\boldsymbol{y}^2) \le 0$, and this contradicts the condition $(\boldsymbol{x}^2) < 0$ and $(\boldsymbol{y}^2) < 0$. □

Similar to the partition of the light cone V into two poles, we can also partition its interior into two parts. Namely, we shall say that timelike vectors $\boldsymbol{e}$ and $\boldsymbol{e}'$ *lie inside* one pole of the light cone V if the inner products $(\boldsymbol{e}, \boldsymbol{x})$ and $(\boldsymbol{e}', \boldsymbol{x})$ have the same sign for all vectors $\boldsymbol{x} \in V$ and *lie inside* different poles if these inner products have opposite signs.

A set $M \subset \mathsf{L}$ is said to be convex if for every pair of vectors $\boldsymbol{e}, \boldsymbol{e}' \in M$, the vectors $\boldsymbol{g}_t = t\boldsymbol{e} + (1-t)\boldsymbol{e}'$ are also in M for all $t \in [0, 1]$. We shall prove that the interior of each pole of the light cone V is convex, that is, the vector $\boldsymbol{g}_t$ lies in the same pole as $\boldsymbol{e}$ and $\boldsymbol{e}'$ for all $t \in [0, 1]$. To this end, let us note that in the expression $(\boldsymbol{g}_t, \boldsymbol{x}) = t(\boldsymbol{e}, \boldsymbol{x}) + (1-t)(\boldsymbol{e}', \boldsymbol{x})$, the coefficients t and $1-t$ are nonnegative, and the inner products $(\boldsymbol{e}, \boldsymbol{x})$ and $(\boldsymbol{e}', \boldsymbol{x})$ have the same sign. Therefore, for every vector $\boldsymbol{x} \in V$, the inner product $(\boldsymbol{g}_t, \boldsymbol{x})$ has the same sign as $(\boldsymbol{e}, \boldsymbol{x})$ and $(\boldsymbol{e}', \boldsymbol{x})$.

Lemma 7.59 *Timelike vectors* $\boldsymbol{e}$ *and* $\boldsymbol{e}'$ *lie inside one pole of the light cone* V *if and only if* $(\boldsymbol{e}, \boldsymbol{e}') < 0$.

Proof If timelike vectors $\boldsymbol{e}$ and $\boldsymbol{e}'$ lie inside one pole, then by definition, we have the inequality $(\boldsymbol{e}, \boldsymbol{x}) \cdot (\boldsymbol{e}', \boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in V$. Let us assume that $(\boldsymbol{e}, \boldsymbol{e}') \ge 0$. As we established above, the vector $\boldsymbol{g}_t = t\boldsymbol{e} + (1-t)\boldsymbol{e}'$ is timelike and lies inside the same pole as $\boldsymbol{e}$ and $\boldsymbol{e}'$ for all $t \in [0, 1]$.

Let us consider the inner product $(\boldsymbol{g}_t, \boldsymbol{e}) = t(\boldsymbol{e}, \boldsymbol{e}) + (1-t)(\boldsymbol{e}, \boldsymbol{e}')$ as a function of the variable $t \in [0, 1]$. It is obvious that this function is continuous and that it assumes for $t = 0$ the value $(\boldsymbol{e}, \boldsymbol{e}') \ge 0$, and for $t = 1$ the value $(\boldsymbol{e}, \boldsymbol{e}) < 0$. Therefore, as is proved in a course in calculus, there exists a value $\tau \in [0, 1]$ such that $(\boldsymbol{g}_\tau, \boldsymbol{e}) = 0$. But this contradicts Corollary 7.58.

Thus we have proved that if vectors $\boldsymbol{e}$ and $\boldsymbol{e}'$ lie inside one pole of the cone V, then $(\boldsymbol{e}, \boldsymbol{e}') < 0$. The converse assertion is obvious. Let $\boldsymbol{e}$ and $\boldsymbol{e}'$ lie inside different poles, for instance, $\boldsymbol{e}$ is within V_+, while $\boldsymbol{e}'$ is within V_-. Then we have by definition that the vector $-\boldsymbol{e}'$ lies inside the pole V_+, and therefore, $(\boldsymbol{e}, -\boldsymbol{e}') < 0$, that is, $(\boldsymbol{e}, \boldsymbol{e}') > 0$. □

7.8 Lorentz Transformations

In this section, we shall examine an analogue of orthogonal transformations for pseudo-Euclidean spaces called *Lorentz* transformations. Such transformations have numerous applications in physics.[9] They are also defined by the condition of preserving the inner product.

[9]For example, a Lorentz transformation of Minkowski space—a four-dimensional pseudo-Euclidean space—plays the same role in the special theory of relativity (which is where the term Lorentz transformation comes from) as that played by the Galilean transformations, which describe the passage from one inertial reference frame to another in classical Newtonian mechanics.

Definition 7.60 A linear transformation $\mathcal{U}$ of a pseudo-Euclidean space L is called a *Lorentz transformation* if the relationship

$$\big(\mathcal{U}(\boldsymbol{x}), \mathcal{U}(\boldsymbol{y})\big) = (\boldsymbol{x}, \boldsymbol{y}) \tag{7.84}$$

is satisfied for all vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{L}$.

As in the case of orthogonal transformations, it suffices that the condition (7.84) be satisfied for all vectors $\boldsymbol{x} = \boldsymbol{y}$ of the pseudo-Euclidean space L. The proof of this coincides completely with the proof of the analogous assertion in Sect. 7.2.

Here, as in the case of Euclidean spaces, we shall make use of the inner product $(\boldsymbol{x}, \boldsymbol{y})$ in order to identify L^* with L (let us recall that for this, we need only the nonsingularity of the bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ and not the positive definiteness of the associated quadratic form $(\boldsymbol{x}^2)$). As a result, for an arbitrary linear transformation $\mathcal{A} : \mathsf{L} \to \mathsf{L}$, we may consider $\mathcal{A}^*$ also as a transformation of the space L into itself. Repeating the same arguments that we employed in the case of Euclidean spaces, we obtain that $|\mathcal{A}^*| = |\mathcal{A}|$. In particular, from definition (7.84), it follows that for a Lorentz transformation $\mathcal{U}$, we have the relationship

$$U^* A U = A, \tag{7.85}$$

where U is the matrix of the transformation $\mathcal{U}$ in an arbitrary basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ of the space L, and $A = (a_{ij})$ is the Gram matrix of the bilinear form $(\boldsymbol{x}, \boldsymbol{y})$, that is, the matrix with elements $a_{ij} = (\boldsymbol{e}_i, \boldsymbol{e}_j)$.

The bilinear form $(\boldsymbol{x}, \boldsymbol{y})$ is nonsingular, that is, $|A| \neq 0$, and from the relationship (7.85) follows the equality $|\mathcal{U}|^2 = 1$, from which we obtain that $|\mathcal{U}| = \pm 1$. As in the case of a Euclidean space, a transformation with determinant equal to 1 is called *proper*, while if the determinant is equal to -1, it is *improper*.

It follows from the definition that every Lorentz transformation maps the light cone V into itself. It follows from Theorem 7.55 that a Lorentz transformation either maps each pole into itself (that is, $\mathcal{U}(V_+) = V_+$ and $\mathcal{U}(V_-) = V_-$), or else interchanges them (that is, $\mathcal{U}(V_+) = V_-$ and $\mathcal{U}(V_-) = V_+$). Let us associate with each Lorentz transformation $\mathcal{U}$ the number $\nu(\mathcal{U}) = +1$ in the first case, and $\nu(\mathcal{U}) = -1$ in the second. Like the determinant $|\mathcal{U}|$, this number $\nu(\mathcal{U})$ is a natural characteristic of the associated Lorentz transformation. Let us denote the pair of numbers $(|\mathcal{U}|, \nu(\mathcal{U}))$ by $\varepsilon(\mathcal{U})$. It is obvious that

$$\varepsilon\big(\mathcal{U}^{-1}\big) = \varepsilon(\mathcal{U}), \qquad \varepsilon(\mathcal{U}_1\mathcal{U}_2) = \varepsilon(\mathcal{U}_1)\varepsilon(\mathcal{U}_2),$$

where on the right-hand side, it is understood that the first and second components of the pairs are multiplied separately. We shall soon see that in an arbitrary pseudo-Euclidean space, there exist Lorentz transformations $\mathcal{U}$ of all four types, that is, with $\varepsilon(\mathcal{U})$ taking all values

$$(+1, +1), \qquad (+1, -1), \qquad (-1, +1), \qquad (-1, -1).$$

This property is sometimes interpreted as saying that a pseudo-Euclidean space has not two (as in the case of a Euclidean space), but *four* orientations.

Like orthogonal transformations of a Euclidean space, Lorentz transformations are characterized by the fact that they map an orthonormal basis of a pseudo-Euclidean space to an orthonormal basis. Indeed, suppose that for the vectors of the orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, the equalities

$$(\boldsymbol{e}_i, \boldsymbol{e}_j) = 0 \quad \text{for } i \neq j, \qquad \big(\boldsymbol{e}_1^2\big) = \cdots = \big(\boldsymbol{e}_{n-1}^2\big) = 1, \qquad \big(\boldsymbol{e}_n^2\big) = -1 \tag{7.86}$$

are satisfied. Then from the condition (7.84), it follows that the images $\mathcal{U}(\boldsymbol{e}_1), \dots, \mathcal{U}(\boldsymbol{e}_n)$ satisfy analogous equalities, that is, they form an orthonormal basis in L. Conversely, if for the vectors $\boldsymbol{e}_i$, the equality (7.86) is satisfied and analogous equalities hold for the vectors $\mathcal{U}(\boldsymbol{e}_i)$, then as is easily verified, for arbitrary vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of the pseudo-Euclidean space L, the relationship (7.84) is satisfied.

Two orthonormal bases are said to have the *same orientation* if for a Lorentz transformation $\mathcal{U}$ taking one basis to the other, $\varepsilon(\mathcal{U}) = (+1, +1)$. The choice of a class of bases with the same orientation is called an *orientation* of the pseudo-Euclidean space L. Taking for now on faith the fact (which will be proved a little later) that there exist Lorentz transformations $\mathcal{U}$ with all theoretically possible $\varepsilon(\mathcal{U})$, we see that in a pseudo-Euclidean space, it is possible to introduce exactly four orientations.

Example 7.61 Let us consider some concepts about pseudo-Euclidean spaces that we encountered in Example 7.49, that is, for $\dim \mathsf{L} = 2$ and $s = 1$. As we have seen, in this space, there exists a basis $\boldsymbol{f}_1, \boldsymbol{f}_2$ for which the relationships $(\boldsymbol{f}_1^2) = (\boldsymbol{f}_2^2) = 0$, $(\boldsymbol{f}_1, \boldsymbol{f}_2) = \frac{1}{2}$, are satisfied, and the scalar square of the vector $\boldsymbol{x} = x\boldsymbol{f}_1 + y\boldsymbol{f}_2$ is equal to $(\boldsymbol{x}^2) = xy$. If $\mathcal{U} : \mathsf{L} \to \mathsf{L}$ is a Lorentz transformation given by the formula

$$x' = ax + by, \qquad y' = cx + dy,$$

then the equality $(\mathcal{U}(\boldsymbol{x}), \mathcal{U}(\boldsymbol{x})) = (\boldsymbol{x}, \boldsymbol{x})$ for the vector $\boldsymbol{x} = x\boldsymbol{f}_1 + y\boldsymbol{f}_2$ takes the form $x'y' = xy$, that is, $(ax + by)(cx + dy) = xy$ for all x and y. From this, we obtain

$$ac = 0, \qquad bd = 0, \qquad ad + bc = 1.$$

In view of the equality $ad + bc = 1$, the values $a = b = 0$ are impossible.

If $a \neq 0$, then $c = 0$, and this implies that $ad = 1$, that is, $d = a^{-1} \neq 0$ and $b = 0$. Thus the transformation $\mathcal{U}$ has the form

$$x' = ax, \qquad y' = a^{-1}y. \tag{7.87}$$

This is a proper transformation.

On the other hand, if $b \neq 0$, then $d = 0$, and this implies that $c = b^{-1}$, $a = 0$. The transformation $\mathcal{U}$ has in this case the form

$$x' = by, \qquad y' = b^{-1}x. \tag{7.88}$$

This is an improper transformation.

If we write the transformation $\mathcal{U}$ in the form (7.87) or (7.88), depending on whether it is proper or improper, then the sign of the number a or respectively b indicates whether $\mathcal{U}$ interchanges the poles of the light cone or preserves each of them. Namely, let us prove that the transformation (7.87) causes the poles to change places if $a < 0$, and preserves them if $a > 0$. And analogously, the transformation (7.88) interchanges the poles if $b < 0$ and preserves them if $b > 0$.

By Theorem 7.55, the partition of the light cone V into two poles V_+ and V_- does not depend on the choice of timelike vector, and therefore, by Lemma 7.59, we need only determine the sign of the inner product $(\boldsymbol{e}, \mathcal{U}(\boldsymbol{e}))$ for an arbitrary timelike vector $\boldsymbol{e}$. Let $\boldsymbol{e} = x\boldsymbol{f}_1 + y\boldsymbol{f}_2$. Then $(\boldsymbol{e}^2) = xy < 0$. In the case that $\mathcal{U}$ is a proper transformation, we have formula (7.87), from which it follows that

$$\mathcal{U}(\boldsymbol{e}) = ax\boldsymbol{f}_1 + a^{-1}y\boldsymbol{f}_2, \qquad \bigl(\boldsymbol{e}, \mathcal{U}(\boldsymbol{e})\bigr) = \bigl(a + a^{-1}\bigr)xy.$$

Since $xy < 0$, the inner product $(\boldsymbol{e}, \mathcal{U}(\boldsymbol{e}))$ is negative if $a + a^{-1} > 0$, and positive if $a + a^{-1} < 0$. But it is obvious that $a + a^{-1} > 0$ for $a > 0$, and $a + a^{-1} < 0$ for $a < 0$. Thus for $a > 0$, we have $(\boldsymbol{e}, \mathcal{U}(\boldsymbol{e})) < 0$, and by Lemma 7.59, the vectors $\boldsymbol{e}$ and $\mathcal{U}(\boldsymbol{e})$ lie inside one pole. Consequently, the transformation $\mathcal{U}$ preserves the poles of the light cone. Analogously, for $a < 0$, we obtain $(\boldsymbol{e}, \mathcal{U}(\boldsymbol{e})) > 0$, that is, $\boldsymbol{e}$ and $\mathcal{U}(\boldsymbol{e})$ lie inside different poles, and therefore, the transformation $\mathcal{U}$ interchanges the poles.

The case of an improper transformation can be examined with the help of formula (7.88). Reasoning analogously to what has gone before, we obtain from it the relationships

$$\mathcal{U}(\boldsymbol{e}) = b^{-1}y\boldsymbol{f}_1 + bx\boldsymbol{f}_2, \qquad \bigl(\boldsymbol{e}, \mathcal{U}(\boldsymbol{e})\bigr) = bx^2 + b^{-1}y^2,$$

from which it is clear that now the sign of $(\boldsymbol{e}, \mathcal{U}(\boldsymbol{e}))$ coincides with the sign of the number b.

Example 7.62 It is sometimes convenient to use the fact that a Lorentz transformation of a pseudo-Euclidean plane can be written in an alternative form, using the hyperbolic sine and cosine. We saw earlier (formulas (7.87) and (7.88)) that in the basis $\boldsymbol{f}_1, \boldsymbol{f}_2$ defined by the relationship (7.78), proper and improper Lorentz transformations are given respectively by the equalities

$$\begin{aligned} \mathcal{U}(\boldsymbol{f}_1) &= a\boldsymbol{f}_1, \qquad & \mathcal{U}(\boldsymbol{f}_2) &= a^{-1}\boldsymbol{f}_2, \\ \mathcal{U}(\boldsymbol{f}_1) &= b\boldsymbol{f}_2, \qquad & \mathcal{U}(\boldsymbol{f}_2) &= b^{-1}\boldsymbol{f}_1. \end{aligned}$$

From this, it is not difficult to derive that in the orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$, related to $\boldsymbol{f}_1, \boldsymbol{f}_2$ by formula (7.78), these transformations are given respectively by the equalities

$$\begin{aligned} \mathcal{U}(\boldsymbol{e}_1) &= \frac{a + a^{-1}}{2}\boldsymbol{e}_1 + \frac{a - a^{-1}}{2}\boldsymbol{e}_2, \\ \mathcal{U}(\boldsymbol{e}_2) &= \frac{a - a^{-1}}{2}\boldsymbol{e}_1 + \frac{a + a^{-1}}{2}\boldsymbol{e}_2, \end{aligned} \tag{7.89}$$

$$\mathcal{U}(\boldsymbol{e}_1) = \frac{b+b^{-1}}{2}\boldsymbol{e}_1 - \frac{b-b^{-1}}{2}\boldsymbol{e}_2,$$
$$\mathcal{U}(\boldsymbol{e}_2) = \frac{b-b^{-1}}{2}\boldsymbol{e}_1 - \frac{b+b^{-1}}{2}\boldsymbol{e}_2. \tag{7.90}$$

Setting here $a = \pm e^{\psi}$ and $b = \pm e^{\psi}$, where the sign $\pm$ coincides with the sign of the number a or b in formula (7.89) or (7.90) respectively, we obtain that the matrices of the proper transformations have the form

$$\begin{pmatrix} \cosh\psi & \sinh\psi \\ \sinh\psi & \cosh\psi \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -\cosh\psi & -\sinh\psi \\ -\sinh\psi & -\cosh\psi \end{pmatrix}, \tag{7.91}$$

while the matrices of the improper transformations have the form

$$\begin{pmatrix} \cosh\psi & \sinh\psi \\ -\sinh\psi & -\cosh\psi \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -\cosh\psi & -\sinh\psi \\ \sinh\psi & \cosh\psi \end{pmatrix}, \tag{7.92}$$

where $\sinh\psi = (e^{\psi} - e^{-\psi})/2$ and $\cosh\psi = (e^{\psi} + e^{-\psi})/2$ are the hyperbolic sine and cosine.

Theorem 7.63 *In every pseudo-Euclidean space there exist Lorentz transformations $\mathcal{U}$ with all four possible values of $\varepsilon(\mathcal{U})$.*

Proof For the case $\dim \mathsf{L} = 2$, we have already proved the theorem: In Example 7.62, we saw that there exist four distinct types of Lorentz transformation of a pseudo-Euclidean space having in a suitable orthonormal basis the matrices (7.91), (7.92). It is obvious that with these matrices, the transformation $\mathcal{U}$ gives all possible values $\varepsilon(\mathcal{U})$.

Let us now move on to the general case $\dim \mathsf{L} > 2$. Let us choose in the pseudo-Euclidean space L an arbitrary timelike vector $\boldsymbol{e}$ and any $\boldsymbol{e}'$ not proportional to it. By Lemma 7.53, the two-dimensional space $\langle \boldsymbol{e}, \boldsymbol{e}' \rangle$ is a pseudo-Euclidean space (therefore nondegenerate), and we have the decomposition

$$\mathsf{L} = \langle \boldsymbol{e}, \boldsymbol{e}' \rangle \oplus \langle \boldsymbol{e}, \boldsymbol{e}' \rangle^{\perp}.$$

From the law of inertia, it follows that the space $\langle \boldsymbol{e}, \boldsymbol{e}' \rangle^{\perp}$ is a Euclidean space. In Example 7.62, we saw that in the pseudo-Euclidean plane $\langle \boldsymbol{e}, \boldsymbol{e}' \rangle$, there exists a Lorentz transformation $\mathcal{U}_1$ with arbitrary value $\varepsilon(\mathcal{U}_1)$. Let us define the transformation $\mathcal{U} : \mathsf{L} \to \mathsf{L}$ as $\mathcal{U}_1$ in $\langle \boldsymbol{e}, \boldsymbol{e}' \rangle$ and $\mathcal{E}$ in $\langle \boldsymbol{e}, \boldsymbol{e}' \rangle^{\perp}$, that is, for a vector $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$, where $\boldsymbol{y} \in \langle \boldsymbol{e}, \boldsymbol{e}' \rangle$ and $\boldsymbol{z} \in \langle \boldsymbol{e}, \boldsymbol{e}' \rangle^{\perp}$, we shall set $\mathcal{U}(\boldsymbol{x}) = \mathcal{U}_1(\boldsymbol{y}) + \boldsymbol{z}$. Then $\mathcal{U}$ is clearly a Lorentz transformation, and $\varepsilon(\mathcal{U}) = \varepsilon(\mathcal{U}_1)$. □

There is an analogue to Theorem 7.24 for Lorentz transformations.

Theorem 7.64 *If a space L' is invariant with respect to a Lorentz transformation $\mathcal{U}$, then its orthogonal complement $(\mathsf{L}')^{\perp}$ is also invariant with respect to $\mathcal{U}$.*

Proof The proof of this theorem is an exact repetition of the proof of Theorem 7.24, since there, we did not use the positive definiteness of the quadratic form $(\boldsymbol{x}^2)$ associated with the bilinear form $(\boldsymbol{x}, \boldsymbol{y})$, but only its nonsingularity. See Remark 7.25 on p. 227. □

The study of a Lorentz transformation of a pseudo-Euclidean space is reduced to the analogous question for orthogonal transformations of a Euclidean space, based on the following result.

Theorem 7.65 *For every Lorentz transformation $\mathcal{U}$ of a pseudo-Euclidean space* L, *there exist nondegenerate subspaces* L_0 *and* L_1 *invariant with respect to* $\mathcal{U}$ *such that* L *has the orthogonal decomposition*

$$\mathsf{L} = \mathsf{L}_0 \oplus \mathsf{L}_1, \quad \mathsf{L}_0 \perp \mathsf{L}_1, \tag{7.93}$$

where the subspace L_0 *is a Euclidean space, and the dimension of* L_1 *is equal to* 1, 2, *or* 3.

It follows from the law of inertia that if $\dim \mathsf{L}_1 = 1$, then L_1 is spanned by a timelike vector. If $\dim \mathsf{L}_1 = 2$ or 3, then the pseudo-Euclidean space L_1 can be represented in turn by a direct sum of subspaces of lower dimension invariant with respect to $\mathcal{U}$. However, such a decomposition is no longer necessarily orthogonal (see Example 7.48).

Proof of Theorem 7.65 The proof is by induction on n, the dimension of the space L. For $n = 2$, the assertion of the theorem is obvious—in the decomposition (7.93) one has only to set $\mathsf{L}_0 = (\mathbf{0})$ and $\mathsf{L}_1 = \mathsf{L}$.[10]

Now let $n > 2$, and suppose that the assertion of the theorem has been proved for all pseudo-Euclidean spaces of dimension less than n. We shall use results obtained in Chaps. 4 and 5 on linear transformations of a vector space into itself. Obviously, one of the following three cases must hold: the transformation $\mathcal{U}$ has a complex eigenvalue, $\mathcal{U}$ has two linearly independent eigenvectors, or the space L is cyclic for $\mathcal{U}$, corresponding to the only real eigenvalue. Let us consider the three cases separately.

Case 1. A linear transformation $\mathcal{U}$ of a real vector space L has a complex eigenvalue λ. As established in Sect. 4.3, then $\mathcal{U}$ also has the complex conjugate eigenvalue $\overline{\lambda}$, and moreover, to the pair $\lambda, \overline{\lambda}$ there corresponds the two-dimensional real invariant subspace $\mathsf{L}' \subset \mathsf{L}$, which contains no real eigenvectors. It is obvious that L' cannot be a pseudo-Euclidean space: for then the restriction of $\mathcal{U}$ to L' would have real eigenvalues, and L' would contain real eigenvectors of the transformation $\mathcal{U}$; see Examples 7.61 and 7.62. Let us show that L' is nondegenerate.

[10]The nondegeneracy of the subspace $\mathsf{L}_0 = (\mathbf{0})$ relative to a bilinear form follows from the definitions given on pages 266 and 195. Indeed, the rank of the restriction of the bilinear form to the subspace $(\mathbf{0})$ is zero, and therefore, it coincides with $\dim(\mathbf{0})$.

Suppose that L' is degenerate. Then it contains a lightlike vector $\boldsymbol{e} \neq \mathbf{0}$. Since $\mathcal{U}$ is a Lorentz transformation, the vector $\mathcal{U}(\boldsymbol{e})$ is also lightlike, and since the subspace L' is invariant with respect to $\mathcal{U}$, it follows that $\mathcal{U}(\boldsymbol{e})$ is contained in L'. Therefore, the subspace L' contains two lightlike vectors: $\boldsymbol{e}$ and $\mathcal{U}(\boldsymbol{e})$. By Lemma 7.53, these vectors cannot be linearly independent, since then L' would be a pseudo-Euclidean space, but that would contradict our assumption that L' is degenerate. From this, it follows that the vector $\mathcal{U}(\boldsymbol{e})$ is proportional to $\boldsymbol{e}$, and that implies that $\boldsymbol{e}$ is an eigenvector of the transformation $\mathcal{U}$, which, as we have observed above, cannot be. This contradiction means that the subspace L' is nondegenerate, and as a consequence, it is a Euclidean space.

Case 2. The linear transformation $\mathcal{U}$ has two linearly independent eigenvectors: $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. If at least one of them is not lightlike, that is, $(\boldsymbol{e}_i^2) \neq 0$, then $\mathsf{L}' = \langle \boldsymbol{e}_i \rangle$ is a nondegenerate invariant subspace of dimension 1. And if both eigenvectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are lightlike, then by Lemma 7.53, the subspace $\mathsf{L}' = \langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle$ is an invariant pseudo-Euclidean plane.

Thus in both cases, the transformation $\mathcal{U}$ has a nondegenerate invariant subspace L' of dimension 1 or 2. This means that in both cases, we have an orthogonal decomposition (7.73), that is, $\mathsf{L} = \mathsf{L}' \oplus (\mathsf{L}')^\perp$. If L' is one-dimensional and spanned by a timelike vector or is a pseudo-Euclidean plane, then this is exactly decomposition (7.93) with $\mathsf{L}_0 = (\mathsf{L}')^\perp$ and $\mathsf{L}_1 = \mathsf{L}'$. In the opposite case, the subspace L' is a Euclidean space of dimension 1 or 2, and the subspace $(\mathsf{L}')^\perp$ is a pseudo-Euclidean space of dimension $n-1$ or $n-2$ respectively. By the induction hypothesis, for $(\mathsf{L}')^\perp$, we have the orthogonal decomposition $(\mathsf{L}')^\perp = \mathsf{L}'_0 \oplus \mathsf{L}'_1$ analogous to (7.93). From this, for L we obtain the decomposition (7.93) with $\mathsf{L}_0 = \mathsf{L}' \oplus \mathsf{L}'_0$ and $\mathsf{L}_1 = \mathsf{L}'_1$.

Case 3. The space L is cyclic for the transformation $\mathcal{U}$, corresponding to the unique real eigenvalue λ and principal vector $\boldsymbol{e}$ of grade $m = n$. Obviously, for $n = 2$, this is impossible: as we saw in Example 7.61, in a suitable basis of a pseudo-Euclidean plane, a Lorentz transformation has either diagonal form (7.87) or the form (7.88) with distinct eigenvalues ± 1. In both cases, it is obvious that the pseudo-Euclidean plane L cannot be a cyclic subspace of the transformation $\mathcal{U}$.

Let us consider the case of a pseudo-Euclidean space L of dimension $n \geq 3$. We shall prove that L can be a cyclic subspace of the transformation $\mathcal{U}$ only if $n = 3$.

As we established in Sect. 5.1, in a cyclic subspace L, there is a basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ defined by formula (5.5), that is,

$$\boldsymbol{e}_1 = \boldsymbol{e}, \qquad \boldsymbol{e}_2 = (\mathcal{U} - \lambda\mathcal{E})(\boldsymbol{e}), \qquad \dots, \qquad \boldsymbol{e}_n = (\mathcal{U} - \lambda\mathcal{E})^{n-1}(\boldsymbol{e}), \tag{7.94}$$

in which relationships (5.6) hold:

$$\mathcal{U}(\boldsymbol{e}_1) = \lambda\boldsymbol{e}_1 + \boldsymbol{e}_2, \qquad \mathcal{U}(\boldsymbol{e}_2) = \lambda\boldsymbol{e}_2 + \boldsymbol{e}_3, \qquad \dots, \qquad \mathcal{U}(\boldsymbol{e}_n) = \lambda\boldsymbol{e}_n. \tag{7.95}$$

In this basis, the matrix of the transformation $\mathcal{U}$ has the form of a Jordan block

$$U=\begin{pmatrix}\lambda & 0 & 0 & \cdots & \cdots & 0\\ 1 & \lambda & 0 & \cdots & \cdots & 0\\ 0 & 1 & \lambda & & & 0\\ \vdots & & \ddots & \ddots & & \vdots\\ \vdots & & & \ddots & \lambda & 0\\ 0 & 0 & 0 & \cdots & 1 & \lambda\end{pmatrix}. \tag{7.96}$$

It is easy to see that the eigenvector $\boldsymbol{e}_n$ is lightlike. Indeed, if we had $(\boldsymbol{e}_n^2)\neq 0$, then we would have the orthogonal decomposition $\mathsf{L}=\langle\boldsymbol{e}_n\rangle\oplus\langle\boldsymbol{e}_n\rangle^{\perp}$, where both subspaces $\langle\boldsymbol{e}_n\rangle$ and $\langle\boldsymbol{e}_n\rangle^{\perp}$ are invariant. But this contradicts the assumption that the space L is cyclic.

Since $\mathcal{U}$ is a Lorentz transformation, it preserves the inner product of vectors, and from (7.95), we obtain the equality

$$\begin{aligned}(\boldsymbol{e}_i,\boldsymbol{e}_n)&=\big(\mathcal{U}(\boldsymbol{e}_i),\mathcal{U}(\boldsymbol{e}_n)\big)=(\lambda\boldsymbol{e}_i+\boldsymbol{e}_{i+1},\lambda\boldsymbol{e}_n)\\&=\lambda^2(\boldsymbol{e}_i,\boldsymbol{e}_n)+\lambda(\boldsymbol{e}_{i+1},\boldsymbol{e}_n)\end{aligned} \tag{7.97}$$

for all $i=1,\dots,n-1$.

If $\lambda^2\neq 1$, then from (7.97), it follows that

$$(\boldsymbol{e}_i,\boldsymbol{e}_n)=\frac{\lambda}{1-\lambda^2}(\boldsymbol{e}_{i+1},\boldsymbol{e}_n).$$

Substituting into this equality the values of the index $i=n-1,\dots,1$, taking into account that $(\boldsymbol{e}_n^2)=0$, we therefore obtain step by step that $(\boldsymbol{e}_i,\boldsymbol{e}_n)=0$ for all i. This means that the eigenvector $\boldsymbol{e}_n$ is contained in the radical of the space L, and since L is a pseudo-Euclidean space (that is, in particular, nondegenerate), it follows that $\boldsymbol{e}_n=\mathbf{0}$. This contradiction shows that $\lambda^2=1$.

Substituting $\lambda^2=1$ into the equalities (7.97) and collecting like terms, we find that $(\boldsymbol{e}_{i+1},\boldsymbol{e}_n)=0$ for all indices $i=1,\dots,n-1$, that is, $(\boldsymbol{e}_j,\boldsymbol{e}_n)=0$ for all indices $j=2,\dots,n$. In particular, we have the equalities $(\boldsymbol{e}_{n-1},\boldsymbol{e}_n)=0$ for $n>2$ and $(\boldsymbol{e}_{n-2},\boldsymbol{e}_n)=0$ for $n>3$. From this it follows that $n=3$. Indeed, from the condition of preservation of the inner product, we have the relationship

$$\begin{aligned}(\boldsymbol{e}_{n-2},\boldsymbol{e}_{n-1})&=\big(\mathcal{U}(\boldsymbol{e}_{n-2}),\mathcal{U}(\boldsymbol{e}_{n-1})\big)=(\lambda\boldsymbol{e}_{n-2}+\boldsymbol{e}_{n-1},\lambda\boldsymbol{e}_{n-1}+\boldsymbol{e}_n)\\&=\lambda^2(\boldsymbol{e}_{n-2},\boldsymbol{e}_{n-1})+\lambda(\boldsymbol{e}_{n-2},\boldsymbol{e}_n)+\lambda\big(\boldsymbol{e}_{n-1}^2\big)+(\boldsymbol{e}_{n-1},\boldsymbol{e}_n),\end{aligned}$$

from which, taking into account the relationships $\lambda^2=1$ and $(\boldsymbol{e}_{n-1},\boldsymbol{e}_n)=0$, we have the equality $(\boldsymbol{e}_{n-2},\boldsymbol{e}_n)+(\boldsymbol{e}_{n-1}^2)=0$. If $n>3$, then $(\boldsymbol{e}_{n-2},\boldsymbol{e}_n)=0$, and from this, we obtain that $(\boldsymbol{e}_{n-1}^2)=0$, that is, the vector $\boldsymbol{e}_{n-1}$ is lightlike.

Let us examine the subspace $\mathsf{L}'=\langle\boldsymbol{e}_n,\boldsymbol{e}_{n-1}\rangle$. It is obvious that it is invariant with respect to the transformation $\mathcal{U}$, and since it contains two linearly independent lightlike vectors $\boldsymbol{e}_n$ and $\boldsymbol{e}_{n-1}$, then by Lemma 7.53, the subspace L' is a pseudo-Euclidean space, and we obtain the decomposition $\mathsf{L} = \mathsf{L}' \oplus (\mathsf{L}')^{\perp}$ as a direct sum of two invariant subspaces. But this contradicts the fact that the space L is cyclic. Therefore, the transformation $\mathcal{U}$ can have cyclic subspaces only of dimension 3.

Putting together cases 1, 2, and 3, and taking into account the induction hypothesis, we obtain the assertion of the theorem. □

Combining Theorems 7.27 and 7.65, we obtain the following corollary.

Corollary 7.66 *For every transformation of a pseudo-Euclidean space, there exists an orthonormal basis in which the matrix of the transformation has block-diagonal form with blocks of the following types*:

1. *blocks of order* 1 *with elements* ± 1;
2. *blocks of order* 2 *of type* (7.29);
3. *blocks of order* 2 *of type* (7.91)–(7.92);
4. *blocks of order* 3 *corresponding to a three-dimensional cyclic subspace with eigenvalue* ± 1.

It follows from the law of inertia that the matrix of a Lorentz transformation can contain not more than one block of type 3 *or* 4.

Let us note as well that a block of type 4 corresponding to a three-dimensional cyclic subspace cannot be brought into Jordan normal form in an orthonormal basis. Indeed, as we saw earlier, a block of type 4 is brought into Jordan normal form in the basis (7.94), where the eigenvector $\boldsymbol{e}_n$ is lightlike, and therefore, it cannot belong to any orthonormal basis.

With the proof of Theorem 7.65 we have established necessary conditions for a Lorentz transformation to have a cyclic subspace—in particular, its dimension must be 3, corresponding to an eigenvalue equal to ± 1, and eigenvector that is lightlike. Clearly, these necessary conditions are not sufficient, since in deriving them, we used the equalities $(\boldsymbol{e}_i, \boldsymbol{e}_k) = (\mathcal{U}(\boldsymbol{e}_i), \mathcal{U}(\boldsymbol{e}_k))$ for only some of the vectors of the basis (7.94). Let us show that Lorentz transformations with cyclic subspaces indeed exist.

Example 7.67 Let us consider a vector space L of dimension $n = 3$. Let us choose in L a basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ and define a transformation $\mathcal{U} : \mathsf{L} \to \mathsf{L}$ using relationships (7.95) with the number $\lambda = \pm 1$. Then the matrix of the transformation $\mathcal{U}$ will take the form of a Jordan block with eigenvalue λ.

Let us choose the Gram matrix for a basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ such that L is given the structure of a pseudo-Euclidean space. With the proof of Theorem 7.65, we have found necessary conditions $(\boldsymbol{e}_2, \boldsymbol{e}_3) = 0$ and $(\boldsymbol{e}_3^2) = 0$. Let us set $(\boldsymbol{e}_1^2) = a$, $(\boldsymbol{e}_1, \boldsymbol{e}_2) = b$, $(\boldsymbol{e}_1, \boldsymbol{e}_3) = c$, and $(\boldsymbol{e}_2^2) = d$. Then the Gram matrix can be written as

$$A = \begin{pmatrix} a & b & c \\ b & d & 0 \\ c & 0 & 0 \end{pmatrix}. \tag{7.98}$$

On the other hand, as we know (see Example 7.51, p. 270), in L there exists an orthonormal basis in which the Gram matrix is diagonal and has determinant -1. Since the sign of the determinant of the Gram matrix is one and the same for all bases, it follows that $|A| = -c^2 d < 0$, that is, $c \neq 0$ and $d > 0$.

The conditions $c \neq 0$ and $d > 0$ are also sufficient for the vector space in which the inner product is given by the Gram matrix A in the form (7.98) to be a pseudo-Euclidean space. Indeed, choosing a basis $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3$ in which the quadratic form associated with the matrix A has canonical form (6.28), we see that the condition $|A| < 0$ is satisfied by, besides a pseudo-Euclidean space, only a space in which $(\boldsymbol{g}_i^2) = -1$ for all $i = 1, 2, 3$. But such a quadratic form is negative definite, that is, $(\boldsymbol{x}^2) < 0$ for all vectors $\boldsymbol{x} \neq \boldsymbol{0}$, and this contradicts that $(\boldsymbol{e}_2^2) = d > 0$.

Let us now consider the equalities $(\boldsymbol{e}_i, \boldsymbol{e}_k) = (\mathcal{U}(\boldsymbol{e}_i), \mathcal{U}(\boldsymbol{e}_k))$ for all indices $i \leq k$ from 1 to 3. Taking into account $\lambda^2 = 1$, $(\boldsymbol{e}_2, \boldsymbol{e}_3) = 0$, and $(\boldsymbol{e}_3^2) = 0$, we see that they are satisfied automatically except for the cases $i = k = 1$ and $i = 1$, $k = 2$. These two cases give the relationships $2\lambda b + d = 0$ and $c + d = 0$. Thus we may choose the number a arbitrarily, the number d to be any positive number, and set $c = -d$ and $b = -\lambda d/2$. It is also not difficult to ascertain that linearly independent vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ satisfying such conditions in fact exist.

Just as in a Euclidean space, the presence of different orientations of a pseudo-Euclidean space determined by the value of $\varepsilon(\mathcal{U})$ for the Lorentz transformation $\mathcal{U}$ is connected with the concept of continuous deformation of a transformation (p. 230), which defines an equivalence relation on the set of transformations.

Let $\mathcal{U}_t$ be a family of Lorentz transformations continuously depending on the parameter t. Then $|\mathcal{U}_t|$ also depends continuously on t, and since the determinant of a Lorentz transformation is equal to ± 1, the value of $|\mathcal{U}_t|$ is constant for all t. Thus Lorentz transformations with determinants having opposite signs cannot be continuously deformed into each other. But in contrast to orthogonal transformations of a Euclidean space, Lorentz transformations $\mathcal{U}_t$ have an additional characteristic, the number $\nu(\mathcal{U}_t)$ (see the definition on p. 276). Let us show that like the determinant $|\mathcal{U}_t|$, the number $\nu(\mathcal{U}_t)$ is also constant.

To this end, let us choose an arbitrary timelike vector $\boldsymbol{e}$ and make use of Lemma 7.59. The vector $\mathcal{U}_t(\boldsymbol{e})$ is also timelike, and moreover, $\nu(\mathcal{U}_t) = +1$ if $\boldsymbol{e}$ and $\mathcal{U}_t(\boldsymbol{e})$ lie inside one pole of the light cone, that is, $(\boldsymbol{e}, \mathcal{U}_t(\boldsymbol{e})) < 0$, and $\nu(\mathcal{U}_t) = -1$ if $\boldsymbol{e}$ and $\mathcal{U}_t(\boldsymbol{e})$ lie inside different poles, that is, $(\boldsymbol{e}, \mathcal{U}_t(\boldsymbol{e})) > 0$. It then remains to observe that the function $(\boldsymbol{e}, \mathcal{U}_t(\boldsymbol{e}))$ depends continuously on the argument t, and therefore can change sign only if for some value of t, it assumes the value zero. But from inequality (7.82) for timelike vectors $\boldsymbol{x} = \boldsymbol{e}$ and $\boldsymbol{y} = \mathcal{U}_t(\boldsymbol{e})$, there follows the inequality

$$\big(\boldsymbol{e}, \mathcal{U}_t(\boldsymbol{e})\big)^2 \geq \big(\boldsymbol{e}^2\big) \cdot \big(\mathcal{U}_t(\boldsymbol{e})^2\big) > 0,$$

showing that $(\boldsymbol{e}, \mathcal{U}_t(\boldsymbol{e}))$ cannot be zero for any value of t.

Thus taking into account Theorem 7.63, we see that the number of equivalence classes of Lorentz transformations is certainly not less than four. Now we shall show that there are exactly four. To begin with, we shall establish this for a pseudo-Euclidean plane, and thereafter shall prove it for a pseudo-Euclidean space of arbitrary dimension.

Example 7.68 The matrices (7.91), (7.92) presenting all possible Lorentz transformations of a pseudo-Euclidean plane can be continuously deformed into the matrices

$$\begin{aligned} E &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad & F_1 &= \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \\ F_2 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad & F_3 &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \tag{7.99}$$

respectively. Indeed, we obtain the necessary continuous deformation if in the matrices (7.91), (7.92) we replace the parameter ψ by $(1-t)\psi$, where $t \in [0,1]$. It is also clear that none of the four matrices (7.99) can be continuously deformed into any of the others: any two of them differ either by the signs of their determinants or in that one of them preserves the poles of the light cone, while the other causes them to exchange places.

In the general case, we have an analogue of Theorem 7.28.

Theorem 7.69 *Two Lorentz transformations $\mathcal{U}_1$ and $\mathcal{U}_2$ of a real pseudo-Euclidean space are continuously deformable into each other if and only if $\varepsilon(\mathcal{U}_1) = \varepsilon(\mathcal{U}_2)$.*

Proof Just as in the case of Theorem 7.28, we begin with a more specific assertion: we shall show that an arbitrary Lorentz transformation $\mathcal{U}$ for which

$$\varepsilon(\mathcal{U}) = \big(|\mathcal{U}|, \nu(\mathcal{U})\big) = (+1, +1) \tag{7.100}$$

holds can be continuously deformed into $\mathcal{E}$. Invoking Theorem 7.65, let us examine the orthogonal decomposition (7.93), denoting by $\mathcal{U}_i$ the restriction of the transformation $\mathcal{U}$ to the invariant subspace L_i, where $i = 0, 1$. We shall investigate three cases in turn.

Case 1. In the decomposition (7.93), the dimension of the subspace L_1 is equal to 1, that is, $\mathsf{L}_1 = \langle \boldsymbol{e} \rangle$, where $(\boldsymbol{e}^2) < 0$. Then to the subspace L_1, there corresponds in the matrix of the transformation $\mathcal{U}$ a block of order 1 with $\sigma = +1$ or -1, and $\mathcal{U}_0$ is an orthogonal transformation that depending on the sign of σ, can be proper or improper, so that the condition $|\mathcal{U}| = \sigma|\mathcal{U}_0| = 1$ is satisfied. However, it is easy to see that for $\sigma = -1$, we have $\nu(\mathcal{U}) = -1$ (since $(\boldsymbol{e}, \mathcal{U}(\boldsymbol{e})) > 0$), and therefore, the condition (7.100) leaves only the case $\sigma = +1$, and consequently, the orthogonal transformation $\mathcal{U}_0$ is proper. Then $\mathcal{U}_1$ is the identity transformation (of a one-dimensional space). By Theorem 7.28, an orthogonal transformation $\mathcal{U}_0$ is continuously deformable into the identity, and therefore, the transformation $\mathcal{U}$ is continuously deformable into $\mathcal{E}$.

Case 2. In the decomposition (7.93), the dimension of the subspace L_1 is equal to 2, that is, L_1 is a pseudo-Euclidean plane. Then as we established in Examples 7.62 and 7.68, in some orthonormal basis of the plane L_1, the matrix of the transformation $\mathcal{U}_1$ has the form (7.92) and is continuously deformable into one of the four matrices (7.99). It is obvious that the condition $\nu(\mathcal{U}) = 1$ is associated with only the matrix E and one of the matrices F_2, F_3, namely the one in which the eigenvalues ± 1 correspond to the eigenvectors $\boldsymbol{g}_\pm$ in such a way that $(\boldsymbol{g}_+^2) < 0$ and $(\boldsymbol{g}_-^2) > 0$. In this case, it is obvious that we have the orthogonal decomposition $\mathsf{L}_1 = \langle \boldsymbol{g}_+ \rangle \oplus \langle \boldsymbol{g}_- \rangle$.

If the matrix of the transformation $\mathcal{U}_1$ is continuously deformable into E, then the orthogonal transformation $\mathcal{U}_0$ is proper, and it follows that it is also continuously deformable into the identity, which proves our assertion.

If the matrix of the transformation $\mathcal{U}_1$ is continuously deformable into F_2 or F_3, then the orthogonal transformation $\mathcal{U}_0$ is improper, and consequently, its matrix is continuously deformable into the matrix (7.32), which has the eigenvalue -1 corresponding to some eigenvector $\boldsymbol{h} \in \mathsf{L}_0$. From the orthogonal decomposition $\mathsf{L} = \mathsf{L}_0 \oplus \langle \boldsymbol{g}_+ \rangle \oplus \langle \boldsymbol{g}_- \rangle$, taking into account $(\boldsymbol{g}_+^2) < 0$, it follows that the invariant plane $\mathsf{L}' = \langle \boldsymbol{g}_-, \boldsymbol{h} \rangle$ is a Euclidean space. The matrix of the restriction of $\mathcal{U}$ to L' is equal to $-E$, and is therefore continuously deformable into E. And this implies that the transformation $\mathcal{U}$ is continuously deformable into $\mathcal{E}$.

Case 3. In the decomposition (7.93), the subspace L_1 is a cyclic three-dimensional pseudo-Euclidean space with eigenvalue $\lambda = \pm 1$. This case was examined in detail in Example 7.67, and we will use the notation introduced there. It is obvious that the condition $\nu(\mathcal{U}) = 1$ is satisfied only for $\lambda = 1$, since otherwise, the transformation $\mathcal{U}_1$ takes the lightlike eigenvector $\boldsymbol{e}_3$ to $-\boldsymbol{e}_3$, that is, it transposes the poles of the light cone. Thus condition (7.100) corresponds to the Lorentz transformation $\mathcal{U}_1$ with the value $\varepsilon(\mathcal{U}_1) = (+1, +1)$ and proper orthogonal transformation $\mathcal{U}_0$.

Let us show that such a transformation $\mathcal{U}_1$ is continuously deformable into the identity. Since $\mathcal{U}_0$ is obviously also continuously deformable into the identity, this will give us the required assertion.

Thus let $\lambda = 1$. We shall fix in L_1 a basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ satisfying the following conditions introduced in Example 7.67:

$$\begin{gathered} \left(\boldsymbol{e}_1^2\right) = a, \qquad (\boldsymbol{e}_1, \boldsymbol{e}_2) = -\frac{d}{2}, \\ (\boldsymbol{e}_1, \boldsymbol{e}_3) = -d, \qquad \left(\boldsymbol{e}_2^2\right) = d, \qquad (\boldsymbol{e}_2, \boldsymbol{e}_3) = \left(\boldsymbol{e}_3^2\right) = 0 \end{gathered} \tag{7.101}$$

with some numbers a and $d > 0$. The Gram matrix A in this basis has the form (7.98) with $c = -d$ and $b = -d/2$, while the matrix U_1 of the transformation $\mathcal{U}_1$ has the form of a Jordan block.

Let $\mathcal{U}_t$ be a linear transformation of the space L_1 whose matrix in the basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ has the form

$$U_t = \begin{pmatrix} 1 & 0 & 0 \\ t & 1 & 0 \\ \varphi(t) & t & 1 \end{pmatrix}, \tag{7.102}$$

where t is a real parameter taking values from 0 to 1, and $\varphi(t)$ is a continuous function of t that we shall choose in such a way that $\mathcal{U}_t$ is a Lorentz transformation. As we know, for this, the relationship (7.85) with matrix $U = U_t$ must be satisfied. Substituting in the equality $U_t^* A U_t = A$ the matrix A of the form (7.98) with $c = -d$ and $b = -d/2$ and matrix U_t of the form (7.102) and equating corresponding elements on the left- and right-hand sides, we obtain that the equality $U_t^* A U_t = A$ holds if $\varphi(t) = t(t-1)/2$. For such a choice of function $\varphi(t)$, we obtain a family of Lorentz transformations $\mathcal{U}_t$ depending continuously on the parameter $t \in [0, 1]$. Moreover, it is obvious that for $t = 1$, the matrix U_t has the Jordan block U_1, while for $t = 0$, the matrix U_t equals E. Thus the family $\mathcal{U}_t$ effects a continuous deformation of the transformation $\mathcal{U}_1$ into $\mathcal{E}$.

Now let us prove the assertion of Theorem 7.69 in general form. Let $\mathcal{W}$ be a Lorentz transformation with arbitrary $\varepsilon(\mathcal{W})$. We shall show that it can be continuously deformed into the transformation $\mathcal{F}$, having in some orthonormal basis the block-diagonal matrix

$$F = \begin{pmatrix} E & 0 \\ 0 & F' \end{pmatrix},$$

where E is the identity matrix of order $n-2$ and F' is one of the four matrices (7.99). It is obvious that by choosing a suitable matrix F', we may obtain the Lorentz transformation $\mathcal{F}$ with any desired $\varepsilon(\mathcal{F})$. Let us select the matrix F' in such a way that $\varepsilon(\mathcal{F}) = \varepsilon(\mathcal{W})$.

Let us select in our space an arbitrary orthonormal basis, and in that basis, let the transformation $\mathcal{W}$ have matrix W. Then the transformation $\mathcal{U}$ having in this same basis the matrix $U = WF$ is a Lorentz transformation, and moreover, by our choice of $\varepsilon(\mathcal{F}) = \varepsilon(\mathcal{W})$, we have the equality $\varepsilon(\mathcal{U}) = \varepsilon(\mathcal{W})\varepsilon(\mathcal{F}) = (+1, +1)$. Further, from the trivially verified relationship $F^{-1} = F$, we obtain $W = UF$, that is, $\mathcal{W} = \mathcal{U}\mathcal{F}$. We shall now make use of a family $\mathcal{U}_t$ that effects the continuous deformation of the transformation $\mathcal{U}$ into $\mathcal{E}$. From the equality $\mathcal{W} = \mathcal{U}\mathcal{F}$, with the help of Lemma 4.37, we obtain the relationship $\mathcal{W}_t = \mathcal{U}_t\mathcal{F}$, in which $\mathcal{W}_0 = \mathcal{E}\mathcal{F} = \mathcal{F}$ and $\mathcal{W}_1 = \mathcal{U}\mathcal{F} = \mathcal{W}$. Thus it is this family $\mathcal{W}_t = \mathcal{U}_t\mathcal{F}$ that accomplishes the deformation of the Lorentz transformation $\mathcal{W}$ into $\mathcal{F}$.

If $\mathcal{U}_1$ and $\mathcal{U}_2$ are Lorentz transformations such that $\varepsilon(\mathcal{U}_1) = \varepsilon(\mathcal{U}_2)$, then by what we showed earlier, each of them is continuously deformable into $\mathcal{F}$ with one and the same matrix F'. Consequently, by transitivity, the transformations $\mathcal{U}_1$ and $\mathcal{U}_2$ are continuously deformable into each other. □

Similarly to what we did in Sects. 4.4 and 7.3 for nonsingular and orthogonal transformations, we can express the fact established by Theorem 7.69 in topological form: the set of Lorentz transformations of a pseudo-Euclidean space of a given dimension has *exactly four* path-connected components. They correspond to the four possible values of $\varepsilon(\mathcal{U})$.

Let us note that the existence of four (instead of two) orientations is not a specific property of pseudo-Euclidean spaces with the quadratic form (7.76), as was the case with the majority of properties of this section. It holds for all vector spaces with a bilinear inner product $(\boldsymbol{x}, \boldsymbol{y})$, provided that it is nonsingular and the quadratic form $(\boldsymbol{x}^2)$ is neither positive nor negative definite. We can indicate (without pretending to provide a proof) the reason for this phenomenon. If the form $(\boldsymbol{x}^2)$, in canonical form, appears as

$$x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_n^2, \quad \text{where } s \in \{1, \ldots, n-1\},$$

then the transformations that preserve it include first of all, the orthogonal transformations preserving the form $x_1^2 + \cdots + x_s^2$ and not changing the coordinates $x_{s+1}, \ldots, x_n$, and secondly, the transformations preserving the quadratic form $x_{s+1}^2 + \cdots + x_n^2$ and not changing the coordinates $x_1, \ldots, x_s$. Every type of transformation is "responsible" for its own orientation.

Linear Algebra and Geometry
Shafarevich, I.R.; Remizov, A.
2013, XXII, 526 p., Hardcover
ISBN: 978-3-642-30993-9