Chapter 12
Hyperbolic Geometry

The discovery of hyperbolic (or Lobachevskian) geometry had an enormous impact on the development of mathematics and on how the relationship between mathematics and the real world was understood. The discussions that swirled around the new geometry also seem to have influenced the views of many in the humanities, who, in this regard, unfortunately were too much taken by a literary image: the contrast between "down-to-earth" Euclidean geometry and the "otherworldly" non-Euclidean geometry invented by learned mathematicians. It seemed that the difference between the two geometries was that in the first geometry, as was clear to everyone, parallel lines did not intersect, while in the second, what to normal intelligence was difficult of comprehension, they do intersect. However, of course, this is exactly the opposite of the truth: in the non-Euclidean geometry of Lobachevsky, given a point external to a given line, it is possible for *infinitely many* lines to pass through the point without intersecting the line. It is this that distinguishes Lobachevsky's geometry from that of Euclid.

Ivan Karamazov, in Dostoevsky's novel *The Brothers Karamazov*, likely sowed confusion among those in the humanities with the following literary image:

> At the same time there were and are even now geometers and philosophers, even some of the most outstanding among them, who doubt that the whole universe, or, even more broadly, the whole of being, was created purely in accordance with Euclidean geometry; they even dare to dream that two parallel lines, which according to Euclid cannot possibly meet on earth, may perhaps meet somewhere in infinity.

Around the time this novel was being written, Friedrich Engels wrote *Anti-Dühring*, where an even more vivid image is used:

> But in higher mathematics, another contradiction is achieved, that lines that intersect before our eyes, nevertheless a mere five or six centimeters from their point of intersection are to be considered parallel, that is, lines that cannot intersect even when extended to infinity.

In this, the author sees the manifestation of some sort of "dialectic."

And even up to the present, it is possible to encounter, in print, such literary images that oppose Euclidean and non-Euclidean geometries by saying that in the former, parallel lines do not intersect, while in the latter, they "intersect somewhere

I.R. Shafarevich, A.O. Remizov, *Linear Algebra and Geometry*,
DOI 10.1007/978-3-642-30994-6_12, © Springer-Verlag Berlin Heidelberg 2013

or other." Usually, by non-Euclidean geometry is meant the hyperbolic geometry of Lobachevsky, which is quite understandable by anyone who has passed a college course in some technical subject, and there are many such people today. To be sure, nowadays, this is presented in mathematics departments in more advanced courses in differential geometry. But hyperbolic geometry is so tightly linked to a first course in linear algebra, that it would be a pity not to say something about it here.

12.1 Hyperbolic Space*

In this chapter we shall be dealing exclusively with *real* vector spaces.

We shall define *hyperbolic space* of dimension n, which we shall hereinafter denote by $\mathbb{L}_n$ or simply $\mathbb{L}$ if we do not need to indicate the dimension, as a part of n-dimensional projective space $\mathbb{P}(\mathsf{L})$, where L is a real vector space of dimension $n+1$. We shall denote the dimension of the space $\mathbb{L}$ by $\dim \mathbb{L}$.

Let us equip L with a pseudo-Euclidean product $(\boldsymbol{x}, \boldsymbol{y})$; see Sect. 7.7. Let us recall that there, the quadratic form $(\boldsymbol{x}^2)$ has index of inertia n, and in some basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{n+1}$ (called orthonormal) for the vector

$$\boldsymbol{x} = \alpha_1 \boldsymbol{e}_1 + \cdots + \alpha_n \boldsymbol{e}_n + \alpha_{n+1} \boldsymbol{e}_{n+1}, \tag{12.1}$$

it takes the form

$$\left(\boldsymbol{x}^2\right) = \alpha_1^2 + \cdots + \alpha_n^2 - \alpha_{n+1}^2. \tag{12.2}$$

In the pseudo-Euclidean space L, let us consider the light cone V defined by the condition $(\boldsymbol{x}^2) = 0$. We say that a vector $\boldsymbol{a}$ lies *inside* the cone V if $(\boldsymbol{a}^2) < 0$ (recall that in Chap. 7, we called such vectors timelike). It is obvious that the same then holds as well for all vectors on the line $\langle \boldsymbol{a} \rangle$, since $((\alpha \boldsymbol{a})^2) = \alpha^2 (\boldsymbol{a}^2) < 0$, and we shall consider this space over the field of real numbers. Such lines are also said to lie *inside* the light cone V.

Points of the projective space $\mathbb{P}(\mathsf{L})$ corresponding to lines of the space L lying inside the light cone V are called *points* of the space $\mathbb{L}$. Consequently, they correspond to those lines $\langle \boldsymbol{x} \rangle$ of the space L that in the form (12.1) satisfy the inequality

$$\alpha_1^2 + \cdots + \alpha_n^2 < \alpha_{n+1}^2. \tag{12.3}$$

In view of condition (12.3), the set $\mathbb{L} \subset \mathbb{P}(\mathsf{L})$ is contained in *one* affine subset $\alpha_{n+1} \neq 0$ (see Sect. 9.1). Indeed, in the case $\alpha_{n+1} = 0$, we would obtain in (12.3) the inequality $\alpha_1^2 + \cdots + \alpha_n^2 < 0$, which is impossible in view of the fact that $\alpha_1, \dots, \alpha_n$ are real. As we did previously in Sect. 9.1, we can identify the affine subset $\alpha_{n+1} \neq 0$ with the affine subspace $E : \alpha_{n+1} = 1$ and hence view $\mathbb{L}$ as a part of E; see Fig. 12.1.

The space of vectors of the affine space E is the vector subspace $\mathsf{E}_0 \subset \mathsf{L}$ defined by the condition $\alpha_{n+1} = 0$. In other words, $\mathsf{E}_0 = \langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_n \rangle$. Let us note that the space of vectors E_0 is not simply a vector space. As a subspace of the pseudo-Euclidean space L, it would seem that it should also be a pseudo-Euclidean space.

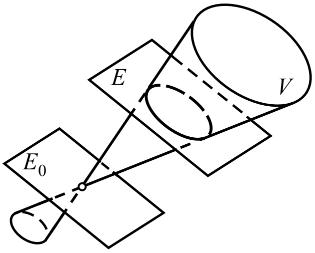


Fig. 12.1 Model of hyperbolic space

But in fact, as can be seen from formula (12.2), the inner product $(\boldsymbol{x}, \boldsymbol{y})$ makes it a *Euclidean* space, in which the vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ form an orthonormal basis. This means that E is an affine Euclidean space, and the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{n+1}$ of the space L forms within it a frame of reference with respect to which a point of the hyperbolic space $\mathbb{L} \subset E$ with coordinates $(y_1, \dots, y_n)$ is characterized by the relationship

$$y_1^2 + \cdots + y_n^2 < 1, \qquad y_i = \frac{\alpha_i}{\alpha_{n+1}}, i = 1, \dots, n. \tag{12.4}$$

This set is called the *interior* of the unit sphere in E and will be denoted by U.

Let us now turn our attention to identifying the subspaces of a hyperbolic space. They correspond to those vector spaces $\mathsf{L}' \subset \mathsf{L}$ that have a common point with the interior of the light cone V, that is, they contain a timelike vector $\boldsymbol{a} \in \mathsf{L}'$. The inner product $(\boldsymbol{x}, \boldsymbol{y})$ defined in L is clearly also defined for all vectors in the subspace $\mathsf{L}' \subset \mathsf{L}$. The space L' contains the timelike vector $\boldsymbol{a}$, and therefore, by Lemma 7.53, it is a pseudo-Euclidean space, and therefore, the associated hyperbolic space $\mathbb{L}' \subset \mathbb{P}(\mathsf{L}')$ is defined. Since $\mathbb{P}(\mathsf{L}') \subset \mathbb{P}(\mathsf{L})$ is a projective subspace, it follows that $\mathbb{L}' \subset \mathbb{P}(\mathsf{L})$. But hyperbolic space $\mathbb{L}'$ is defined by the condition $(\boldsymbol{x}^2) < 0$ both in $\mathbb{P}(\mathsf{L})$ and in $\mathbb{P}(\mathsf{L}')$, and therefore, $\mathbb{L}' \subset \mathbb{L}$. Here by definition, $\dim \mathbb{L}' = \dim \mathbb{P}(\mathsf{L}') = \dim \mathsf{L}' - 1$. The hyperbolic space $\mathbb{L}'$ thus constructed is called a *subspace* in $\mathbb{L}$.

In particular, if L' is a hyperplane in L, then $\dim \mathbb{L}' = \dim \mathbb{L} - 1$, and then the subspace $\mathbb{L}' \subset \mathbb{L}$ is called a *hyperplane* in $\mathbb{L}$.

In the sequel we shall require the partition of $\mathbb{L}$ into two parts by the hyperplane $\mathbb{L}' \subset \mathbb{L}$:

$$\mathbb{L} \setminus \mathbb{L}' = \mathbb{L}^+ \cup \mathbb{L}^-, \qquad \mathbb{L}^+ \cap \mathbb{L}^- = \varnothing, \tag{12.5}$$

similar to how in Sect. 3.2, the partition of the vector space L into two half-spaces was accomplished with the help of the hyperplane $\mathsf{L}' \subset \mathsf{L}$.

The partition (12.5) of the space $\mathbb{L}$ cannot be accomplished by an analogous partition of the projective space $\mathbb{P}(\mathsf{L})$. Indeed, if we use the definition of the subsets L^+ and L^- from Sect. 3.2, then we see that for a vector $\boldsymbol{x} \in \mathsf{L}^+$, the vector $\alpha\boldsymbol{x}$ is in L^- if $\alpha < 0$, so that the condition $\boldsymbol{x} \in \mathsf{L}^+$ does not hold for the line $\langle \boldsymbol{x} \rangle$. But such a partition is possible for the affine Euclidean space E; it was constructed in Sect. 8.2 (see p. 299).

Let us recall that the partition of the affine space E by the hyperplane $E' \subset E$ was defined via the partition of the space of vectors E_0 of the affine space E with

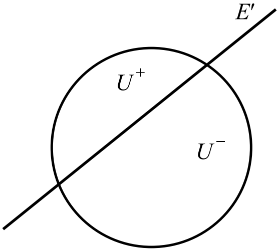


Fig. 12.2 Hyperbolic half-spaces

the aid of the hyperplane $\mathsf{E}_0' \subset \mathsf{E}_0$ corresponding to the affine hyperplane E', that is, consisting of vectors $\overrightarrow{AB}$, where A and B are all possible points of E'. If we are given a partition $\mathsf{E}_0 \setminus \mathsf{E}_0' = \mathsf{E}_0^+ \cup \mathsf{E}_0^-$, then we must choose an arbitrary point $O \in E'$ and define E^+ as the collection of all points $A \in E$ such that $\overrightarrow{OA} \in \mathsf{E}_0^+$ (E^- is defined analogously). The sets E^+ and E^- thus obtained are called *half-spaces*, and they do not depend on the choice of point $O \in E'$. Thus we have partitioned the set $E \setminus E'$ into two half-spaces: $E \setminus E' = E^+ \cup E^-$.

Let L' be a hyperplane in the pseudo-Euclidean space L having nonempty intersection with the interior of the light cone V, and let E' be the associated hyperplane in the affine space E, that is, $E' = E \cap \mathbb{P}(\mathsf{L}')$. Then E' has nonempty intersection with the interior of the unit sphere U, given by relationship (12.4), and for the set $\mathbb{L} \subset E$, we obtain the partition (12.5), where

$$\mathbb{L}' = \mathbb{L} \cap E', \qquad \mathbb{L}^+ = E^+ \cap \mathbb{L}, \qquad \mathbb{L}^- = E^- \cap \mathbb{L}. \tag{12.6}$$

The sets $\mathbb{L}^+$ and $\mathbb{L}^-$ defined by relationships (12.6) are called *half-spaces* of the space $\mathbb{L}$.

To put it more simply, the hyperplane E' divides the interior of the sphere $U \subset E$ identified with the space $\mathbb{L}$ into two parts, U^+ and U^- (see Fig. 12.2), which correspond to the half-spaces $\mathbb{L}^+$ and $\mathbb{L}^-$.

Let us show that both half-spaces $\mathbb{L}^+$ and $\mathbb{L}^-$ are nonempty, although Fig. 12.2 is sufficiently convincing by itself. We give the proof for $\mathbb{L}^+$ (for $\mathbb{L}^-$, the proof is similar).

Let us consider an arbitrary point $O \in E' \cap \mathbb{L}$. It corresponds to the vector $\boldsymbol{a} = \alpha_1 \boldsymbol{e}_1 + \cdots + \alpha_n \boldsymbol{e}_n + \boldsymbol{e}_{n+1}$ with $(\boldsymbol{a}^2) < 0$ (see the definition of the affine space E on p. 434). Let $\boldsymbol{c} \in \mathsf{E}_0^+$ and $B \in E^+$ be points such that $\overrightarrow{OB} = \boldsymbol{c}$. Let us consider vectors $\boldsymbol{b}_t = \boldsymbol{a} + t\boldsymbol{c} \in \mathsf{L}$ and points $B_t \in E$ for which $\overrightarrow{OB_t} = \boldsymbol{b}_t$ for varying values of $t \in \mathbb{R}$. Let us note that if $t > 0$, then $B_t \in E^+$, and if here $(\boldsymbol{b}_t^2) < 0$, then $B_t \in E^+ \cap \mathbb{L} = \mathbb{L}^+$. As can be seen without difficulty, the scalar square $(\boldsymbol{b}_t^2)$ is a quadratic trinomial in t:

$$\bigl(\boldsymbol{b}_t^2\bigr) = \bigl((\boldsymbol{a} + t\boldsymbol{c})^2\bigr) = \bigl(\boldsymbol{a}^2\bigr) + 2t(\boldsymbol{a}, \boldsymbol{c}) + t^2\bigl(\boldsymbol{c}^2\bigr) = P(t). \tag{12.7}$$

By our selection, the vector $\boldsymbol{c} \neq 0$ belongs to the Euclidean space E_0, and therefore, $(\boldsymbol{c}^2) > 0$. On the other hand, by assumption, we have $(\boldsymbol{a}^2) < 0$. This yields that the discriminant of the quadratic trinomial $P(t)$ on the right-hand side of relationship (12.7) is positive, and therefore, $P(t)$ has two real roots, t_1 and t_2, and from the condition $(\boldsymbol{a}^2) < 0$ it follows that they have different signs, that is, $t_1 t_2 < 0$. Then, as is easy to see, $P(t) < 0$ for every t between the roots t_1 and t_2. We will choose a positive such number t.

Since the hyperbolic space $\mathbb{L}$ can be viewed as a part of the affine space E, then from E we can transfer onto $\mathbb{L}$ the notion of line segment, the notion of *lying between* for three points on a line segment, and the notion of convexity. An easy verification (analogous to what we did at the end of Sect. 8.2) shows that the subsets $\mathbb{L}^+$ and $\mathbb{L}^-$ introduced earlier of the set $\mathbb{L} \setminus \mathbb{L}'$ are characterized by the property of convexity: if two points A, B are in $\mathbb{L}^+$, then all points lying on the segment $[A, B]$ are also in $\mathbb{L}^+$ (the same clearly holds for the subset $\mathbb{L}^-$).

Let us consider linear transformations $\mathcal{A}$ of a vector space L that are Lorentz transformations with respect to a symmetric bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y})$ corresponding to the quadratic form $(\boldsymbol{x}^2)$ and the associated projective transformations $\mathbb{P}(\mathcal{A})$. The latter transformations obviously take the set $\mathbb{L}$ to itself: given that a transformation $\mathcal{A}$ is a Lorentz transformation and from the condition $(\boldsymbol{x}^2) < 0$, it follows that $(\mathcal{A}(\boldsymbol{x})^2) = (\boldsymbol{x}^2) < 0$. The transformations of the set $\mathbb{L}$ that arise in this way are called *motions* of the hyperbolic space $\mathbb{L}$.

Thus motions of the space $\mathbb{L}$ are projective transformations of the projective space $\mathbb{P}(\mathsf{L})$ containing $\mathbb{L}$ and taking the quadratic form $(\boldsymbol{x}^2)$ into itself. By what we have said thus far, the definition of the interior of the light cone V can be written in homogeneous coordinates in the form

$$x_1^2 + \cdots + x_n^2 - x_{n+1}^2 < 0, \tag{12.8}$$

and in inhomogeneous coordinates $y_i = x_i / x_{n+1}$ in the form

$$y_1^2 + \cdots + y_n^2 < 1. \tag{12.9}$$

We consider motions of a hyperbolic space as transformations of the set $\mathbb{L}$, that is, as transformations taking the interior of the unit sphere given by condition (12.9) into itself.

Let us write down some simple properties of motions:

Property 12.1 The sequential application (composition) of two motions f_1 and f_2 (as transformations of the set $\mathbb{L}$) is again a motion.

This follows at once from the fact that the composition of nonsingular transformations $\mathcal{A}_1$ and $\mathcal{A}_2$ is a nonsingular transformation, and this holds as well for the corresponding projective transformations $\mathbb{P}(\mathcal{A}_1)$ and $\mathbb{P}(\mathcal{A}_2)$. Moreover, if $\mathcal{A}_1$ and $\mathcal{A}_2$ are Lorentz transformations with respect to the bilinear form $\varphi(\boldsymbol{x}, \boldsymbol{y})$, then the result of their composition has the same property.

Property 12.2 A motion is a bijection of $\mathbb{L}$ to itself.

This assertion follows from the fact that the corresponding transformations $\mathcal{A} : \mathsf{L} \to \mathsf{L}$ and $\mathbb{P}(\mathcal{A}) : \mathbb{P}(\mathsf{L}) \to \mathbb{P}(\mathsf{L})$ are bijections. But by the definition of a hyperbolic space, it is also necessary to verify that every line contained in the interior of the light cone V is the image of a similar such line. If we have the line $\langle \boldsymbol{a} \rangle$ with a timelike vector $\boldsymbol{a}$, then we know already that there exists a vector $\boldsymbol{b}$ such that $\mathcal{A}(\boldsymbol{b}) = \boldsymbol{a}$. Since $\mathcal{A}$ is a Lorentz transformation of a pseudo-Euclidean space L, we have the relationship $(\boldsymbol{b}^2) = (\mathcal{A}(\boldsymbol{b})^2) = (\boldsymbol{a}^2) < 0$, from which it follows that the vector $\boldsymbol{b}$ is also timelike. Thus the transformation $\mathcal{A}$ takes the line $\langle \boldsymbol{b} \rangle$ lying inside V into the line $\langle \boldsymbol{a} \rangle$, also inside V.

Property 12.3 Like every bijection, a motion f has an inverse transformation f^{-1}. It is also a motion.

The verification of this property is trivial.

At first glance, it is not obvious that there are "sufficiently many" motions of a hyperbolic space. We shall establish this a bit later, but for now, we shall point out some important types of motions.

A transformation g is of type (a) if $g = \mathbb{P}(\mathcal{A})$, where $\mathcal{A}$ is a Lorentz transformation of the space L such that $\mathcal{A}(\boldsymbol{e}_{n+1}) = \boldsymbol{e}_{n+1}$.

Since the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{n+1}$ of the pseudo-Euclidean space L is orthonormal, we have the decomposition

$$\mathsf{L} = \langle \boldsymbol{e}_{n+1} \rangle \oplus \langle \boldsymbol{e}_{n+1} \rangle^{\perp}, \quad \langle \boldsymbol{e}_{n+1} \rangle^{\perp} = \langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_n \rangle, \tag{12.10}$$

and all transformations $\mathcal{A} : \mathsf{L} \to \mathsf{L}$ with the indicated property take the subspace $\mathsf{E}_0 = \langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_n \rangle$ into itself.

Conversely, if we define $\mathcal{A} : \mathsf{L} \to \mathsf{L}$ as an orthogonal transformation of the Euclidean subspace E_0 and set $\mathcal{A}(\boldsymbol{e}_{n+1}) = \boldsymbol{e}_{n+1}$, then $\mathbb{P}(\mathcal{A})$ will of course be a motion of the hyperbolic space. In other words, these transformations can be described as orthogonal transformations of inhomogeneous coordinates. All thus constructed motions of the space $\mathbb{L}$ have the fixed point O corresponding to the line $\langle \boldsymbol{e}_{n+1} \rangle$ in L, or in other words, the point $O = (0, \dots, 0)$ in the inhomogeneous system of coordinates $(y_1, \dots, y_n)$.

From the point of view of hyperbolic space, the constructed motions precisely *coincide* with those motions that leave the point $O \in \mathbb{L}$ fixed. Indeed, as we have seen, the point O corresponds to the line $\langle \boldsymbol{e}_{n+1} \rangle$, and the motion g is equal to $\mathbb{P}(\mathcal{A})$, where $\mathcal{A}$ is a Lorentz transformation of the space L. The condition $g(O) = O$ means that $\mathcal{A}(\langle \boldsymbol{e}_{n+1} \rangle) = \langle \boldsymbol{e}_{n+1} \rangle$, that is, $\mathcal{A}(\boldsymbol{e}_{n+1}) = \lambda \boldsymbol{e}_{n+1}$. From the fact that $\mathcal{A}$ is a Lorentz transformation, it follows that $\lambda = \pm 1$. By multiplying $\mathcal{A}$ by ± 1, which obviously does not change the transformation $g = \mathbb{P}(\mathcal{A})$, we can obtain that the conditions $\mathcal{A}(\boldsymbol{e}_{n+1}) = \boldsymbol{e}_{n+1}$ are satisfied, whence by definition, it follows that g is a transformation of type (a).

Type (b) is connected with a certain line $\mathbb{L}_1 \subset \mathbb{L}$ of a hyperbolic space. By definition, the line $\mathbb{L}_1$ is determined by the plane $\mathsf{L}' \subset \mathsf{L}$, $\dim \mathsf{L}' = 2$. Since by assumption, the plane L' must contain at least one timelike vector $\boldsymbol{x}$, it follows by Lemma 7.53 (p. 271) that it is a pseudo-Euclidean space. From formula (6.28) and Theorem 6.17

(law of inertia), it follows that all such spaces of a given dimension are isomorphic. Therefore, we can choose a basis in L' with any convenient Gram matrix, provided only that it defines a pseudo-Euclidean plane. We have seen (in Example 7.49, p. 269) that it is convenient to choose as such a basis the lightlike vectors $\boldsymbol{f}_1, \boldsymbol{f}_2$, for which

$$\bigl(\boldsymbol{f}_1^2\bigr) = \bigl(\boldsymbol{f}_2^2\bigr) = 0, \qquad (\boldsymbol{f}_1, \boldsymbol{f}_2) = \frac{1}{2},$$

and this means that for every vector $\boldsymbol{x} = x\boldsymbol{f}_1 + y\boldsymbol{f}_2$, its scalar square $(\boldsymbol{x}^2)$ is equal to xy. In Example 7.61 (p. 277), we found explicit formulas for the Lorentz transformations of a pseudo-Euclidean plane in such a basis:

$$\mathcal{U}(\boldsymbol{f}_1) = \alpha \boldsymbol{f}_1, \qquad \mathcal{U}(\boldsymbol{f}_2) = \alpha^{-1} \boldsymbol{f}_2 \tag{12.11}$$

or

$$\mathcal{U}(\boldsymbol{f}_1) = \alpha \boldsymbol{f}_2, \qquad \mathcal{U}(\boldsymbol{f}_2) = \alpha^{-1} \boldsymbol{f}_1, \tag{12.12}$$

where α is an arbitrary nonzero number. In the sequel we shall need only transformations given by formula (12.11).

Since L' is a nondegenerate space, it follows that by Theorem 6.9, we have the decomposition $\mathsf{L} = \mathsf{L}' \oplus (\mathsf{L}')^{\perp}$. Let us now define a linear transformation $\mathcal{A}$ of the space L by the condition

$$\mathcal{A}(\boldsymbol{x} + \boldsymbol{y}) = \mathcal{U}(\boldsymbol{x}) + \boldsymbol{y}, \quad \text{where } \boldsymbol{x} \in \mathsf{L}', \boldsymbol{y} \in \bigl(\mathsf{L}'\bigr)^{\perp}, \tag{12.13}$$

where $\mathcal{U}$ is one of the Lorentz transformations of the pseudo-Euclidean plane L' defined by formulas (12.11) and (12.12). It is clear that then $\mathcal{A}$ is a Lorentz transformation of the space L.

A motion of type (b) of the space $\mathbb{L}$ is a transformation $\mathbb{P}(\mathcal{A})$ obtained in the case that in formula (12.13), we take as $\mathcal{U}$ the transformation given by relationships (12.11). All motions thus constructed have a *fixed line* $\mathbb{L}_1$ corresponding to the plane L'.

It is quite obvious that motions of types (a) and (b) do not exhaust all motions of the hyperbolic plane, even if in the definition of motions of type (b), as $\mathcal{U}$ in formula (12.13) we were to use transformations $\mathcal{U}$ given not only by relationships (12.11), but also by (12.12). For example, they certainly do not include motions associated with Lorentz transformations that have a three-dimensional cyclic subspace (see Corollary 7.66 and Example 7.67). However, for our further purposes, it will suffice to use only motions of these two types.

Example 12.4 In the sequel we are going to require explicit formulas for transformations of type (b) in the case of the hyperbolic plane (that is, for $n = 2$). In this case, L is a three-dimensional pseudo-Euclidean space, and in the orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$, such that

$$\bigl(\boldsymbol{e}_1^2\bigr) = 1, \qquad \bigl(\boldsymbol{e}_2^2\bigr) = 1, \qquad \bigl(\boldsymbol{e}_3^2\bigr) = -1,$$

the scalar square of the vector $\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3$ is equal to $(\boldsymbol{x}^2) = x_1^2 + x_2^2 - x_3^2$. The points of the hyperbolic plane $\mathbb{L}$ are contained in the affine plane $x_3 = 1$, have inhomogeneous coordinates $x = x_1/x_3$ and $y = x_2/x_3$, and satisfy the relationship $x^2 + y^2 < 1$.

For writing the transformation $\mathcal{A}$, let us consider the pseudo-Euclidean plane $\mathsf{L}' = \langle \boldsymbol{e}_1, \boldsymbol{e}_3 \rangle$ and let us choose in it a basis consisting of lightlike vectors $\boldsymbol{f}_1, \boldsymbol{f}_2$ associated with vectors $\boldsymbol{e}_1, \boldsymbol{e}_3$ by the relationships

$$\boldsymbol{f}_1 = \frac{\boldsymbol{e}_1 + \boldsymbol{e}_3}{2}, \qquad \boldsymbol{f}_2 = \frac{\boldsymbol{e}_1 - \boldsymbol{e}_3}{2}, \tag{12.14}$$

from which we also obtain the inverse formulas $\boldsymbol{e}_1 = \boldsymbol{f}_1 + \boldsymbol{f}_2$ and $\boldsymbol{e}_3 = \boldsymbol{f}_1 - \boldsymbol{f}_2$.

Let us note that the orthogonal complement $(\mathsf{L}')^{\perp}$ equals $\langle \boldsymbol{e}_2 \rangle$, and by Theorem 6.9, we have the decomposition $\mathsf{L} = \mathsf{L}' \oplus \langle \boldsymbol{e}_2 \rangle$. Then in accord with formula (12.13), for the vector $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$, where $\boldsymbol{x} \in \mathsf{L}'$ and $\boldsymbol{y} \in \langle \boldsymbol{e}_2 \rangle$, we obtain the value $\mathcal{A}(\boldsymbol{z}) = \mathcal{U}(\boldsymbol{x}) + \boldsymbol{y}$, where $\mathcal{U} : \mathsf{L}' \to \mathsf{L}'$ is the Lorentz transformation defined in the basis $\boldsymbol{f}_1, \boldsymbol{f}_2$ by formula (12.11). From this, taking into account expression (12.14), we obtain

$$\mathcal{U}(\boldsymbol{e}_1) = \frac{\alpha + \alpha^{-1}}{2}\boldsymbol{e}_1 + \frac{\alpha - \alpha^{-1}}{2}\boldsymbol{e}_3, \qquad \mathcal{U}(\boldsymbol{e}_3) = \frac{\alpha - \alpha^{-1}}{2}\boldsymbol{e}_1 + \frac{\alpha + \alpha^{-1}}{2}\boldsymbol{e}_3.$$

Let us set

$$a = \frac{\alpha + \alpha^{-1}}{2}, \qquad b = \frac{\alpha - \alpha^{-1}}{2}. \tag{12.15}$$

Then $a + b = \alpha$ and $a^2 - b^2 = 1$. It is obvious that any numbers a and b satisfying these relationships can be defined in terms of the number $\alpha = a + b$ by formulas (12.15). Therefore, we obtain the linear transformation $\mathcal{A} : \mathsf{L} \to \mathsf{L}$, for which

$$\mathcal{A}(\boldsymbol{e}_1) = a\boldsymbol{e}_1 + b\boldsymbol{e}_3, \qquad \mathcal{A}(\boldsymbol{e}_2) = \boldsymbol{e}_2, \qquad \mathcal{A}(\boldsymbol{e}_3) = b\boldsymbol{e}_1 + a\boldsymbol{e}_3.$$

It is easy to see that for such a transformation, the vector $\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3$ is carried to the vector

$$\mathcal{A}(\boldsymbol{x}) = (ax_1 + bx_3)\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + (bx_1 + ax_3)\boldsymbol{e}_3.$$

In inhomogeneous coordinates, $x = x_1/x_3$ and $y = x_2/x_3$. This means that a point with coordinates (x, y) is carried to the point with coordinates (x', y'), where

$$x' = \frac{ax + b}{bx + a}, \qquad y' = \frac{y}{bx + a}, \quad a^2 - b^2 = 1. \tag{12.16}$$

This particular type of motion yields, however, an important general property:

Theorem 12.5 *For every pair of points of a hyperbolic space there exists a motion taking one point into the other.*

Proof Let the first point correspond to the line $\langle \boldsymbol{a} \rangle$, and the second to the line $\langle \boldsymbol{b} \rangle$, where $\boldsymbol{a}, \boldsymbol{b} \in \mathsf{L}$. If the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are proportional, that is, $\langle \boldsymbol{a} \rangle = \langle \boldsymbol{b} \rangle$, then our requirements will be satisfied by the identity transformation of the space $\mathbb{L}$ (which can be obtained in the form $\mathbb{P}(\mathcal{E})$, where $\mathcal{E}$ is the identity transformation of the space L).

But if $\langle \boldsymbol{a} \rangle \neq \langle \boldsymbol{b} \rangle$, that is, $\dim\langle \boldsymbol{a}, \boldsymbol{b} \rangle = 2$, then let us set $\mathsf{L}' = \langle \boldsymbol{a}, \boldsymbol{b} \rangle$. Let us consider the Lorentz transformation $\mathcal{U} : \mathsf{L}' \to \mathsf{L}'$ of type (b) given by formula (12.11), the corresponding Lorentz transformation $\mathcal{A} : \mathsf{L} \to \mathsf{L}$ defined by formula (12.13), and the projective transformation $\mathbb{P}(\mathcal{A}) : \mathbb{P}(\mathsf{L}) \to \mathbb{P}(\mathsf{L})$.

Let us show that the constructed projective transformation $\mathbb{P}(\mathcal{A})$ takes a point corresponding to the line $\langle \boldsymbol{a} \rangle$ to a point corresponding to the line $\langle \boldsymbol{b} \rangle$, that is, the linear transformation $\mathcal{A} : \mathsf{L} \to \mathsf{L}$ takes the line $\langle \boldsymbol{a} \rangle$ to the line $\langle \boldsymbol{b} \rangle$. Since vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are contained in the plane L', then by definition, it suffices for us to prove that for an appropriate choice of number α, the transformation $\mathcal{U} : \mathsf{L}' \to \mathsf{L}'$ given by formula (12.11) takes the line $\langle \boldsymbol{a} \rangle$ to the line $\langle \boldsymbol{b} \rangle$.

This is easily verified by a simple calculation using the basis $\boldsymbol{f}_1, \boldsymbol{f}_2$, given by formula (12.14), in the pseudo-Euclidean plane L'. Let us consider the timelike vectors $\boldsymbol{a} = a_1 \boldsymbol{f}_1 + a_2 \boldsymbol{f}_2$ and $\boldsymbol{b} = b_1 \boldsymbol{f}_1 + b_2 \boldsymbol{f}_2$. Since in the chosen basis, the scalar square of a vector is equal to the product of its coordinates, it follows that $(\boldsymbol{a}^2) = a_1 a_2 < 0$ and $(\boldsymbol{b}^2) = b_1 b_2 < 0$. From this, it follows in particular that all numbers a_1, a_2, b_1, b_2 are nonzero.

We obtain from formula (12.11) that $\mathcal{U}(\boldsymbol{a}) = \alpha a_1 \boldsymbol{f}_1 + \alpha^{-1} a_2 \boldsymbol{f}_2$, and the condition $\langle \mathcal{U}(\boldsymbol{a}) \rangle = \langle \boldsymbol{b} \rangle$ means that $\mathcal{U}(\boldsymbol{a}) = \mu \boldsymbol{b}$ for some $\mu \neq 0$. This yields the relationships $\alpha a_1 = \mu b_1$ and $\alpha^{-1} a_2 = \mu b_2$, that is,

$$\mu = \frac{\alpha a_1}{b_1}, \qquad a_2 = \alpha \mu b_2 = \frac{\alpha^2 a_1 b_2}{b_1}, \qquad \alpha^2 = \frac{a_2 b_1}{a_1 b_2} = \frac{a_1 a_2 b_1 b_2}{(a_1 b_2)^2}.$$

It is obvious that the latter relationship can be solved for a real number α if $a_1 a_2 b_1 b_2 > 0$, and this inequality is satisfied, since by assumption, $a_1 a_2 < 0$ and $b_1 b_2 < 0$. □

Let us note that we have thus far not used motions of type (a). We shall need them to strengthen the theorem we have just proved. To do so, we shall make use of the notion of a flag, analogous to that introduced in Sect. 3.2 for real vector spaces.

Definition 12.6 A *flag* in a space $\mathbb{L}$ is a sequence of subspaces

$$\mathbb{L}_0 \subset \mathbb{L}_1 \subset \cdots \subset \mathbb{L}_n = \mathbb{L} \tag{12.17}$$

such that:

(a) $\dim \mathbb{L}_i = i$ for all $i = 0, 1, \ldots, n$;
(b) each pair of subspaces $(\mathbb{L}_{i+1}, \mathbb{L}_i)$ is directed.

A subspace $\mathbb{L}_i$ is a hyperplane in $\mathbb{L}_{i+1}$, and as we have seen (see formula (12.5)), it defines a partition $\mathbb{L}_{i+1}$ into two half-spaces: $\mathbb{L}_{i+1} \setminus \mathbb{L}_i = \mathbb{L}_{i+1}^+ \cup \mathbb{L}_{i+1}^-$. And as earlier, the pair $(\mathbb{L}_{i+1}, \mathbb{L}_i)$ is said to be *directed* if the order of the half-spaces is indicated, for example by denoting them by $\mathbb{L}^+_{i+1}$ and $\mathbb{L}^-_{i+1}$. Let us note that in a flag defined by the sequence (12.17), the subspace $\mathbb{L}_0$ has dimension 0, that is, it consists of a single point. We shall call this point the *center* of the flag (12.17).

Theorem 12.7 *For any two flags of a hyperbolic space, there exists a motion taking the first flag to the second. Such a motion is unique.*

Proof In the space $\mathbb{L}$, let us consider two flags Φ and Φ' with centers at the points $P \in \mathbb{L}$ and $P' \in \mathbb{L}$, respectively. Let $O \in \mathbb{L}$ be the point corresponding to the line $\langle \boldsymbol{e}_{n+1} \rangle$ in L, that is, the point with coordinates $y_1 = 0, \ldots, y_n = 0$ in relationship (12.4). By Theorem 12.5, there exist motions f and f' taking P to O and P' to O. Then the flags $f(\Phi)$ and $f'(\Phi')$ have their centers at the point O. Each flag is by definition a sequence of subspaces (12.17) in $\mathbb{L}$ to which correspond the subspaces of the vector space L. Thus to the flags $f(\Phi)$ and $f'(\Phi')$ there correspond two sequences of vector subspaces,

$$\langle \boldsymbol{e}_{n+1} \rangle = \mathsf{L}_0 \subset \mathsf{L}_1 \subset \cdots \subset \mathsf{L}_n = \mathsf{L} \quad \text{and} \quad \langle \boldsymbol{e}_{n+1} \rangle = \mathsf{L}'_0 \subset \mathsf{L}'_1 \subset \cdots \subset \mathsf{L}'_n = \mathsf{L},$$

where $\dim \mathsf{L}_i = \dim \mathsf{L}'_i = i + 1$ for all $i = 0, 1, \ldots, n$.

Let us recall that the space $\mathbb{L}$ is identified with a part of the affine Euclidean space E, namely with the interior of the unit sphere $U \subset E$ given by relationship (12.4). To investigate $\mathbb{L}$ as a part of E (see Fig. 12.1), it will be convenient for us to associate with each subspace $\mathsf{M} \subset \mathsf{L}$ containing the vector $\boldsymbol{e}_{n+1}$, the affine subspace $N \subset E$ of dimension one less containing the point O. To this end, let us first associate with each subspace $\mathsf{M} \subset \mathsf{L}$ containing the vector $\boldsymbol{e}_{n+1}$, the vector subspace $\mathsf{N} \subset \mathsf{M}$ determined by the decomposition $\mathsf{M} = \langle \boldsymbol{e}_{n+1} \rangle \oplus \mathsf{N}$. Employing notation introduced earlier, we obtain that

$$\mathsf{N} = \big(\langle \boldsymbol{e}_{n+1} \rangle^{\perp} \cap \mathsf{M}\big) = \big(\langle \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n \rangle \cap \mathsf{M}\big) \subset \langle \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n \rangle = \mathsf{E}_0,$$

that is, N is contained in the space of vectors of the affine space E. Consequently, the vector subspace $\mathsf{N} \subset \mathsf{E}_0$ determines a set of parallel affine subspaces in E that are characterized by their spaces of vectors coinciding with N. Such affine subspaces can be mapped to each other by a translation (see p. 296), and to determine one of them uniquely, it suffices simply to designate a point contained in this subspace. As such a point, we shall choose O. Then the vector subspace $\mathsf{N} \subset \mathsf{E}_0$ uniquely determines the affine subspace $N \subset E$, where clearly, $\dim N = \dim \mathsf{N} = \dim \mathsf{M} - 1$.

Thus we have established a bijection between k-dimensional vector subspaces $\mathsf{M} \subset \mathsf{L}$ containing the vector $\boldsymbol{e}_{n+1}$ and $(k-1)$-dimensional affine subspaces $N \subset E$ containing the point O. Here clearly, the notions of directedness for the pair $\mathsf{M}' \subset \mathsf{M}$ and $N' \subset N$ coincide. In particular, flags $f(\Phi)$ and $f'(\Phi')$ of the space $\mathbb{L}$ with center O correspond to two particular flags of the affine Euclidean space E with center at the point O.

By Theorem 8.40 (p. 316), in an affine Euclidean space, there exists for every pair of flags, a motion that takes the first flag to the second. Since in our case, both flags have a common center O, it follows that this motion has the fixed point O, and by Theorem 8.39, it is an orthogonal transformation $\mathcal{A}$ of the Euclidean space E_0. Let us consider $g = \mathbb{P}(\mathcal{A})$, the motion of type (a) of the space $\mathbb{L}$ corresponding to this orthogonal transformation $\mathcal{A}$. Clearly, it takes the flag $f(\Phi)$ to $f'(\Phi')$, that is, $gf(\Phi) = f'(\Phi')$. From this, we obtain that $f'^{-1}gf(\Phi) = \Phi'$, as asserted in the theorem.

It remains to prove the assertion about uniqueness in the statement of the theorem. Let f_1 and f_2 be two motions taking some flag Φ with center at the point P to the same flag, that is, such that $f_1(\Phi) = f_2(\Phi)$. Then $f = f_1^{-1} f_2$ is a motion, and $f(\Phi) = \Phi$. If we prove that f is the identity transformation, then the required equality $f_1 = f_2$ will follow.

By Theorem 12.5, there exists a motion g taking the point P to O. Let us set $\Phi' = g(\Phi)$. Then Φ' is a flag with center at the point O. From the equalities $f(\Phi) = \Phi$ and $g(\Phi) = \Phi'$ it follows that $gfg^{-1}(\Phi') = \Phi'$. Let us denote the motion gfg^{-1} by h. It clearly takes the flag Φ' to itself, and in particular, has the property that $h(O) = O$. From what we said on p. 438, it follows that h is a motion of type (a), that is, $h = \mathbb{P}(\mathcal{A})$, where $\mathcal{A}$ is a Lorentz transformation of the space L that in turn, is determined by a certain orthogonal transformation $\mathcal{U}$ of the Euclidean space E_0.

Let Φ'' be the flag in the Euclidean space E_0 corresponding to the flag Φ' of the space $\mathbb{L}$. Then from the condition $h(\Phi') = \Phi'$, it follows that $\mathcal{U}(\Phi'') = \Phi''$, that is, the transformation $\mathcal{U}$ takes the flag Φ'' to itself. Consequently (see p. 225), the transformation $\mathcal{U}$ is the identity, which yields that the motion h that it defines is the identity. From the relationship $h = gfg^{-1}$, it then follows that $gf = g$, that is, f is the identity transformation. □

Thus motions of a hyperbolic space possess the same property as that established in Sect. 8.4 (p. 317) for motions of affine Euclidean spaces. It is this that explains the special place of hyperbolic spaces in geometry. The Norwegian mathematician Sophus Lie called this property "free mobility." There exists a theorem (which we shall not only not prove, but not even formulate precisely) showing that other than the space of Euclid and the hyperbolic space of Lobachevsky, there is only one space that exhibits this property, called a Riemann space (we shall have a bit to say about this in Sect. 12.3). This assertion is called the *Helmholtz–Lie theorem*. For its formulation, it would be necessary first of all to define just what we mean here by "space," but we are not going to delve into this.

The property that we have deduced (Theorem 12.7) suffices for discussing the axiomatic foundations of hyperbolic geometry.

12.2 The Axioms of Plane Geometry*

Hyperbolic geometry arose historically as a result of the analysis of the axiomatic systems of Euclidean geometry. The viewpoint toward geometry as based on a small number of postulates from which all the remaining results are derived by way of formal proof arose in ancient Greece approximately in the sixth century B.C.E. Tradition connects this viewpoint with the name Pythagoras. An account of geometry with this point of view is contained in Euclid's *Elements* (third century B.C.E.). This point of view was accepted during the development of science in the modern era, and for a long time, geometry was taught directly from Euclid's books, and then later, there appeared simplified accounts. Moreover, this same point of view came to permeate all of mathematics and physics. In this spirit were written, for example, Newton's *The Mathematical Principles of Natural Philosophy*, known as the *Principia*. In physics and generally in the natural sciences, "laws of nature" played the role of axioms.

In mathematics, this direction of thought led to a more thorough working out of the axiom system of Euclidean geometry. Euclid divides the assertions on which his exposition is based into three types. One he calls "definitions"; another, "axioms"; and the third, "postulates" (the principle separating the last two of these is unclear to modern researchers). Many of his "definitions" also seem questionable. For example, the following: "A line is a length without width" (definitions of "length" and "width" are not given). Some "axioms" and "postulates" (we shall call all of these axioms) are simple corollaries of others, so that they could as well have been discarded. But what attracted the most attention was the "fifth postulate," which in Euclid is formulated thus:

> That if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

This axiom differs from the others in that its formulation is notably more complex. Therefore, the following question arose (probably already in antiquity): can this assertion be proved as a theorem derived from the other axioms? An enormous number of "proofs of the fifth postulate" appeared, in which, however, there was always found a logical error. These investigations nevertheless helped in clarifying the situation. For example, it was proved that in the context of the other axioms, the fifth postulate is equivalent to the following assertion about parallel lines that is now usually presented as this postulate: through every point A not lying on a line a, it is possible to construct exactly one line b parallel to a (lines a and b are said to be parallel if they do not intersect). Here the *existence* of a line b parallel to a and passing through the point A can easily be proved. The entire content of the fifth postulate is reduced to the assertion about its *uniqueness*.

Finally, at the beginning of the nineteenth century, a number of researchers, one of whom was Nikolai Ivanovich Lobachevsky (1792–1856), came up with the idea that a proof of the fifth postulate is impossible, and so its *negation* leads to a new geometry, logically no less perfect than the geometry of Euclid, even though it contains in some respects some unusual propositions and relationships.

The question could be posed more precisely as a result of the development of the axiomatic method. This was done by Moritz Pasch (1843–1930), Giuseppe Peano (1858–1932), and David Hilbert (1862–1943) at the end of the nineteenth century. In his work on the foundations of geometry, Hilbert formulated in particular the principles on which an axiomatic system is constructed. Today, such an approach has become commonplace; we used it to define vectors and Euclidean spaces. The general principle consists in fixing a certain set of *objects*, which remain undefined (for example, in the case of the definition of a vector space, these were scalars and vectors), and also in fixing certain *relations* that are to exist among these objects, which are likewise undefined (in the case of the definition of a vector space, these were addition of vectors and multiplication of a vector by a scalar). Finally, axioms are introduced that establish the specific properties of the introduced concepts (in the case of the definition of a vector space, these were enumerated in Sect. 3.1). With such a formulation, there remains only the question of *consistency* of the theory, that is, whether it is possible from the given axioms to derive simultaneously some statement as well as its negation. In the sequel, we shall introduce an axiom system for hyperbolic geometry (restriction to the case of dimension 2) and discuss the question of its consistency.

Let us begin with a discussion of axioms. The lists of axioms that Hilbert and his predecessors introduced in their early work turned out to possess certain logical defects. For example, in deduction, it turned out to be necessary to use certain assertions that were not contained among the axioms. Hilbert then supplemented his system of axioms. Later, this system of axioms was simplified for the sake of clarity. We shall use the axiom system proposed by the German geometer Friedrich Schur (1856–1932).[1] Here we shall restrict our attention (exclusively for the sake of brevity) to the axiomatics of the plane.

A *plane* is a certain set Π, whose elements A, B, and so on, are called *points*. Certain bijective mappings $f : \Pi \to \Pi$ are called *motions*. These are the fundamental objects. The *relationships* among them are expressed as follows:

(A) Certain distinguished subsets l, l', and so on, of the set Π are called *lines*. That an element $A \in \Pi$ belongs to the subset l is expressed by saying that "the point A lies on the line l" or "the line l passes through the point A."

(B) For three given points A, B, C lying on a given line l, it is specified when the point C is considered to *lie between* the points A and B. This must be specified for every line l and for every three points lying on it.

These objects and relations satisfy the conditions called *axioms*, which it is convenient to collect into several groups:

I. Axioms of relationship
   1. For every two points, there exists a line passing through them.
   2. If these points are distinct, then such a line is unique.
   3. On every line there lie at least two points.
   4. For every line, there exists a point not lying on it.

II. Axioms of order
   1. If on some line l, the point C lies between points A and B, then it is distinct from them and also lies between points B and A.

[1] Here we shall follow the ideas of Boris Nikolaevich Delaunay, or Delone (1890–1980), in his pamphlet *Elementary Proof of the Consistency of Hyperbolic Geometry*, 1956.

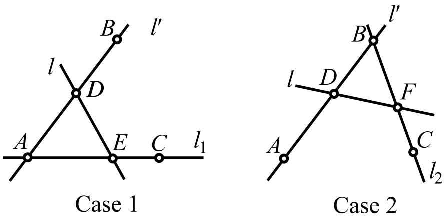


Fig. 12.3 Intersection of the *sides of a triangle* by a *line*

2. If A and C are two distinct points on some line, then on this line there is at least one point B such that C lies between points A and B.
3. Among three points A, B, and C lying on a given line, not more than one of the points lies between the two others.

Before formulating the last axiom of this group, let us give some new definitions. The set of all points C on a given line l passing through the points A and B that lie between them (including the points A and B themselves) is called a *segment* with endpoints A and B, and is denoted by $[A, B]$. Axiom 2 of group II can be reformulated thus: $[A, C] \neq l \setminus (A \cup C)$, with the inequality here being understood as an inequality of sets. That a segment $[A, B]$ contains points other than A and B is proved on the basis of the axioms of group I and the last axiom of group II, to the formulation of which we now turn. Three points A, B, C not all lying on any one line are called a *triangle*, and this relationship is denoted by $[A, B, C]$. The segments $[A, B]$, $[B, C]$, and $[C, A]$ are called the *sides* of the triangle $[A, B, C]$.

4. Pasch's axiom. If points A, B, C do not all lie on the same line, none of them belong to the line l, and the line l intersects one side of the triangle $[A, B, C]$, then it also intersects another side of the triangle.

In other words, if a line l has a point D in common with the line l' passing through points A and B, with D lying between A and B on l', then the line l either has a common point E with the line l_1 passing through B and C, with E lying between them on l_1, or has a common point F with the line l_2 passing through A and C, with F lying between them on l_2. The two cases discussed in this last axiom are depicted in Fig. 12.3.

III. Axioms of motion
1. For every motion f, the inverse mapping f^{-1} (which exists by the definition of a motion as a bijective mapping of the set Π) is also a motion.
2. The composition of two motions is a motion.
3. A motion preserves the order of points. That is, a motion f takes a line l to a line $f(l)$, and if the point C on the line l lies between points A and B on this line, then the point $f(C)$ of the line $f(l)$ lies between points $f(A)$ and $f(B)$.

The formulation of the fourth axiom of motion requires certain results that can be obtained as corollaries of the axioms of relationship and order. We shall not prove these here, but let us give only the formulations.[2]

Let us begin with properties of lines. Let us choose a point O on a line l. Points A and B on this same line, both of them different from O, are said to line *on one side* of O if O does not lie between A and B. If we select some point A different from O, then points B different from O and lying together with A on one side of O form a subset of the set of points of the line l called a *half-line* and denoted by l^+. It can be proved that if we choose in this subset another point A', then the half-line formed with it will be the same as before. Here what is important is only the choice of the point O. If we choose a point A_1 such that O lies between A and A_1, then the point A_1 determines another half-line, denoted by l^-. The half-lines l^+ and l^- determined by the points A and A_1 do not intersect, and their union is $l \setminus O$, that is, $l^+ \cap l^- = \varnothing$ and $l^+ \cup l^- = l \setminus O$.

One can verify analogous properties for a line l in the plane $\varPi$. Let us consider two points A and B that do not belong to the line l. One says that they lie *on one side* of l if either the line l' passing through them does not intersect the line l, or the lines l and l' intersect in a point C that does not lie between points A and B of the line l'. The set of points not lying on the line l and lying on the same side of l as the point A is called a *half-plane*. Again, it is possible to prove that with the choice of another point A' instead of A in this half-plane, we define the same set. There exist two points A and A' that do not belong to the same half-plane. However we select these points (given a fixed line l), we will always obtain two subsets $\varPi^+$ and $\varPi^-$ of the plane $\varPi$ such that $\varPi^+ \cap \varPi^- = \varnothing$ and $\varPi^+ \cup \varPi^- = \varPi \setminus l$.

Suppose we are given a point O and a line l passing through it. If in the partition of $l \setminus O$ into two half-lines, one of them is distinguished, and in the partition $\varPi \setminus l$ into two half-planes, one of them is distinguished (for example, let us denote them by l^+ and $\varPi^+$, respectively), then the pair (O, l) is called a *flag* and is denoted by $\varPhi$. As follows from what was discussed in Sect. 12.1, this is a special case (for $n = 2$) of the notion of a flag introduced earlier.

Every motion takes a flag to a flag, that is, if f is a motion and $\varPhi$ is the flag (O, l), then the sets $f(l)^+$ and $f(l)^-$, whose union is $f(l) \setminus f(O)$, coincide with $f(l^+)$ and $f(l^-)$, where l^+ and l^- are the half-lines on the line l determined by the point O. Here their order can change. Analogously, a pair of half-planes $f(\varPi)^+$ and $f(\varPi)^-$ defined by the line $f(l)$ coincide with the pair $f(\varPi^+)$ and $f(\varPi^-)$, where $\varPi^+$ and $\varPi^-$ are the half-planes determined by the line l. Their order also can change.

We can now formulate the last (fourth) axiom of motion:

4. Axiom of free mobility. For any two flags $\varPhi$ and $\varPhi'$, there exists a motion f taking the first flag to the second, that is, $f(\varPhi) = \varPhi'$. Such a motion is unique, and it is uniquely determined by the flags $\varPhi$ and $\varPhi'$.

[2] Some of these are proved in first courses in geometry, and in any case, elementary proofs of all of these results can be found in Chap. 2 of the book *Higher Geometry*, by N.V. Efimov (Mir, 1953).

IV. Axiom of continuity

1. Let a set of points of some line l be represented arbitrarily as the union of two sets M_1 and M_2, where no point of the set M_1 lies between two points of the set M_2, and conversely. Then there exists a point O on the line l such that M_1 and M_2 coincide with the half-lines of l determined by the point O, to either of which the point O can be joined.

This axiom is also called *Dedekind's axiom.*

Axioms I–IV that we have presented are called axioms of "absolute geometry." They hold for both Euclidean and hyperbolic geometry. These two geometries are distinguished by the addition of one axiom that deals with parallel lines. Let us recall that parallel lines are lines having no points in common. Thus in both cases, one more axiom is added:

V. Axiom of parallel lines

1. In Euclidean geometry: For every line l and every point A not lying on it, there exists at most one line l' passing through the point A and parallel to l.

1′. In hyperbolic geometry: For every line l and every point A not lying on it, there exist at least two distinct lines l' and l'' parallel to l.

The justified interest in precisely these two axioms is due to the fact that already in absolute geometry (that is, with only the axioms from groups I–IV), it is possible to prove that for every line l and every point A not on l, there exists at least one line l' passing through A and parallel to l.

It is now possible to formulate more precisely the goal that mathematics set for itself in the attempt to "prove the fifth postulate," that is, to derive assertion 1 in group V of axioms from axioms in groups I–IV. But Lobachevsky (and other researchers of the same epoch) came to the conclusion that this was impossible, and this meant that the system comprising groups I–IV and axiom 1′ was consistent.

Strictly speaking, we could have posed such questions even earlier, in connection with any of the theories that we encountered based on some system of axioms, such as the theory of vector spaces or that of Euclidean spaces. The question of the consistency of the concepts of vector spaces or Euclidean spaces is easily answered: it suffices to show (in the case of real spaces) examples of vector spaces over $\mathbb{R}^n$ of any finite dimension or Euclidean spaces with inner product $(\boldsymbol{x}, \boldsymbol{y}) = x_1y_1 + \cdots + x_ny_n$. Of course, this assumes the construction and proof of the consistency of the theory of the real numbers, but that lies outside the scope of our investigation, and we shall not consider it here. However, assuming as given that the properties of real numbers are defined and do not raise any doubts, we may, for example, say that if the system of axioms of a real vector space given in Sect. 3.1 were inconsistent, then we would be able to derive two mutually contradictory assertions about the space $\mathbb{R}^n$. However, any assertion about the space $\mathbb{R}^n$ can be reduced by definition to an assertion about the real numbers, and then we would obtain a contradiction in the domain of real numbers.

The same question could be posed in relationship to Euclidean geometry, that is, with respect to the system of axioms consisting of axioms of groups I–IV and axiom 1 of group V. Here the answer is in fact already known, since we have constructed the theory of affine Euclidean spaces (even in arbitrary dimension n). It is easily ascertained that for $n = 2$, all the axioms of Euclidean geometry that we introduced are satisfied. Some refinements are perhaps necessary only in connection with the axioms of order.

These axioms do not require an inner product on the space and are formulated for an arbitrary real affine space V in Sect. 8.2. All the assertions constituting the axioms of order now follow directly from the properties of order of the real numbers, except only Pasch's axiom. Its idea is that if a line "enters" a triangle, then it must "exit" from it. Intuitively, this is quite convincing, but with our approach, we must derive this assertion from the properties of affine spaces. It is a very simple argument, whose details we leave to the reader.

Specifically, by what is given, points A and B (we shall use the same notation as in the formulation of the axioms) lie in different half-planes into which the line l divides the plane Π. Everything depends on the half-plane to which the point C belongs: to the same one as A, or to the same one as B. In the first case, the line l has a common point with the line l_2, which lies on it between B and C, while in the second case, the common point is with the line l_1, which lies between A and C; see Fig. 12.3. In each of these two cases, the assertion of Pasch's axiom is easy to verify if we recall the definitions.

We in fact checked in one form or another that the remaining axioms are satisfied even as assertions that relate to arbitrary dimension.

We shall now turn to the axioms of hyperbolic geometry, that is, the axioms of groups I–IV and axiom $1'$ of group V. We shall prove that they are consistent, based on the consistency of the usual properties (which likewise are easily reduced to certain axioms) of the set of real numbers $\mathbb{R}$ and based on the theory of Euclidean spaces of dimension 2 and 3 constructed on this basis. On this foundation, we shall prove the following result.

Theorem 12.8 *The system of axioms of hyperbolic geometry is consistent.*

Proof We shall consider in the Euclidean plane L the open disk K (given, for example, in some coordinate system by the condition $x^2 + y^2 < 1$). We shall call the set of its points a "plane" (denoted by $\overline{\Pi}$), and we shall call "points" only the points of this disk. The intersection of every line l of the plane L with the disk K that has at least one point in common with this disk is the interior of some segment (this was proved in the previous section). We shall call such nonempty intersections $l \cap K$ "lines," denoted by $\overline{l}, \overline{l}'$, and so on. Finally, we shall call a projective transformation of the plane L taking the disk K into itself a "motion."

Since the definition of projective transformation assumes a study of the projective plane, and a projective space of dimension n and its projective transformations were defined in Chap. 9 in terms of a vector space of dimension $n + 1$, it follows that for the analysis of the hyperbolic plane, we must use here a notion connected with a three-dimensional vector space. However, it would not be difficult to give a formulation appealing only to properties of the Euclidean plane.

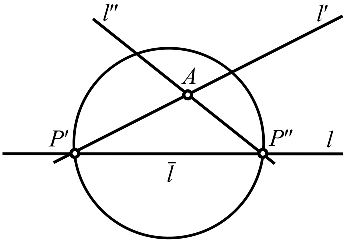


Fig. 12.4 "*Lines*" and "*points*" of the *hyperbolic plane*

Now let us define the fundamental relationships between "lines" and "points." That a "line" $\bar{l}$ passes through a "point" $A \in \overline{\Pi}$ will be understood to mean the condition that the line l passes through the point A. Thus an arbitrary "line" $\bar{l}$ is the set of "points" that lie on it. Let "points" A, B, C lie on the "line" $\bar{l}$. We shall say that a "point" C *lies between* "points" A and B if such is the case for A, B, and C as points on the Euclidean line l that contains $\bar{l}$ (this makes sense, since l is contained in Euclidean space).

It remains to verify that the notions and relationships presented satisfy the axioms of hyperbolic geometry, that is, the axioms of groups I–IV and axiom $1'$ of group V. The verification of this for the axioms of groups I, II, and IV is trivial, since the corresponding objects and relationships are defined exactly as in the surrounding Euclidean plane. For the axioms of group III (axioms of motion), the required properties were proved in the previous section (indeed, for the case of a space of arbitrary dimension n). It remains only to consider axiom $1'$ of group V.

Let $\bar{l}$ be the "line" associated with the line l in the Euclidean plane L. Then the line l intersects the boundary S of the disk K in two different points: P' and P''. Let A be a "point" of the "plane" $\overline{\Pi}$ (that is, a point of the disk K) not lying on the line l. By the axioms of Euclidean geometry, through the points A and P' in the plane L, there passes some line l'. It determines the "line" $\bar{l}' = l' \cap K$ of the "plane" $\overline{\Pi}$. Similarly, the point P'' determines the "line" $\bar{l}'' = l'' \cap K$; see Fig. 12.4.

The lines l' and l'' are distinct, since they pass through different points P' and P'' of the plane L. Therefore, by the axioms of Euclidean geometry, they have no common points other than A. But the "lines" $\bar{l}'$ and $\bar{l}''$, as nonempty segments of Euclidean lines excluding the endpoints, contain infinitely many points and in particular, the "points" $B' \in \bar{l}'$ and $B'' \in \bar{l}''$, with $B' \neq B''$. This means that the "lines" $\bar{l}'$ and $\bar{l}''$ are distinct. On the other hand, in the sense of our definitions, both of them are parallel to the "line" $\bar{l}$, that is, they have no common "points" with it (points of the disk K). For example, the line l' has with l the common point P' in the Euclidean plane L, which means that by the axioms of Euclidean geometry, they have no other common points, and in particular, no common points in the disk K.

We see that assertion $1'$ holds for every "line" $\bar{l} \subset \overline{\Pi}$ and every "point" $A \notin \bar{l}$. Let us now assume that from the axioms of hyperbolic geometry there could be derived an inconsistency (that is, some assertion and its negation). Then we could apply the same reasoning to the notions that earlier, with the proof of Theorem 12.8, we wrote in quotation marks: "point," "plane," "line," and "motion." Since they, as we have seen, satisfy all the axioms of hyperbolic geometry, we would again arrive at a contradiction. But the notions “plane,” “line,” and “motion,” and also the relationship “lies between” for three points on a line were defined in terms of Euclidean geometry. Thus we would arrive at a contradiction to Euclidean geometry itself. □

Let us focus attention on this fine logical construction: we construct objects in some domain that satisfy a certain system of axioms, and thus we prove the consistency of this system if the consistency of the domain from which the necessary objects are taken has been accepted. Today, one says that a *model* of this axiom system has thereby been constructed in another domain. In particular, we earlier constructed a model of hyperbolic geometry in the theory of vector spaces. Only by constructing such a model was the question of the provability of the “fifth postulate” decided in mathematics.

In conclusion, it is of interest to dwell a bit on the history of this question. Independent of Lobachevsky, a number of researchers came to the conclusion that a negation of the “fifth postulate” leads to a meaningful and consistent branch of mathematics, a “new geometry,” eventually given the designation “non-Euclidean geometry.” There is no question here of priority. All the researchers clearly worked independently of one another (Gauss’s correspondence from the 1820s, Lobachevsky’s publication of 1829, and János Bolyai’s of 1832). Most of these who became known later were amateurs, not professional mathematicians. But there were some exceptions: outside of Lobachevsky, there was the greatest mathematician of that epoch—Gauss. The majority of such researchers known to us who clearly arrived at the same conclusions independently became known precisely because of their correspondence with Gauss, which was published along with other of Gauss’s papers after his death. It is clear from these publications that in his youth, Gauss had attempted to prove the fifth postulate, but later concluded that there existed a meaningful and consistent geometry that did not include this postulate. In his letters, Gauss discussed the similar views of his correspondents with great interest.

He clearly received the work of Lobachevsky with sympathetic understanding when it began to appear in translation, and on Gauss’s recommendation, Lobachevsky was elected a member of the Göttingen Academy of Sciences.

In one of Gauss’s diaries can be seen the name Nikolai Ivanovich Lobachevsky, written in Cyrillic letters:

НИКОЛАЙ ИВАНОВИЧ ЛОБАЧЕВСКИЙ

But it is surprising that Gauss himself, throughout his entire life, published not a line on this subject. Why was that? The usual explanation is that Gauss was afraid of not being understood. Indeed, in one letter in which he touched on the question of the “fifth postulate” and non-Euclidean geometry, he wrote, “since I fear the clamor of the Boeotians.” But it seems that this cannot be the full explanation of his mysterious silence. In his other works, Gauss did not fear being misunderstood by his readers.[3] It is possible, however, that there is another explanation for Gauss's silence. He was one of the few who realized that however many interesting theorems of non-Euclidean geometry might be deduced, this would prove nothing definitively; there would always remain the theoretical possibility that future derivations would yield a contradictory assertion. And perhaps Gauss understood (or sensed) that at the time (first half of the nineteenth century), the mathematical concepts had not yet been developed to pose and solve this question rigorously.

Apparently, Lobachevsky was among the small number of mathematicians in addition to Gauss who understood this. For him, as with Gauss, there stood the question of "incomprehensibility." First of all, for Lobachevsky, there was the lack of comprehension among Russian mathematicians, especially analysts, who totally failed to accept his work. In any case, he constantly attempted to find a consistent foundation for his geometry. For example, he discovered its striking parallel with *spherical geometry* and expressed the idea that it was the "geometry of the sphere with imaginary radius." His geometry could indeed have been realized in the form of some other model if the very notion of model had been sufficiently developed at that time.

Beyond this (as noted by the French mathematician André Weil (1906–1998)), here we have the simplest case of *duality* between compact and noncompact symmetric spaces, discovered in the twentieth century by Élie Cartan.

Moreover, Lobachevsky proved that in three-dimensional hyperbolic space, there is a surface (called today a *horosphere*) such that if we consider only the set of its points and take as lines the curves of a specific type lying on it (called today *horocycles*), then all the axioms of Euclidean geometry are satisfied. From this it follows that if hyperbolic geometry is consistent, then Euclidean geometry is also consistent. Even if we accept the hypothesis that the "fifth postulate" does not hold, Euclidean geometry is still realized on the horosphere. Thus in principle, Lobachevsky came very close to the concept of a model. But he did not succeed in constructing a model of hyperbolic geometry in the framework of Euclidean geometry. Such a construction was not easily granted to mathematicians.

The following paragraph offers only a hint, and not a precise formulation, of the corresponding assertions.

First, in 1868, Eugenio Beltrami (1835–1899) constructed in three-dimensional Euclidean space a certain surface called a *pseudosphere* or *Beltrami surface*, whose Gaussian curvature (see the definition on p. 265) at every point is the same negative number. Hyperbolic geometry can be realized on the pseudosphere, where the role of lines is played by so-called *geodesic lines*.[4] However, here we are talking about only a piece of the pseudosphere and a piece of the hyperbolic plane. Here the posing of the question must be radically changed, since the majority of the axioms that we have given assume (as in, for example, Euclidean geometry) the possibility

[3] For example, his first published book, *Disquisitiones Arithmeticae*, was considered for a long time to be quite inaccessible.

[4] More about this can be found, for example, in the book *A Course of Differential Geometry and Topology*, by A. Mishchenko and A. Fomenko (Mir, 1988).

of continuing lines to infinity. The coincidence of two bounded pieces is understood in the sense of the coincidence of the measures of lengths and angles, about which, in the case of hyperbolic geometry, more will be said in the following section. Moreover, Hilbert later proved that the hyperbolic plane cannot in this sense be completely identified with any surface in three-dimensional space (much later it was proved that it is possible for some surface in five-dimensional space).

The model of hyperbolic geometry that we gave for the proof of Theorem 12.8 was constructed by Felix Klein (1849–1925) in 1870. The history of its appearance was also astounding. Formally speaking, this model was constructed in 1859 by the English mathematician Arthur Cayley (1821–1895). But he considered it only as a certain construction in projective geometry and apparently did not notice the connection with non-Euclidean geometry. In 1869, the young (twenty-year-old) Klein became acquainted with his work. He recalled that in 1870, he gave a talk on the work of Cayley at the seminar of the famous mathematician Weierstrass, and, as he writes, "I finished with a question whether there might exist a connection between the ideas of Cayley and Lobachevsky. I was given the answer that these two systems were conceptually widely separated." As Klein puts it, "I allowed myself to be convinced by these objections and put aside this already mature idea." However, in 1871, he returned to this idea, formulated it mathematically, and published it. But then his work was not understood by many. In particular, Cayley himself was convinced as long as he lived that there was some logical error involved. Only after several years were these ideas fully understood by mathematicians.

Of course, one can ask not only about the existence of Euclidean and hyperbolic geometries, but also about a number of different (in a certain sense) geometries. Here we shall formulate only the results that are relevant to the current discussion.[5]

First of all, we must give a precise sense to what we mean by "different" or "identical" geometries. This can be done with the help of the notion of *isomorphism* of geometries, which is analogous to the notion of isomorphism of vector spaces introduced earlier. Within the framework of a system of axioms used in this section, this can be done as follows. Let Π and Π' be two planes satisfying the axioms of groups I–IV, and let G and G' be sets of motions of the respective planes. Mappings $\varphi : \Pi \to \Pi'$ and $\psi : G \to G'$ define an isomorphism (φ, ψ) of these geometries if the following conditions are satisfied:

(1) Both mappings φ and ψ are bijections.

(2) The mapping φ takes every line l in the plane Π to some line $\varphi(l)$ in the plane Π'.

(3) The mapping φ preserves the relationship "lies between." This means that if points A, B, and C lie on the line l, with C lying between A and B, then the point $\varphi(C)$ lies between $\varphi(A)$ and $\varphi(B)$ on the line $\varphi(l)$.

(4) The mappings φ and ψ agree in the following sense: for every motion $f \in G$, its image $\psi(f)$ is equal to $\varphi f \varphi^{-1}$. This means that for every point $A \in \Pi$, the equality $(\psi(f))(\varphi(A)) = \varphi(f(A))$ holds.

[5] Their proofs are given in every course in higher geometry, for example, in the book *Higher Geometry*, by N.V. Efimov, mentioned earlier.

(5) For every motion $f \in G$, the equality $\psi(f^{-1}) = \psi(f)^{-1}$ holds, and for every pair of motions $f_1, f_2 \in G$, we have $\psi(f_1 f_2) = \psi(f_1)\psi(f_2)$.

Let us note that some of these conditions can be derived from the others, but for brevity, we shall not do this.

We shall consider geometries up to isomorphism as just described, that is, we shall consider two geometries the same if there exists an isomorphism between them. In particular, geometries with respective axioms 1 and $1'$ in group V are clearly not isomorphic to each other, that is, they are two different geometries. From this point of view, geometries (in the plane) satisfying axioms 1 and $1'$ are fundamentally different from each other. Namely, it has been proved that all geometries satisfying axiom 1 in group V are isomorphic.[6] But geometries that satisfy axiom $1'$ in group V are characterized up to isomorphism by a certain real number c called their *curvature*. This number is usually assumed to be negative, and then it can take on any value $c < 0$.

Klein suggested that Euclidean geometry can be viewed as the limiting case of hyperbolic geometry as the curvature c approaches zero.[7] As Klein further observed, if axiom 1 (of Euclid) is satisfied in our world, then we shall never know it. Since every physical measurement is taken with a certain degree of error, to establish the precise equality $c = 0$ is impossible, for there always remains the possibility that the number c is less than zero, but it is so small in absolute value that it lies beyond the limits of our measurements.

12.3 Some Formulas of Hyperbolic Geometry*

First of all, we shall define the distance between points in the hyperbolic plane using its definition as the set of points of the projective plane $\mathbb{P}(\mathsf{L})$ corresponding to the lines of the three-dimensional pseudo-Euclidean space L lying within the light cone and its interpretation as the set of points on the unit circle U in the affine Euclidean plane E; see Sect. 12.1.

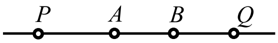
The meaning of the notion of distance is that it should be preserved under motions of the hyperbolic plane. But we have defined a motion as a certain special projective transformation $\mathbb{P}(\mathcal{A})$ of the projective plane $\mathbb{P}(\mathsf{L})$. Theorem 9.16 shows that in general, it is impossible to associate a number that does not change under an arbitrary projective transformation not only with two points, but even with three points of the projective line. But we shall use the fact that motions of the hyperbolic plane are not arbitrary projective transformations $\mathbb{P}(\mathsf{L})$, but only those that take the light cone in the space L into itself.

Namely, to two arbitrary points A and B correspond the lines $\langle \boldsymbol{a} \rangle$ and $\langle \boldsymbol{b} \rangle$, lying inside the light cone. We shall show that they determine two additional points, P

[6] Of course, here we are assuming that they all satisfy the axioms of groups I–IV.

[7] Felix Klein. *Nicht-Euklidische Geometrie*, Göttingen, 1893. Reprinted by AMS Chelsea, 2000.

Fig. 12.5 The *segment* $[PQ]$



and Q, that correspond to lines lying on the light cone. But *four points* of a projective space lying on a line already determine a number that does not change under arbitrary projective transformations, namely their cross ratio (defined in Sect. 9.3). We shall use this number for defining the distance between points A and B. This definition has the special feature that it uses points corresponding to lines lying on the light cone (P and Q), which are thus not points of the hyperbolic plane.

We shall assume that the points A and B are distinct (if they coincide, then the distance between them is zero by definition). This means that the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are linearly independent. It is obvious that then a unique projective line l passes through these points; it corresponds to the plane $\mathsf{L}' = \langle \boldsymbol{a}, \boldsymbol{b} \rangle$. The line l determines a line l' in the affine Euclidean space E, depicted in Figs. 12.1 and 12.2. Since the line l' contains the points A and B, which lie inside the circle U, it intersects its boundary in *two* points, which we shall take as P and Q. This was in fact already proved in Sect. 12.1, but we shall now repeat the corresponding argument.

The points of l are the lines $\langle \boldsymbol{x} \rangle$ consisting of all vectors proportional to the vectors $\boldsymbol{x} = \overrightarrow{OA} + t\overrightarrow{AB}$, where t is an arbitrary real number. Here the vector $\overrightarrow{OA}$ equals $\boldsymbol{a}$, and the vector $\overrightarrow{AB} = \boldsymbol{c}$ belongs to the subspace E_0 if we assume that the points A, B and the line l lie in the affine space E. This means that $\boldsymbol{x} = \boldsymbol{a} + t\boldsymbol{c}$, where the vector $\boldsymbol{c}$ can be taken as fixed, and the number t as variable. Points $\boldsymbol{x}$ at the intersection of the line l' with the light cone $V \subset \mathsf{L}$ are given by the condition $(\boldsymbol{x}^2) = 0$, that is,

$$\bigl((\boldsymbol{a} + t\boldsymbol{c})^2\bigr) = \bigl(\boldsymbol{a}^2\bigr) + 2(\boldsymbol{a}, \boldsymbol{c})t + \bigl(\boldsymbol{c}^2\bigr)t^2 = 0. \tag{12.18}$$

We know that $(\boldsymbol{a}^2) < 0$, and the vector $\boldsymbol{c}$ belongs to E_0. Since E_0 is a Euclidean space and the points A and B are distinct, it follows that $(\boldsymbol{c}^2) > 0$. From this it follows that the quadratic equation (12.18) in the unknown t has two real roots t_1 and t_2 of opposite signs. Suppose for the sake of definiteness that $t_1 < t_2$. Then for $t_1 < t < t_2$, the value of $((\boldsymbol{a} + t\boldsymbol{c})^2)$ is negative, and all points of the line l' corresponding to the values t in this interval belong to $\mathbb{L}$. We see that the line l intersects the light cone V in two points corresponding to the values $t = t_1$ and $t = t_2$, while the values $t_1 < t < t_2$ are associated with the points of the line $\mathbb{L}_1$ (that is, one-dimensional hyperbolic space) passing through A and B. Thus the line $\mathbb{L}_1$ coincides with the line segment $l \subset E$ whose endpoints are P and Q, which correspond to the values $t = t_1$ and $t = t_2$; see Fig. 12.5.

It is clear that point A is contained in the interval (P, Q). Applying the same argument to the point B, we obtain that the point B is also in the interval (P, Q).

Let us label the points P and Q in such a way that P will denote the endpoint of the interval (P, Q) that is closer (in the sense of Euclidean distance) to the point A, and by Q the endpoint that is closer to B, as depicted in Fig. 12.5.

Now it is possible to give a definition of the distance between points A and B, which we shall denote by $r(A, B)$:

$$r(A, B) = \log \mathrm{DV}(A, B, Q, P), \tag{12.19}$$

where $\mathrm{DV}(A, B, Q, P)$ is the cross ratio (see p. 337). Let us note that in the definition (12.19), we have not indicated the base of the logarithm. We could take any base greater than 1, since a change in base results simply in multiplying all distances by some fixed positive constant. But in any case, the length of a segment AB can be defined only up to a multiplicative factor that corresponds to the arbitrariness in the selection of a unit length on a line.

We shall explain a bit later why the logarithm appears in definition (12.19). The reason for using the cross ratio is explained by the following theorem.

Theorem 12.9 *The distance $r(A, B)$ does not change under any motion f of the hyperbolic plane, that is, $r(f(A), f(B)) = r(A, B)$.*

Proof The assertion of the theorem follows at once from the fact that a motion f of the hyperbolic plane is determined by a certain projective transformation $\mathbb{P}(\mathcal{A})$. This transformation $\mathbb{P}(\mathcal{A})$ carries the line l' passing through points A and B to the line passing through the points $\mathbb{P}(\mathcal{A})(A)$ and $\mathbb{P}(\mathcal{A})(B)$. This means that the transformation takes the points P and Q, the intersection of the line l' with the boundary of the disk U, to the points P' and Q', the intersection of the line $\mathbb{P}(\mathcal{A})(l')$ with this boundary. That is, $P' = \mathbb{P}(\mathcal{A})(P)$ and $Q' = \mathbb{P}(\mathcal{A})(Q)$, or conversely, $Q' = \mathbb{P}(\mathcal{A})(P)$ and $P' = \mathbb{P}(\mathcal{A})(Q)$. Moreover, the transformation $\mathbb{P}(\mathcal{A})$ preserves the cross ratio of four points on a line (Theorem 9.17). □

To explain the role of the cross ratio, we jumped a bit ahead and skipped the verification that the argument of the logarithm in formula (12.19) was a number greater than 1 and also that in the definition of $r(A, B)$, all the conditions entering into the definition of a distance (p. xvii) were satisfied. We now return to this.

Let us assume that the points P, A, B, Q are arranged in the order shown in Fig. 12.5. For the cross product, we may use formula (9.28),

$$\mathrm{DV}(A, B, Q, P) = \frac{|AQ| \cdot |PB|}{|BQ| \cdot |PA|} > 1, \tag{12.20}$$

since clearly, $|AQ| > |BQ|$ and $|PB| > |PA|$. Therefore, the argument of the logarithm in formula (12.19) is a number greater than 1, and so the logarithm is a positive real number. Therefore, $r(A, B) > 0$ for all pairs of distinct points A and B.

Let us note that it would be possible to make do without the order of the points P and Q that we chose. For this, it would be sufficient to verify (this follows directly from the definition of the cross ratio) that under a transposition of the points P and Q, the cross ratio d is converted into $1/d$. Thus the logarithm (12.19) that gives the distance is defined up to sign, and we can define the distance as the absolute value.

If we interchange the positions of A and B, then the points P and Q defined in the agreed-upon way also exchange places. It is easy to verify that the cross ratio determines a distance according to formula (12.19) that will not change. In other words, we have the equality

$$r(B,A)=r(A,B). \tag{12.21}$$

For any third point C collinear with A and B and lying between them, the condition

$$r(A,B)=r(A,C)+r(C,B) \tag{12.22}$$

is satisfied. It follows from the fact that (in the notation we have adopted)

$$\mathrm{DV}(A,B,Q,P)=\frac{|AQ|\cdot|BP|}{|BQ|\cdot|AP|}=\mathrm{DV}(A,C,Q,P)\cdot\mathrm{DV}(C,B,Q,P), \tag{12.23}$$

since

$$\mathrm{DV}(A,C,Q,P)=\frac{|AQ|\cdot|CP|}{|CQ|\cdot|AP|},\qquad \mathrm{DV}(C,B,Q,P)=\frac{|CQ|\cdot|BP|}{|BQ|\cdot|CP|}. \tag{12.24}$$

For the verification, it remains only to substitute the expressions (12.24) into formula (12.23).

In any sufficiently complete course in geometry, it is proved without using the parallel postulate (that is, in the framework of "absolute geometry") that there exists a function $r(A,B)$ of a pair of points A and B that satisfies the following conditions:

1. $r(A,B)>0$ if $A\neq B$, and $r(A,B)=0$ if $A=B$;
2. $r(B,A)=r(A,B)$ for all points A and B;
3. $r(A,B)=r(A,C)+r(C,B)$ for every point C collinear with A and B and lying between them;

and most importantly,

4. the function $r(A,B)$ is invariant under motions.

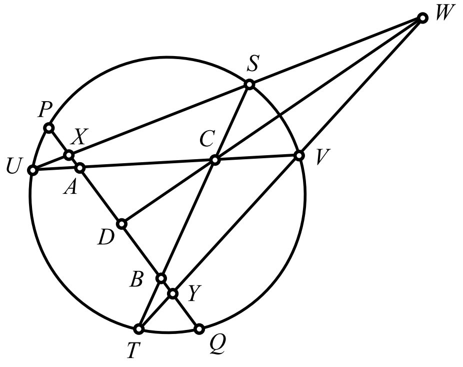
Using the definitions given at the beginning of this book, we may say in short that $r(A,B)$ is a metric on the set of points in the plane under consideration and motions are isometries of this metric space.

Such a function is unique if we fix two distinct points A_0 and B_0 for which $r(A_0,B_0)=1$ ("unit of measurement"). This means that these assertions also hold in hyperbolic geometry, and formula (12.19) defines this distance (and the base of the logarithm in (12.19) is chosen in correspondence with the chosen "unit of measurement").

Every triple of points A,B,C satisfies the condition

$$r(A,B)\le r(A,C)+r(B,C). \tag{12.25}$$

Fig. 12.6 The triangle inequality



This is the familiar *triangle inequality*, and in many courses in geometry, it is derived without use of the parallel postulate, that is, as a theorem of "absolute geometry." Thus inequality (12.25) holds as well in hyperbolic geometry. But we shall now give a direct (that is, resting directly on formula (12.19)) proof of this due to Hilbert.

Let us recall that in the model that we have considered, the points of the hyperbolic plane are points of the disk K in the Euclidean plane L, and the lines of the hyperbolic plane are the line segments of the plane L that lie inside the disk K.

Let us consider three points A, B, C in the disk K. We shall denote the points of intersection of a line passing through A and B with the boundary of the disk K by P and Q, and the analogous points for the line passing through A and C will be denoted by U and V, and for the line passing through B and C, by S and T. See Fig. 12.6.

Let us denote the point of intersection of the line AB and the line SU by X, and the point of intersection of the line AB and the line TV by Y. Then we have the inequality

$$\mathrm{DV}(A, B, Y, X) \geq \mathrm{DV}(A, B, Q, P). \tag{12.26}$$

Indeed, the left-hand side of (12.26) is equal by definition to

$$\mathrm{DV}(A, B, Y, X) = \frac{|AY| \cdot |BX|}{|BY| \cdot |AX|}, \tag{12.27}$$

and its right-hand side is given by the relationship (12.20). Therefore, inequality (12.26) follows from the fact that

$$\frac{|AY|}{|BY|} > \frac{|AQ|}{|BQ|} \quad \text{and} \quad \frac{|BX|}{|AX|} > \frac{|BP|}{|AP|}. \tag{12.28}$$

Let us prove the first of inequalities (12.28). Let us define $a = |AB|$, $t_1 = |BQ|$, and $t_2 = |BY|$. Then we obviously obtain the expressions $|AQ|/|BQ| = (a+t_1)/t_1$ and $|AY|/|BY| = (a + t_2)/t_2$. For $a > 0$, the function $(a + t)/t$ in the variable t decreases monotonically with increasing t, and therefore, from the fact that $t_2 < t_1$ (which is obvious from Fig. 12.6) follows the first of inequalities (12.28). Defining

$a = |AB|$, $t_1 = |AX|$, and $t_2 = |AP|$, using completely analogous arguments, we may prove the second inequality of (12.28).

Let us denote the intersection of the lines SU and TV by W, let us connect this line with the point C, and let us denote the point of intersection of the line thus obtained with the line AB by D. Then the points X, A, D, Y and points U, A, C, V are obtained from each other by a perspective mapping just as was done for the points Y, B, D, X and T, B, C, S. Then in view of Theorem 9.19, we have the relationships

$$\frac{|AY| \cdot |DX|}{|DY| \cdot |AX|} = \frac{|AV| \cdot |CU|}{|CV| \cdot |AU|}, \qquad \frac{|BX| \cdot |DY|}{|DX| \cdot |AY|} = \frac{|BS| \cdot |CT|}{|CS| \cdot |BT|}.$$

Multiplying these equalities, we have

$$\frac{|AY| \cdot |BX|}{|BY| \cdot |AX|} = \frac{|AV| \cdot |CU|}{|CV| \cdot |AU|} \cdot \frac{|BS| \cdot |CT|}{|CS| \cdot |BT|}.$$

Taking the logarithm of the last equality, and taking into account (12.27) for $\mathrm{DV}(A, B, Y, X)$, the analogous expression for $\mathrm{DV}(A, C, U, V)$ and that for $\mathrm{DV}(B, C, S, T)$, and definition (12.19), we obtain the relationship

$$\log \mathrm{DV}(A, B, Y, X) = r(A, C) + r(B, C),$$

from which, taking into account (12.26), we obtain the required inequality (12.25).

Let us note that if the point B approaches Q along the segment PQ (see Fig. 12.6), then $|BQ|$ approaches zero, and consequently, $r(A, B)$ approaches infinity. This means that despite that fact that the line passing through the points A and B is represented in our figure by a segment of finite length, its length in the hyperbolic plane in infinite.

The measurement of angles is similar to that of line segments. As we know, an arbitrary point O on a line l partitions it into two half-lines. One half-line together with the point O is called a *ray* h with center O. Two rays h and k with common center O are called an *angle*; we shall assume that the ray h is obtained from k by a counterclockwise rotation. This angle is denoted by $\angle(h, k)$.

In "absolute geometry," it is proved that for each angle with vertex at the point O, there is a unique real number $\measuredangle(h, k)$ satisfying the following four conditions:

1. $\measuredangle(h, k) > 0$ for all $h \neq k$;
2. $\measuredangle(k, h) = \measuredangle(h, k)$;
3. if f is a motion and $f(h) = h'$, $f(k) = k'$, and $O' = f(O)$ is the vertex of the angle $\angle(h', k')$, then $\measuredangle(h', k') = \measuredangle(h, k)$.

To formulate the fourth property, we must introduce some additional concepts. Let the rays h and k forming the angle $\angle(h, k)$ lie on lines l_1 and l_2. The points in the plane lying on the same side of the line l_1 as the points of the half-line k and on the same side of the line l_2 as the points of the half-line h are called *interior points* of the angle $\angle(h, k)$. A ray l with the same center O as the rays h and k is said to be an *interior ray* of the angle $\angle(h, k)$ if it consists of interior points of this angle.

We can now formulate the last property:

4. If l is an interior ray of the angle $\angle(h, k)$, then $\measuredangle(h, l) + \measuredangle(l, k) = \measuredangle(h, k)$.

As in the case of distance between points, the measure of an angle is defined uniquely if we choose a “unit measurement,” that is, if we take a particular angle $\angle(h_0, k_0)$ as the “unit angle measure.”

We shall point out an explicit method of defining the measure of angles in hyperbolic geometry that is realized in the disk K given by the relationship $x^2 + y^2 < 1$ in the Euclidean plane L with coordinates x, y.

Let $\angle(h', k')$ be the angle with center at the point O', and let f be an arbitrary motion taking the point O' to the center O of the disk K. From the definitions, it is obvious that f takes the half-lines h' and k' to some half-lines h and k with center at the point O. Let us set the measure of $\measuredangle(h', k')$ equal to the Euclidean angle between the half-lines h and k. The main difficulty in this definition is that it uses a motion f, and therefore, we must prove that the measure of the angle thus obtained does not depend on the choice of the motion f (of course, with the condition $f(O') = O$).

Let g be another motion with the same property that $g(O') = O$. Then $g^{-1}(O) = O'$, and this means that $fg^{-1}(O) = O$, that is, the motion fg^{-1} leaves the point O fixed. As we saw in Sect. 12.1 (p. 438), a motion possessing such a property is of type (a), which means that fg^{-1} corresponds to an orthogonal transformation of the Euclidean plane L; that is, the angle $\angle(\overline{h}, \overline{k})$ is taken to the angle $\angle(h, k)$ via the orthogonal transformation fg^{-1}, which preserves the inner product in L and therefore does not change the measure of angles. This proves the correctness of the definition of angle measure that we have introduced. Equally easy are the verifications of properties 1–3.

The best-known property of angles in hyperbolic geometry is the following.

Theorem 12.10 *In hyperbolic geometry, the sum of the angles of a triangle is less than two right angles, that is, less than* π.

Since we are talking about a triangle, we can restrict our attention to the plane in which this triangle lies and assume that we are working in the hyperbolic plane. The key result is related to the fact that an angle $\angle(h, k)$ in hyperbolic geometry also determines a Euclidean angle, and we may then compare the measures of these angles. We shall denote the measure of the angle $\angle(h, k)$ in hyperbolic geometry, as before, by $\measuredangle(h, k)$, and its Euclidean measure by $\measuredangle_{\mathrm{E}}(h, k)$.

Lemma 12.11 *If one ray of the angle* $\angle(h, k)$ (*for example*, h) *passes through the center* O *of the disk* K, *then the measure of this angle in the sense of hyperbolic geometry is less than the Euclidean measure, that is,*

$$\measuredangle(h, k) < \measuredangle_{\mathrm{E}}(h, k). \tag{12.29}$$

First, we shall show how easily Theorem 12.10 follows from the lemma, and then we shall prove the lemma itself.

Proof of Theorem 12.10 Let us denote the vertices of the triangle in question by A, B, C. Since the measure of an angle is invariant under a motion, it follows by

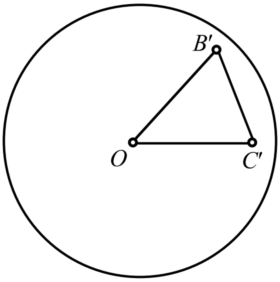


Fig. 12.7 A *triangle* in the *hyperbolic plane*

Theorem 12.5 that we can choose a motion taking one of the vertices of the triangle (for example, A) to the center O of the disk K. Let the vertices B and C be taken to B' and C'. See Fig. 12.7.

It suffices to prove the theorem for the triangle $OB'C'$. But for the angle $\angle B'OC'$, we have by definition the equality

$$\measuredangle B'OC' = \measuredangle_{\mathrm{E}} B'OC',$$

and for the two remaining angles, we have by the lemma, the inequalities

$$\measuredangle OB'C' < \measuredangle_{\mathrm{E}} OB'C', \qquad \measuredangle OC'B' < \measuredangle_{\mathrm{E}} OC'B'.$$

Adding, we obtain for the sum of the angles of triangle $OB'C'$ the inequality

$$\measuredangle B'OC' + \measuredangle OB'C' + \measuredangle OC'B' < \measuredangle_{\mathrm{E}} B'OC' + \measuredangle_{\mathrm{E}} OB'C' + \measuredangle_{\mathrm{E}} OC'B'.$$

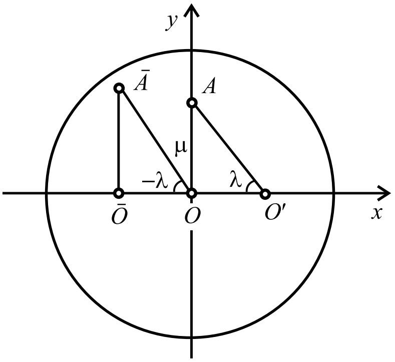
By a familiar theorem of Euclidean geometry, the sum on the right-hand side is equal to π, and this proves Theorem 12.10. □

Proof of Lemma 12.11 We shall have to use the explicit form of the definition of the measure of an angle. Let the ray h of the angle $\angle(h,k)$ pass through the point O. To describe the disk K, we shall introduce a Euclidean rectangular system of coordinates (x,y) and assume that the vertex of angle $\angle(h,k)$ is located at the point O' with coordinates $(\lambda,0)$, where $\lambda \neq 0$. For this, it is necessary to execute a rotation about the center of the disk in such a way that the point O' passes through some point of the line $y=0$ and use the fact that angles are invariant under such a rotation.

Now we must write down explicitly a motion f of the hyperbolic plane taking the point O to O'. We already constructed such a motion in Sect. 12.1; see Example 12.4 on p. 439. There, we proved that there exists a motion of the hyperbolic plane that takes the point with coordinates (x,y) to the point with coordinates (x',y'), given by the relationships

$$x' = \frac{ax+b}{bx+a}, \qquad y' = \frac{y}{bx+a}, \quad a^2 - b^2 = 1. \tag{12.30}$$

Fig. 12.8 *Angles* in the *hyperbolic plane*



If we want the point $O' = (\lambda, 0)$ to be sent to the origin $O = (0, 0)$, then we should set $a\lambda + b = 0$, or equivalently, $\lambda = -b/a$. It is not difficult to verify that it is possible to represent any number λ in this form. Thus the mapping (12.30) has the form

$$x' = \frac{x - \lambda}{1 - \lambda x}, \qquad y' = \frac{y}{a(1 - \lambda x)}. \tag{12.31}$$

Let the ray k intersect the y-axis at the point A with coordinates $(0, \mu)$; see Fig. 12.8. (We note that this point is not required to be in the disk K.)

From formula (12.31), it is clear that our transformation takes a vertical line $x = c$ to a vertical line $x = c'$. The point O is taken to the point $\overline{O} = (-\lambda, 0)$, the point $A = (0, \mu)$ to the point $\overline{A} = (-\lambda, \mu/a)$, and the vertical line OA to the vertical line $\overline{OA}$. By the definition of an angle in hyperbolic geometry, $\measuredangle OO'A = \measuredangle_{\mathrm{E}} \overline{O} O \overline{A}$. The tangents of the Euclidean angles are known to us:

$$\tan\bigl(\measuredangle_{\mathrm{E}} OO'A\bigr) = \frac{\mu}{\lambda}, \qquad \tan(\measuredangle_{\mathrm{E}} \overline{O} O \overline{A}) = \frac{\overline{OA}}{\lambda} = \frac{\mu}{\lambda a};$$

see Fig. 12.8. Since $a^2 = 1 + b^2$, we have $a > 1$, and we see that in Euclidean geometry, we have the inequality $\tan(\measuredangle_{\mathrm{E}} \overline{O} O \overline{A}) < \tan(\measuredangle_{\mathrm{E}} OO'A)$. The tangent is a strictly increasing function, and therefore we have the inequality $\measuredangle_{\mathrm{E}} \overline{O} O \overline{A} < \measuredangle_{\mathrm{E}} OO'A$ for angles that are Euclidean. But $\measuredangle OO'A = \measuredangle_{\mathrm{E}} \overline{O} O \overline{A}$, and this means that $\measuredangle OO'A < \measuredangle_{\mathrm{E}} OO'A$. □

It is of interest to compare Theorem 12.10 with the analogous result for *spherical geometry*. We have not yet encountered spherical geometry in this course, even though it was developed in detail much earlier than hyperbolic geometry, indeed in antiquity. In spherical geometry, the role of lines in played by great circles on the sphere, that is, sections of the sphere obtained by all possible planes passing through its center. The analogy between great circles on the sphere and lines in the plane consists in the fact that the arc of the great circle joining points A and B has length no greater than that of any other curve on the sphere with endpoints A and B. This arc length of a great circle (which, of course, depends also on the radius R of the sphere) is called the *distance* on the sphere from point A to point B.

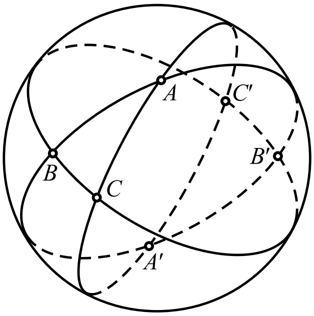


Fig. 12.9 A *triangle* on the *sphere*

The measurement of lengths and angles on the sphere can generally be defined in exactly the same way as in Euclidean or hyperbolic geometry. Here the angle between two "lines" (that is, great circles) is equal to the value of the dihedral angle formed by the planes passing through these great circles. We have the following result.

Theorem 12.12 *The sum of the angles of a triangle on the sphere is greater than two right angles*, *that is*, *greater than* π.

Proof Let there be given a triangle with vertices A, B, C on a sphere of radius R. Let us draw all the great circles whose arcs are the sides AB, AC, and BC of triangle ABC. See Fig. 12.9.

Let us denote by $\varSigma_A$ the part of the sphere enclosed between the great circle passing through the points A, B and the great circle passing through A, C. We introduce the analogous notation $\varSigma_B$ and $\varSigma_C$. Let us denote by $\widehat{A}$ the measure of the dihedral angle $\widehat{BAC}$ and similarly for $\widehat{B}$ and $\widehat{C}$. Then the assertion of the theorem is equivalent to asserting that $\widehat{A}+\widehat{B}+\widehat{C}>\pi$.

But it is easy to see that the area of $\varSigma_A$ is the same fraction of the area of the sphere as $2\widehat{A}$ is of 2π. Since the area of the sphere is equal to $4\pi R^2$, it follows that the area of $\varSigma_A$ is equal to

$$4\pi R^2 \cdot \frac{2\widehat{A}}{2\pi} = 4R^2\widehat{A}.$$

Similarly, we obtain expressions for the areas $\varSigma_B$ and $\varSigma_C$; they are equal to $4R^2\widehat{B}$ and $4R^2\widehat{C}$ respectively. Let us now observe that the regions $\varSigma_A$, $\varSigma_B$, and $\varSigma_C$ together cover the entire sphere. Here each point of the sphere not part of triangle ABC or of triangle $A'B'C'$ symmetric to it on the sphere belongs to only one of the regions $\varSigma_A$, $\varSigma_B$, and $\varSigma_C$, and every point in triangle ABC or the symmetric triangle $A'B'C'$ is contained in all three regions. We therefore have

$$4R^2(\widehat{A}+\widehat{B}+\widehat{C}) = 4\pi R^2 + 2S_{\triangle ABC} + 2S_{\triangle A'B'C'} = 4\pi R^2 + 4S_{\triangle ABC}.$$

From this we obtain the relationship

$$\widehat{A}+\widehat{B}+\widehat{C}=\pi+\frac{S_{\triangle ABC}}{R^2}, \tag{12.32}$$

from which it follows that $\widehat{A}+\widehat{B}+\widehat{C}>\pi$. □

Formula (12.32) gives an example of a series of relationships systematically developed by Lobachevsky: if we were to assume that $R^2<0$ (that is, R is a purely imaginary number), then clearly, we would obtain from (12.32) the inequality

$$\widehat{A}+\widehat{B}+\widehat{C}<\pi,$$

which is Theorem 12.10 of hyperbolic geometry. This is why Lobachevsky considered that his geometry is realized "on a sphere of imaginary radius." However, the analogy between theorems obtained on the basis of the negation of the "fifth postulate" and formulas obtained from those of spherical geometry by replacing R^2 with a negative number had been already noted by many mathematicians working on these questions (some even as early as the eighteenth century).

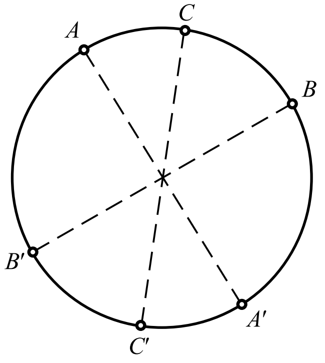
The reader should be warned that spherical geometry is entirely inconsistent with the system of axioms that we considered in Sect. 12.2. That system does not include one of the fundamental axioms of relationship: several different lines can pass through two distinct points. Indeed, infinitely many great circles pass through any two antipodal points on the sphere. In connection with this, Riemann proposed another geometry less radically different from Euclidean geometry. We shall describe it in the two-dimensional case.

For this, we shall use a description of the projective plane Π as the collection of all lines in three-dimensional space passing through some point O. Let us consider the sphere S with center at O. Every point $P\in S$ together with the center O of the sphere determines a line l, that is, some point Q of the projective plane Π. The association $P\to Q$ defines a mapping of the sphere S to the projective plane Π whereby great circles on the sphere are taken precisely to lines of Π. Clearly, exactly two points of the sphere are mapped to a single point $Q\in\Pi$: together with the point P, there is also the second point of the intersection of the line l with the sphere, that is, the antipodal point P'. But Euclidean motions taking the sphere S into itself (we might call them *motions of spherical geometry*) give certain transformations defined on the projective plane Π and satisfying the axioms of motion. It is possible as well to transfer the measures of lengths and angles from the sphere S to the projective plane Π. Then we have the analogue of Theorem 12.12 from spherical geometry.

This branch of geometry is called *elliptic geometry*.[8] In elliptic geometry, every pair of lines intersect, since such is the case in the projective plane. Thus there are no parallel lines. However, in "absolute geometry," it is proved that there exists at least

[8] Elliptic geometry is sometimes called *Riemannian geometry*, but that term is usually reserved for the branch of differential geometry that studies Riemannian manifolds.

Fig. 12.10 Elliptic geometry



one line passing through any given point A not lying on a given line l that is parallel to l. This means that in elliptic geometry, not all the axioms of "absolute geometry" are satisfied. The reason for this is easily ascertained: in elliptic geometry, there in no natural concept of "lying between." Indeed, a great circle of the sphere S is mapped to a line l of the projective plane Π, where two antipodal points of the sphere (A and A', B and B', C and C', and so on) are taken to one point of the plane Π. See Fig. 12.10. It is clear from the figure that in elliptic geometry, we may assume equally well that the point C does or does not lie between A and B.

Nevertheless, elliptic geometry possesses the property of "free mobility." Moreover, one can prove (Helmholtz–Lie theorem) that among all geometries (assuming some rigorous definition of this term), only three of them—Euclidean, hyperbolic, and elliptic—possess this property.