Notes from a course at the Nesin Mathematics Village, July 15–21, 2019, given jointly with Ayşe Berkman

Hyperbolic Geometry

David Pierce

August 5, 2019

Matematik Bölümü
Mimar Sinan Güzel Sanatlar Üniversitesi
mat.msgsu.edu.tr/~dpierce/
polytropy.com

Contents

List of Figures

Introduction

In **Chapter 1,** we start with the **Poincaré half-plane,** which we shall denote by

$$\mathbb{H}^2(\mathbb{R}).$$

This is an example of a **hyperbolic plane,** namely a set of points with designated lines and circles—**hyperbolic lines** and **hyperbolic circles**—having certian properties. In $\mathbb{H}^2(\mathbb{R})$, the points are those of $\mathbb{R}^2$ (or $\mathbb{C}$) with positive y-coordinate or (imaginary part). A segment of a hyperbolic line is the path of the shortest journey between the endpoints, if speed is proportional to y-coordinate. We can compute the time of the journey with an integral. A hyperbolic circle is the set of points, reaching any of which takes the same time.

The hyperbolic lines turn out to be semicircles and vertical rays; the hyperbolic circles, ordinary circles. We can define these in the upper half-plane associated with any **Euclidean field,** namely an ordered field K whose every positive element is a square. Thus we obtain the hyperbolic plane that we call

$$\mathbb{H}^2(K).$$

Here a hyperbolic segment may not have a length that is a number; but we can still tell when two segments are equal. An **isometry** of the plane takes every segment to an equal segment. We work out what the isometries are.

In **Chapter 2,** we show that every model of Hilbert's axioms for a hyperbolic plane is isomorphic to a Poincaré half-plane over some Euclidean field. This work is based on Hilbert's "New Development of Bolyai-Lobachevskian Geometry," which, having been translated from *Mathematische Annalen* **57,** is Appendix III in the 1971 English edition of Hilbert's *Foundations of Geometry* [10]. In Hilbert's words, we show

> that it is possible to develop Bolyai-Lobachevskian geometry in the plane exclusively with the plane axioms without the use of the continuity axioms.

Our specific aim in Chapter 2 is what Hartshorne [9, p. 388] calls

> Hilbert's tour de force, the creation of an abstract field out of the geometry of a hyperbolic plane.

We set out just enough axioms for this. Thus we differ from Borsuk and Szmielew in *Foundations of Geometry* [3] (one of the "additional references" in a note added to the end of Hilbert's paper); for they

> develop Euclidean and Bolyai-Lobachevskian geometry on the basis of an axiom system due, in principle, to Hilbert. It should be noted at once, however, that the authors develop these geometries, in principle, as far as necessary to be able to prove them categorical, i.e., to show that the Cartesian space known from analytic geometry is up to isomorphism the only model of Euclidean geometry, and Klein space (constructed with the help of notions known from the analytic geometry of projective space) is up to isomorphism the only model of Bolyai-Lobachevskian geometry.

The field interpreted in these models will be the field of real numbers. However, in the most general sense, any field (or even skew-field) yields a "Cartesian space"; any Euclidean field yields a "Klein space."

In a footnote on the passage quoted above, Hilbert refers to similar work in plane elliptic geometry, by Dehn and Hessenberg, then in absolute geometry ("even without any assumption about intersecting or nonintersecting lines"), by Hjelmslev, in *Mathematische Annalen* **60, 61,** and **64.**

Like the Ancients (as Netz [12] points out), dispensing with verbal descriptions, we may let diagrams alone define certain points.

In the past, I have taught hyperbolic geometry axiomatically, as in Chapter 2, but through Lobachevski [11]. I have then introduced the Poincaré half-plane as a model, and used this model for drawing diagrams; but students have had trouble thinking of curved lines as straight. Chapter 1 is an attempt to solve this problem.

The general problem of presenting hyperbolic geometry may be insoluble, if only because ill defined. Different courses will have different emphases for different purposes. The present course concerns the logical interactions between geometry and algebra. We use analysis at the beginning, only to justify thinking of certain curved lines as straight. The notes are still a fairly rough draft.

1 Field

1.1 Lines

In the words at the head of Euclid's *Elements* [4, 5, 6, 7, 8], "a **line** is breadthless length" (Γραμμὴ μῆκος ἀπλατές). Such a line may be curved: the circumference of a circle will be an example.

The English word "line" is the noun corresponding to the adjective "linen." This means flaxen and is related to the Latin LINVM, meaning flax. A line is a thread or cord, originally flaxen. A cord may lie in a heap, or be wound around a spool, not straight at all; or it may be pulled taut, into a *straight line.*[1]

In the Cartesian plane $\mathbb{R}^2$, the **straight lines** are the solution-sets of equations

$$ax + by = c, \tag{1.1}$$

where at least one of a and b is not 0, but each, along with c, is an element of $\mathbb{R}$. If A and B are distinct points in the plane and have coordinates (a_0, a_1) and (b_0, b_1) respectively,

[1] After the definition of a line in the *Elements,* there is an obscure definition of a straight line. In *The Forgotten Revolution* [13, pp. 323–4], Lucio Russo argues plausibly that this definition is only a later addition to the *Elements,* its origin being a student's crib-sheet.

then there is a unique straight line passing through them, and the segment denoted by AB has length given by the rule

$$|AB| = \sqrt{(a_0 - b_0)^2 + (a_1 - b_1)^2}. \tag{1.2}$$

We may also use the notation AB for the whole straight line that contains A and B.

We consider now lengths of other lines. If (f_0, f_1) or f is an injective, continuously differentiable function into $\mathbb{R}^2$ from an interval $[0, T]$, then we can understand f as a **journey,** which proceeds

- from *time* 0 to time T,
- from *position* $f(0)$ to position $f(T)$.

Suggestively, we may say that

- the domain of f is the **period,** and
- the range of f is the **path,**

of the journey. The **speed** of the journey at each time t is given by

$$|f'(t)| = \sqrt{{f_0}'(t)^2 + {f_1}'(t)^2}.$$

The **length** of the journey, or the **distance** from its beginning to its end, is ambiguous: it could be the length of the path or the period.

- The length of the period is just T.
- The length of the path of f is given by

$$L = \int_0^T |f'(t)| \, \mathrm{d}\, t = \int_0^T |f'|.$$

The path of f may also be the range of a (still injective and continuously differentiable) function γ on an interval $[a, b]$, where

$$\gamma(a) = f(0), \qquad \gamma(b) = f(T).$$

With the substitution

$$\gamma(u) = f(t), \qquad \gamma'(u)\,\mathrm{d}\,u = f'(t)\,\mathrm{d}\,t, \tag{1.3}$$

we obtain the equation

$$\int_0^T |f'| = \int_a^b |\gamma'|.$$

It may be that the speed of f is constrained by the rule

$$|f'| = \sigma \circ f, \tag{1.4}$$

where σ is a continuous function into $\mathbb{R}$, defined on a subset of $\mathbb{R}^2$ that includes the path of f (which is the range of γ). We are going to work with a nontrivial example of such a constraint in the next section and beyond.

Meanwhile, rewriting (1.4) as

$$\frac{|f'|}{\sigma \circ f} = 1,$$

we obtain

$$T = \int_0^T \frac{|f'|}{\sigma \circ f}.$$

As distance is the time integral of speed, so time is the distance integral of the inverse of speed. Again using the substitution (1.3), we obtain

$$T = \int_a^b \frac{|\gamma'|}{\sigma \circ \gamma}. \tag{1.5}$$

Thus, knowing only a path from a point A to a point B in the domain of σ, we know how long a journey takes on that path under the constraint (1.4).

The constrained journey from A to B that has the shortest *period* may not have a straight path. We call it a **geodesic segment** (with respect to σ). The length of that shortest period is then the **geodesic distance** between A and B. If a point D is at the same distance from a point C that B is from A, then the two geodesic segments AB and CD are **equal.** An unbounded line whose segments are all geodesic will be a **geodesic line.**

In another of the definitions (ὅροι) at the head of Euclid's *Elements,*

> A **circle** is a plane figure bounded by one line, all straight lines extending to which from one of the points lying inside the figure are equal to one another; and the **center** is that point of the circle. // Κύκλος ἐστὶ σχῆμα ἐπίπεδον ὑπὸ μιᾶς γραμμῆς περιεχόμενον, πρὸς ἣν ἀφ' ἑνὸς σημείου τῶν ἐντὸς τοῦ σχήματος κειμένων πᾶσαι αἱ προσπίπτουσαι εὐθεῖαι ἴσαι ἀλλήλαις εἰσίν. Κέντρον δὲ τοῦ κύκλου τὸ σημεῖον καλεῖται.

Now we have the notion of a **geodesic circle,** whose circumference is the locus of points whose geodesic distances from a given point are all the same. We may confuse a circle with its circumference.

Let us define a **geometry** to be an ordered triple of three sets, the sets consisting respectively of points, straight lines, and circles. A standard example then is

$$\mathbb{A}^2(\mathbb{R}),$$

whose

- points compose $\mathbb{R}^2$,
- straight lines are defined by equations (1.1) as above,
- circles are defined by equations

$$|z-a|^2 = r^2,$$

where the variable z ranges over $\mathbb{C}$, and also $a \in K(\mathbf{i})$ and r in K.

We shall make use of this example, but also another, which we are about to define, where the "straight lines" will be geodesic lines; circles, geodesic circles.

1.2 Poincaré Half-plane

The points of our new geometry will compose the upper half-plane,

$$\{(x, y) \in \mathbb{R}^2 \colon y > 0\},$$

which is also $\{z \in \mathbb{C} \colon \Im(z) > 0\}$. Geodesics will be determined as in §1.1 by one of the equivalent rules

$$\sigma(x, y) = y, \qquad \sigma(z) = \Im(z).$$

The geometry will be the **Poincaré half-plane,** which we may denote by

$$\mathbb{H}^2(\mathbb{R}).$$

It will be an example of a *hyperbolic plane.* In such a geometry, to refer to geodesics, we may use the adjective **hyperbolic;** however, we intend no allusion to the conic section called an hyperbola. For example, the length of the geodesic segment bounded by A and B is the *hyperbolic distance* from A to B. When we do not apply the adjective geodesic or hyperbolic, we mean a term in the usual sense: the Euclidean sense, to be considered further in §1.4.

Theorem 1.1. *Between two points of* $\mathbb{H}^2(\mathbb{R})$*, the geodesic segment is*

1) *the straight segment joining the points, if they are on the same vertical line;*

2) *otherwise, the arc that they bound on a circle whose center is on the x-axis.*

Proof. 1. Suppose the points are (c, a) and (c, b). Along the straight path from one to the other, (1.5) takes the form

$$T = \int_a^b \frac{\mathrm{d}\, y}{y}.$$

On an arbitrary path γ, we compute

$$\int_\gamma \frac{\sqrt{\mathrm{d}\, x^2 + \mathrm{d}\, y^2}}{y} = \int_a^b \frac{\sqrt{(\mathrm{d}\, x/\, \mathrm{d}\, y)^2 + 1}}{y}\, \mathrm{d}\, y \geqslant T.$$

2. The points being A and B, we let the perpendicular bisector of AB cut the x-axis at O. Letting I be the projection of A on the x-axis, we define

$$\alpha = IOA, \qquad \beta = IOB, \tag{1.6}$$

assuming then

$$0 < \alpha < \beta < \pi.$$

Using a system of polar coordinates centered at O, for travel along the circular arc from A to B centered at O, if the arc has radius r, from (1.5) we compute

$$T = \int_\alpha^\beta \frac{r\, \mathrm{d}\, \vartheta}{r \sin \vartheta} = \int_\alpha^\beta \frac{\mathrm{d}\, \vartheta}{\sin \vartheta}.$$

This is independent of r. If we follow an arbitrary path,

$$\int_\gamma \frac{\sqrt{\mathrm{d}\, r^2 + (r\, \mathrm{d}\, \vartheta)^2}}{r \sin \vartheta} = \int_\alpha^\beta \frac{\sqrt{(\mathrm{d}\, r/r\, \mathrm{d}\, \vartheta)^2 + 1}}{\sin \vartheta}\, \mathrm{d}\, \vartheta \geqslant T. \qquad \square$$

A geodesic circle in $\mathbb{H}^2(\mathbb{R})$ will be an ordinary circle, but its center will be displaced so that its distance from the x-axis will be, not the arithmetic, but the geometric mean of the distances of the nearest and farthest points of the circle from the axis. (Distances from the x-axis can only be meant in the ordinary, Euclidean sense, since the points of the axis do not belong to $\mathbb{H}^2(\mathbb{R})$.)

The following theorem, complicated to state in words, is illustrated by Figg. 1.1 and 1.2.

Theorem 1.2. *In $\mathbb{H}^2(\mathbb{R})$, the geodesic circle whose circumference passes through a point E and whose geodesic center is a point B is the (Euclidean) circle*

- *whose diameter is AC,*
- *where A and C are on the same vertical line with B,*
- *this line cuts the x-axis at O,*
- $|OA| \cdot |OC| = |OB|^2$,
- *and moreover*
    1) *A is E, if this lies on OB;*
    2) *otherwise, A is where DE cuts OB,*
        - *D being where the circle with center X and passing through E cuts the x-axis,*
        - *X being where the perpendicular bisector of BE cuts the x-axis.*

Proof. In Figg. 1.1 and 1.2,

- the altitudes AO, DE', and $D'E$ of triangle ADD' meet (as the altitudes of a triangle always do) at C;
- therefore the circle with diameter DD' passes through E and E';
- the circle also cuts AO at B and B'.

As a result, the points E and E' lie on the circle with diameter AC. Together with the points B and O,

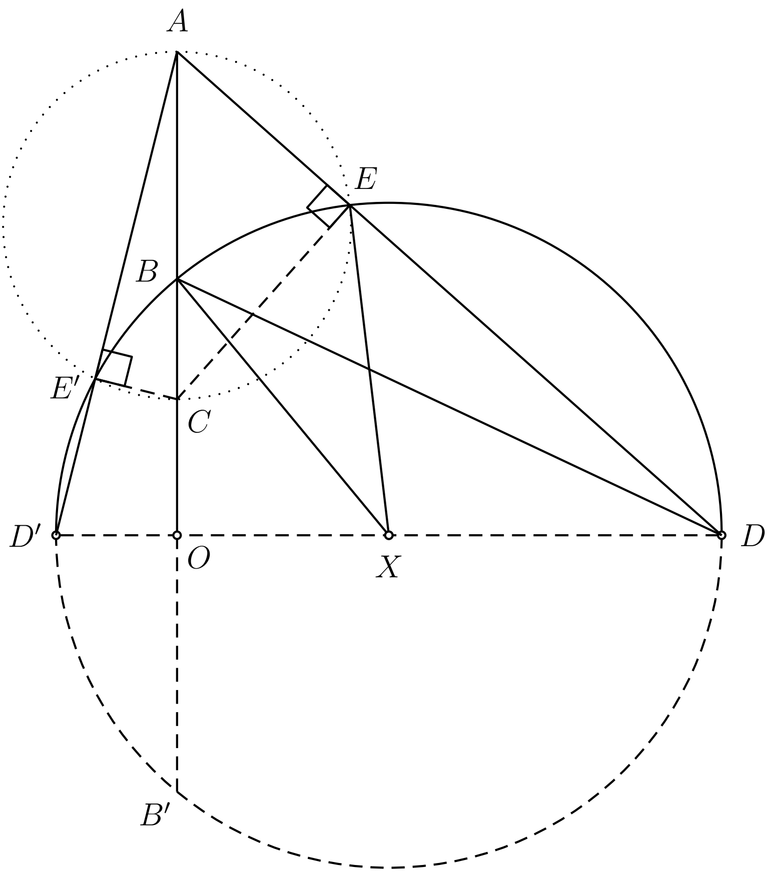


Figure 1.1: A hyperbolic circle

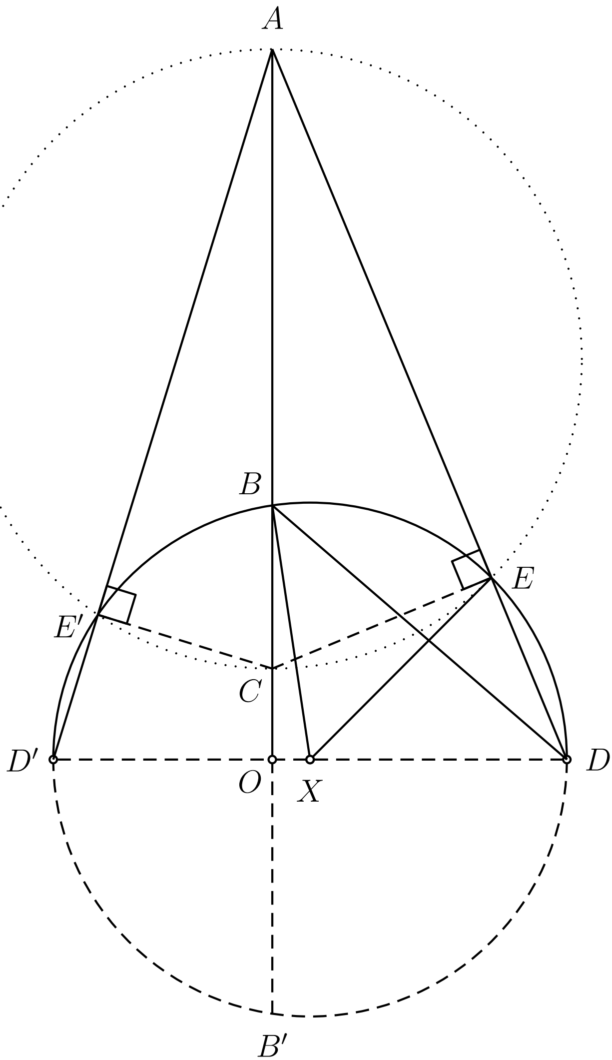


Figure 1.2: Another hyperbolic circle

1) the point E determines A, and
2) the point A determines C;

we show this now.

1. Given E not on BO, we let angle BOX be right, but also require X to be equidistant from B and E. Letting the circle with center X passing through B (and therefore E) cut OX at D, we let A be the intersection of DE and BO.

2. Given A on BO, we choose X arbitrarily so that angle BOX is right, and then we draw the circle with center X passing through B, letting the circle cut

- OX at D and D',
- AD and AD' at E and E' respectively,
- OB also at B'.

We let C be the intersection of AO and DE' (and therefore of DE' as well). By similarity of triangles AEC and AOD, as well as the theorem (found in Euclid) on secants of circles, and since $|OB'| = |OB|$,

$$|AC| \cdot |AO| = |AE| \cdot |AD| = |AB| \cdot |AB'| = |AO|^2 - |OB|^2.$$

Thus C is independent of the choice of X.

It is now enough to show that A, C, and E are all at the same geodesic distance from B.

- AB and BC are equal as geodesic segments, because, if A and B have y-coordinates a and b respectively, then, as in the proof of Theorem 1.1, the geodesic length of AB is given by

$$\int_a^b \frac{\mathrm{d}\,y}{y} = \log y\Big|_a^b = \log b - \log a = \log \frac{b}{a}.$$

- AB and BE are equal as geodesic segments, because, letting

$$\alpha = OXB, \qquad \beta = OXE,$$

so that

$$\frac{\alpha}{2} = ODB, \qquad \frac{\beta}{2} = ODE,$$

we have the geodesic length of EB as given by

$$\int_{\alpha}^{\beta} \frac{\mathrm{d}\,\vartheta}{\sin\vartheta} = \log\frac{\tan(\beta/2)}{\tan(\alpha/2)} = \frac{a}{b},$$

where we use

$$\int \frac{\mathrm{d}\,\vartheta}{\sin\vartheta} = \log\left|\tan\frac{\vartheta}{2}\right| + C, \tag{1.7}$$

which we can derive as follows.

$$\int_{\alpha}^{\beta} \frac{\mathrm{d}\,\vartheta}{\sin\vartheta} = \int_{\alpha}^{\beta} \frac{\sin\vartheta\,\mathrm{d}\,\vartheta}{\sin^2\vartheta} = \int_{\alpha}^{\beta} \frac{\sin\vartheta\,\mathrm{d}\,\vartheta}{1-\cos^2\vartheta} = -\int_{\cos\alpha}^{\cos\beta} \frac{\mathrm{d}\,t}{1-t^2}$$

by the substitution

$$t = \cos\vartheta, \qquad \mathrm{d}\,t = -\sin\vartheta\,\mathrm{d}\,\vartheta.$$

Since

$$\frac{1}{1-t^2} = \frac{1}{(1+t)(1-t)} = \frac{1}{2}\left(\frac{1}{1+t} + \frac{1}{1-t}\right),$$

so that

$$\begin{aligned} -\int \frac{\mathrm{d}\,t}{1-t^2} &= \frac{1}{2}\left(\log|1-t| - \log|1+t|\right) + C \\ &= \frac{1}{2}\log\left|\frac{1-t}{1+t}\right| + C = \frac{1}{2}\log\frac{|1-t^2|}{(1+t)^2} + C, \end{aligned}$$

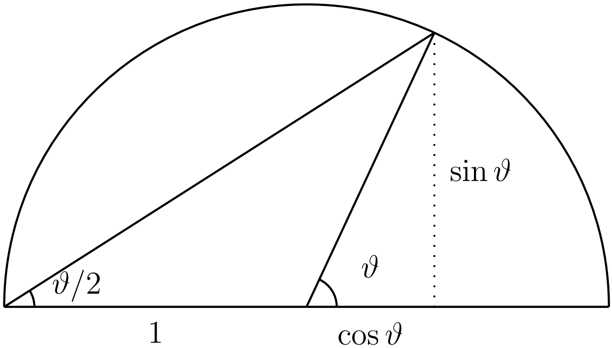


Figure 1.3: Angle and its half

we conclude

$$\int \frac{\mathrm{d}\,\vartheta}{\sin\vartheta} = \frac{1}{2}\log\frac{1-\cos^2\vartheta}{(1+\cos\vartheta)^2} + C = \log\frac{|\sin\vartheta|}{1+\cos\vartheta} + C.$$

We now obtain (1.7) from the identity (read from Fig. 1.3)

$$\tan\frac{\vartheta}{2} = \frac{\sin\vartheta}{1+\cos\vartheta}. \qquad \square$$

We note in passing that some sources may use the formula

$$\tan\frac{\vartheta}{2} = \frac{1-\cos\vartheta}{\sin\vartheta} = \csc\vartheta - \cot\vartheta.$$

1.3 Euclidean Fields

The field $\mathbb{R}$ of real numbers is an example of a **Euclidean field,** because it is an ordered field whose every positive element has a square root. Thus a Euclidean field is an ordered field whose ordering is definable by the rule

$$x \leqslant y \Leftrightarrow \exists z\; x + z^2 = y.$$

Examples of ordered fields are $\mathbb{Q}$ and $\mathbb{R}(t)$, but these are not Euclidean. The latter field consists of ratios

$$\frac{a_0 + a_1 t + \cdots + a_m t^m}{b_0 + b_1 t + \cdots + b_n t^n},$$

where a_i and b_j are from $\mathbb{R}$, and not all b_j are 0. The ordering is determined by

$$t > 0, \qquad\qquad 1 + tx > 0$$

for all x in $\mathbb{R}(t)$. This means t is a positive infinitesimal; in particular, letting x be $-n$ for some positive integer n, we have

$$0 < t < \frac{1}{n}.$$

The presence of infinitesimals (and therefore infinite elements, such as $1/t$) means $\mathbb{R}(t)$ is not *Archimedean.*

Every ordered field K is included in a smallest Euclidean ordered field,

$$K^{\text{Euc}}.$$

We obtain it by continually replacing K with

$$K[X]/(X^2 - a),$$

where a is a positive element of K with no square root. The field K^{Euc} is countable if K is. Since $\mathbb{R}$ is not countable, it is not $\mathbb{Q}^{\text{Euc}}$. The latter is the smallest Euclidean field; it contains $\sqrt{2}$, $\sqrt{\sqrt{2} + \sqrt{3}}$, and so forth.

If K is a Euclidean field, we can form the field $K(\mathbf{i})$, where as usual $\mathbf{i}$ satisfies

$$z^2 + 1 = 0.$$

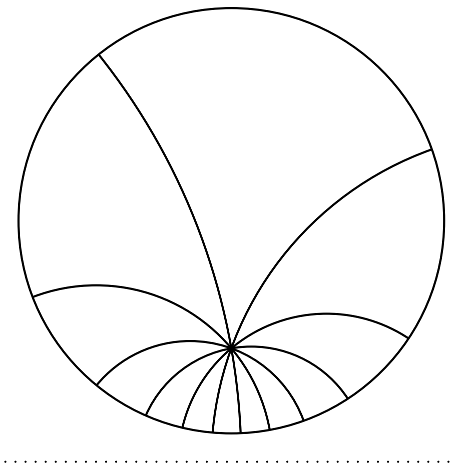


Figure 1.4: Hyperbolic circle and radii

We turn the upper half-plane over K,

$$\{(x, y) \in K^2 : y > 0\},$$

which is

$$\{z \in K(\mathbf{i}) : \Im(z) > 0\},$$

into a hyperbolic plane, as we did in case K was $\mathbb{R}$; the result will be a geometry that we can denote by

$$\mathbb{H}^2(K).$$

However, since we may no longer have the analytic properties of $\mathbb{C}$, we have not got an obvious theoretical way to define geodesic lines. We just declare that the analogues of Theorem 1.1 and 1.2 shall be true. By fiat then, the curves emanating from the point in Fig. 1.4 are equal radii of a circle.

If B lies between A and C on the x-axis in K^2, then the vertical line though B actually cuts the circle with diameter

AC, as it would in case K were $\mathbb{R}$. There are two solutions to the system of equations

$$\left(x - \frac{a+c}{2}\right)^2 + y^2 = \left(\frac{a-c}{2}\right)^2, \qquad x = b$$

over K, because positive elements of K have square roots.

1.4 Geometric Axioms

The first four of the thirteen books of Euclid's *Elements* are the original exposition of *Euclidean plane geometry.*[2] The exposition begins with some definitions, two of which—for line and circle—we have seen in §1.1. Here is another Euclidean definition, in a fairly literal translation, complete with subjunctive mood:

> Whenever a straight line, erected on a straight line, make the adjacent angles equal to one another, a **right** angle is either of the equal angles, and **perpendicular** is called the former line to that on which it stands. // Ὅταν δὲ εὐθεῖα ἐπ' εὐθεῖαν σταθεῖσα τὰς ἐφεξῆς γωνίας ἴσας ἀλλήλαις ποιῇ, ὀρθὴ ἐκατέρα τῶν ἴσων γωνιῶν ἐστι, καὶ ἡ ἐφεστηκυῖα εὐθεῖα κάθετος καλεῖται, ἐφ' ἣν ἐφέστηκεν.

Next are five *postulates* (αἰτήματα). Their precise meaning is controversial; we paraphrase them for our purposes:

1. Any two points can be joined by a segment of a straight line. That Euclid intends this segment to be unique is clear from his proof of *SAS,* discussed presently.
2. Any segment of a straight line can be extended.

[2]Books v and vi develop a theory of proportion; vii–ix, of numbers; x, of irrationality; xi–xii, of solid geometry.

3. A circle can be drawn through any point with any center.
4. All right angles are equal.
5. If a straight line crosses two others, making the internal angles on the same side less than two right angles, then the two lines, extended on that side, will intersect.

There are five additional "common notions" (κοιναὶ ἔννοιαι):[3]

1. What are equal to the same are equal to one another.
2. If equals be added to equals, the wholes are equal.
3. If equals be subtracted from equals, the remainders are equal.
4. What are congruent to one another are equal to one another.
5. The whole is greater than the part.

Modern mathematicians may ignore any distinction between postulates and common notions, but call them all axioms. They may find in Euclid's axioms various inadequacies, which Hilbert attempts to remedy in *The Foundations of Geometry* [10].

A point on a straight line divides the line into two opposite **rays.** Hilbert accepts the following as an axiom, because his system does not have circles.

Theorem 1.3. *By Euclid's first three postulates and first three common notions, from a given ray, a segment equal to any given segment can be cut.*

Proof. Let the given segment be AB in Fig. 1.5, and let the given ray begin at C and pass through D. All segments in the

[3]Heiberg [4] lists nine common notions, but brackets the four whereby, (1) if unequals be added to equals, the wholes are equal; (2) doubles and (3) halves of the same are equal to one another; (4) two straight lines cannot bound a space. Mourmouras [6] omits to bracket the last.

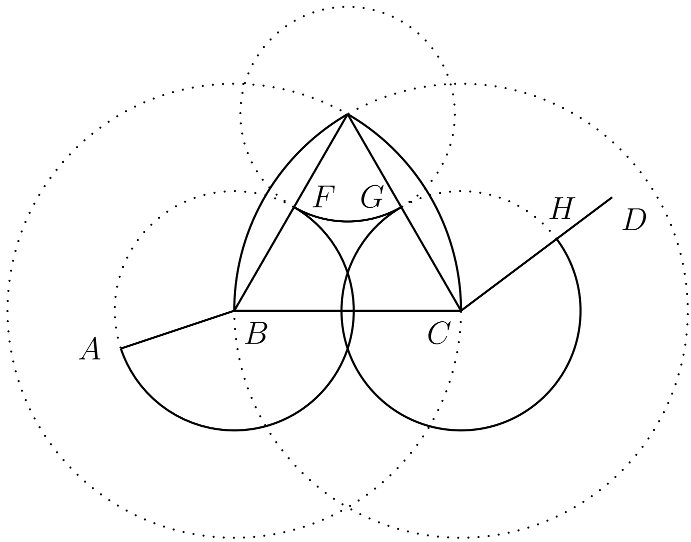


Figure 1.5: Cutting AB from CD

following are drawn and extended with Postulates 1 and 2.

$$
\begin{aligned}
BE = BC \ \ \& \ \ BC = EC, &\qquad \text{[Postulate 3]}\\
BE = EC, &\qquad \text{[Common Notion 1]}\\
BF = BA, &\qquad \text{[Postulate 3]}\\
EG = EF, &\qquad \text{[Postulate 3]}\\
CG = BF, &\qquad \text{[Common Notion 2 or 3]}\\
CH = CG. &\qquad \text{[Postulate 3]} \qquad \square
\end{aligned}
$$

Likewise, Hilbert accepts as an axiom that, on either side of a given ray, an angle equal to a given angle can be drawn. Euclid does the drawing, again by means circles, now using, in justification, the Side-Side-Side (SSS) theorem.

Hilbert accepts as an axiom **Side-Angle-Side,** or **SAS:** If two sides and the included angle in one triangle are equal re-

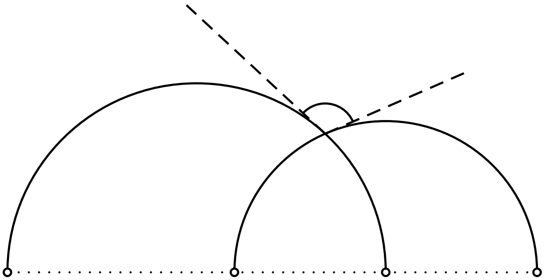


Figure 1.6: Geodesic angle

spectively to two sides and the included angle in another triangle, then the remaining angles are equal respectively. (It follows that the remaining sides are equal too.)

Hilbert uses "equal" and "congruent" as synonyms for a relation whose properties, even transitivity, are defined by axioms, as above. For Euclid, equality

- is present in the radii of a circle,
- is *explained* as congruence.

Congruent things can be made to coincide. Thus Side-Angle-Side is a theorem for Euclid, because the equations in the hypothesis mean that one triangle can be laid exactly on top of the other.

Being congruent is not the same as having the same *measurement,* although today we may take the latter to determine equality. This is how we proved Theorem 1.2.

In $\mathbb{H}^2(\mathbb{R})$, two hyperbolic rays emanating from the same point form a **hyperbolic angle,** whose size is that of the angle formed by the rays tangent to the geodesic rays at their common endpoint, as in Fig. 1.6. Note that the latter angle always has positive size, if the hyperbolic rays are distinct, since such rays are never tangent to one another.

If we are working in $\mathbb{H}^2(K)$ for an arbitrary Euclidean field K, a circle still has tangents, namely straight lines, the equation of each of which has a single common solution with the equation of the circle. Thinking of K^2 as a vector space, understanding a point X of the space as (x_0, x_1), defining

$$X \cdot Y = x_0 y_0 + x_1 y_1,$$

we define angles ABC and DEF to be **equal,** provided

$$\frac{|(A - B) \cdot (C - B)|}{|A - B| \cdot |C - B|} = \frac{|(D - E) \cdot (F - E|}{|D - E| \cdot |F - E|}.$$

The angle ABC is **right,** provided $(A - B) \cdot (C - B) = 0$. We now have the first three of Euclid's postulates in our setting, as follows.

Theorem 1.4. *In $\mathbb{H}^2(K)$ for an arbitrary Euclidean field K,*

1) *for any two points, there is a unique geodesic segment between them;*
2) *any geodesic segment can be extended to a longer one,*
3) *a geodesic circle can be drawn passing through any point with any center.*
4) *all geodesic right angles are equal to one another.*

In $\mathbb{H}^2(K)$, though we can always extend a hyperbolic *segment,* a hyperbolic *line* has at least one "bound" or "limit" in K. Given the semicircle in the upper half-plane bounded on the x-axis by $(\alpha, 0)$ and $(\beta, 0)$, let us say, as Hilbert does, that α and β are the **ends** of the corresponding hyperbolic line. We may denote the line by

$$\{\alpha, \beta\}.$$

(See also §2.2.) The ends of the vertical hyperbolic line starting from $(\alpha, 0)$ are α and ∞, and the line itself is

$$\{\alpha, \infty\}.$$

Thus

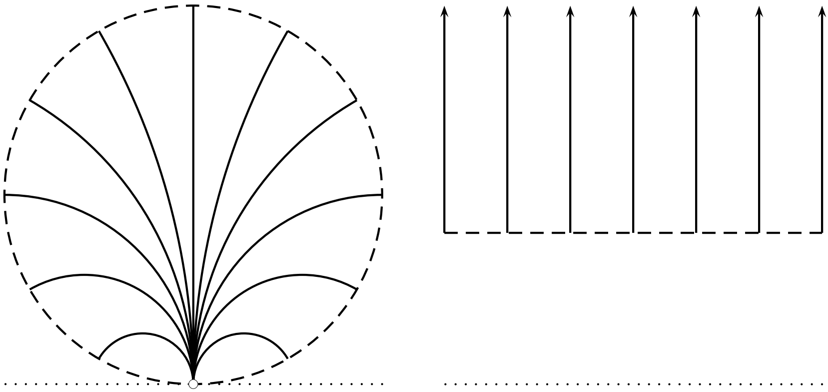


Figure 1.7: Parallel geodesic rays from horocycles

- a hyperbolic line has two ends;
- a hyperbolic segment, two endpoints;
- a hyperbolic ray, a single end and a single endpoint.

Two hyperbolic rays are **parallel** if they have the same end. Then a hyperbolic ray is parallel to a hyperbolic line, and two hyperbolic lines are parallel, if they share an end. Parallel hyperbolic rays do not intersect; but some non-intersecting hyperbolic rays are not parallel.

In each part of Fig. 1.7, geodesic rays are

- parallel to one another,
- at right angles to a **horocycle,** which is a line (not geodesic), the perpendicular bisectors of whose chords all have a common end.

In place of Euclid's fifth postulate, we have the following.

Theorem 1.5. *In $\mathbb{H}^2(K)$ for any K, for any hyperbolic line and any point not lying on it, there are two distinct hyperbolic rays, emanating from that point, each parallel to the given line,*

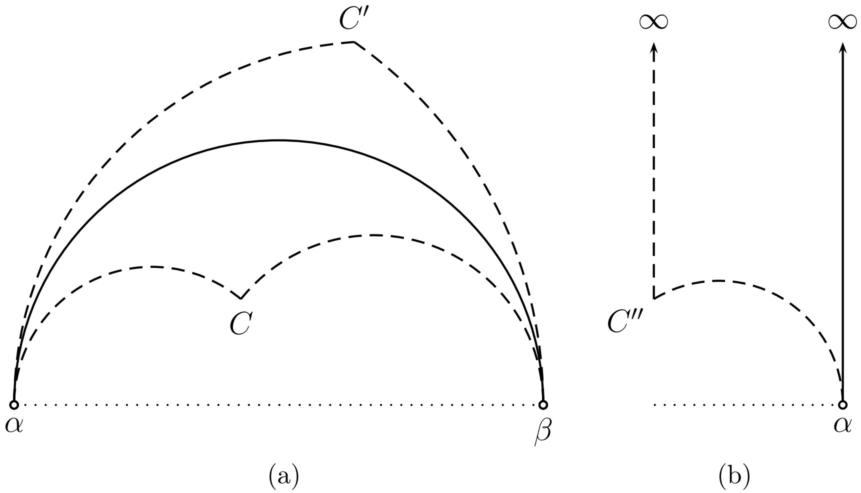


Figure 1.8: Hyperbolic parallel postulate

but together forming an acute angle, within which any hyperbolic ray emanating from the same point cuts the given line.

Proof. If the geodesic line is $\{\alpha, \beta\}$, and the point is C, then the two rays are those that we may write as $C\alpha$ and $C\beta$. See Fig. 1.8. □

1.5 Hyperbolic Automorphisms

An **automorphism** of a geometry in the sense of §1.1 is a permutation of the points that induces a permutation of the straight lines and a permutation of the circles, sending centers to centers. We seek automorphisms of $\mathbb{H}^2(K)$.

Theorem 1.6. *Every automorphism of* $\mathbb{H}^2(K)$ *takes equal segments to equal segments.*

Proof. An automorphism preserves the diagram of Fig. 1.5 used for the proof of Theorem 1.3. □

For all z in $K(\mathbf{i}) \cup \{\infty\}$, we define

$$\left.\begin{aligned} z + \infty = \infty = \infty + z, \\ z \neq 0 \implies \frac{z}{0} = \infty = z \cdot \infty = \infty \cdot z, \\ z \neq \infty \implies \frac{z}{\infty} = 0. \end{aligned}\right\} \tag{1.8}$$

We leave $0/0$, $0 \cdot \infty$, $\infty \cdot 0$, and ∞/∞ undefined; but we do say

$$\overline{\infty} = \infty.$$

We define three permutations of $K(\mathbf{i}) \cup \{\infty\}$ by

$$\mathrm{r}(z) = -\overline{z}, \qquad \iota(z) = \frac{1}{\overline{z}}, \qquad \rho(z) = -\frac{1}{z}.$$

They are indeed permutations, since each is its own inverse. We shall call the first two **reflection** and **inversion** respectively; the third, **rotation,** for reasons that will become apparent. Each of the three permutations is its own inverse and is the composite of the other two; since they have order 2, they commute with one another.

We define also two families of operations on $K(\mathbf{i}) \cup \{\infty\}$ by

$$\mathrm{d}_a(z) = az, \qquad \tau_b(z) = z + b,$$

where a and b are in K, and $a > 0$. The operations are also permutations, since

$$\mathrm{d}_a{}^{-1} = \mathrm{d}_{a^{-1}}, \qquad \tau_b^{-1} = \tau_{-b}.$$

Here d_a is **dilation** by a, and τ_b is **translation** by b.

Each of the permutations of $K(\mathbf{i}) \cup \{\infty\}$ that we have defined is also a permutation of either of

- the upper half-plane $\{z \in K(\mathbf{i}) : \Im(z) > 0\}$,
- $K \cup \{\infty\}$.

Except "rotation," our terms allude to what the operations do in $\mathbb{A}^2(K)$. In $\mathbb{H}^2(K)$,

- d_a is a translation along $\{0, \infty\}$;
- r is reflection about $\{0, \infty\}$;
- ι is reflection about $\{-1, 1\}$;
- ρ is rotation by π about $\mathbf{i}$.

Theorem 1.7. *τ_b, d_a, r, ι, and ρ are automorphisms of $\mathbb{H}^2(K)$.* $\blacksquare$

Proof. Obviously τ_b, d_a, and r preserve hyperbolic linearity and circularity. It is now enough to show that ι does the same, since $\rho = \mathrm{r} \circ \iota$. We show that ι

1) fixes $\{0, \infty\}$,
2) interchanges
   a) $\{\alpha, \infty\}$ and $\{\alpha^{-1}, 0\}$, if $\alpha \neq 0$;
   b) $\{\alpha - \rho, \alpha + \rho\}$ and $\{(\alpha - \rho)^{-1}, (\alpha + \rho)^{-1}\}$, if $\alpha \neq \rho$.

To do so, we note that in $\{z \in K(\mathbf{i}) : \Im(z) > 0\}$,

- $\{\alpha, \infty\}$ is defined by $z + \overline{z} = 2\alpha$;
- $\{\alpha - \rho, \alpha + \rho\}$, by $|z - \alpha|^2 = \rho^2$.

We show that ι

1) fixes $z + \overline{z} = 0$,
2) interchanges
   a) $z + \overline{z} = 2\alpha$ and $\left|z - \dfrac{1}{2\alpha}\right|^2 = \left(\dfrac{1}{2\alpha}\right)^2$,
   b) $|z - \alpha|^2 = \rho^2$ and $\left|z - \dfrac{\alpha}{\alpha^2 - \rho^2}\right| = \left(\dfrac{\rho}{\alpha^2 - \rho^2}\right)^2$.

We do this by replacing z with $\overline{z}^{-1}$ in one equation, then deriving the other. It is clear that $z + \overline{z} = 0$ remains unchanged. In the other cases,

a) the line is sent to the curve given by

$$\begin{aligned}\overline{z}^{-1} + z^{-1} &= 2\alpha,\\ z + z^{-1} &= 2\alpha|z|^2,\\ |z|^2 - \frac{1}{2\alpha}z - \frac{1}{2\alpha}\overline{z} &= 0,\end{aligned}$$

and thus to the indicated circle;

b) the first circle is sent to the curve given by

$$\begin{aligned}|\overline{z}^{-1} - \alpha|^2 &= \rho^2,\\ |1 - \alpha\overline{z}|^2 &= \rho^2|z|^2,\\ \left(\alpha^2 - \rho^2\right)|z|^2 - \alpha\overline{z} - \alpha z &= -1,\\ |z|^2 - \frac{\alpha\overline{z}}{\alpha^2 - \rho^2} - \frac{\alpha z}{\alpha^2 - \rho^2} &= \frac{1}{\rho^2 - \alpha^2},\end{aligned}$$

and thus to the second circle.

So ι permutes hyperbolic lines. In particular, it permutes the circles not passing through 0 *and* that have centers on the x-axis. Therefore it does the same without the last condition. Moreover, when the circles are in the upper half-plane, ι also takes hyperbolic centers to hyperbolic centers, since

$$ac = b^2 \implies \frac{1}{a} \cdot \frac{1}{c} = \left(\frac{1}{b}\right)^2. \qquad \square$$

By definition,

$$\mathrm{SL}_2(K) = \left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1\right\}, \qquad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

An **action** of a group (such as $\mathrm{SL}_2(K)$) on a structure (such as a geometry) is a homomorphism from the group into the group of automorphisms of the structure.

Theorem 1.8. $\mathrm{SL}_2(K)$ *acts on* $\mathbb{H}^2(K)$ *by the rule*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}.$$

The range of the action is generated by ρ *and the* τ_b *and* d_a.

Proof. The listed automorphisms are the images of

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & b \\ 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} \sqrt{a} & 0 \\ 0 & \sqrt{a^{-1}} \end{pmatrix}$$

respectively. Composition of automorphisms will correspond to multiplication of matrices, since

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix},$$

$$\frac{a\dfrac{ez+f}{gz+h}+b}{c\dfrac{ez+f}{gz+h}+d} = \frac{(ae+bg)z+af+bh}{(ce+dg)z+cf+dh}.$$

Finally, if $ad-bc=1$, then

$$c \neq 0 \implies \frac{az+b}{cz+d} = \frac{a}{c} + \frac{bc-ad}{c(cz+d)} = (\tau_{a/c} \circ \rho \circ \tau_{cd} \circ \mathrm{d}_{c^2})z,$$

$$c = 0 \implies \frac{az+b}{cz+d} = \frac{az+b}{d} = \frac{b}{d} + \frac{a}{d} \cdot z = (\tau_{b/d} \circ \mathrm{d}_{a/d})z. \quad \square$$

Lemma 1.1. *The action of* $\mathrm{SL}_2(K)$ *on* $\mathbb{H}^2(K)$ *takes any point to any point.*

Proof. $c + d\mathbf{i} = (\tau_c \circ \mathrm{d}_{d/b} \circ \tau_{-a})(a + b\mathbf{i})$. $\square$

Theorem 1.9. *The action of* $\mathrm{SL}_2(K)$ *on* $\mathbb{H}^2(K)$ *takes any ray to any ray.*

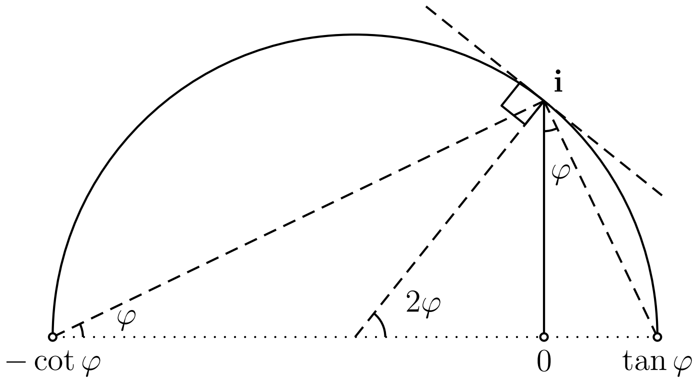


Figure 1.9: Hyperbolic line through **i**

Proof. By Lemma 1.1, it is enough to consider rays with endpoint **i**. Such a ray has the end $\tan\varphi$ for some φ, and then the normal to the ray at **i** makes angle 2φ to the x-axis, as in Fig. 1.9. Since

$$\frac{a\mathbf{i}+b}{c\mathbf{i}+d}=\mathbf{i}\iff a\mathbf{i}+b=d\mathbf{i}-c\iff d=a\ \&\ c=-b,$$

an element of $\mathrm{SL}_2(K)$ that fixes **i** is precisely a matrix

$$\begin{pmatrix}\cos\vartheta & \sin\vartheta\\ -\sin\vartheta & \cos\vartheta\end{pmatrix}$$

for some ϑ. We may denote the resulting operation on $\mathbb{H}^2(K)$ by

$$\rho_{2\vartheta},$$

since, applying the matrix to the end $\tan\varphi$, we compute

$$\frac{\cos\vartheta\tan\varphi+\sin\vartheta}{-\sin\vartheta\tan\varphi+\cos\vartheta}=\frac{\cos\vartheta\sin\varphi+\sin\vartheta\cos\varphi}{-\sin\vartheta\sin\varphi+\cos\vartheta\cos\varphi}$$
$$=\frac{\sin(\vartheta+\varphi)}{\cos(\vartheta+\varphi)}=\tan(\vartheta+\varphi).$$

Thus $\rho_{2\vartheta}$ effects rotation about $\mathbf{i}$ by 2ϑ. □

Porism. *The kernel of the action of* $\mathrm{SL}_2(K)$ *on* $\mathbb{H}^2(K)$ *is the two-element subgroup generated by*

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

We denote by

$$\mathrm{PSL}_2(K)$$

the quotient of $\mathrm{SL}_2(K)$ by the kernel in the porism. We may consider the quotient as a group of automorphisms of $\mathbb{H}^2(K)$.

Theorem 1.10. $\mathrm{PSL}_2(K)$ *is closed under conjugation by* r.

Proof. By Theorem 1.8, we need only note

$$\mathrm{r}\circ\tau_b\circ\mathrm{r}=\tau_{-b},\qquad \mathrm{r}\circ\mathrm{d}_a\circ\mathrm{r}=\mathrm{d}_a,\qquad \mathrm{r}\circ\rho\circ\mathrm{r}=\rho.$$ □

We can now understand the group of automorphisms of $\mathbb{H}^2(K)$ generated by all of the elements listed in Theorem 1.7 as the semidirect product

$$\mathrm{PSL}_2(K)\rtimes\langle\mathrm{r}\rangle.$$

Theorem 1.11. *For any one point and two ends in* $\mathbb{H}^2(K)$, *the group* $\mathrm{PSL}_2(K)\rtimes\langle\mathrm{r}\rangle$ *of automorphisms of* $\mathbb{H}^2(K)$ *has an element fixing the point and interchanging the ends.*

Proof. By Theorem 1.9, there is at most one such element, and we may assume the point is $\mathbf{i}$. Letting the ends be $\tan\vartheta_0$ and $\tan\vartheta_1$, we have

$$(\rho_{2(\vartheta_0+\vartheta_1)} \circ \mathrm{r}) \tan\vartheta_j = \tan\vartheta_{1-j}. \qquad \square$$

Theorem 1.12. *In $\mathbb{H}^2(K)$, if two segments are equal, then an automorphism in $\mathrm{PSL}_2(K)$ takes one segment to the other.*

Proof. By Theorem 1.9 and Fig. 1.5, used for the proof of Theorem 1.3, if two segments are equal, some automorphism in $\mathrm{PSL}_2(K)$ takes one to the other. □

Theorem 1.13. *The automorphisms in $\mathrm{PSL}_2(K) \rtimes \langle \mathrm{r} \rangle$ preserve lengths in $\mathbb{H}^2(K)$.*

Proof. In the proof of Theorem 1.9, since

$$\tan(\vartheta + \varphi) = \tan(\vartheta' + \varphi) \iff \rho_{2\vartheta} = \rho_{2\vartheta'},$$

$\rho_{2\vartheta}$ and is the unique automorphism in $\mathrm{PSL}_2(K)$ to take the given ray to the ray with the same endpoint $\mathbf{i}$ and end $\tan(\vartheta + \varphi)$. Hence, when a segment AB and a ray with endpoint C are given, a unique element σ of $\mathrm{PSL}_2(K)$ sends A to C and B to a point D of the ray. By Theorem 1.3, there is a point D' of the ray such that $AB = CD'$. Then σ must be the automorphism, which exists by Theorem 1.12, that takes B to D'; so D' is D. □

Theorem 1.14. *The automorphisms in $\mathrm{PSL}_2(K) \rtimes \langle \mathrm{r} \rangle$ preserve angle measurements in $\mathbb{H}^2(K)$.*

Proof. This being clear for the other generators, we need only show it for ι. Since

$$\iota = \mathrm{d}_a \circ \iota \circ \mathrm{d}_a,$$

it is enough to consider angles whose vertices are on the unit circle. More than that, if rot_ϑ is rotation in $\mathbb{A}^2(K)$ by ϑ about 0, then this preserves angles, and

$$\mathrm{rot}_{-\vartheta} \circ \iota \circ \mathrm{rot}_\vartheta,$$

so it is enough to consider angles with vertices at $\mathbf{i}$. Since $\iota = \mathrm{r} \circ \rho$, it is enough to observe that, by the proof of Theorem 1.9, ρ preserves angles at $\mathbf{i}$. □

Theorem 1.15. *SAS holds in* $\mathbb{H}^2(K)$.

In $\mathbb{H}^2(K)$ we can now prove all results of so-called *absolute* geometry, namely Euclidean geometry without a parallel postulate (either Euclid's or our Theorem 1.5). For example, we have the following two theorems.

Theorem 1.16. *SSS holds in* $\mathbb{H}^2(K)$.

Theorem 1.17. *In* $\mathbb{H}^2(K)$, *three points are collinear if each is equidistant from the same two points.*

An **isometry** of a geometry that satisfies the absolute axioms is a permutation of the points that preserves lengths. An isometry is automatically an automorphism.

- An isometry of $\mathbb{H}^2(K)$ preserves angle measurements, by Theorem 1.16.
- Every element of $\mathrm{PSL}_2(K) \rtimes \langle \mathrm{r} \rangle$ is an isometry of $\mathbb{H}^2(K)$, by Theorem 1.13.

Theorem 1.18. *Every isometry of* $\mathbb{H}^2(K)$ *is an automorphism in* $\mathrm{PSL}_2(K) \rtimes \langle \mathrm{r} \rangle$.

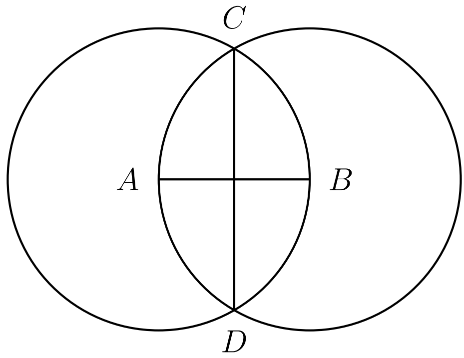


Figure 1.10: Perpendicular bisector construction

Proof. By Theorem 1.17, only the identical isometry fixes each of three non-collinear points. Thus every isometry is determined by where it sends such points. By Theorems 1.16, 1.9, 1.11, and 1.13, some automorphism in $\mathrm{PSL}_2(K) \rtimes \langle \mathrm{r} \rangle$ agrees with the isometry at those points, and therefore everywhere. □

It is not clear whether $\mathbb{H}^2(K)$ has automorphisms other than those that we have found. For example, $\mathbb{A}^2(K)$ has automorphisms, such as dilations, that are not isometries.

In $\mathbb{H}^2(K)$ and $\mathbb{A}^2(K)$, every automorphism preserves right angles, because it preserves the diagram in Fig. 1.10.

We shall not know till §2.5 that automorphisms of $\mathbb{H}^2(K)$ preserve parallelism.

2 Plane

2.1 Axioms

Hilbert's axioms for **hyperbolic geometry** (strictly, hyperbolic *plane geometry*) are as follows. (The naming and numeration are his, not the exact words.)

I. Incidence.

- **1, 2.** Two points lie on exactly one line.
- **3.** For every line, at least two points lie on it, but one does not.

II. Order.

- **1.** If B lies between A and C, then
  - all three are distinct points of one line, and
  - B lies between C and A.
- **2.** Between any two points lies a third.
- **3.** Of three points of a line, no more than one lies between the other two.
- **4.** If ABC is a proper triangle, a line crossing AB crosses either AC or CB (if not C itself).

III. Congruence.

- **1.** On a given line, on either side of a given point (this uses order), we can mark off a segment congruent to a given segment.
- **2.** Congruence of segments is transitive.
- **3.** Sums of congruent segments are congruent.

4. A given ray can always serve as the side of an angle congruent to a given angle.
5. Side-Angle-Side.

IV. Parallels. For any line b and point A not on it, there are two rays emanating from A that do not cut b, but form an acute angle, inside of which, any ray from A cuts b.

In the Parallel Axiom, any point of b divides it into two rays, each **parallel** to one of the rays emanating from A. According to Hilbert, that parallelism is symmetric and transitive "follows immediately." Apparently this does not mean the proofs are trivial. For symmetry, a note refers to "a method due to Gauss. Cf. Bonola-Liebmann, *Die nichteuklidische Geometrie* (Leipzig, 1908 and 1921)" and Bonola [2, pp. 70–1] (Hilbert himself gives no page references). Gauss's proof has two cases; Lobachevski's [11], with one, seems simpler.

2.2 Ends

A ray and the rays parallel to it compose an **end.** (This is consistent with the usage initiated in §1.4.) A point A and an end α determine the ray $A\alpha$. A line with ends α and β is unique; we shall denote such a line by

$$\{\alpha, \beta\},$$

since the ends of a line have no intrinsic order. However, Hilbert calls the line (α, β). We shall construct this line in Lemma 2.3, using Lemma 2.2, which needs Lemma 2.1. (The lemmas are Hilbert's, along with Lemmas 2.4 and 2.5; the labelling of theorems as such is our addition.)

Lemma 2.1. *Lines making equal alternate angles with a third are not parallel in any direction.*

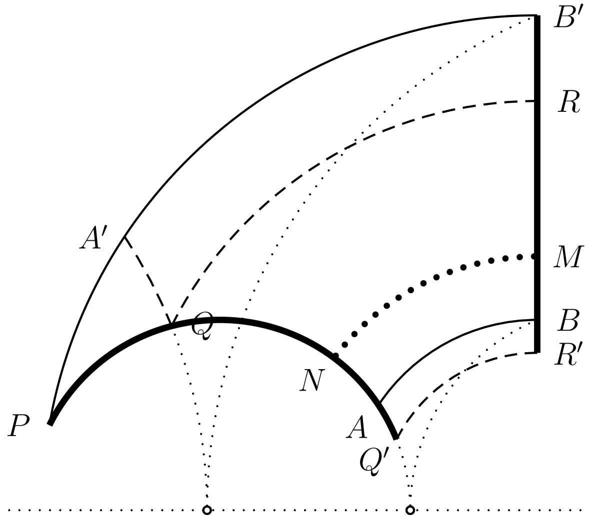


Figure 2.1: Common perpendicular

Proof. If parallel in one direction, then in the other, by symmetry (which can be spelled out). □

Lemma 2.2. *Two lines, as AP and BB' in Fig. 2.1, that neither intersect nor are parallel have a common perpendicular, MN.*

Proof. The construction and demonstration is as follows (reading by columns):

$AB, PB' \perp BB'$,	$QR \perp BB'$,
$B'P > BA = B'A'$,	$BR' = B'R$,
$B'A'Q = BAQ'$,	$B'A'QR \cong BAQ'R'$,
$AQ' = A'Q$,	$Q'R' = QR$,

and we let MN join the midpoints of RR' and QQ'. In the simpler case, where $B'P = BA$, then MN joins the midpoints

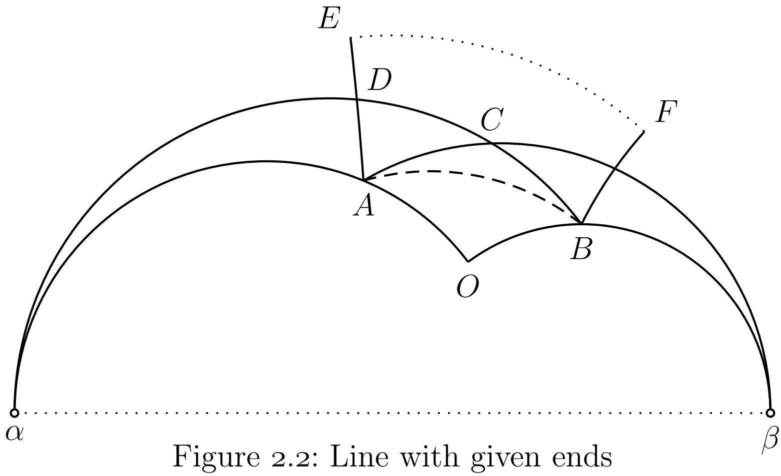


Figure 2.2: Line with given ends

of $B'B$ and PA. Otherwise, the point Q exists on AP because the parallel to $A'Q$ from B' must cut PA, because that parallel neither cuts nor, by Lemma 2.1, is parallel to the parallel to PA drawn from B. □

We shall use now an infinitary version of Side-Angle-Side: if $AB = A'B'$, and $BA\alpha$ and $B'A'\alpha'$ are equal as angles, then so are $AB\alpha$ and $A'B'\alpha'$.

Lemma 2.3. *For any distinct ends α and β as in Fig. 2.2, the line $\{\alpha, \beta\}$ is the common perpendicular EF of the bisectors of angles $CA\alpha$ and $CB\beta$, where $OA = OB$.*

Proof. We shall show

1) the bisectors neither intersect nor are parallel;
2) EF lies on $\{\alpha, \beta\}$.

To establish the latter claim, we observe (usually by symmetry)

$$\begin{aligned} \alpha OB &\cong \beta OA, \\ OBC &= OAC, \\ \alpha AE &= \alpha BF, \\ AE &= BF, \\ \alpha AE &\cong \alpha BF, \end{aligned}$$

so the rays from E and F respectively having end α make the same angle with EF, and therefore they coincide with EF by Lemma 2.1.

Suppose if possible the angle bisectors meet at a point M. Then $AM = BM$ and $\alpha AM = \alpha BM$, so

$$\alpha AM \cong \alpha BM,$$

and in particular $\alpha MA = \alpha MB$, which is absurd.

Suppose if possible the angle bisectors have the common end μ. If DB', equal to DA, is marked off along DB, then

$$\alpha AD \cong \mu B'D,$$

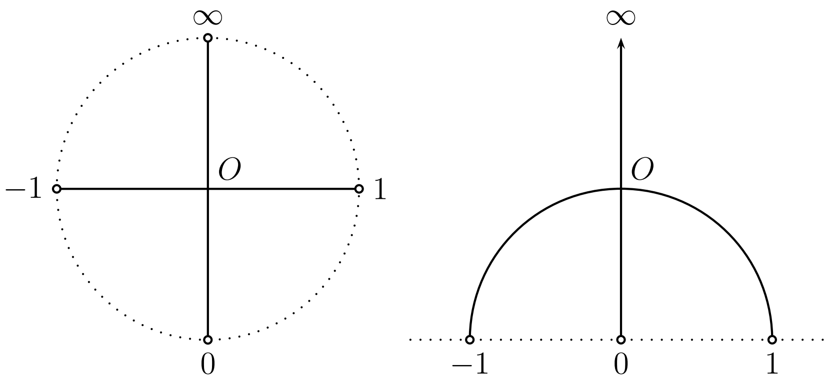
and so $\mu B'D = \alpha AD = \mu BD$, which is absurd by Lemma 2.1 unless B' coincides with B. In this case we have another absurdity,

$$DAB = DBA = CBA = CAB. \qquad \square$$

2.3 Euclidean Field

Now let us denote

- some particular end by ∞,



(a) Disk model (b) Half-plane model

Figure 2.3: Line of ends

- the set of remaining ends by K,
- a particular element of K by 0,
- a particular point on $\{0, \infty\}$ by O,
- one end of the perpendicular to $\{0, \infty\}$ at O by 1.

We shall make K into a Euclidean field, as suggested by Fig. 2.3.

2.3.1 Ordering

On the set of all ends, we define a quaternary relation Q so that $Q(\alpha, \beta, \gamma, \delta)$ if and only if $\{\alpha, \gamma\}$ crosses $\{\beta, \delta\}$. If α, β, γ, and δ are distinct ends, then

$$Q(\alpha, \beta, \gamma, \delta) \iff Q(\alpha, \beta, \delta, \gamma),$$

and also

$$Q(\alpha, \beta, \gamma, \delta) \iff Q(\alpha^\sigma, \beta^\sigma, \gamma^\sigma, \delta^\sigma),$$

where σ is an even permutation of the set $\{\alpha, \beta, \gamma, \delta\}$. We use Q to order K by defining

$$\alpha < 0 \iff Q(\alpha, 0, 1, \infty).$$

If neither of α and β is 0, we define $\alpha < \beta$ if and only if one of the following holds:

$$\begin{gathered}\alpha < 0 \ \ \& \ \ Q(\infty, \alpha, \beta, 0),\\ \alpha < 0 \ \ \& \ \ 0 < \beta,\\ 0 < \alpha \ \ \& \ \ Q(0, \alpha, \beta, \infty).\end{gathered}$$

The **positive** ends are those finite ends α for which $0 < \alpha$.

2.3.2 Addition

Our definition of addition on K will be motivated by an analogy. In $\mathbb{C}$, when $\alpha \in \mathbb{R}$, we can define r_α as reflection about $z + \overline{z} = 2\alpha$, so that

$$\mathrm{r}_\alpha(z) = \alpha - (z - \alpha) = 2\alpha - z.$$

Then

$$\begin{aligned}\mathrm{r}_\alpha \mathrm{r}_0 \mathrm{r}_\beta(z) &= \mathrm{r}_\alpha \mathrm{r}_0(2\beta - z)\\ &= \mathrm{r}_\alpha(z - 2\beta)\\ &= 2\alpha + 2\beta - z\\ &= \mathrm{r}_{\alpha+\beta}(z).\end{aligned}$$

In our hyperbolic plane, when α in K, we define r_α to be reflection about $\{\alpha, \infty\}$, so that, if

$$\mathrm{r}_\alpha(P) = P'$$

this means either

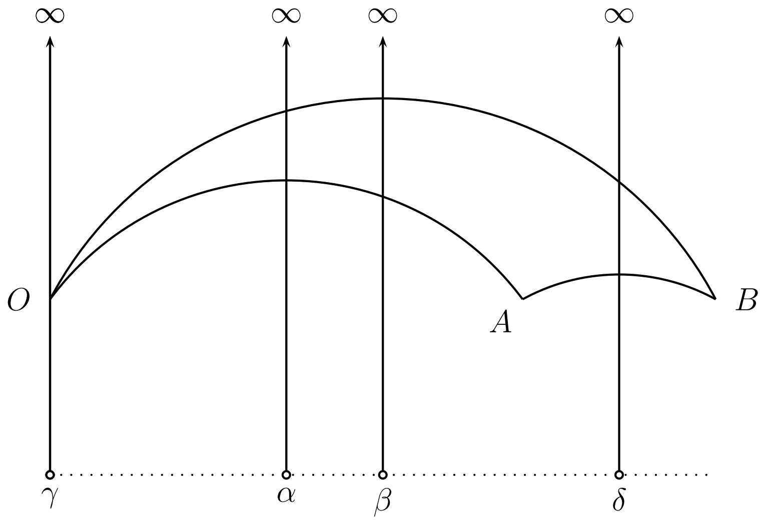


Figure 2.4: Parallel perpendicular bisectors

- P is on $\{0, \infty\}$, and $P' = P$, or
- $\{0, \infty\}$ is the perpendicular bisector of PP'.

Theorem 2.1. *Reflection about a line is an automorphism of the hyperbolic plane.*

Proof. We have to show that reflection preserves lengths and angles. This follows from Side-Angle-Side. □

The following corresponds to Lobachevski's Theorem 30 and is illustrated by Fig. 2.4.

Lemma 2.4. *In any triangle AOB, if the perpendicular bisectors of AO and OB are parallel, then they are parallel to that of AB.*

Proof. Hartshorne [9] improves on the proofs of Lobachevski and Hilbert by using Lemma 2.2, as follows. If the perpendicular bisectors of AO and AB

meet, that of OB meets them at the same point, by the usual proof;

neither meet nor are parallel, the perpendiculars dropped to their common perpendicular from O, A, and B are equal, so the perpendicular bisector of OB must also be perpendicular to that common perpendicular;

are parallel, OB can only be parallel to them. □

Lemma 2.5. *If α, β, and γ belong to K, then for some δ in K,*

$$\mathrm{r}_\beta \mathrm{r}_\gamma \mathrm{r}_\alpha = \mathrm{r}_\delta.$$

Proof. Hilbert has three cases, but we need only two, by the method of Hartshorne. (Hartshorne omits the easier case.)

1. If the three finite ends are distinct as in Fig. 2.4, let O be a point on $\{\gamma, \infty\}$, and let

$$\mathrm{r}_\alpha O = A, \qquad \mathrm{r}_\beta O = B.$$

The perpendicular bisector of AB is $\{\delta, \infty\}$ for some δ, by Lemma 2.4. The operation

$$\mathrm{r}_\delta \mathrm{r}_\beta \mathrm{r}_\gamma \mathrm{r}_\alpha$$

fixes A and the line through it with end ∞; being also the product of an even number of reflections, the operation is the identity.

2. If the three ends are not distinct, the only possibility that is not completely trivial is that α and β coincide. In this case, we let

$$\delta = \mathrm{r}_\alpha \gamma.$$

□

We can now define addition in K by the rule

$$\mathrm{r}_\alpha \mathrm{r}_0 \mathrm{r}_\beta = \mathrm{r}_{\alpha+\beta}. \tag{2.1}$$

Theorem 2.2. *The addition just defined makes K an ordered abelian group.*

Proof. Addition is

- associative, since composition is associative;
- commutative, since composition of automorphisms having order 2 is commutative.

The additive inverse $-\alpha$ is defined so that the image of $\{\alpha, \infty\}$ under r_0 is $\{-\alpha, \infty\}$. That the sum of positive ends is positive can be read from Fig. 2.4 when $\gamma = 0$, so that $\delta = \alpha + \beta$. □

For a finite end α we define

$$\infty + \alpha = \infty = \alpha + \infty.$$

It is not now clear whether an automorphism of the hyperbolic plane must preserve parallelism and therefore induce a permutation of the ends. However, some automorphisms do this.

Theorem 2.3. *Addition of a finite end α is a permutation of $K \cup \{\infty\}$ induced by an automorphism of the plane.*

Proof. That automorphism is $\mathrm{r}_{\alpha/2}\mathrm{r}_0$. □

2.3.3 Multiplication and Inversion

The choice of O on $\{0, \infty\}$ makes the points of this line into an abelian group under addition. We make the positive ends into a multiplicative abelian group isomorphic to this one under the map that takes α to the point where $\{\alpha, -\alpha\}$ intersects $\{0, \infty\}$; the intersection is at right angles. If A is a point on $\{0, \infty\}$ corresponding to a positive end α, then the segment OA has a midpoint B, corresponding to a positive end β, and in this case

$$\beta = \sqrt{\alpha}.$$

We extend the multiplication to all of $K \smallsetminus \{0\}$ by the usual rules of signs. We extend multiplication partially to $K \cup \{\infty\}$ as in (1.8) in §1.5.

Theorem 2.4. 1. *Multiplication by a positive end is a permutation of $K \cup \{\infty\}$ induced by an automorphism of the plane.*
2. *Multiplication by -1 is induced by* r_0.
3. *Inversion of ends is induced by reflection about* $\{-1, 1\}$.

Proof. Multiplication by a positive end γ is induced by the automorphism such that, as in Fig. 2.5, $O \cdot \gamma$ is the intersection of $\{0, \infty\}$ and $\{\gamma, -\gamma\}$, and if Q is the foot of the perpendicular from a point P dropped to $\{0, \infty\}$, then $Q \cdot \gamma$ is the foot of the perpendicular from $P \cdot \gamma$ dropped to $\{0, \infty\}$, and, as segments,

$$(P \cdot \gamma)(Q \cdot \gamma) = PQ, \qquad Q(Q \cdot \gamma) = O(O \cdot \gamma). \qquad \square$$

As a special case of (2.1),

$$\mathrm{r}_{\alpha \cdot \gamma + \beta \cdot \gamma} = \mathrm{r}_{\alpha \cdot \gamma} \mathrm{r}_0 \mathrm{r}_{\beta \cdot \gamma}.$$

Applying the automorphism of Theorem 2.4 to (2.1), we obtain

$$\mathrm{r}_{\alpha \cdot \gamma} \mathrm{r}_0 \mathrm{r}_{\beta \cdot \gamma} = \mathrm{r}_{(\alpha + \beta) \cdot \gamma}.$$

Thus multiplication distributes over addition on K.

Given a hyperbolic plane, we have obtained an ordered field K. It is straightforward that, as Hartshorne shows [9, 43.1, p. 416–8],

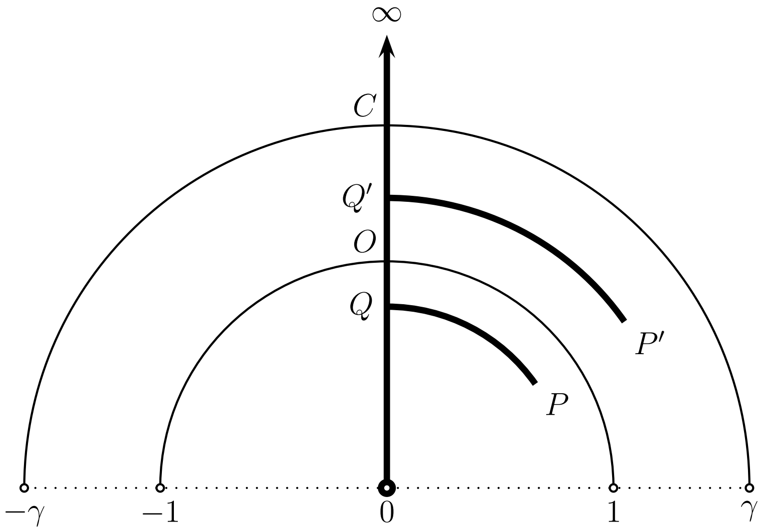


Figure 2.5: Automorphism

1) K is independent (up to isomorphism) of the choice of $(\infty, 0, 1)$, since an automorphism of the plane will take one choice to another;
2) two hyperbolic planes with isomorphic fields are themselves isomorphic, since an isomorphism of lines determines an isomorphism of points.

2.4 Coordinates

Having obtained an ordered field K from a hyperbolic plane and a choice of ends denoted respectively ∞, 0, and 1, we shall describe the plane in terms of K. First we note which lines pass through O.

Theorem 2.5. *In the hyperbolic plane,* *1) the line* $\{\xi, \infty\}$ *contains* O *if and only if*

$$\xi = 0,$$

2) the line $\{\xi, \eta\}$*, neither end infinite, contains* O *if and only if*

$$\xi \cdot \eta = -1.$$

Proof. 1. Clear.

2. Reflection of the plane about $\{-1, 1\}$, and then $\{0, \infty\}$, fixes

- O, and only this, of all points of the plane;
- $\{\alpha, -\alpha^{-1}\}$, and only this, of all lines $\{\alpha, \beta\}$, whenever $\alpha \in K \setminus \{0\}$.

Since the line does cross $\{0, \infty\}$, it must cross at the point. □

Theorem 2.6. *The point* $O \cdot \alpha$ *lies on* $\{0, \infty\}$ *and on every line* $\{-\alpha \cdot \omega, \alpha \cdot \omega^{-1}\}$*, but no other line.*

Proof. The point is, by Theorem 2.4, the intersection of $\{0, \infty\}$ and $\{-\alpha, \alpha\}$. The rest follows also from Theorem 2.5. □

We want to know which lines pass through an arbitrary point P. For this, we seek an automorphism of the plane that takes O to P.

Theorem 2.7. *In the hyperbolic plane, a line* $\{\xi, \eta\}$ *that is not* $\{0, \infty\}$ *contains the intersection of* $\{-\nu, \nu\}$ *and* $\{\mu, \infty\}$ *if and only if the point*

$$(\xi\eta, \xi + \eta)$$

in K^2 *lies on the line defined by*

$$x - \mu y + \nu^2 = 0. \tag{2.2}$$

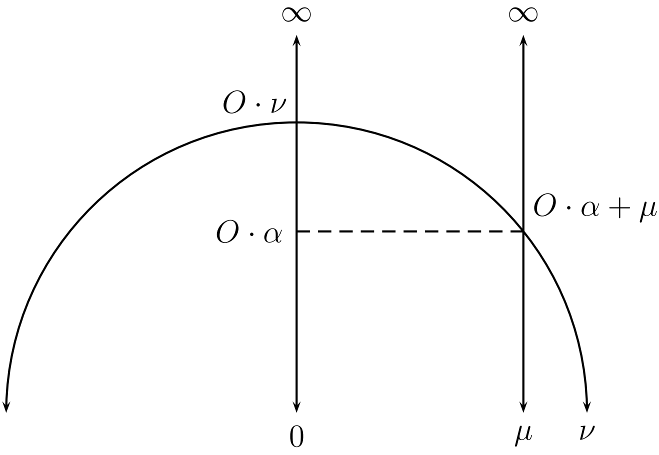


Figure 2.6: Lines through an arbitrary point

Proof. We show first that the intersection is $O \cdot \alpha + \mu$ for *some* α, as in Fig. 2.6. By applying the automorphisms of Theorems 2.3 and 2.4 to Theorem 2.6, we know that any point $O \cdot \alpha + \mu$ is the intersection with $\{\mu, \infty\}$ of no other line, but every line

$$\{-\alpha \cdot \omega + \mu, \alpha \cdot \omega^{-1} + \mu\}.$$

To find which of these is orthogonal to $\{0, \infty\}$, we solve

$$-(-\alpha \cdot \omega + \mu) = \alpha \cdot \omega^{-1} + \mu,$$
$$\alpha \cdot \omega^2 - 2\mu \cdot \omega - \alpha = 0,$$
$$\omega = \frac{\mu \pm \sqrt{\mu^2 + \alpha^2}}{\alpha}, \qquad \omega^{-1} = \omega - \frac{2\mu}{\alpha}.$$

Thus $O \cdot \alpha + \mu$ is the intersection of $\{\mu, \infty\}$ and

$$\{-\sqrt{\mu^2 + \alpha^2}, \sqrt{\mu^2 + \alpha^2}\}.$$

Again by Theorem 2.6, $O \cdot \alpha + \mu$ lies on $\{\xi, \eta\}$ if and only if

$$\frac{\xi - \mu}{\alpha} \cdot \frac{\eta - \mu}{\alpha} = -1,$$
$$\xi\eta - \mu(\xi + \eta) + \mu^2 + \alpha^2 = 0. \qquad \square$$

Three lines $\{\alpha, \beta\}$, different from $\{0, \infty\}$, have a common point if and only if the three points $(\alpha\beta, \alpha + \beta)$ are collinear in K^2. Thus we recover the hyperbolic plane from K.

Every line defined by an equation (2.2) corresponds to a point of the hyperbolic plane, provided

$$-\nu < \mu < \nu.$$

Hilbert refers to $\big(\xi\eta, (\xi + \eta)/2\big)$ as the **coordinates** of the line $\{\xi, \eta\}$, and he concludes,

> Having seen that the equation of a point in line coordinates is linear it is easy to deduce the special case of Pascal's Theorem for a pair of lines and Desargues' Theorem for perspectively situated triangles as well as the other theorems of projective geometry.

"Pascal's Theorem for a pair of lines" is Pappus's Theorem, which, along with Desargues's Theorem, holds in the projective plane over K. Therefore the duals also hold, and so the theorems themselves hold in the hyperbolic plane.

Hilbert continues:

> The familiar formulas of Bolyai-Lobachevskian geometry can then also be derived with no difficulty and the development of this geometry has been thus completed with the aid of Axioms I-IV alone.

Hartshorne remarks [9, p. 394],

> we have found that calculations seem to work out better if we continue to think of a line as given by coordinates . . . and a point as given by an equation. This is the opposite of the analytic geometry we are used to . . .

He derives the construction of parallels in the next section, and some formulas, though without discussing Pascal and Desargues, as far as I can tell, except for a brief mention of the former [9, p. 426],

1) as proved by Hjelmslev in the projective plane, in which he embeds an arbitrary Hilbert plane;
2) as shown by Hilbert to be equivalent to commutativity of the field of segment arithmetic.

This is in the context of an historical sketch of doing geometry without continuity, which "reached its modern form in the book of Artin [1]."

2.5 Bolyai's Parallel Construction

We establish a ruler-and-compass construction of parallels, assuming they exist.

Theorem 2.8. *In the hyperbolic plane, if*

- *perpendiculars PS and QR to PQ are erected,*
- *the perpendicular dropped from R to PS meets this at S,*
- *the parallel to QR from P meets SR at T,*

then

$$PT = QR.$$

Proof. Since $PQ \perp PS$, we may assume, as in Fig. 2.7,

$$PQ = \{0, \infty\}, \qquad PS = \{-1, 1\}, \qquad P = O.$$

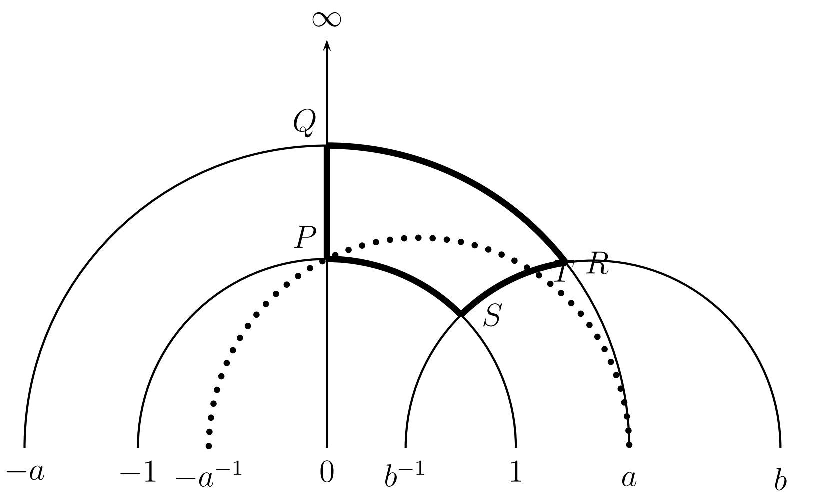


Figure 2.7: Parallel construction

Since also $PQ \perp QR$, for some finite end a,

$$QR = \{-a, a\}.$$

Since $PS \perp RS$, for some finite end b,

$$RS = \{b, b^{-1}\}.$$

Then finally, since PT passes through O and is parallel to QR, by Theorem 2.6,

$$PT = \{-a^{-1}, a\}.$$

Rotation of the plane about P so that a goes to 0 is effected by the automorphism

$$x \mapsto \frac{x-a}{ax+1},$$

since
$$\frac{-x^{-1}-a}{-ax^{-1}+1} = -\frac{1+ax}{-a+x}.$$
First translating Q to P, then rotating about P so that 1 goes to 0, is effected by the automorphism
$$x \mapsto \frac{x-a}{x+a}.$$
These send $\{b, b^{-1}\}$ to lines that cross PQ at the same point, since the products of their ends are the same:
$$\begin{aligned}\frac{x-a}{ax+1} \cdot \frac{x^{-1}-a}{ax^{-1}+1} &= \frac{x-a}{ax+1} \cdot \frac{1-ax}{a+x} \\ &= \frac{x-a}{a+x} \cdot \frac{1-ax}{1+ax} = \frac{x-a}{x+a} \cdot \frac{x^{-1}-a}{x^{-1}+a}. \quad \square\end{aligned}$$

According to Hartshorne, by the classification of Hilbert planes (satisfying Axioms I–III) by Pejas, the construction must work in any such plane satisfying Archimedes's Axiom and also the axiom of intersection of two circles (they intersect if each encircles a point of the other).

Bibliography

[1] E. Artin. *Geometric Algebra.* Interscience, New York, 1957.

[2] Roberto Bonola. *Non-Euclidean Geometry: A Critical and Historical Study of its Development.* Dover, New York, 1955. Authorized English translation with additional appendices by H. S. Carslaw, and with an Introduction by Federigo Enriques, first published by Open Court, 1912. With a supplement containing the Dr. George Bruce Halstead translations of The Science of Absolute Space by John Bolyai (from the fourth, 1896 edition) and The Theory of Parallels by Nicholas Lobachevski (1891; from the Open Court edition of 1914).

[3] Karol Borsuk and Wanda Szmielew. *Foundations of Geometry: Euclidean and Bolyai-Lobachevskian Geometry, Projective Geometry.* North-Holland, Amsterdam, 1960. Revised English translation.

[4] Euclid. *Euclidis Elementa*, volume I of *Euclidis Opera Omnia.* Teubner, Leipzig, 1883. Edited with Latin interpretation by I. L. Heiberg. Books I–IV.

[5] Euclid. *The Thirteen Books of Euclid's* Elements. Dover Publications, New York, 1956. Translated from the text of Heiberg with introduction and commentary by Thomas L. Heath. In three volumes. Republication of the second edition of 1925. First edition 1908.

[6] Euclid. *Στοιχεία Εὐκλείδου.* `users.ntua.gr/dimour/euclid/`, 1999. Edited by Dimitrios E. Mourmouras. Accessed December 24, 2014.

[7] Euclid. *The Bones: A handy where-to-find-it pocket reference companion to Euclid's* Elements. Green Lion Press, Santa Fe, NM, 2002. Conceived, designed, and edited by Dana Densmore.

[8] Euclid. *Euclid's Elements.* Green Lion Press, Santa Fe, NM, 2002. All thirteen books complete in one volume. The Thomas L. Heath translation, edited by Dana Densmore.

[9] Robin Hartshorne. *Geometry: Euclid and beyond.* Undergraduate Texts in Mathematics. Springer-Verlag, New York, 2000.

[10] David Hilbert. *Foundations of Geometry.* Open Court, La Salle, Illinois, 1971. Second English edition, translated by Leo Unger from the Tenth German edition, revised and enlarged by Paul Bernays.

[11] Nicholas Lobachevski. The theory of parallels. In Bonola [2]. Translated by George Bruce Halstead, 1891. Open Court edition, 1914.

[12] Reviel Netz. *The Shaping of Deduction in Greek Mathematics*, volume 51 of *Ideas in Context.* Cambridge University Press, Cambridge, 1999. A study in cognitive history.

[13] Lucio Russo. *The Forgotten Revolution: How Science Was Born in 300* BC *and Why It Had to Be Reborn.* Springer-Verlag, Berlin, 2004. Translated from the 1996 Italian original by Silvio Levy.