Introduction

It seems natural that a course entitled Geometry should begin with the question:

What is geometry?

Right now, I would like to answer this question in the form of a short historic overview of the subject. Geometry is, after all, something that people have been doing for a very long time. The following brief history of geometry will be incomplete, inaccurate (true history is much more complicated) and biased (we will ignore what happened in India or China, for example). It is a shortened, smoothed out version of history that is meant only as a rough explanation of how the material that will be covered in this course came into being.

The word geometry comes from the Greek word γεωμετρία, which is a composite of the words for earth and measure. Geometry began as the science of measuring the earth, or surveying, and it began around 2000 BC in Egypt and Mesopotamia (Babylon, in today’s Iraq). These were among the first great civilizations and they depended on agriculture along the rivers Nile in Egypt, and Tigris and Euphrates in Mesopotamia. These rivers would periodically inundate and fertilize the surrounding land, which made periodic surveying necessary to delimit the fields. The science of geometry developed from this, with applications also in construction and astronomy. The Egyptians and Babylonians could compute areas and volumes of simple geometric figures, they had some approximations for \( \pi \), and they already knew Pythagoras’ theorem. Strangely though, no records of general theorems or proofs have survived from this period. Egyptian papyri and Babylonian clay tables with their cuneiform script contain only worked exercises. Maybe they did not state general theorems, maybe they just did not write them down, or maybe they did but these documents did not survive. Basically we have no idea how they conducted their research.

This changed with the period of Greek geometry (Thales ~600 BC to Euclid ~300 BC). They clearly stated general theorems for which they gave proofs. That is, they deduced more complicated statements from simpler ones by logical reasoning. This suggests putting all statements in order so that each statement is proved using only statements that have previously been proved. By necessity, one must begin with a few (as few as possible) hopefully very simple statements that are accepted without proof. In Euclid’s Elements, geometry (most or even all of what was known at the time) is presented in this form. It begins with a few definitions and postulates (today we say axioms) from which all theorems are deduced one by one. These postulates were simple statements like “there is a unique straight line through two points” and “two lines intersect in a unique point or they are parallel”. But one of the postulates was considered more complicated and less obvious than the others, the parallel postulate: “Given a line and a point not on the line, there is a unique parallel to the line through the point.” For centuries to come, people tried to prove this one postulate using the other, simpler ones, so that it could be eliminated from the unproved postulates. One way people tried to prove the parallel postulate was to assume instead that there are many parallels and derive a contradiction. But even though some strange theorems could be deduced from this alternative parallel postulate (like that there is an upper bound for the area of triangles) no true contradiction would appear. This finally lead to the realization that the alternative parallel postulate did not contradict the other postulates. Instead it leads to a logically equally valid version of geometry which is now called hyperbolic geometry (Lobachevsky 1829, Bolyai 1831). Later it was realized that one may also assume that there are no parallels (the other postulates also have to be changed a little for this), and the resulting geometry is called elliptic geometry. This is simply the geometry on the sphere, where pairs of opposite points are considered as one point, and lines are great circles. Both hyperbolic and elliptic geometry are called non-Euclidean geometries, because their axioms are different from Euclid’s.

Another important development in geometry was the introduction of coordinates by Descartes and Fermat in the first half of the 17th century. One could then describe geometric figures and prove theorems using numbers. This way of doing geometry was called analytic geometry, as opposed to the old way beginning with geometric axioms, which was called synthetic geometry. By the late 19th/early 20th century, it was proved that both approaches are in fact equivalent: One can either start with axioms for numbers and use them to define the objects of geometry, or one can start with axioms of geometry and define numbers geometrically, one gets the same theorems.
The study of the rules of perspective in painting (da Vinci & Dürer ~1500) lead to the development of projective geometry (Poncelet, 1822), dealing with the question: Which properties of geometric figures do not change under projections? For example, straight lines remain straight lines, but parallels do not remain parallel.

Another type of geometry is Möbius geometry, which deals with properties that remain unchanged under transformations mapping circles to circles (such as inversion on a circle). Then there is also Lie geometry (about which I will say nothing now) and there are other types of geometry. Klein’s Erlangen Program (1871) provided a systematic treatment of all these different kinds of geometry and their interrelationships. It also provided a comprehensive and maybe surprising answer to the question: What is geometry? We will come back to this.

Contents of this course

- spherical geometry
- hyperbolic geometry
- projective geometry
- Möbius geometry
- Lie geometry
- some other stuff

Spherical geometry

The $n$-dimensional unit sphere is $S^n = \{ x \in \mathbb{R}^{n+1} \mid \|x\| = 1 \} \subset \mathbb{R}^{n+1}$, where $\|x\| = \sqrt{x \cdot x}$ and $\langle x, y \rangle = \sum x_i y_i$ is the standard Euclidean scalar product. We will consider mainly $S^2 \subset \mathbb{R}^3$.

A great circle in $S^2$ is the intersection of $S^2$ with a plane through the origin, $E = \{ x \in \mathbb{R}^{n+1} \mid \langle x, n \rangle = 0 \}, \|n\| = 1$. (The intersection with a plane not through 0 is called a small circle.) The points $\pm n \in S^2$ are called the poles of the great circle. For $x, y \in S^2, x \neq \pm y$, there is a unique great circle through $x$ and $y$. (If $x = \pm y$, there is a one parameter family.) Two great circles intersect in two diametrically opposite points.

Theorem. The shortest continuously differentiable curve connecting two points $x, y \in S^2$ is the shorter arc of the great circle through $x$ and $y$. Its length is $d(x, y) := \arccos(x \cdot y)$.

Proof. The second sentence of the theorem is clear. To prove the first sentence, let $\gamma : [t_0, t_1] \to S^2$ be a continuously differentiable curve from $x$ to $y$. We have to show $\text{length}(\gamma) \geq d(x, y)$, with equality only if $\gamma$ is an arc of a great circle. We may assume that $\gamma(t) = x$ only for $t = t_0$ and $\gamma(t) = y$ only for $t = t_1$. Let $f(p) = \arccos(x \cdot p)$, which is defined where $|\langle x, p \rangle| \leq 1$ and differentiable where strict inequality holds. For $p \in S^2 \setminus \{x, -x\}$ let $v(p) = \text{grad} f(p) - \langle \text{grad} f(p), p \rangle p$. A direct calculation shows that $\|v\| = 1$. Then

$$d(x, y) = f(y) - f(x) = \int_{t_0}^{t_1} \langle \text{grad} f(\gamma(t)), \gamma'(t) \rangle dt = \int_{t_0}^{t_1} \langle v(\gamma(t), \gamma'(t)) \rangle dt \leq \int_{t_0}^{t_1} \|\gamma'(t)\| dt = \text{length}(\gamma),$$

where we have used $\langle \gamma', \gamma \rangle = 0$ and the Cauchy-Schwarz inequality. Equality only holds if $\gamma'$ is always in the direction of $v(\gamma)$. This can be used to show that $\gamma$ must be an arc of a great circle.

In fact, this proof works also if $\gamma$ is only assumed piecewise $C^1$.

Corollaries. (i) The shortest curves connecting two opposite points are halves of great circles. (ii) The function $d(x, y)$ is a metric on the sphere.
Hemispheres and digons

A hemisphere is the intersection of $S^2$ with a half-space

$$H = \{ x \in \mathbb{R}^3 \mid \langle x, n \rangle \geq 0 \}, \quad \|n\| = 1.$$  

A hemisphere is bounded by a great circle. One of its poles, $n$, lies inside the hemisphere, the other, $-n$, outside.

A spherical digon is the intersection of two hemispheres.

The interior angle $\alpha$ and the exterior angle $\hat{\alpha}$ of the digon satisfy $\alpha + \hat{\alpha} = \pi$ and $\cos \hat{\alpha} = \langle m, n \rangle$, so $\cos \alpha = -\langle m, n \rangle$. The exterior angle is the spherical distance between the poles $m$ and $n$.

The area of the digon is $2\alpha$.

(One could alternatively define a spherical digon as a spherical region bounded by two half great circles. Then the interior angle $\alpha$ could exceed $\pi$. The formula for the area would still hold.)

Spherical triangles

There are actually several different sensible definitions for spherical triangles:

1. The intersection of three hemispheres with poles not on one great circle. Such triangles are called Euler triangles.
2. A spherical region bounded by three great circular arcs. Such triangles are either an Euler triangle, the outside of an Euler triangle, the union of an Euler triangle and a hemisphere, or the difference of a hemisphere and an Euler triangle.
3. An oriented closed curve on the sphere consisting of three great circular arcs. (The arcs may intersect each other.) Such spherical triangles were considered by Möbius. To measure the angles of such a triangle sensibly, the sphere must be given an orientation.
4. Finally, a spherical triangle in the sense of Study consists of (i) three points $A, B, C \in S^2$, (ii) the three great circles $g_{AB}, g_{BC}, g_{CA}$ through $A$ and $B$, $B$ and $C$, $C$ and $A$, each of which is given an arbitrary orientation, and (iii) numbers $a, b, c, \hat{\alpha}, \hat{\beta}, \hat{\gamma} \in \mathbb{R}$ such that: (a) If one moves from $A$ along $g_{AB}$ the signed distance $c$, then one reaches $B$. Similarly for the other points and great circles. (b) If one rotates $g_{AB}$ around $A$ by the signed angle $\hat{\alpha}$, it is moved into $g_{CA}$ with correct orientation. Similarly for the other points and great circles. Triangles are considered different even if the corresponding numbers differ only by integer multiples of $\pi$. 
We will now consider Euler triangles, and “spherical triangle” will mean “Euler triangle”, unless stated otherwise.

In the definition above, an Euler triangle is defined in terms of hemispheres. But an Euler triangle is also determined by its vertices:

Suppose \( A, B, C \in S^2 \) do not all lie on one great circle. (That is, \( A, B, C \) are linearly independent unit vectors in \( \mathbb{R}^3 \)). Then the intersection of \( S^2 \) with the cone

\[
\mathcal{C} = \{ \lambda A + \mu B + \nu C \in \mathbb{R}^3 \mid \lambda, \mu, \nu \geq 0 \}
\]

is an Euler triangle. Indeed, there are unique unit vectors \( A', B', C' \in S^2 \) with

\[
\begin{align*}
\langle A', A \rangle &= 0, & \langle A', B \rangle &= 0, & \langle A', C \rangle &= 0, \\
\langle B', A \rangle &= 0, & \langle B', B \rangle &= 0, & \langle B', C \rangle &= 0, \\
\langle C', A \rangle &= 0, & \langle C', B \rangle &= 0, & \langle C', C \rangle &= 0,
\end{align*}
\]

and \( \mathcal{C} \) is the intersection of the half-spaces \( H_{A'}, H_{B'}, H_{C'} \), where \( H_X = \{ x \in \mathbb{R}^3 \mid \langle X, x \rangle \geq 0 \} \).

The side lengths \( a, b, c \) and exterior angles \( \hat{\alpha}, \hat{\beta}, \hat{\gamma} \) (see figure above) are determined by

\[
\cos a = \langle B, C \rangle, \quad \cos \hat{\alpha} = \langle B', C' \rangle, \quad \text{etc.}
\]

The interior angles are \( \alpha = \pi - \hat{\alpha} \), etc.

The triangle with vertices \( A', B', C' \) is called the polar triangle of the triangle with vertices \( A, B, C \).

The polar triangle of the polar triangle is the original triangle.

The side lengths of the polar triangle are the exterior angles of the original triangle and vice versa.

The side lengths \( a, b, c \) of a spherical triangle satisfy the inequalities

\[-a + b + c > 0, \quad a - b + c > 0, \quad a + b - c > 0, \quad a + b + c < 2\pi.\]

The first three inequalities follow directly from the triangle inequality of the spherical metric \( d \).

The third follows from

\[
d(\mathbf{A}, \mathbf{B}) + d(\mathbf{B}, \mathbf{C}) + d(\mathbf{C}, \mathbf{A}) < d(\mathbf{A}, \mathbf{B}) + d(\mathbf{B}, \mathbf{A}) + d(\mathbf{C}, \mathbf{A}) = 2\pi.
\]

Applying the same reasoning to the polar triangle, we get for the exterior angles of a spherical triangle

\[-\hat{\alpha} + \hat{\beta} + \hat{\gamma} > 0, \quad \hat{\alpha} - \hat{\beta} + \hat{\gamma} > 0, \quad \hat{\alpha} + \hat{\beta} - \hat{\gamma} > 0, \quad \hat{\alpha} + \hat{\beta} + \hat{\gamma} < 2\pi,
\]

and hence for the interior angles

\[-\alpha + \beta + \gamma < \pi, \quad \alpha - \beta + \gamma < \pi, \quad \alpha + \beta - \gamma < \pi, \quad \alpha + \beta + \gamma > \pi.
\]

**Theorem.** The area of a spherical triangle with interior angles \( \alpha, \beta, \gamma \) is \( \alpha + \beta + \gamma - \pi \).

**Proof.** The three great circles through \( A \) and \( B \) and \( C \) and \( A \) divide the sphere into eight triangles. Two of them are the triangle \( A, B, C \) and the triangle \(-A, -B, -C\). They have the same area because they are symmetric with respect to the origin. The other six triangles all have a side in common with either triangle \( A, B, C \) or triangle \(-A, -B, -C\) and complement these to six digons, two with angle \( \alpha \), two with angle \( \beta \), two with angle \( \gamma \). Altogether, the six digons cover each of the triangles \( A, B, C \) and \(-A, -B, -C\) three times, and the rest of the sphere once. So

\[
4(\alpha + \beta + \gamma) = \text{sum of area of the six digons} = \text{area}(S^2) + 2\text{area}(\triangle A, B, C) + 2\text{area}(\triangle -A, -B, -C) = 4\pi + 4\text{area}(\triangle A, B, C).
\]

\[\square\]
Theorem. Let \( a, b, c \in \mathbb{R} \). Then

(i) \( a, b, c \) satisfy the inequalities

\[
-a + b + c > 0, \quad a - b + c > 0, \quad a + b - c > 0, \quad a + b + c < 2\pi \quad (*)
\]

if and only if there is a spherical triangle with side lengths \( a, b, c \).

(ii) If it exists, this triangle is unique up to an orthogonal transformation of \( \mathbb{R}^3 \).

Proof. (i) The “if” part was shown in the last lecture. To see converse, assume \( a, b, c \) satisfy \((*)\).

Note first that this implies \( 0 < a, b, c < \pi \). Now ponder the following picture:

(ii) Suppose \( A, B, C \) and \( \tilde{A}, \tilde{B}, \tilde{C} \) are the vertices of two triangles with corresponding sides of equal length. Because corresponding scalar products are equal \( \langle A, B \rangle = \langle \tilde{A}, \tilde{B} \rangle \), etc., the linear map \( T : \mathbb{R}^3 \to \mathbb{R}^3 \) with \( T(A) = \tilde{A}, T(B) = \tilde{B}, T(B) = \tilde{B} \) is orthogonal. □

Remarks. (i) There is of course an analogous theorem regarding the angles of a spherical triangle.

(ii) The region \( D := \left\{ \left( \frac{a}{b} \right) \in \mathbb{R}^3 \mid a, b, c \text{ satisfy (*)} \right\} \) is the interior of the tetrahedron with vertices \( \left( \frac{0}{0} \right), \left( \frac{\pi}{0} \right), \left( \frac{\pi}{\pi} \right) \). The region of possible vectors of exterior angles \( \left( \frac{\alpha}{\beta} \right) \) is also \( D \).

By the previous theorem (and its analogue for angles), the side lengths determine the angles and vice versa. This gives a bijection \( D \to D \). If the side lengths approach a face (vertex) of \( D \), then the corresponding angles approach a vertex (face) of \( D \).

Theorem. Side lengths and exterior angles of a spherical triangle satisfy the equations

\[
\cos \hat{\alpha} = \frac{-\cos a + \cos b \cos c}{\sin b \sin c}, \quad \text{side cosine theorem}
\]

\[
\cos a = \frac{-\cos \hat{\alpha} + \cos \hat{\beta} \cos \hat{\gamma}}{\sin \hat{\beta} \sin \hat{\gamma}}, \quad \text{angle cosine theorem}
\]

and four more equations obtained by simultaneous permutations of \( a, b, c \) and \( \hat{\alpha}, \hat{\beta}, \hat{\gamma} \).

The following proof using Gram matrices exemplifies a general method which will be useful again later. The Gram matrix for a (finite) sequence of vectors \( v_1, \ldots, v_k \) is the symmetric matrix \( (\langle v_i, v_j \rangle)_{i,j=1}^k \) of pairwise scalar products.

Proof. Let \( V = (A \ B \ C) \in \mathbb{R}^{3 \times 3} \) be the matrix whose columns are the vertices of the spherical triangle, considered as column vectors. Then the Gram matrix for \( A, B, C \) is

\[
G = V^t V = \begin{pmatrix}
\langle A, A \rangle & \langle A, B \rangle & \langle A, C \rangle \\
\langle B, A \rangle & \langle B, B \rangle & \langle B, C \rangle \\
\langle C, A \rangle & \langle C, B \rangle & \langle C, C \rangle
\end{pmatrix} = \begin{pmatrix}
1 & \cos c & \cos b \\
\cos c & 1 & \cos a \\
\cos b & \cos a & 1
\end{pmatrix}.
\]
Their Gram matrix is

\[ G' = W^t W = \begin{pmatrix}
\langle A', A' \rangle & \langle A', B' \rangle & \langle A', C' \rangle \\
\langle B', A' \rangle & \langle B', B' \rangle & \langle B', C' \rangle \\
\langle C', A' \rangle & \langle C', B' \rangle & \langle C', C' \rangle \\
\end{pmatrix} = \begin{pmatrix}
1 & \cos \gamma & \cos \beta \\
\cos \gamma & 1 & \cos \alpha \\
\cos \beta & \cos \alpha & 1 \\
\end{pmatrix}. \]

Also,

\[ W^t V = \begin{pmatrix}
\langle A', A' \rangle & \langle A', B' \rangle & \langle A', C' \rangle \\
\langle B', A' \rangle & \langle B', B' \rangle & \langle B', C' \rangle \\
\langle C', A' \rangle & \langle C', B' \rangle & \langle C', C' \rangle \\
\end{pmatrix} = \begin{pmatrix}
\langle A', A \rangle & 0 & 0 \\
0 & \langle B', B \rangle & 0 \\
0 & 0 & \langle C', C \rangle \\
\end{pmatrix} =: D \]

is a diagonal matrix with positive entries. So \( W^t = DV^{-1} \) and \( W = (V^t)^{-1} D \), and

\[ G' = DV^{-1}(V^t)^{-1} D = D(V^t)V^{-1} D = DG^{-1}D. \quad (**) \]

The inverse of \( G \) is

\[ G^{-1} = \frac{1}{\det G} \begin{pmatrix}
\sin^2 a & -\cos c + \cos a \cos b & -\cos b + \cos c \cos a \\
-\cos c + \cos a \cos b & \sin^2 b & -\cos a + \cos b \cos c \\
-\cos b + \cos c \cos a & -\cos a + \cos b \cos c & \sin^2 c \\
\end{pmatrix}. \]

Substitute this into (**) and consider diagonal elements: One finds \( 1 = D_{11}^{-1} \frac{1}{\det G} \sin^2 a \), therefore \( D_{11} = \frac{\det G}{\sin^2 a} \), and similarly \( D_{22} = \frac{\det G}{\sin^2 b} \), \( D_{33} = \frac{\det G}{\sin^2 c} \). Now consider for example element \((3,2)\) in (**) \( \cos \alpha = D_{33}^{-1} \frac{1}{\det G} (-\cos a + \cos b \cos c)D_{22} \). This is the side cosine theorem.

The angle cosine theorem is the side cosine theorem applied to the polar triangle. \( \square \)

**Theorem.** Side lengths and interior angles of a spherical triangle satisfy

\[ \frac{\sin a}{\sin \alpha} = \frac{\sin b}{\sin \beta} = \frac{\sin c}{\sin \gamma}, \]

(sine theorem)

\[ \tan \frac{\alpha}{2} = \frac{\sin \frac{a+b+c}{2} \sin \frac{a+b-c}{2}}{\sin \frac{a+b+c}{2} \sin \frac{a-b+c}{2}}, \]

(half-angle theorem)

\[ \tan \frac{\alpha}{2} = \frac{\cos \frac{a+b+c}{2} \cos \frac{a-b+c}{2}}{\cos \frac{a+b+c}{2} \cos \frac{a-b-c}{2}}, \]

(half-side theorem)

**Proof.** In terms of the interior angle \( \alpha \), the side cosine theorem says \( \cos \alpha = \frac{\cos a - \cos b \cos c}{\sin b \sin c} \).

Using \( \cos \alpha = 2 \cos^2 \frac{\alpha}{2} - 1 = 1 - 2 \sin^2 \frac{\alpha}{2} \) and other trigonometric identities, one obtains the equations

\[ \cos \frac{\alpha}{2} = \sqrt{\frac{\sin(-a+b+c)}{2} \sin \frac{a+b+c}{2}}, \]

\[ \sin \frac{\alpha}{2} = \sqrt{\frac{\sin(a-b+c)}{2} \sin \frac{a+b-c}{2}}, \]

which are also of independent interest. Dividing one by the other, one obtains the half-angle theorem. The half-side theorem is the half-angle theorem for the polar triangle. To prove the sine theorem, consider

\[ \sin \alpha = 2 \sin \frac{\alpha}{2} \cos \frac{\alpha}{2} = \frac{P}{\sin b \sin c}, \]

where

\[ P = 2 \left( \sin \frac{a+b+c}{2} \sin \frac{a-b+c}{2} \sin \frac{a+b-c}{2} \sin \frac{a-b-c}{2} \right). \]

So \( \sin a \sin b \sin c = P \). But since the expression for \( P \) is symmetric in \( a, b, c \), one has equally

\[ \sin \alpha \sin b \sin c = \sin \beta \sin c \sin a = \sin \gamma \sin a \sin b = P. \]

Divide by \( \sin a \sin b \sin c \) to obtain the sine theorem. \( \square \)
In principle, all relations between side lengths and angles of a spherical triangle can be derived from the inequalities for the sides and the side cosine theorem (or alternatively from the inequalities for the angles and the angle cosine theorem) using only algebra and analysis, without further recourse to geometry. We have derived the half-angle theorem and the sine theorem in this way. On the other hand, the following proof of Napier’s rule, which is due to Napier himself, uses a remarkable geometric construction.

**Theorem** (Napier’s rule). Consider a right-angled spherical triangle with $\gamma = \frac{\pi}{2}$, and let $\bar{a} = \frac{\pi}{2} - a$, $\bar{b} = \frac{\pi}{2} - b$. Then

\[
\begin{align*}
\cos c &= \sin b \sin \bar{a} = \cot \alpha \cot \beta, \\
\cos \beta &= \sin \alpha \sin \bar{b} = \cos \gamma - \cos a \cos b \\
\cos \bar{a} &= \sin \gamma \sin \alpha = \cot \bar{c} \cot \alpha, \\
\cos \bar{b} &= \sin \beta \sin \bar{c} = \cot \bar{c} \cot \alpha, \\
\cos \alpha &= \sin \bar{a} \sin \beta = \cot \beta \cot \bar{b}.
\end{align*}
\]

That is: “The cosine of any part is equal to the product of sines of opposite parts and to the product of cotangents of adjacent parts.” (The “parts” are $\bar{a}$, $\bar{b}$, $\alpha$, $\beta$, in this cyclic order.)

**Proof.** The first line of equations follows directly from the cosine rules $\cos \gamma = \cos c - \cos a \cos b$ and $\cos c = \frac{\cos \gamma \cos \alpha \cos \beta}{\sin \alpha \sin \beta}$. To prove the remaining equations, assume first that $a, b, c, \alpha, \beta < \frac{\pi}{2}$. Consider the following construction. Draw the two great circles that have $A$ and $B$ as poles. Together with the extended sides of the original triangle, they form 4 other right angled triangles:

(The five triangles form a right-angled pentagram called the *pentagramma mirificum*. In its center there is a spherical pentagon each vertex of which is the pole of the opposite side.) Note that $a_1 = \frac{\pi}{2} - c$, $b_1 = \frac{\pi}{2} - \beta$, $\alpha_1 = \frac{\pi}{2} - a$, $c_1 = \frac{\pi}{2} - b$, $\beta_1 = \alpha$, so that

\[\bar{a}_1 = c, \quad \bar{b}_1 = \beta, \quad \alpha_1 = \bar{a}, \quad \beta_1 = \bar{b} \quad \Rightarrow \alpha.
\]

This proves the other equations of Napier’s rule under the assumption of acute angles and sides. Now suppose that a side length $a, b, c$ or an angle $\alpha, \beta$ is greater than $\frac{\pi}{2}$. (The remaining cases where one is equal to $\frac{\pi}{2}$ consist of doubly or triply right-angled triangles for which Napier’s rule can easily be checked.) It can be shown that, first, for one of the neighbor triangles (into which the sphere is separated by the same great circles) $a, b, c, \alpha, \beta < \frac{\pi}{2}$, and, second, if Napier’s rule holds for one of the neighbor triangles, it holds for all of them. \(\square\)
Stereographic projection

Project the unit sphere \( S^2 \subset \mathbb{R}^3 \) from the “north pole” \( e_3 = \binom{0}{0}{1} \) to the plane \( x_3 = 0 \). This map \( \sigma : S^2 \setminus \{e_3\} \to \mathbb{R}^2 \) is called stereographic projection. One easily derives the following equations for \( \sigma \) and its inverse:

\[
\sigma \left( \begin{array}{c} x_1 \\ x_2 \\ x_3 
\end{array} \right) = \frac{1}{1-x_3} \left( \begin{array}{c} x_1 \\ x_2 
\end{array} \right),
\]

\[
\sigma^{-1} \left( \begin{array}{c} u_1 \\ u_2 
\end{array} \right) = \frac{1}{u_1^2 + u_2^2 + 1} \left( \begin{array}{c} 2u_1 \\ 2u_2 \\ u_1^2 + u_2^2 - 1 
\end{array} \right).
\]

**Theorem.** Stereographic projection \( \sigma \) maps circles in \( S^2 \) which contain \( e_3 \) to lines in \( \mathbb{R}^2 \) and all other circles in \( S^2 \) to circles in \( \mathbb{R}^2 \). All circles and lines in \( \mathbb{R}^2 \) are images of circles in \( S^2 \).

**Remark.** The fact that circles through \( e_3 \) are mapped to lines is geometrically clear: A circle through \( e_3 \) is the intersection of \( S^2 \) with a plane through \( e_3 \). So all projection rays lie in this plane, and the circle is mapped to the line in which it intersects the image plane \( x_3 = 0 \).

**Proof.** A circle in \( S^2 \) is the intersection of \( S^2 \) with a plane

\[ E = \{ x \in \mathbb{R}^3 \mid \langle x, n \rangle = d \}, \quad \text{where} \quad \|n\| = 1, \quad 0 \leq d < 1. \]

It contains \( e_3 \) iff \( d = \langle e_3, n \rangle = n_3 \). A point \( u \in \mathbb{R}^2 \) in the image of the circle iff \( \sigma^{-1}(u) \in E \), that is, iff

\[
d = \frac{1}{u_1^2 + u_2^2 + 1} \left( 2u_1 n_1 + 2u_2 n_2 + (u_1^2 + u_2^2 - 1)n_3 \right),
\]

or equivalently,

\[
0 = (n_3 - d)(u_1^2 + u_2^2) + 2u_1 n_1 + 2u_2 n_2 - (n_3 + d).
\]

If \( n_3 = d \), this is the equation for a line. Otherwise, divide by \( n_3 - d \), complete the squares, and use \( \|n\|^2 = 1 \) to obtain

\[
\left( u_1 + \frac{n_1}{n_3 - d} \right)^2 + \left( u_2 + \frac{n_2}{n_3 - d} \right)^2 - \frac{1-d^2}{(n_3 - d)^2} = 0.
\]

This is the equation for a circle with center \( c = -\frac{1}{n_3 - d} \binom{n_1}{n_2} \) and radius \( r = \sqrt{\frac{1-d^2}{n_3 - d}} \).

To show (without further calculations) that every circle and every line in \( \mathbb{R}^2 \) is the image of a circle in \( S^2 \), you can make an argument using the fact that three points in \( \mathbb{R}^2 \) uniquely determine a line or circle through them and three points in \( S^2 \) uniquely determine a circle through them. \( \square \)
Theorem. Stereographic projection is conformal, that is, it preserves angles: If two curves in $S^2 \setminus \{e_3\}$ intersect at some angle, then their image curves in $\mathbb{R}^2$ intersect at the same angle.

Proof. Let $\tilde{\gamma}, \tilde{\eta} : (-\varepsilon, \varepsilon) \to S^2 \setminus \{e_3\}$ be two curves with $\tilde{\gamma}(0) = \tilde{\eta}(0) = \tilde{p}$, and let $\gamma = \sigma \circ \tilde{\gamma}$, $\eta = \sigma \circ \tilde{\eta}$ be their images curves in $\mathbb{R}^2$. Let $\tilde{v} = \tilde{\gamma}'(0)$, $\tilde{w} = \tilde{\eta}'(0)$, $v = \gamma'(0)$, $w = \eta'(0)$, and let $p = \gamma(0) = \eta(0) = \sigma(\tilde{p})$. The intersection angles $\hat{\alpha}$ and $\alpha$ between $\tilde{\gamma}$, $\tilde{\eta}$ and $\gamma$, $\eta$, are determined by

$$\cos \hat{\alpha} = \frac{\langle \hat{v}, \hat{w} \rangle}{\sqrt{\langle \hat{v}, \hat{v} \rangle \langle \hat{w}, \hat{w} \rangle}}, \quad \cos \alpha = \frac{\langle v, w \rangle}{\sqrt{\langle v, v \rangle \langle w, w \rangle}}.$$ 

We want to show that $\hat{\alpha} = \alpha$. In fact, we will show that

$$\langle \hat{v}, \hat{w} \rangle = \frac{4}{(p, p) + 1} \langle v, w \rangle, \quad \langle \hat{v}, \tilde{v} \rangle = \frac{4}{(p, p) + 1} \langle v, v \rangle, \quad \langle \hat{w}, \tilde{w} \rangle = \frac{4}{(p, p) + 1} \langle w, w \rangle,$$

and this implies $\cos \hat{\alpha} = \cos \alpha$, and hence $\hat{\alpha} = \alpha$.

First we derive equations for $\hat{v}$, $\hat{w}$ in terms of $v$, $w$. By the equation for $\sigma^{-1}$ we have

$$\gamma = \frac{1}{(\gamma, \gamma) + 1} \left( \begin{array}{c} 2 \gamma_1 \\ 2 \gamma_2 \end{array} (\gamma, \gamma) - 1 \right),$$

so

$$\gamma' = \frac{1}{(\gamma, \gamma) + 1} \left( \begin{array}{cc} 2 \gamma_1 & 2 \gamma_2 \\ 2 \gamma_2 & 2 \gamma_1 \end{array} \right) - \frac{2(\gamma, \gamma')}{(\gamma, \gamma) + 1} \left( \begin{array}{c} 2 \gamma_1 \\ 2 \gamma_2 \end{array} (\gamma, \gamma) - 1 \right),$$

$$= \frac{2}{(\gamma, \gamma) + 1} \left( \begin{array}{cc} \gamma_1^2 & \gamma_2^2 \\ \gamma_2 \gamma_1 & \gamma_1 \gamma_2 \end{array} \right) - \frac{(\gamma, \gamma')}{(\gamma, \gamma) + 1} \left( \begin{array}{c} 2 \gamma_1 \\ 2 \gamma_2 \end{array} (\gamma, \gamma) - 1 \right),$$

and hence

$$\hat{v} = \gamma'(0) = \frac{2}{(p, p) + 1} \left( \begin{array}{c} v_1 \\ v_2 \end{array} \right) - \frac{2}{(p, p) + 1} \left( \begin{array}{c} p_1 \\ p_2 \end{array} \right),$$

In the same way one gets

$$\hat{w} = \gamma'(0) = \frac{2}{(p, p) + 1} \left( \begin{array}{c} w_1 \\ w_2 \end{array} \right) - \frac{2}{(p, p) + 1} \left( \begin{array}{c} p_1 \\ p_2 \end{array} \right),$$

so

$$\langle \hat{v}, \hat{w} \rangle = \frac{4}{(p, p) + 1} \left( \langle v, w \rangle + \langle p, v \rangle \langle p, w \rangle - \frac{(p, v)(p, w) + 2(p, v)(p, w) - 1}{(p, v)(p, p) + 1} \right),$$

and similarly for $\langle \hat{v}, \tilde{v} \rangle$ and $\langle \hat{w}, \tilde{w} \rangle$.

All of this works in the same way also for the $n$-dimensional sphere $S^n \subset \mathbb{R}^{n+1}$. (But for $n = 1$, the projection from the circle to the line, the two theorems are trivial.) Stereographic projection from $e_{n+1}$ is $\sigma : S^n \setminus \{e_{n+1}\} \to \mathbb{R}^n$,

$$\sigma(x) = \frac{1}{1 - x_{n+1}} \left( \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array} \right),$$

$$\sigma^{-1}(u) = \frac{1}{\langle u, u \rangle + 1} \left( \begin{array}{c} 2u_1 \\ \vdots \\ 2u_n \\ \langle u, u \rangle - 1 \end{array} \right).$$

It maps $(n - 1)$-dimensional spheres in $S^n$ (intersections of $S^n$ with affine hyperplanes in $\mathbb{R}^{n+1}$) to $(n - 1)$-dimensional spheres and planes in $\mathbb{R}^n$, and it is conformal.
Bilinear and quadratic forms

We will define $n$-dimensional hyperbolic space as

$$ H^n = \{ x \in \mathbb{R}^{n+1} \mid x_1^2 + x_2^2 + \ldots + x_n^2 - x_{n+1}^2 = -1, \ x_{n+1} > 0 \} $$

$$ = \{ x \in \mathbb{R}^{n+1} \mid \langle x, x \rangle = -1, \ x_{n+1} > 0 \}, $$

where now $\langle x, y \rangle = x_1y_1 + \ldots + x_ny_n - x_{n+1}y_{n+1}$, and lengths of curves in $H^n$ and angles between them are measured using this scalar product instead of the normal Euclidean scalar product. We will need to be familiar with such indefinite scalar products, and since we will deal with general bilinear and quadratic forms later, it may be a good idea to refresh some material from linear algebra.

Let $V$ be an $n$-dimensional vector space over a field $K$. (We will be interested in the cases $K = \mathbb{R}$ and $K = \mathbb{C}$.) A bilinear form on $V$ is a function $b : V \times V \to K$, which is linear in each argument. If $e_1, \ldots, e_n$ is a basis of $V$, then the matrix of the bilinear form $b$ is $B \in K^{n \times n}$ with $B_{ij} = b(e_i, e_j)$. If $x, y \in K^n$ are the coordinate vectors for $v, w \in V$, that is $v = \sum x_i e_i$ and $w = \sum y_i e_i$, then $b(v, w) = x^t By = \sum_{i,j} B_{ij} x_i y_j$. If $f_1, \ldots, f_n$ is another basis with $f_j = \sum T_{ij} e_i$, and $p, q \in K^n$ are the coordinate vectors of $v, w$ in this new basis, then $x = Tp$, $y = Tq$, and the matrix of $b$ with respect to the new basis is $B = T^t BT$.

The bilinear form $B$ is symmetric if $b(v, w) = b(w, v)$ for all $v, w \in V$. A bilinear form is symmetric if its matrix with respect to one (and hence every) basis is symmetric. A quadratic form on $V$ is a function $q : V \to K$ for which there exists a bilinear form $b$ such that $q(v) = b(v, v)$ for all $v \in V$. (\textup{\textcircled{a}})

If $q$ is a quadratic form, then there exists a unique symmetric bilinear form satisfying (\textup{\textcircled{a}}). Hence symmetric bilinear forms and quadratic forms are in one-to-one correspondence. Quadratic forms are homogeneous polynomials of degree 2 in (any) coordinates.

Example. $x_1^2 + x_1 x_2 + x_2^2$ is a quadratic form on $K^2$. For example, it comes from the bilinear form $x_1 y_1 + x_1 y_2 + x_2 y_2$, which is not symmetric. But it also comes from the symmetric bilinear form $x_1 y_1 + \frac{1}{2} x_1 y_2 + \frac{1}{2} x_2 y_1 + x_2 y_2$.

Symmetric bilinear forms

Suppose $b$ is symmetric and $q(v) = b(v, v)$. The kernel of $b$ is

$$ \ker b = \{ v \in V \mid b(v, w) = 0 \text{ for all } w \in V \}. $$

This is a linear subspace of $V$. The bilinear and quadratic forms $b, q$ are called degenerate if $\ker b \neq \{ 0 \}$, and non-degenerate if $\ker b = \{ 0 \}$. The form $b$ is degenerate iff its matrix with respect to one (hence every) basis has determinant 0. Let $U_0 = \ker b$ and let $U$ be any complementary subspace, so that $V = U \oplus U_0$. Then the restrictions $b|_U$ and $q|_U$ are non-degenerate bilinear/quadratic forms on $U$.

There exist bases $\hat{e}_1, \ldots, \hat{e}_n$ of $V$ such that

$$ b(\hat{e}_i, \hat{e}_j) = 0 \text{ if } i \neq j. $$

(\textup{\textcircled{a}})

The basis vectors $e_i$ with $b(\hat{e}_i, \hat{e}_i) = 0$ form a basis for $U_0 = \ker b$. Assume the basis is ordered so that these come last. Then the matrix of $b$ with respect to this basis is diagonal with diagonal elements $\lambda_1, \ldots, \lambda_r, 0, \ldots, 0$, where $\lambda_i \neq 0$ and $r = n - \dim \ker b$. In the coordinates $u_1, \ldots, u_n$ with respect to this basis, $q$ is a sum of squares:

$$ \lambda_1 u_1^2 + \lambda_2 u_2^2 + \ldots + \lambda_r u_r^2. $$

A basis $\hat{e}_1, \ldots, \hat{e}_n$ satisfying (\textup{\textcircled{a}}) and the corresponding coordinates $u_1, \ldots, u_n$ can be found using the \textit{generalized Gram-Schmidt orthogonalization procedure} or by completing the squares.
Generalized Gram-Schmidt orthogonalization procedure

The ordinary Gram-Schmidt orthogonalization procedure takes as input a basis $e_1, \ldots, e_n$ of $V$ and a positive definite symmetric bilinear form $b$. It works like this:

\[
\text{for } i = 1 \text{ to } n-1 \text{ do} \\
\quad \text{for } j = i + 1 \text{ to } n \text{ do} \\
\quad \quad e_j \leftarrow e_j - \frac{b(e_j, e_i)}{b(e_i, e_i)} e_i \\
\quad \text{end for} \\
\text{end for}
\]

This may not work if $b$ is not positive definite because, for some $i$, $b(e_i, e_i)$ may be 0. If that is the case, do the following: If there is among the $e_{i+1}, \ldots, e_n$ a basis vector, say $e_k$, with $b(e_k, e_k) \neq 0$, then swap $e_i$ and $e_k$ and continue. If that that is not possible because $b(e_{i+1}, e_{i+1}), \ldots, b(e_n, e_n)$ are all 0, then there must be some $e_k$ ($i+1 \leq k \leq n$) with $b(e_i, e_k) \neq 0$, because $b$ was assumed to be non-degenerate. Assign $e_i + e_k$ and $e_i - e_k$ to $e_i$ and $e_k$ and continue.

If $b$ is degenerate, find a basis for $\ker b$ and a basis for a complementary subspace and apply the orthogonalization procedure to the latter.

Warning: This algorithm is not numerically stable. Its main purpose is to prove the existence of a diagonalizing basis.

Completing the squares

**Examples.** Suppose in some coordinates $q$ is

\[
x_1^2 + x_2^2 + 2x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 = (x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3)^2 + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2 + \frac{1}{2}x_2x_3 \\
= (x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3)^2 + \frac{1}{2}(x_2 + \frac{1}{2}x_3)^2 + 2 = (x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3)^2 + \frac{20}{12} x_3^2 \\
= u_1^2 + \frac{3}{4} u_2^2 + \frac{50}{12} u_3^2.
\]

It may happen that there is no square to complete:

\[
x_1x_2 + x_2x_3 = \frac{1}{4}(x_1 + x_2)^2 - \frac{1}{4}(x_1 - x_2)^2 + x_2x_3 = \frac{1}{4}y_1^2 + \frac{1}{4}y_2^2 + \frac{1}{2}(y_1 - y_2)x_3,
\]

continue as before.

**Case $K = \mathbb{C}$**

We can always make all $\lambda_1, \ldots, \lambda_r$ equal 1 by dividing $e_k$ ($k = 1, \ldots, r$) by $\sqrt{b(e_k, e_k)}$ (arbitrary choice of square roots). Thus, any symmetric bilinear form on a complex vector space is in suitable coordinates $\sum_{k=1}^r x_k y_k$.

**Case $K = \mathbb{R}$**

Let

\[
i_+ = \text{(number of } \hat{e}_k \text{ for which } b(\hat{e}_k, \hat{e}_k) > 0), \\
i_- = \text{(number of } \hat{e}_k \text{ for which } b(\hat{e}_k, \hat{e}_k) < 0), \\
i_0 = \text{(number of } \hat{e}_k \text{ for which } b(\hat{e}_k, \hat{e}_k) = 0) = \dim \ker b.
\]

The numbers $i_+, i_-, i_0$ do not depend on the particular basis, but only on $b$. (Why?) The numbers $i_+$ and $i_-$ are the positive and negative index of $b$. The signature of $b$ is $(i_+, i_-, i_0)$, also written $(i_+, i_-, i_0)$ if $i_0 = 0$. We can normalize the $\hat{e}_k$ ($k = 1, \ldots, r$) by dividing by $\sqrt{|b(\hat{e}_k, \hat{e}_k)|}$.

Thus, one obtains a basis with $b(\hat{e}_i, \hat{e}_i) = \pm 1$ or 0, and any symmetric bilinear form on a real vector space is in suitable coordinates

\[
\sum_{k=1}^{i_+} x_k y_k - \sum_{k=i_+}^{i_+ + i_-} x_k y_k.
\]
Scalar products

A scalar product is a non-degenerate symmetric bilinear form. Scalar products are often written $\langle \cdot, \cdot \rangle$ or $\langle \cdot, \cdot \rangle$. Vectors $v, w$ with $\langle v, w \rangle = 0$ are called orthogonal to each other.

Any $n$-dimensional complex vector space with a scalar product can be identified with $\mathbb{C}^n$ with the scalar product $\sum_k x_k y_k$.

$\mathbb{R}^{p,q}$ denotes the real vector space $\mathbb{R}^{p+q}$, equipped with the scalar product

$$\langle x, y \rangle = \sum_{k=1}^p x_k y_k - \sum_{k=p+1}^{p+q} x_k y_k,$$

which has signature $(p, q)$. By an appropriate choice of basis, any real vector space with a scalar product with signature $(p, q)$ can be identified with $\mathbb{R}^{p,q}$. A vector $v$ in $\mathbb{R}^{p,q}$ (or any other real vector space with scalar product) is called spacelike, timelike, or lightlike if $\langle v, v \rangle > 0$, $\langle v, v \rangle < 0$, or $\langle v, v \rangle = 0$.

The set of lightlike vectors is called the light cone. A Euclidean scalar product is a scalar product with negative index $i_- = 0$. A Lorentz scalar product is a scalar product with negative index $i_- = 1$. A vector space with a Euclidean or Lorentz scalar product is a Euclidean or Lorentz vector space, respectively. An orthonormal basis for $\mathbb{R}^{p,q}$ is a basis $e_1, \ldots, e_n$ with $\langle e_i, e_j \rangle = 0$ if $i \neq j$, $\langle e_i, e_i \rangle = 1$ for $i = 1, \ldots, p$, and $\langle e_i, e_i \rangle = -1$ for $i = p+1, \ldots, n$.

**Theorem.** Let $V$ be a real vector space with scalar product $\langle \cdot, \cdot \rangle$, let $e_1, \ldots, e_n$ be a basis, let $B_{ij} = \langle e_i, e_j \rangle$, and let

$$a_0 = 1, \quad a_1 = B_{11}, \quad a_2 = \det \left( \begin{array}{cc} B_{11} & B_{12} \\ B_{21} & B_{22} \end{array} \right), \ldots \quad a_k = \det \left( \begin{array}{ccc} B_{11} & \cdots & B_{1k} \\ \vdots & \ddots & \vdots \\ B_{k1} & \cdots & B_{kk} \end{array} \right), \ldots \quad a_n = \det B.$$

Suppose that none of the $a_k$ are 0. Then the negative index of the scalar product $\langle \cdot, \cdot \rangle$ is equal to the number of sign changes in the sequence $a_0, a_1, \ldots, a_n$.

Orthogonal transformations

Let $V$ be a vector space with scalar product $\langle \cdot, \cdot \rangle$. An orthogonal transformation on $V$ is a linear map $T : V \rightarrow V$ with $\langle Tv, Tw \rangle = \langle v, w \rangle$ for all $v, w \in V$. The orthogonal transformations form a group, the orthogonal group of $V$, $\langle \cdot, \cdot \rangle$. The orthogonal group of $\mathbb{C}^n$ with standard scalar product is denoted by $O(n, \mathbb{C})$. The orthogonal group of $\mathbb{R}^{p,q}$ is denoted by $O(p,q)$, or if $q = 0$ also by $O(p)$ or $O(p, \mathbb{R})$. The corresponding matrix groups are also denoted by $O(n, \mathbb{C})$, $O(p,q)$. The columns of a matrix in $O(p,q)$ form an orthonormal basis of $\mathbb{R}^{p,q}$. The determinant of an orthogonal transformation is $\pm 1$. The orthogonal transformations with determinant $+1$ form subgroups, the special orthogonal groups, denoted by $SO(n, \mathbb{C})$ and $SO(p,q)$.

Lorentz vector spaces

$\mathbb{R}^{1,1}$ is $\mathbb{R}^2$ with $\langle x, y \rangle = x_1 y_1 - x_2 y_2$. So $\langle x, x \rangle = x_1^2 - x_2^2$ and this is $\geq 0$ if $|x_1|^2 \geq |x_2|$. The figure also shows another orthonormal basis.

$\mathbb{R}^{2,1}$ is $\mathbb{R}^3$ with $\langle x, y \rangle = x_1 y_1 + x_2 y_2 - x_3 y_3$. So $\langle x, x \rangle = x_1^2 + x_2^2 - x_3^2$ and this is $\geq 0$ if $\sqrt{x_1^2 + x_2^2} \geq |x_3|$.

In a Lorentz vector space, any non-zero vector which is orthogonal to a timelike vector is spacelike.

An orthogonal transformation $T \in O(n,1)$ either maps each connected component of the hyperboloid $\langle x, x \rangle = -1$ onto itself, or it interchanges the two components. Those that map each component onto itself form a subgroup, $O^+(n,1)$ Also, $SO^+(n,1) = O^+(n,1) \cap SO(n,1)$. In all, $O(n,1)$ has 4 connected components, each consisting of the transformations with determinant either $+1$ or $-1$ and either fixing or interchanging the components of the hyperboloid $\langle x, x \rangle = -1$.
Hyperbolic geometry

Hyperbolic space of $n$ dimensions is

$$H^n = \{ x \in \mathbb{R}^{n,1} \mid \langle x, x \rangle = -1, \ x_{n+1} > 0 \},$$

where $\langle x, y \rangle = x_1 y_1 + \ldots + x_n y_n - x_{n+1} y_{n+1}$ is the Lorentz scalar product of $\mathbb{R}^{n,1}$. Lengths of curves and angles between them are measured using this scalar product (and not the Euclidean scalar product of $\mathbb{R}^{n+1}$). The length of a curve $\gamma : [t_1, t_2] \to H^n$ is

$$\text{length}(\gamma) = \int_{t_1}^{t_2} \sqrt{\langle \gamma'(t), \gamma'(t) \rangle} \, dt$$

and the angle $\alpha$ between two curves $\gamma, \eta : (-\varepsilon, \varepsilon) \to H^n$ intersecting in $p = \gamma(0) = \eta(0)$ with non-zero velocities $v = \gamma'(0)$, $w = \eta'(0)$ is determined by

$$\cos \alpha = \frac{\langle v, w \rangle}{\sqrt{\langle v, v \rangle \langle w, w \rangle}}.$$

Why is this well defined? Why is $\langle \gamma', \gamma' \rangle$ never negative and why is $\left| \frac{\langle v, w \rangle}{\sqrt{\langle v, v \rangle \langle w, w \rangle}} \right| \leq 1$? For $v \in \mathbb{R}^{n,1}$ let $v^\perp = \{ w \in \mathbb{R}^{n,1} \mid \langle v, w \rangle = 0 \}$. If $\langle v, v \rangle < 0$ ($v$ is timelike) then the restriction of the scalar product $\langle \cdot, \cdot \rangle$ to $v^\perp$ is a Euclidean scalar product. Now $\langle \gamma, \gamma \rangle = -1$ implies $\langle \gamma, \gamma' \rangle = 0$, so $\gamma'(t) \in (\gamma(t))^\perp$. In the formulas for lengths and angles, the Lorentz scalar product is therefore applied to vectors in a subspace of $\mathbb{R}^{n,1}$ on which it is Euclidean.

One-dimensional hyperbolic space

$H^1$ is one branch of the hyperbola $x_1^2 - x_2^2 = -1$ in $\mathbb{R}^{1,1}$.

This is a curve which can be parameterized as $\gamma(s) = \left( \frac{\sinh s}{\cosh s} \right)$. (Because $\cosh^2 s - \sinh^2 s = 1$.) Now $\gamma'(s) = \left( \frac{\cosh s}{\sinh s} \right)$, so $\langle \gamma'(s), \gamma'(s) \rangle = \cosh^2 s - \sinh^2 s = 1$ and

$$\text{length}(\gamma|_{s_1, s_2}) = \int_{s_1}^{s_2} \sqrt{\langle \gamma'(s), \gamma'(s) \rangle} = s_2 - s_1.$$

On the other hand,

$$\langle \gamma(s_1), \gamma(s_2) \rangle = \sinh s_1 \sinh s_2 - \cosh s_1 \cosh s_2 = -\cosh(s_1 - s_2).$$

(Because $\cosh(x + y) = \cosh x \cosh y + \sinh x \sinh y$.) Hence, the hyperbolic distance $d(p_1, p_2)$ of two points $p_1, p_2 \in H^1$ is given by

$$-\cosh d(p_1, p_2) = \langle p_1, p_2 \rangle.$$

Hyperbolic lines

A hyperbolic line in $n$-dimensional hyperbolic space $H^n$ is a non-empty intersection of $H^n$ with a 2-dimensional linear subspace $U$ of $\mathbb{R}^{n,1}$.
Proposition. If \( U \) is a 2-dimensional linear subspace of \( \mathbb{R}^{n,1} \) with \( U \cap H^n \neq \emptyset \), then the restriction \( (\cdot,\cdot)|_U \) has signature \((1,1)\).

Proof. By assumption, there is a \( u \in U \) with \( u(n) = -1 \). Extend \( u \) to a basis \( u, v \) of \( U \), and let \( \tilde{v} = v + (u,v)u \). Then \( u, \tilde{v} \) is a basis of \( U \) with \( (u, \tilde{v}) = 0 \). Because any non-zero vector which is orthogonal to a timelike vector is spacelike, \( (\tilde{v}, \tilde{v}) > 0 \). So \( (\cdot,\cdot)|_U \) has signature \((1,1)\).

By an appropriate choice of basis, any 2-dimensional \( U \) intersecting \( H^n \) can therefore be identified with \( \mathbb{R}^{1,1} \). The hyperbolic line \( U \cap H^n \) is thus identified with 1-dimensional hyperbolic space \( H^1 \in \mathbb{R}^{1,1} \).

Remark. In the same way, any non-empty intersection of \( H^n \) with a \((k+1)\)-dimensional subspace can be identified with \( H^k \).

For two points \( p_1, p_2 \in H^n \), there is a unique hyperbolic line containing them: \( \text{span}(p_1, p_2) \cap H^n \).

Theorem. The shortest piecewise continuously differentiable curve connecting two points \( p_1, p_2 \in H^n \) is the hyperbolic line segment between them. Its hyperbolic length is
\[
d(p_1, p_2) = \text{arcosh}(-\langle p_1, p_2 \rangle).
\]

This can be proved in the same way as we proved the corresponding theorem for the sphere.

Two-dimensional hyperbolic space

The hyperbolic plane \( H^2 \) is one component of the hyperboloid of two sheets \( x_1^2 + x_2^2 - x_3^2 = -1 \) in \( \mathbb{R}^{2,1} \). Any 2-dimensional subspace \( U \) of \( \mathbb{R}^{2,1} \) which intersects \( H^2 \) is
\[
U = \{ x \in \mathbb{R}^{2,1} \mid \langle x, n \rangle = 0 \}
\]
for some \( n \in \mathbb{R}^{2,1} \) with \( n(n) = 1 \). The vector \(-n\) would give the same subspace, but up to sign, the unit normal \( n \) of \( U \) is unique. Thus, the spacelike unit vectors in \( \mathbb{R}^{2,1} \) are in 2-to-1 correspondence with the hyperbolic lines in \( H^2 \). They are in 1-to-1 correspondence with the hyperbolic half-planes
\[
\{ x \in \mathbb{R}^{2,1} \mid \langle x, n \rangle \geq 0 \} \cap H^2.
\]

Proposition. Let \( n_1, n_2 \in \mathbb{R}^{2,1} \) with \( \langle n_1, n_1 \rangle = \langle n_2, n_2 \rangle = 1 \), and let \( l_1, l_2 \) be the corresponding hyperbolic lines, \( l_i = \{ x \in H^2 \mid \langle x, n_i \rangle = 0 \} \). Assume \( n_1 \neq \pm n_2 \), so that \( l_1 \) and \( l_2 \) are different lines. Then the following statements are equivalent.

(i) The lines \( l_1 \) and \( l_2 \) intersect.
(ii) The restriction of \( (\cdot,\cdot) \) to \( \text{span}(n_1, n_2) \) has signature \((2,0)\).
(iii) \( \langle n_1, n_2 \rangle < 1 \).

Proof. (i)\(\Rightarrow\)(ii): If \( l_1 \cap l_2 \neq \emptyset \) then there is an \( x \in H^2 \) with \( \langle x, n_1 \rangle = \langle x, n_2 \rangle = 0 \). So \( x^\perp = \text{span}(n_1, n_2) \), and (ii) follows because any non-zero vector orthogonal to a timelike vector is spacelike.

(ii)\(\Rightarrow\)(i): If the restriction of the scalar product to \( \text{span}(n_1, n_2) \) has signature \((2,0)\), then its restriction to the orthogonal complement \( \text{span}(n_1, n_2)^\perp \) must have signature \((0,1)\), so the complement intersects \( H^2 \).

(ii)\(\Rightarrow\)(iii): The vectors \( n_1, n_2 \) form a basis of \( \text{span}(n_1, n_2) \). In this basis, the matrix of the restriction of \( (\cdot,\cdot) \) is
\[
B = \begin{pmatrix}
\langle n_1, n_1 \rangle & \langle n_1, n_2 \rangle \\
\langle n_2, n_1 \rangle & \langle n_2, n_2 \rangle
\end{pmatrix} = \begin{pmatrix}
1 & \langle n_1, n_2 \rangle \\
\langle n_2, n_1 \rangle & 1
\end{pmatrix}.
\]

So \( B_{11} = 1 \) and \( \det B = 1 - \langle n_1, n_2 \rangle^2 \). The equivalence of (ii) and (iii) follows from last lecture’s signature theorem (and the fact that the matrix of a non-degenerate bilinear form has non-zero determinant). \(\square\)
Now suppose the hyperbolic lines $l_1, l_2$ intersect in $x \in H^2$, and let $h_i = \{ y \in H^2 \mid \langle x, y \rangle \geq 0 \}$ be half-planes bounded by $l_1, l_2$. Since the Lorentz scalar product of $\mathbb{R}^{2,1}$ is Euclidean on the subspace $x^+ = \text{span}(n_1, n_2)$, we can measure angles between vectors in $x^+$ in the usual way.

The exterior angle $\hat{\alpha}$ of the half-planes $h_1, h_2$ at $x$ is determined by

$$\cos \hat{\alpha} = \langle n_1, n_2 \rangle.$$ 

The interior angle $\alpha$ is $\pi - \hat{\alpha}$, so

$$-\cos \alpha = \langle n_1, n_2 \rangle.$$ 

In particular, $l_1$ and $l_2$ intersect orthogonally if $\langle n_1, n_2 \rangle = 0$.

**Remark.** A hyperbolic rotation (of $H^2$) with center $x \in H^2$ is a map $T \in O(2,1)$ with $T(x) = x$ and which is a Euclidean rotation on $x^+$. The exterior angle between the two half-spaces $h_1, h_2$ is the angle of the hyperbolic rotation mapping one to the other.

**Proposition.** (i) If $x \in H^2$ and $l$ is a hyperbolic line, then there is a unique hyperbolic line through $x$ which intersects $l$ orthogonally.

(ii) If $l_1, l_2$ are two hyperbolic lines with unit normals $n_1, n_2$ such that $\langle \langle n_1, n_2 \rangle \rangle > 1$ (so the lines do not intersect), then there is a unique line $l_3$ intersecting both $l_1$ and $l_2$ orthogonally.

**Proof.** Exercise.

**Hyperbolic triangles**

Let $A, B, C \in H^2$ be three points in the hyperbolic plane. Assume that they do not all lie on one hyperbolic line (this is equivalent to assuming $A, B, C$ to be linearly independent). The hyperbolic triangle with vertices $A, B, C$ is the intersection of $H^2 \subset \mathbb{R}^{2,1}$ with the set of non-negative linear combinations

$$\{ \lambda A + \mu B + \nu C \mid \lambda, \mu, \nu \in \mathbb{R}_{\geq 0} \}.$$ 

The side lengths $a = d(B, C)$, $b = d(C, A)$, $c = d(A, B)$ satisfy

$$-\cosh a = \langle B, C \rangle, \quad -\cosh b = \langle C, A \rangle, \quad -\cosh c = \langle A, B \rangle.$$ 

Let $A', B', C'$ be the spacelike unit vectors such that the half-plane bounded by the line through $B, C$ containing $A$ is

$$h_{A'} = \{ x \in H^2 \mid \langle A', x \rangle \geq 0 \},$$ 

and analogously for $B'$ and $C'$. Then the hyperbolic triangle with vertices $A, B, C$ is also the intersection $h_{A'} \cap h_{B'} \cap h_{C'}$. The interior angles $\alpha, \beta, \gamma$ at $A, B, C$ satisfy

$$-\cos \alpha = \langle B', C' \rangle, \quad -\cos \beta = \langle C', A' \rangle, \quad -\cos \gamma = \langle A', B' \rangle.$$ 

**Theorem.** The side lengths and interior angles of a hyperbolic triangle satisfy

$$\cos \alpha = \frac{-\cosh a + \cosh b \cosh c}{\sinh b \sinh c}, \quad (\text{hyperbolic side cosine theorem})$$ 

$$\cos a = \frac{\cos \alpha + \cos \beta \cos \gamma}{\sin \beta \sin \gamma}. \quad (\text{hyperbolic angle cosine theorem})$$

**Proof** (sketch). This can be proved in the same way as we proved the spherical cosine theorems using the Gram matrices $G, G'$ of $A, B, C$ and $A', B', C'$. Only now the scalar product is

$$\langle x, y \rangle = x^t E y \quad \text{with} \quad E = \left( \begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{array} \right).$$

So if $V = (A \ B \ C)$ and $W = (A' \ B' \ C')$, then

$$G = V^t E V = \left( \begin{array}{ccc} -1 & -\cosh c & -\cosh b \\ -\cosh c & -1 & -\cosh a \\ -\cosh b & -\cosh a & -1 \end{array} \right), \quad G' = W^t E W = \left( \begin{array}{ccc} 1 & -\cos \gamma & -\cos \beta \\ -\cos \gamma & -1 & -\cos \alpha \\ -\cos \beta & -\cos \alpha & -1 \end{array} \right),$$

and $D = W^t E V$ is a diagonal matrix with positive elements on the diagonal. Continue as in the spherical case . . .
Remark. The spherical cosine theorems for a sphere of radius $R$ (instead of 1) are

$$
\cos \alpha = \frac{\cos \frac{a}{R} - \cos \frac{b}{R} \cos \frac{c}{R}}{\sin \frac{b}{R} \sin \frac{c}{R}}, \quad \cos \frac{a}{R} = \frac{\cos \alpha + \cos \beta \cos \gamma}{\sin \beta \sin \gamma}.
$$

One gets the hyperbolic cosine theorems by setting $R = i$. That’s why it is sometimes said that hyperbolic geometry is the geometry on a sphere with imaginary radius.

From the hyperbolic cosine theorems, one can derive

$$
\sinh \frac{a}{\sin \alpha} = \sinh \frac{b}{\sin \beta} = \sinh \frac{c}{\sin \gamma}
$$

(hyperbolic sine theorem) in the same way in which we derived the spherical sine theorem from the spherical cosine theorems. One can also derive hyperbolic versions of the half-angle and half-side theorems, and other formulas of spherical trigonometry.

**Theorem.** (i) A hyperbolic triangle with side lengths $a, b, c \in \mathbb{R}_{>0}$ exists if and only if the triangle inequalities are satisfied. (ii) A hyperbolic triangle with angles $\alpha, \beta, \gamma \in (0, \pi)$ exists if and only if

$$
\alpha + \beta + \gamma < \pi.
$$

**Remark.** We will see later that the area of a hyperbolic triangle is $\pi - (\alpha + \beta + \gamma)$.

**Proof (sketch).** 1. Show that a symmetric $(3 \times 3)$-matrix $G$ is the Gram matrix of 3 linearly independent vectors in $\mathbb{R}^2$ if and only if the bilinear form $x^t G y$ has signature $(2, 1)$.

2. Consider

$$
G = \left( \begin{array}{ccc}
-1 & -\cosh c & -\cosh b \\
-\cosh c & -1 & -\cosh a \\
-\cosh b & -\cosh a & -1
\end{array} \right).
$$

Because

$$
G_{11} = -1 < 0, \quad \det \left( \begin{array}{cc}
G_{11} & G_{12} \\
G_{21} & G_{22}
\end{array} \right) = -\sinh^2 c < 0,
$$

the signature of $G$ is $(2, 1)$ if and only if $\det G < 0$ (by the theorem from Lecture 6).

3. Show the remarkable identity

$$
\det G = -4 \sinh \left( \frac{-a+b+c}{2} \right) \sinh \left( \frac{a-b+c}{2} \right) \sinh \left( \frac{a+b-c}{2} \right) \sinh \left( \frac{a+b+c}{2} \right),
$$

and use it to prove part (i) of the theorem.

Part (ii) can be shown in the same way by considering $G' = \left( \begin{array}{ccc}
1 & -\cos \gamma & -\cos \beta \\
-\cos \gamma & 1 & -\cos \alpha \\
-\cos \beta & -\cos \alpha & 1
\end{array} \right)$ and using the identity

$$
\det G' = -4 \cos \left( \frac{-a+b+c}{2} \right) \cos \left( \frac{a-b+c}{2} \right) \cos \left( \frac{a+b-c}{2} \right) \cos \left( \frac{a+b+c}{2} \right).
$$

$\square$
The Klein model of the hyperbolic plane

Project $H^2 \subset \mathbb{R}^{2,1}$ to the plane $x_3 = 1$ with $0 \in \mathbb{R}^{2,1}$ as center of projection. This is central projection, because 0 is the center of the hyperboloid of which $H^2$ is one sheet. A point $\left(\frac{z_1}{x_2}, \frac{z_2}{x_3}\right) \in H^2$ is mapped to $\frac{1}{x_3} \left(\frac{z_1}{x_2}, \frac{z_2}{x_3}\right)$, and all of $H^2$ is mapped to the inside of the unit circle in the image plane $x_3 = 1$. Hyperbolic lines are mapped to secants of this unit circle. If we forget about the $x_3$-coordinate, we get an image of the hyperbolic plane which is called the Klein model. Thus, the Klein model of the hyperbolic plane is the unit disk

$$D^2 = \{u \in \mathbb{R}^2 \mid u_1^2 + u_2^2 < 1\},$$

where $\left(\frac{z_1}{x_2}, \frac{z_2}{x_3}\right) \in H^2$ corresponds to $\frac{1}{x_3} \left(\frac{z_1}{x_2}, \frac{z_2}{x_3}\right) \in D^2$, and, inversely, $\left(\frac{u_1}{u_2}\right) \in D^2$ corresponds to $1 \sqrt{1 - u_1^2 - u_2^2} \left(\frac{u_1}{u_2}\right) \in H^2$. Angles and lengths in are measured in $H^2$.

For two hyperbolic lines $l_1, l_2$ with unit normals $n_1, n_2 \in \mathbb{R}^{2,1}$, there are three possibilities:

1. $|\langle n_1, n_2 \rangle| < 1$.
2. $|\langle n_1, n_2 \rangle| = 1$.
3. $|\langle n_1, n_2 \rangle| > 1$.

The lines intersect. The lines do not intersect and their images in the Klein model intersect on the unit circle. The lines do not intersect and their images in the Klein model intersect outside the unit circle.

In cases 2 and 3 the lines do not intersect, thus they are parallel. However, to distinguish the two cases, parallel is sometimes used to mean only lines in case 2, and lines in case 3 are then called ultra-parallel.

Angle of parallelism

If $l \subset H^2$ is a hyperbolic line and $x \in H^2$ is a point not on $l$, then there exist two parallels (in the narrow sense) to $l$ through $x$ and infinitely many ultra-parallels. The angle $\alpha$ between one of the parallels and the perpendicular through $x$ is called the angle of parallelism. It depends only on the distance $b$ from $x$ to $l$. In fact, $\alpha = 2 \arctan e^{-b}$ (exercise).

More distance formulas

Let $x \in H^2$ and let $l = \{y \in H^2 \mid \langle y, n \rangle = 0\}$ be a hyperbolic line.

**Proposition.** The point on $l$ that is closest to $x$ is the intersection $x_p$ of $l$ with the line $l_p$ through $x$ that is perpendicular to $l$.

**Proof.** The case $x \in l$ is trivial, so assume $x \notin l$. Let $\gamma : \mathbb{R} \to l \subset H^2$ be a parameterization of $l$ with unit speed. So $\langle \gamma, \gamma \rangle = -1$, $\langle \gamma, n \rangle = 0$, and $\langle \gamma', \gamma' \rangle = 1$. Let $f(s) = d(x, \gamma(s)) = \text{arcosh}(-\langle x, \gamma(s) \rangle)$. Convince yourself that $f(s) \to \infty$ as $s \to \pm\infty$. So $f$ must attain a minimum, say at $s = s_0$. Then $f'(s_0) = 0$, and this implies $\langle x, \gamma'(s_0) \rangle = 0$. Let $x_p = \gamma(s_0)$, $n_p = \gamma'(s_0)$, and let $l_p = \{y \in H^2 \mid \langle y, n_p \rangle = 0\}$. Then

- $x_p$ is the point on $l$ closest to $x$,
- $\langle x, n_p \rangle = 0$ so $x \in l_p$,
- $\langle x_p, n_p \rangle = 0$ because $\langle \gamma, \gamma' \rangle = 0$, so $x_p \in l_p$,
- $\langle n, n_p \rangle = 0$ because $\langle \gamma', n \rangle = 0$, so $l \perp l_p$. □
Proposition. The distance \(d(x, l)\) from \(x\) to \(l\) satisfies \(|\langle x, n \rangle| = \sinh d(x, l)\).

Proof. Let \(V = (x, x_p, n, n_p)\) and let \(E\) be the diagonal matrix with 1, 1, -1 on the diagonal. Then
\[
0 = \det V^t EV = \det \begin{pmatrix} 
\langle x, x \rangle & \langle x, x_p \rangle & \langle x, n \rangle & \langle x, n_p \rangle \\
\langle x_p, x \rangle & \langle x_p, x_p \rangle & \langle x_p, n \rangle & \langle x_p, n_p \rangle \\
\langle n, x \rangle & \langle n, x_p \rangle & \langle n, n \rangle & \langle n, n_p \rangle \\
\langle n_p, x \rangle & \langle n_p, x_p \rangle & \langle n_p, n \rangle & \langle n_p, n_p \rangle 
\end{pmatrix} = \det \begin{pmatrix} 
-1 & -\cosh d(x, x_p) & 0 & 0 \\
-\cosh d(x_p, x) & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 
\end{pmatrix} 
\]
\[
= 1 + \langle x, n \rangle^2 - \cosh^2 d(x, x_p) = -\sinh^2 d(x, x_p) + \langle x, n \rangle^2 
\]
and \(d(x, x_p) = d(x, l)\) by the previous proposition. □

Of course the sign of \(\langle x, n \rangle\) depends on whether or not \(x\) is contained in the half-plane \(\langle x, n \rangle \geq 0\).

Proposition. The distance \(d(l_1, l_2)\) between two lines \(l_i = \{y \in \mathbb{H}^2 \mid \langle y, n_i \rangle = 0\}\) with \(|\langle n_1, n_2 \rangle| > 1\) satisfies \(\langle n_1, n_2 \rangle = \cosh d(l_1, l_2)\).

Proof (sketch). This can be proved in the same way as the previous proposition, but letting \(V = (x, n_1, n_2, n_p)\), where \(l_p = \{y \in \mathbb{H}^2 \mid \langle y, n_p \rangle = 0\}\) is the common perpendicular to \(l_1, l_2\), and \(x = l_1 \cap l_p\).

The sign of \(\langle n_1, n_2 \rangle\) depends on whether or not one of the half-planes \(\langle y, n_i \rangle \geq 0\) contains the other.

Summary. Let \(x_1, x_2 \in \mathbb{H}^2\), let \(n, n_1, n_2\) be unit spacelike vectors and let \(l_1, l_2\) be the corresponding lines. The two sides of each line are marked + and − according to the sign of the scalar product of points in \(\mathbb{H}^2\) with the chosen normal on that side.

\[
\begin{align*}
\langle x_1, x_2 \rangle &= -\cosh d(x_1, x_2) \\
\langle n_1, n_2 \rangle &= \cosh d(l_1, l_2) \\
\langle x_1, n \rangle &= \sinh d(x_1, l) \\
\langle x_2, n \rangle &= -\sinh d(x_2, l) \\
\langle n_1, n_2 \rangle &= \cosh \alpha = -\cos \alpha \\
\langle n_1, n_2 \rangle &= -\cosh d(l_1, l_2)
\end{align*}
\]

Hyperbolic “trilaterals”

Let us define a hyperbolic trilateral as a non-empty intersection of three half-planes, of which none is contained in another. If we only consider the generic cases where pairs of boundary lines intersect or are ultra-parallel then there are four types of trilaterals according to the number of vertices. In the figures, the common perpendiculars of ultra-parallel lines are drawn in yellow.

The first case is the case of triangles. In the other cases, one can also derive trigonometric formulas in the same way as we did for triangles. Of particular interest are trilaterals of type (4), which correspond to right-angled hexagons. For those one obtains the cosine and sine theorems
\[
\cosh a' = \frac{\cosh a + \cosh b \cosh c}{\sinh b \sinh c} \quad \text{and} \quad \frac{\sinh a'}{\sinh a} = \frac{\sinh b'}{\sinh b} = \frac{\sinh c'}{\sinh c}.
\]
Intersections of $H^2 \subset \mathbb{R}^{2,1}$ with planes

Any plane in $\mathbb{R}^{2,1}$ is determined by an equation of the form $\langle x, v \rangle = b$ with $v \in \mathbb{R}^{2,1} \setminus \{0\}$ and $b \in \mathbb{R}$. There are three types of planes according to whether $v$ is spacelike, timelike or lightlike.

- If $\langle v, v \rangle < 0$, we may assume that $\langle v, v \rangle = -1$ and $v_3 > 0$, so $v \in H^2$. A plane of this type which intersects $H^2$ is of the form $\langle v, x \rangle = - \cosh r$. The intersection is therefore a circle with center $v$ and radius $r$.
- If $\langle v, v \rangle > 0$, we may assume that $\langle v, v \rangle = 1$ and $b > 0$. A plane of this type always intersects $H^2$ and is of the form $\langle v, x \rangle = \sinh r$. The intersection with $H^2$ is therefore a curve of constant distance from the line $\langle v, x \rangle = 0$. (For $r = 0$, it is the line itself.)
- If $\langle v, v \rangle = 0$, a non-empty intersection with $H^2$ is called a horocircle.

The Poincaré disk model of the hyperbolic plane

Project $H^2$ to the plane $x_3 = 0$ with $e_3 = \left( \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right)$ as center of projection. This is stereographic projection of $H^2$. It maps $H^2$ to the unit disk of the plane $x_3 = 0$. Analytically, it is the map $\sigma_{H^2}: H^2 \to \mathbb{D}^2$,

$$\sigma_{H^2} \left( \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right) = \frac{1}{x_4 + 1} \left( \begin{array}{c} x_1 \\ x_2 \end{array} \right),$$

$$\sigma_{H^2}^{-1} \left( \begin{array}{c} u_1 \\ u_2 \end{array} \right) = \frac{1}{1 - u_1^2 - u_2^2} \left( \begin{array}{c} 2u_1 \\ 2u_2 \\ 1 + u_1^2 + u_2^2 \end{array} \right).$$

**Theorem.** Stereographic projection of $\sigma_{H^2}: H^2 \to \mathbb{D}^2$ maps intersections of $H^2 \subset \mathbb{R}^{2,1}$ with planes in $\mathbb{R}^{2,1}$ to intersections of $\mathbb{D}^2$ with circles and lines.

In particular, hyperbolic lines are mapped to intersections of $\mathbb{D}^2$ with circles and lines that intersect the unit circle $\partial \mathbb{D}^2$ orthogonally.

This can be shown by a calculation like in the case of $S^2$.

**Theorem.** Consider two curves $\tilde{\gamma}, \tilde{\eta} : (-\varepsilon, \varepsilon) \to H^2$ in the hyperbolic plane with $\tilde{\gamma}(0) = \tilde{\eta}(0) = \tilde{p}$. Let $\gamma = \sigma_{H^2} \circ \tilde{\gamma}$ and $\eta = \sigma_{H^2} \circ \tilde{\eta}$ be their images in $D^2$ under stereographic projection, and let $\hat{v} = \gamma'(0)$, $\hat{w} = \eta'(0)$, $v = \gamma'(0)$, $w = \eta'(0)$. Then

$$\langle \hat{v}, \hat{w} \rangle_{\mathbb{D}^2} = \frac{4}{(1 - p_1^2 - p_2^2)^2} \langle v, w \rangle_{\mathbb{R}^2},$$

where $\langle \cdot, \cdot \rangle_{\mathbb{R}^2}$ is the Lorentz scalar product of $\mathbb{R}^{2,1}$ and $\langle \cdot, \cdot \rangle_{\mathbb{D}^2}$ is the standard Euclidean scalar product of $\mathbb{R}^2$.

This, too, can be shown by a calculation like in the case of $S^2$.

Hence $\sigma_{H^2}$ is conformal in the sense that curves in $H^2$ intersecting at some angle are mapped to curves inside the unit circle of the Euclidean plane intersecting at the same angle.

One can measure hyperbolic lengths and angles directly in $D^2$ by using the variable scalar product

$$g_p(v, w) = \frac{4}{(1 - p_1^2 - p_2^2)^2} \langle v, w \rangle_{\mathbb{R}^2}. \quad (*)$$

For example, the hyperbolic length of a curve in $H^2 \subset \mathbb{R}^{2,1}$ which $\sigma_{H^2}$ maps to $\gamma : [t_1, t_2] \to D^2$ is

$$\int_{t_1}^{t_2} \sqrt{g_p(\gamma'(t), \gamma'(t))} \, dt.$$
In the image of $H^2$ under stereographic projection, lengths appear scaled down by the variable factor $\frac{1}{2}(1 - p_1^2 - p_2^2)$. The image in $D^2$ of an object in the hyperbolic plane gets smaller and smaller as it moves towards the boundary circle $\partial D^2$.

A Riemannian metric on an open set $U \subseteq \mathbb{R}^n$ is a variable Euclidean scalar product. More precisely, it is a $C^\infty$ map $g : U \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, $(p, v, w) \mapsto g_p(v, w)$, such that for each $p \in U$, $g_p(\cdot, \cdot)$ is a Euclidean scalar product on $\mathbb{R}^n$. (Thus, $g_p(v, w) = v^t G(p) w$ with a matrix $G$ depending on $p \in U$.) One can then measure lengths of curves and angles in $U$ using the Riemannian metric, and this is called Riemannian geometry.

A Riemannian metric $g$ on $U \subseteq \mathbb{R}^n$ is called conformal if

$$g_p(u, v) = \lambda(p)(u, v)_{\mathbb{R}^n}$$

for some function $\lambda : U \rightarrow \mathbb{R}_{>0}$. If this is the case, angles measured using $g$ are equal to the Euclidean angles measured using $(\cdot, \cdot)_{\mathbb{R}^n}$.

Thus, $g_p$ as defined by equation (\ast) is a conformal Riemannian metric on $D^2$. The unit disk $D^2$ with this Riemannian metric is called the Poincaré disk model of the hyperbolic plane.

From Klein model to Poincaré disk model via the hemisphere model

We have encountered two ways to map $H^2 \in \mathbb{R}^{2,1}$ to the unit disk. Central projection gives the Klein model and stereographic projection gives the Poincaré disk model. The composition

$$D^2 \xrightarrow{\text{central projection}} H^2 \xrightarrow{\text{stereographic projection}} D^2$$

is a peculiar self-map of the unit disk $D^2$, which maps secants of $D^2$ to circles orthogonal to the boundary $\partial D^2$.

**Proposition.** The same map $D^2 \rightarrow D^2$ is also the result of the following construction: First, project $D^2 \subset \mathbb{R}^2 \subset \mathbb{R}^3$ orthogonally down to the lower hemisphere of $S^2 \subset \mathbb{R}^3$. Then project stereographically back to $D^2$.

**Proof.** Let $x = \left( \frac{x_1}{x_3}, \frac{x_2}{x_3} \right) \in H^2$. Via central projection, this corresponds to $\frac{1}{x_3}(\frac{x_1}{x_2})$ in the Klein model. Orthogonally projecting down to the lower hemisphere of $S^2$ gives

$$\left( \frac{x_1}{x_3}, \frac{x_2}{x_3} \right) = \frac{1}{\sqrt{1 - \left( \frac{x_1}{x_3} \right)^2 - \left( \frac{x_2}{x_3} \right)^2}} \left( \frac{x_1}{x_3}, \frac{x_2}{x_3} \right) = \frac{1}{x_3} \left( \frac{x_1}{x_2}, \frac{x_2}{x_3} \right)$$

Projecting this point back to the unit disk (by stereographic projection of $S^2$) results in

$$\sigma \left( \frac{1}{x_3} \left( \frac{x_1}{x_2}, \frac{x_2}{x_3} \right) \right) = \frac{1}{1 + \frac{\sqrt{3}}{x_3}} = \frac{1}{x_3 + 1} \left( \frac{x_1}{x_2} \right),$$

which is the same as $\sigma_{H^2}(x)$.

The calculation in the proof shows that $\left( \frac{x_1}{x_3}, \frac{x_2}{x_3} \right) \mapsto \frac{1}{x_3} \left( \frac{x_1}{x_2}, \frac{x_2}{x_3} \right)$ maps $H^2$ to the lower hemisphere of $S^2$. Equally $\left( \frac{x_2}{x_3}, \frac{x_1}{x_3} \right) \mapsto \frac{1}{x_3} \left( \frac{x_2}{x_3}, \frac{x_1}{x_2} \right)$ maps $H^2$ to the upper hemisphere of $S^2$. These images of $H^2$ are called the (lower and upper) hemisphere models. The hemisphere models are also conformal, and circles, horospheres, lines, and curves of constant distance from lines are represented by (parts of) circles in $S^2$. 

---

Boris Springborn  
Geometry I Lecture 10  
Winter Semester 07/08
The Poincaré half-plane model

One obtains the Poincaré half-plane model of $H^2$ by projecting the upper hemisphere model stereographically from a point on the equator, say from $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ to the plane $x_1 = 0$. This maps the equator (minus $e_1$) to the $x_2$-axis and the upper hemisphere to the upper half-plane of the the $x_2, x_3$-plane which we identify with the upper half-plane

$$H^2_+ = \{ \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \mid u_2 > 0 \} \subset \mathbb{R}^2.$$ 

Since stereographic projection is conformal and maps circles to circles and lines, hyperbolic lines are represented in $H^2_+$ by half circles meeting the $u_1$-axis orthogonally and vertical lines. You can show by a direct calculation that hyperbolic lengths and angles can be measured in $H^2_+$ by using the Riemannian metric

$$g_p(v, w) = \frac{1}{u_2^2} \langle v, w \rangle_{\mathbb{R}^2}.$$ 

The half-plane $H^2_+$ with this Riemannian metric is called the Poincaré half-plane model of the hyperbolic plane. In $H^2_+$, hyperbolic lengths appear scaled by the variable factor $u_2$ which is the Euclidean distance to the boundary.

Two examples for length calculations in the half-plane model (and some remarks)

1. Consider the curve $\gamma : [0, \alpha] \to H^2_+$, $\gamma(t) = r \begin{pmatrix} \sin t \\ \cos t \end{pmatrix}$. Its hyperbolic length is

$$\int_0^\alpha \sqrt{g_\gamma(\gamma'(t), \gamma'(t))} \, dt = \int_0^\alpha \sqrt{\frac{1}{r^2 \cos^2 t} \left\langle r \begin{pmatrix} \cos t \\ -\sin t \end{pmatrix}, r \begin{pmatrix} \cos t \\ -\sin t \end{pmatrix} \right\rangle} \, dt$$

$$= \int_0^\alpha \frac{1}{\cos t} \, dt = \frac{1}{2} \log \frac{1 + \sin \alpha}{1 - \sin \alpha} = \log \tan \left( \frac{\alpha}{2} + \frac{\pi}{4} \right).$$ (Check it out!)

2. Consider the curve $\eta : [t_1, t_2] \to H^2_+$, $\eta(t) = \begin{pmatrix} 0 \\ t \end{pmatrix}$. Its hyperbolic length is

$$\int_{t_1}^{t_2} \sqrt{g_{\eta(t)}(\eta'(t), \eta'(t))} \, dt = \int_{t_1}^{t_2} \sqrt{\frac{1}{r^2} \left\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle} \, dt = \int_{t_1}^{t_2} \frac{1}{t} \, dt$$

$$= \log t_2 - \log t_1 = \log \frac{t_2}{t_1}.$$ 

Note that in the first example, the length does not depend on $r$, and in the second example, the length depends only on the quotient $t_2/t_1$. In fact, scaling transformations $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \mapsto \lambda \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ (with $\lambda > 0$) of the upper half-plane represent isometries of the hyperbolic plane. For example, scaling
by the factor 2 makes all objects in the upper half-plane look twice as large. At the same time, all distances from the boundary also double. In effect, hyperbolic lengths stay the same.

Horizontal translations \((u_1, u_2) \mapsto (u_1 + c, u_2)\) of the upper half-plane also represent isometries of \(H^2\), and so do reflections on vertical lines.

**Calculating hyperbolic areas in the half-plane model**

If Euclidean lengths in the upper half-plane model have to be scaled by a variable factor of \(1/u_2^2\) to get the hyperbolic length, then Euclidean area has to be scaled by the factor \((1/u_2^2)^2\). So the hyperbolic area of a region \(R \subset H_2^+\) is

\[
\text{area}(R) = \int_R \frac{1}{u_2^2} \, du_1 \, du_2.
\]

**Theorem.** The area of a hyperbolic triangle with interior angles \(\alpha, \beta, \gamma\) is

\[
\pi - \alpha - \beta - \gamma.
\]

**Remark.** In Lecture 8, I sketched a purely analytic proof for the fact that the angle sum in a hyperbolic triangle is always less than \(\pi\). A more visual argument can be made using the Poincaré disk model and moving one vertex of the triangle to the center of the disk.

To prove the theorem, consider first a hyperbolic triangle \(T(\alpha, \beta, 0)\) with one vertex "at infinity". (Strictly speaking, this is not a triangle but what we called a trilateral.) The figure on the right shows such a triangle in the Klein model, in the Poincaré disk model, and in the Poincaré half-plane (where the infinite vertex was used to project from the hemisphere model).

\[
\text{area}(T(\alpha, \beta, 0)) = \int_{T(\alpha, \beta, 0)} \frac{1}{u_2^2} \, du_1 \, du_2
\]

\[
= \int_{u_1=\cos(\pi-\alpha)}^{\cos \beta} \left( \int_{u_2=\sqrt{1-u_1^2}}^{\infty} \frac{1}{u_2^2} \, du_2 \right) \, du_1
\]

\[
= \int_{u_1=\cos(\pi-\alpha)}^{\cos \beta} \left( -\frac{1}{u_2} \right)_{u_2=\sqrt{1-u_1^2}}^{\infty} \, du_1
\]

\[
= \left[ -\arccos u_1 \right]_{\cos(\pi-\alpha)}^{\cos \beta} = \pi - \alpha - \beta.
\]

For \(\beta \to \infty\) one obtains the area of a triangle with two vertices "at infinity":

\[
\text{area}(T(\alpha, 0, 0)) = \pi - \alpha.
\]

Now we can calculate the area for a triangle \(T(\alpha, \beta, \gamma)\) with angles \(\alpha, \beta, \gamma\):

\[
\text{area}(T(\alpha, \beta, \gamma)) = \text{area}(T(\alpha, 0, 0)) - \text{area}(T(\beta_1, \gamma_1, 0)) - \text{area}(T(\beta_2, 0, 0)
\]

\[
= (\pi - \alpha) - (\pi - \beta_1 - \gamma_1) - (\pi - \beta_2)
\]

\[
= \pi - \alpha - \beta_1 - \beta_2 - \gamma_1 = \pi - \alpha - \beta - \gamma.
\]

This proves the theorem.

**Concluding remarks**

(1) All the models we have discussed exist also for higher dimensional hyperbolic space.

(2) We have defined hyperbolic space as one sheet of a hyperboloid and then derived the other models from it. Actually, any metric space isometric to our \(H^n\) is called hyperbolic space and \(H^n\). The hyperboloid is just a model like the others, called the hyperboloid model.
Projective Geometry

Introduction

Consider projecting a plane $E$ to another plane $E'$ from a point $P$ not on $E$ or $E'$.

Every point in $E$ has an image in $E'$ except points on the vanishing line of $E$, which is the intersection of $E$ with the plane parallel to $E'$ through $P$. Every point in $E'$ has a preimage in $E$ except points on the vanishing line of $E'$, which is the intersection of $E'$ with the plane parallel to $E$ through $P$.

The projection maps lines to lines. A family of parallel lines in $E$ is mapped to a family of lines in $E'$ which intersect in a point on the vanishing line.

Idea: Introduce, in addition to the ordinary points of $E$, new points which correspond to points on the vanishing line of $E'$. In the same way, introduce new points of $E'$ which are images of the vanishing line of $E$. These new points are called points at infinity, and the extended planes are called projective planes. The projection becomes a bijection between projective planes. Parallel lines in $E$ intersect in a point at infinity. The points at infinity of $E$ form a line called the line at infinity which corresponds to the vanishing line of $E'$.

Drawing a floor tiled with square tiles

Suppose you have already drawn the first tile. (I don’t want to go into the details of how one can construct the image of the first tile, even though that is interesting and not difficult.) The figure shows how the other tiles can then be constructed.

Analytic treatment

Suppose $E$ is the $x_1 x_2$-plane, $E'$ is the $x_2 x_3$-plane, and $P = \left( \frac{-1}{0} \right)$. A point $A = \left( \frac{x_1}{x_2} \right) \in E$ is mapped a point $A' = \left( \frac{0}{y_3} \right) \in E'$, and by solving $A' = P + t(A - P)$ for $t$ one finds that $y_1 = \frac{x_1}{x_2}$ and $y_2 = \frac{x_1}{x_2} + 1$. So in terms of the coordinates $x_1, x_2$ of plane $E$ and $y_1, y_2$ of plane $E'$, the projection is the function

$$f(x_1, x_2) := \frac{1}{x_1 + 1} \left( \frac{x_1}{x_2} \right).$$

The vanishing line of $E$ is the line $x_1 = -1$, and the vanishing line of $E'$ is the line $y_2 = 1$.

Introduce homogeneous coordinates: Instead of using two numbers $(\frac{x_1}{x_2})$ to describe a point in $E$, use three numbers $(\frac{u_1}{u_2}, \frac{v_1}{v_2})$ such that $x_1 = \frac{u_1}{u_2}$ and $x_2 = \frac{v_1}{v_2}$. The homogeneous coordinates for a point are not unique: $(\frac{x_1}{x_2})$ are homogeneous coordinates for the point $(\frac{x_1}{x_2})$, but for any $\lambda \neq 0$, $(\frac{\lambda x_1}{\lambda x_2})$ are also homogeneous coordinates for the same point. In the same way, use homogeneous coordinates $(\frac{u_1}{u_2})$ with $u_1 = \frac{v_1}{v_2}$ and $u_2 = \frac{v_2}{v_3}$ to describe a point $(\frac{u_1}{u_2}) \in E'$. Let us write the projection $f$ in terms of homogeneous coordinates. Let $(\frac{u_1}{u_2})$ be homogeneous coordinates for $(\frac{x_1}{x_2})$ and let $(\frac{v_1}{v_2})$ be homogeneous coordinates for $(\frac{y_1}{y_2}) = f(\frac{x_1}{x_2})$. Then

$$\frac{v_1}{v_3} = y_1 = \frac{x_2}{x_1 + 1} = \frac{u_2}{u_3 + 1} = \frac{u_2}{u_1 + u_3},$$

so we may choose

$$\left( \begin{array}{c} v_1 \\ v_2 \\ v_3 \end{array} \right) = \left( \begin{array}{c} u_2 \\ u_1 \\ u_1 + u_3 \end{array} \right) = \left( \begin{array}{ccc} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{array} \right) \left( \begin{array}{c} u_1 \\ u_2 \\ u_3 \end{array} \right) =: f \left( \begin{array}{c} u_1 \\ u_2 \\ u_3 \end{array} \right).$$
Using homogeneous coordinates, the projection may thus be written as a linear map \( \tilde{f} : \mathbb{R}^3 \to \mathbb{R}^3 \). Moreover, \( \tilde{f} \) is bijective! Points on the vanishing line \( x_1 = -1 \) of \( E \) have homogeneous coordinates \( \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \) with \( u_1 + u_3 = 0 \), and \( \tilde{f} \) maps these to vectors \( \begin{pmatrix} u_1' \\ u_2' \end{pmatrix} \), which are not homogeneous coordinates for any point in \( E' \). But we can interpret them as homogeneous coordinates for a point at infinity of the extended plane.

Thus, each non-zero vector \( u \in \mathbb{R}^3 \) represents a point in a projective plane (of which it is the vector of homogeneous coordinates) and two vectors \( u, u' \) represent the same point if and only if \( u' = \lambda u \) for some \( \lambda \neq 0 \).

**Projective geometry**

Projective geometry deals with the properties of figures that remain unchanged under projections. An example for a theorem of projective geometry is Pappus’ theorem. It talks only about points, lines, and the incidence relation between points and lines.

We will see that a curve being a conic section is a projective property. But the distinction between circles, ellipses, parabolas and hyperbolas is not. The photograph shows a circle being projected to a parabola.

Also the distinction between ordinary points and points at infinity is not a projective property, because as we have seen, a projection can map ordinary points to points at infinity and vice versa. So from the point of view of projective geometry, points at infinity of a projective plane are not distinguished from ordinary points and the line at infinity is a line like any other.

**Basic definitions**

Let \( V \) be a vector space over a field \( F \). The projective space of \( V \) is the set \( P(V) \) of 1-dimensional subspaces of \( V \). If the dimension of \( V \) is \( n+1 \), then the dimension of the projective space \( P(V) \) is \( n \). A 1-dimensional projective space is called a projective line and a 2-dimensional one is called a projective plane. An element of \( P(V) \) (that is, a 1-dimensional subspace of \( V \)) is called a point of the projective space.

If \( v \in V \setminus \{0\} \), then we write \( [v] := \text{span} v \). So \( [v] \) is a point in \( P(V) \), and \( v \) is called a representative vector for this point. If \( \lambda \neq 0 \) then \( \lambda [v] = [v] \) and \( \lambda v \) another representative vector for the same point.

Suppose we chose a basis \( v_1, \ldots, v_{n+1} \) of \( V \). This gives an identification of \( V \) with \( F^{n+1} \) and of \( P(V) \) with \( P(F^{n+1}) \). A vector \( v \in V \) has a basis representation

\[
v = \sum_{j=1}^{n+1} x_j v_j
\]

and \( x_1, \ldots, x_{n+1} \in F \) are the coordinates of \( v \) with respect to the basis. These coordinates of \( v \in V \) are the homogeneous coordinates of the point \( [v] \in P(V) \). If \( \lambda \neq 0 \), then \( \lambda x_1, \ldots, \lambda x_{n+1} \) are also homogeneous coordinates of \( [v] \).

Let \( U \subset P(V) \) be the subset of points for which a particular homogeneous coordinate, say \( x_{n+1} \), does not vanish: \( U = \{ [v] \in P(V) \mid v = \sum_{j=1}^{n+1} x_j v_j \text{ with } x_{n+1} \neq 0 \} \). Then the map \( U_{n+1} \to F^n \),

\[
v \mapsto \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} x_1/x_{n+1} \\ x_2/x_{n+1} \\ \vdots \\ x_n/x_{n+1} \end{pmatrix}
\]

is a bijection with inverse

\[
\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \mapsto \sum_{j=1}^{n+1} u_j v_j + v_{n+1},
\]

and \( u_1, \ldots, u_n \) are called affine coordinates of \( v \in U_{n+1} \).

We will mostly consider the case where the base field \( F \) of the vector space is the field of real numbers \( \mathbb{R} \). In this case, the concepts of a point, line or curve, etc., have their intuitively geometric meaning. But many theorems of projective geometry hold for arbitrary base fields. In particular, when dealing with curves and surfaces defined by algebraic equations, it is natural to use the base field \( \mathbb{C} \). Finite fields are used in elliptic curve cryptography.

One usually writes \( \mathbb{R}P^n \) for \( P(\mathbb{R}^{n+1}) \) and \( \mathbb{C}P^n \) for \( P(\mathbb{C}^{n+1}) \). More generally, if \( V \) is any real or complex vector space, then \( P(V) \) is called an \( \mathbb{R}P^n \) or \( \mathbb{C}P^n \), respectively.
Basic examples

The points of $\mathbb{RP}^1$, the one-dimensional real projective space or the real projective line, are the 1-dimensional subspaces

$$[\frac{x_1}{x_2}] = \mathbb{R} \left( \frac{x_1}{x_2} \right) \subset \mathbb{R}^2,$$

where $\left( \frac{x_1}{x_2} \right) \neq 0$.

The points with homogeneous coordinate $x_2 \neq 0$ are described by one affine coordinate $x_1/x_2 \in \mathbb{R}$. On the other hand, all representative vectors $\left( \frac{x_1}{x_2} \right) \in \mathbb{R}^2 \setminus \{0\}$ represent the same point $[1,0] \in \mathbb{RP}^1$.

So one can think of $\mathbb{RP}^1$ as $\mathbb{R}$ plus one additional point (which is reasonably denoted by $\infty$).

The points of $\mathbb{RP}^2$, the two-dimensional real projective space or the real projective plane, are the 1-dimensional subspaces

$$[\frac{x_1}{x_2}] = \mathbb{R} \left( \frac{x_1}{x_2} \right) \subset \mathbb{R}^3,$$

where $\left( \frac{x_1}{x_2} \right) \neq 0$.

The points with homogeneous coordinate $x_3 \neq 0$ are described by two affine coordinates $x_1/x_3$, $x_2/x_3$. On the other hand, the points $\left[ \frac{x_1}{x_2} \right]$ form the 1-dimensional real projective space $P(U)$, where $U \subset \mathbb{R}^3$ is the subspace $U = \left( \left\{ \frac{x_1}{x_2} \right\} \right) \in \mathbb{R}^3$. So $\mathbb{RP}^2$ can be thought of as $\mathbb{R}^2$ plus a projective line.

In general one can think of $\mathbb{RP}^n$ as $\mathbb{R}^n$ plus an additional $\mathbb{RP}^{n-1}$.

In the same way, the complex projective line $\mathbb{CP}^1$ is $\mathbb{C}$ plus one additional point $\infty$. In complex analysis, $\mathbb{CP}^1 = \mathbb{C} \cup \{ \infty \}$ is called the extended complex plane and denoted by $\hat{\mathbb{C}}$.

Projective subspaces

A projective subspace of the projective space $P(V)$ is a projective space $P(U)$, where $U$ is a vector subspace of $V$. If $k$ is the dimension of $P(U)$ (that is, $k + 1$ is the dimension of $U$), then $P(U)$ is called a $k$-plane in $P(V)$. In particular, for $k = 1$ it is called a line, for $k = 2$ a plane and for $k = n - 1$ a hyperplane in $P(V)$.

Exercise. How many points are there in the projective plane $P(\mathbb{Z}_3^2)$? How many lines? How many points does each line contain? How many lines pass through each point?

Proposition. Through any two distinct points in a projective space there passes a projective line.

Proposition. Two distinct lines in a projective plane intersect in a unique point.

(Proofs by linear algebra.)

In general, if $P(U_1)$ and $P(U_2)$ are two projective subspaces of $P(V)$, then the intersection $P(U_1) \cap P(U_2)$ is the projective subspace $P(U_1 \cap U_2)$. The projective span or join of $P(U_1)$ and $P(U_2)$ is the projective subspace $P(U_1 + U_2)$.

Exercise. Show that a point is in the join of two projective subspaces if and only if it is on a line joining a point in one of the subspaces with a different point in the other. (Actually, there one a very degenerate case in which this is not true.)

From the dimension formula of linear algebra,

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2),$$

one obtains

$$\dim(P(U_1 + U_2)) = \dim P(U_1) + \dim P(U_2) - \dim(P(U_1) \cap P(U_2)).$$
Desargues’ theorem

**Theorem** (Desargues). Let \( A, A', B, B', C, C' \) be points of a projective plane such that the lines \( AA', BB' \) and \( CC' \) intersect in one point \( P \).

Then the intersection points \( C'' = AA' \cap BB', A'' = BB' \cap CC' \) and \( B'' = CC' \cap AA' \) lie on one line.

If this theorem is considered in an affine plane instead of a projective plane, it breaks up into several special cases which have to be considered separately. For example: **Under the same conditions, if \( AB \parallel A'B' \) and \( BC \parallel B'C' \) then \( CA \parallel C'A' \).**

Some preparation is necessary before the following proof. Let \( P(V) \) be an \( n \)-dimensional projective space. Then \( n + 2 \) points in \( P(V) \) are said to be in general position if one (hence both) of the following equivalent conditions are satisfied:

(i) No \( n + 1 \) of the points are contained in an \((n - 1)\)-dimensional projective subspace.

(ii) Any \( n + 1 \) of the points have linearly independent representative vectors.

So three points on a line are in general position if they are distinct, four points in a plane are in general position if no three of them lie on a line, and five points in a 3-dimensional projective space are in general position if no four of them lie in a plane.

**Lemma.** Let \( P(V) \) be an \( n \)-dimensional projective space and suppose \( P_1, \ldots, P_{n+2} \in P(V) \) are in general position. Then representative vectors \( v_1, \ldots, v_{n+1} \in V \) may be chosen so that

\[
v_1 + v_2 + \cdots + v_{n+1} = v_{n+2}.
\]

This choice is unique up to a common factor. That is, if \( \tilde{v}_1, \ldots, \tilde{v}_{n+1} \) is another choice of representative vectors with \( \tilde{v}_1 + \tilde{v}_2 + \cdots + \tilde{v}_{n+1} = \tilde{v}_{n+2} \), then \( \tilde{v}_k = \lambda v_k \) for some \( \lambda \neq 0 \).

**Proof of the lemma.** Let \( w_1, \ldots, w_{n+2} \) be any representative vectors for the points \( P_1, \ldots, P_{n+2} \). They are linearly dependent because \( \dim V = n + 1 \). So

\[
\sum_{j=1}^{n+2} a_j w_j = 0
\]

for some \( a_j \) which are not all zero. In fact, no \( a_k \) can be zero, because that would mean that there are \( n + 1 \) among the \( w_j \) which are linearly dependent. Hence we may choose

\[
v_1 = a_1 w_1, \quad v_2 = a_2 w_2, \quad \ldots \quad v_{n+1} = a_{n+1} w_{n+1}, \quad v_{n+2} = -a_{n+1} w_{n+1}.
\]

To see the uniqueness claim, suppose \( \lambda_1 v_1, \ldots, \lambda_{n+2} v_{n+2} \) is another choice of representative vectors with \( \sum_{k=1}^{n+1} \lambda_k v_k = v_{n+2} \). This amounts to a system of equations of rank \( n + 1 \) for the \( n + 2 \) variables \( \lambda_k \). So the solution space is 1-dimensional and hence \( \lambda_1 = \lambda_2 = \ldots = \lambda_{n+2} \).

**Proof of Desargues’ theorem.** If \( A, A', P \) are not distinct the statement of the theorem is obvious. (Check this.) So we may assume that \( A, A', P \) are distinct and also \( B, B', P \) and \( C, C', P \). But then \( A, A', P \) are three points on a line in general position. So by the lemma we may choose representative vectors \( a, a', p \in V \) with \( a + a' = p \). For the same reason we may also choose representative vectors \( b, b' \) and \( c, c' \) so that \( b + b' = p \) and \( c + c' = p \). Then

\[
a + a' = b + b' = c + c'.
\]

This implies \( a - b = b' - a' \). Obviously, the vector \( a - b = b' - a' \) is in the span of \( a \) and \( b \) and also in the span of \( a' \) and \( b' \). So the point \( [a - b] = [b' - a'] \in P(V) \) lies on the line \( AB \) and on the line \( A'B' \), hence it is the point of intersection, \( C'' \). Similarly, \( A'' = [b - c] = [c' - b'] \) and \( B'' = [c - a] = [a' - c'] \). But

\[
(a - b) + (b - c) + (c - a) = 0,
\]

which means that vectors \( (a - b), (b - c), (c - a) \) are linearly dependent and so they span a subspace of dimension at most 2. Therefore, \( C'', A'' \) and \( B'' \) lie on a line.
Desargues’ theorem says:

*If the lines joining corresponding points of two triangles meet in one point, then the intersections of corresponding sides lie on one line.*

The converse is also true:

*If the intersections of corresponding sides of two triangles lie on one line, then the lines joining corresponding points meet in one point.*

Surprisingly, the converse statement is in fact equivalent to the original statement after a permutation of the point labels (see figure right).

The Desargues configuration turns out to be very symmetric in the sense that there are many permutations of the labels $A, B, C, A', B', C', A'', B'', C''$ and $P$ preserving the relevant incidences.

Desargues’ theorem also holds for triangles in two different planes of a 3-dimensional projective space $P(V)$. In this case, it can be proved without any calculations: The intersection points of corresponding sides lie on the line in which the planes of the two triangles intersect.

The planar version of Desargues’ theorem can also be proved without any calculations if the third dimension is used:

**Proof** (a 3d proof of Desargues' theorem). Let $E$ be the plane of the two triangles $ABC$, $A'B'C'$ and the point $P$. Choose a line through $P$ which is not in $E$ and two points $X$ and $Y$ on it. The lines $XA$ and $YA'$ lie in one plane, so they intersect in a point $\tilde{A}$. Similarly, let $\tilde{B} = XB \cap YB'$ and $\tilde{C} = XC \cap YC'$. Now the intersection of the line $\tilde{A}B$ and the plane $E$ lies on the line $AB$, because the plane $X\tilde{A}\tilde{B}$ intersects $E$ in $AB$. Similarly, $\tilde{A}\tilde{B} \cap E$ also lies on the line $A'B'$, so $\tilde{A}\tilde{B} \cap E = C''$. In the same way, $\tilde{B}\tilde{C} \cap E = A''$ and $\tilde{C}\tilde{A} \cap E = B''$. Hence $A''$, $B''$ and $C''$ lie on the line where $E$ intersects the plane $\tilde{A}\tilde{B}\tilde{C}$. $\square$
The preceding proof also suggests the following 3-dimensional way to generate any planar Desargues configuration. This construction also reflects the high degree of combinatorial symmetry of the configuration. Let \( P_1, P_2, P_3, P_4, P_5 \) be five points in general position in a 3-dimensional projective space, and let \( E \) be a plane that contains none of these points. Let \( l_{ij} = l_{ji} \) be the 10 lines joining \( P_i \) and \( P_j \) (\( i \neq j \)). The 10 points \( P_{ij} \) where these lines intersect \( E \) form a Desargues configuration. If \((i, j, k, r, s)\) is any permutation of \((1, 2, 3, 4, 5)\), then the points \( P_{ij}, P_{jk}, P_{ki} \) always lie on a line (the intersection of the plane \( P_iP_jP_k \) with \( E \)), which we denote by \( g_{rs} \). Any one of the points \( P_{uv} \) lies on the line \( g_{xy} \) if the four indices \( uvxy \) are different. So there are three lines through each point and three points on each line. Corresponding points of the triangles \( P_{tr}, P_{jr}, P_{kr} \) and \( P_{ts}, P_{js}, P_{ks} \) are joined by the lines \( g_{jk}, g_{kj}, g_{ij}, \) which all pass through \( P_{rs} \). The intersection points of corresponding sides all lie on the line \( g_{rs} \).

The same Desargues figure contains therefore \( 5 \cdot 4 / 2 = 10 \) pairs of triangles satisfying the condition of Desargues’ theorem.

**Pappus’ theorem**

**Theorem** (Pappus). Let \( A, B, C \) be points on one line in a projective plane \( P(V) \), and let \( A', B', C' \) be points on another line. Then the points

\[
C'' = AB' \cap A'B, \quad A'' = BC' \cap B'C, \quad B'' = CA' \cap C'A
\]

lie on a line.

**Proof.** At most one of the points \( A', B', C' \) may lie on the line through \( A, B, C \), so we may assume without loss of generality that \( A, B, B', C' \) are in general position. Then we may choose representative vectors \( a, b, b', c' \in V \) with \( a + b + b' = c' \). Now \( a, b, b' \) form a basis for \( V \) in which the homogeneous coordinate vectors for \( A, B, B', C' \) are \( \left( \frac{1}{0} \right), \left( \frac{0}{1} \right), \left( \frac{0}{1} \right), \left( \frac{1}{1} \right) \). So we may assume without loss of generality that

\[
A = \left[ \begin{array}{c} 1 \\ 0 \end{array} \right], \quad B = \left[ \begin{array}{c} 0 \\ 1 \end{array} \right], \quad B' = \left[ \begin{array}{c} 0 \\ 1 \end{array} \right], \quad C' = \left[ \begin{array}{c} 1 \\ 1 \end{array} \right].
\]

Then

\[
C = \left[ \begin{array}{c} 1 \\ c \\ 0 \end{array} \right] \quad \text{and} \quad A' = \left[ \begin{array}{c} 1 \\ 1 \\ a \end{array} \right]
\]

for some \( c \) and \( a \). Now

\[
AB' = \left\{ \left[ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right] : x_2 = 0 \right\} \quad \text{and} \quad A'B = \left\{ \left[ \begin{array}{c} 1 \\ \frac{t}{a} \\ 0 \end{array} \right] : t \in \mathbb{R} \right\} \quad \text{so} \quad C'' = \left[ \begin{array}{c} 1 \\ 0 \\ a \end{array} \right],
\]

\[
BC' = \left\{ \left[ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right] : x_1 = x_3 = 0 \right\} \quad \text{and} \quad B'C = \left\{ \left[ \begin{array}{c} 0 \\ 1 \\ \frac{t}{a} \end{array} \right] : t \in \mathbb{R} \right\} \quad \text{so} \quad A'' = \left[ \begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right],
\]

\[
C'A = \left\{ \left[ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right] : x_2 = x_3 = 0 \right\} \quad \text{and} \quad CA' = \left\{ \left[ \begin{array}{c} 1 \\ 0 \\ \frac{t}{a} \end{array} \right] : t \in \mathbb{R} \right\} \quad \text{so} \quad B'' = \left[ \begin{array}{c} 1 \\ 1 \\ 1 \\ \frac{a + c - 1}{ac} \end{array} \right]
\]

(where \( B'' \) is obtained by a quick calculation on the side). Now

\[
(c - 1) \left( \frac{1}{a} \right) + a \left( \frac{c}{1} \right) - \left( \frac{a + c - 1}{ac} \right) = 0,
\]

so the vectors are linearly dependent and hence \( A'' \), \( B'' \), \( C'' \) lie on a line. \( \Box \)
The synthetic approach to projective geometry

We have defined a projective space $P(V)$ of a vector space $V$ over a field as the set of 1-dimensional subspaces of $V$. Of course this definition is based on basic axioms of algebra: the field axioms and the vector space axioms. This section provides a rough outline of how the theory of projective spaces $P(V)$ can be based on geometric axioms.

The following definition of a projective space in terms of geometric axioms (due to Oswald Veblen & John W. Young, 1908) is not equivalent to our definition of a projective space of a vector space over a field. (One obtains an equivalent definition if Pappus’ theorem is added as an independent axiom; see the structure theorem below.)

A projective space $P = (\mathcal{P}, \mathcal{L})$ is a set $\mathcal{P}$, the elements of which are called points, together with a set $\mathcal{L}$ of subsets of $\mathcal{P}$, which are called lines, such that the following axioms are satisfied.

**Axiom 1.** For any two distinct points there exists a unique line which contains both points.

**Axiom 2.** If $A, B, C \in P$ are three distinct points and $l \in \mathcal{L}$ is a line that intersects the lines $AB$ and $AC$ in distinct points, then $l$ intersects the line $BC$.

**Axiom 3.** Every line contains at least three points.

Axiom 1 implies that two lines intersect in at most one point. Axiom 2 is a clever way of saying that two lines in one plane always intersect without first defining what a plane is.

A projective subspace of $P$ is a subset $\mathcal{U} \subseteq \mathcal{P}$ of points such that the line through any two points of $\mathcal{U}$ is contained in $\mathcal{U}$. Together with the subset of lines $\{l \in \mathcal{L} | l \subseteq U\}$, the subspace is a projective space in its own right. The intersection of two projective subspaces is a projective subspace.

If $S \subseteq \mathcal{P}$ is any set of points, then the projective span of $S$ is the smallest projective subspace containing $S$, or equivalently, the intersection of all projective subspaces which contain $S$.

The dimension of the projective space $P$ is the smallest number $n$ for which there exist $n + 1$ points $P_1, \ldots, P_{n+1} \in \mathcal{P}$ such that $\mathcal{P}$ is the projective span of $\{P_1, \ldots, P_{n+1}\}$.

The Axioms 1–3 together with the assertion that the dimension of $P$ is 2 are equivalent to the following axioms for a projective plane. (Can you prove this equivalence? It is a little tricky.)

**Axiom P1.** Same as Axiom 1.

**Axiom P2.** Any two lines have non-empty intersection.

**Axiom P3.** Same as Axiom 3.

**Axiom P4.** There are at least two different lines.

If the dimension of $P$ is at least 3, then Desargues’ theorem can be deduced from Axioms 1–3. The 3D proof of the last lecture works in this setting, it uses only the incidence relations between points, lines and planes, and does not involve any calculations. However, there are projective planes in which Desargues’ theorem does not hold. The purely 2-dimensional proof does not work here because it is based on calculations.

A projective plane in which Desargues’ theorem holds is called a Desarguesian plane.

**Theorem** (Veblen & Young). Any projective space in which Desargues’ theorem holds (that is, any projective space of dimension $\geq 3$ and any Desarguesian plane) is isomorphic to a projective space $P(V)$ of a vector space $V$ over a skew field $F$. If Pappus’ theorem also holds in $P$, then $F$ is a field.

(Two projective spaces are isomorphic if there is a bijection between their points that maps lines to lines. A skew field satisfies all field axioms except that the multiplication may not be commutative. You may check that our computational proof of Pappus’ theorem does not work if multiplication is not commutative.)

In a projective plane, Desargues’ theorem can be deduced from Pappus’ theorem. (This was demonstrated by Hessenberg in 1905. It is not obvious at all.)

Thus, any theorem that holds in any projective space $P(V)$ of a vector space $V$ over a filed can also be deduced from Axioms 1–3 together with Pappus’ theorem as independent axiom, and vice versa. Further axioms of order and of continuity have to be added to single out the real projective spaces $\mathbb{R}P^n$ (just like further axioms have to be added to the general field axioms to single out field of reals).
Projective transformations

Let $V, W$ be $(n + 1)$-dimensional vector spaces, and let $f : V \to W$ be an invertible linear map. Since $f$ maps any 1-dimensional subspace $[v] \subseteq V$ to a 1-dimensional subspace $[f(v)] \subseteq W$, it defines an invertible map $P(V) \to P(W)$. A map between projective spaces which arises in this way is called a projective transformation:

A map $\hat{f} : P(V) \to P(W)$ is a projective transformation if there is an invertible linear map $f : V \to W$ such that $\hat{f}([v]) = [f(v)]$.

A projective transformation maps lines in $P(V)$ to lines in $P(W)$ and generally $k$-planes to $k$-planes.

In homogeneous coordinates, a projective transformation is represented by matrix multiplication:

A point in $P(V)$ with homogeneous coordinates $x = \left( \frac{x_1}{\cdots} \right)$ is mapped to the point in $P(W)$ with homogeneous coordinates $y = Ax$ for some invertible $(n + 1) \times (n + 1)$ matrix $A$. In affine coordinates $u_i = x_i/x_{n+1}$, $w_i = y_i/y_{n+1}$ ($i = 1, \ldots, n$) the map is a so-called fractional linear transformation:

$$w_i = \frac{\sum_{j=1}^n a_{ij} u_j + a_{i,n+1}}{\sum_{j=1}^n a_{n+1,j} u_j + a_{n+1,n+1}}.$$ 

Each $w_i$ is the quotient of two affine linear functions of the $u_j$, where the denominator is the same for all $i$.

**Proposition.** Two invertible linear maps $f, g : V \to W$ give rise to the same projective transformation $P(V) \to P(W)$ if and only if $g = \lambda f$ for some scalar $\lambda \neq 0$.

**Proof.** “$\Rightarrow$” If $g = \lambda f$, then $[g(v)] = [\lambda f(v)] = [f(v)]$.

“$\Leftarrow$” Suppose $[g(v)] = [f(v)]$ for all $v \in V \setminus \{0\}$. This implies $g(v) = \lambda(v)f(v)$ for some non-zero scalar $\lambda(v)$ which may a priori depend on $v$. We have to show that it does not. Suppose $v, w \in V \setminus \{0\}$. If $v, w$ are linearly dependent, then it is obvious from the definition of $\lambda(v)$ that $\lambda(v) = \lambda(w)$. So assume $v, w$ are linearly independent. Now

$$g(v + w) = g(v) + g(w) = \lambda(v) f(v) + \lambda(w) f(w)$$

but also

$$g(v + w) = \lambda(v + w) f(v + w) = \lambda(v + w)(f(v) + f(w)).$$

Since $f(v)$ and $f(w)$ are also linearly independent this implies $\lambda(v) = \lambda(v + w) = \lambda(w)$.

The projective transformations $P(V) \to P(V)$ form a group called the projective linear group $PGL(V)$. It is the quotient of the general linear group $GL(V)$ of invertible linear maps $V \to V$ by the normal subgroup of non-zero multiples of the identity: $PGL(V) = GL(V)/\{\lambda I\}_{\lambda \neq 0}$.

**Example 1.** Affine transformations

Consider an affine transformation $\mathbb{R}^n \to \mathbb{R}^n, u \mapsto Mu + b$ with $M \in GL(n, \mathbb{R})$ and $b \in \mathbb{R}^n$. In homogeneous coordinates $x_1, \ldots, x_{n+1}$ with $u_i = \frac{x_i}{x_{n+1}}$, this map can be written

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{pmatrix} \mapsto \begin{pmatrix} M \\ 0 & \cdots & 0 \\ 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{pmatrix} = Ax$$

The affine map $x \mapsto Mx + b$ can be extended to the projective transformation

$$\mathbb{RP}^n \to \mathbb{RP}^n, \ [x] \mapsto [Ax].$$

It maps the plane at infinity $x_{n+1} = 0$ to the plane at infinity. Conversely, any projective transformation which maps the plane $x_{n+1} = 0$ to itself corresponds in the affine coordinates $u$ to an affine map $\mathbb{R}^n \to \mathbb{R}^n$. 

30
Example 2. Central projection

Let $E_1 = P(U_1)$ and $E_2 = P(U_2)$ be two hyperplanes in a projective space $P(V)$ (for example, two lines in a projective plane) and let $W \in P(V)$ be a point not in $E_1$ or $E_2$. Then the central projection from $E_1$ to $E_2$ with center $W$ is the map $\hat{f} : E_1 \to E_2$ that maps a point $A \in E_1$ to the intersection of $E_2$ with the line through $W$ and $A$.

**Proposition.** The central projection $\hat{f}$ is a projective transformation $E_1 \to E_2$.

**Proof.** We have to show that $\hat{f}$ comes from an invertible linear map $f : U_1 \to U_2$. Note that $W$, as point in $P(V)$, is a 1-dimensional subspace of $V$. Since it does not lie in $E_2$, $W \cap U_2 = \{0\}$. This means that $V$ is the direct sum $V = W \oplus U_2$, and there are two linear maps $p_W : V \to W$ and $p_{U_2} : V \to U_2$ (the projections onto $W$ and $U_2$) such that for any $v \in V$, $p_W(v)$ and $p_{U_2}(v)$ are the unique vectors in $W$ and $U_2$ such that $v = p_W(v) + p_{U_2}(v)$.

Claim: The central projection $\hat{f}$ comes from the linear map $p_{U_2}|_{U_1}$, the restriction of $p_{U_2}$ to $U_1$.

To see this, let $a \in U_1$ be a representative vector of $A \in E_1$. Then $p_{U_2}(a) \neq 0$, because $p_{U_2}(a) = 0$ would mean $a \in W$, but $U_1 \cap W = \{0\}$ because by assumption $E_1$ does not contain $W$. This shows that $p_{U_2}|_{U_1}$ is invertible, because it an injective linear map $U_1 \to U_2$ and $\dim U_1 = \dim U_2$. Now $p_{U_2}(a) \in U_2$, so $p_{U_2}(a) \in E_2$. Also $a = p_W(a) + p_{U_2}(a)$, or $p_{U_2}(a) = a - p_W(a)$, so $p_{U_2}(a) \in [a] + W$, which means that $[p_{U_2}(a)]$ is in the (projective) line through $A \in P(V)$ and $W \in P(V)$. Hence $[p_{U_2}(a)]$ is the intersection of $E_2$ with the line through $W$ and $A$, so it is the image of $A$ under the central projection.

One can consider also more general types of projections. For example let $l_1$ and $l_2$ be two lines in a 3-dimensional projective space $P(V)$, and let $l_0$ be a line that does not intersect $l_1$ or $l_2$. Then the projection $l_1 \to l_2$ with the line $l_0$ as center of projection is defined as follows: A point $A \in l_1$ is mapped to the intersection of $l_2$ with the plane spanned by $l_0$ and $A$. This map $l_1 \to l_2$ is also a projective transformation, and the proof is the same (apart from obvious modifications).

Most generally, in an $n$-dimensional projective space $P(V)$, one can project one $k$-plane $E_1$ to another $k$-plane $E_2$ from any $(n - k - 1)$-plane $E_C$ which does not intersect $E_1$ or $E_2$ as center of projection; and this is a projective transformation $E_1 \to E_2$.

**Theorem.** Let $P(V)$ and $P(W)$ be two $n$-dimensional projective spaces and suppose $A_1, \ldots, A_{n+2} \in P(V)$ and $B_1, \ldots, B_{n+2} \in P(W)$ are in general position. Then there exists a unique projective transformation $\hat{f} : P(V) \to P(W)$ with $\hat{f}(A_i) = B_i$ for $i = 1, \ldots, n + 2$.

**Proof.** Existence: By the lemma of Lecture 13, we may choose representative vectors $a_1, \ldots, a_{n+2}$ for $A_1, \ldots, A_{n+2}$ and $b_1, \ldots, b_{n+2}$ for $B_1, \ldots, B_{n+2}$ such that $\sum_{i=1}^{n+1} a_i = a_{n+2}$ and $\sum_{i=1}^{n+1} b_i = b_{n+2}$. Also by the general position assumption, $a_1, \ldots, a_{n+1}$ and $b_1, \ldots, b_{n+1}$ are bases of $V$ and $W$, respectively. Hence there is an invertible linear map $f : V \to W$ with $f(a_i) = b_i$ for $i = 1, \ldots, n + 1$. But then also $f(a_{n+2}) = f(\sum_{i=1}^{n+1} a_i) = \sum_{i=1}^{n+1} f(a_i) = \sum_{i=1}^{n+1} b_i = b_{n+2}$.

So $f$ maps the 1-dimensional subspaces $A_i = [a_i] \subseteq V$ to $B_i = [b_i] \subseteq W$ for $i = 1, \ldots, n + 2$.

Uniqueness: Let $g : V \to W$ be another invertible linear map with $g(a_i) \in B_i$ for $i = 1, \ldots, n + 2$. Then $b_i = g(a_i)$ would be another set of representative vectors for the $B_i$ with $\tilde{b}_{n+2} = g(a_{n+2}) = g(\sum_{i=1}^{n+1} a_i) = \sum_{i=1}^{n+1} g(a_i) = \sum_{i=1}^{n+1} b_i$.

By the uniqueness part of the lemma from Lecture 13, this implies $b_i = \lambda b_i$ for some $\lambda \neq 0$, so $g = \lambda f$, and $g$ and $f$ induce the same projective transformation $P(V) \to P(W)$.
Theorem. Let \( l_1, l_2 \) be two different lines in a projective plane \( P(V) \). A projective transformation \( l_1 \to l_2 \) is a central projection if and only if it maps the intersection \( l_1 \cap l_2 \) to itself. Otherwise it is the composition of two central projections \( l_1 \to l \to l_2 \).

Proof.

One can classify projective transformations \( \hat{f} : P(V) \to P(V) \) according to the normal forms of the corresponding linear maps \( f : V \to V \). Note that fixed points of \( \hat{f} \) correspond to 1-dimensional eigenspaces of \( f \). Let us consider a projective transformation \( \hat{f} : \mathbb{R}P^1 \to \mathbb{R}P^1 \) corresponding to the linear map \( f : \mathbb{R}^2 \to \mathbb{R}^2 \). The following cases are possible:

1. \( f \) has two distinct real eigenvalues. Then it has two linearly independent eigenvectors. So the projective transformation \( \hat{f} \) has two fixed points. The eigenvectors of \( f \) form a basis of \( \mathbb{R}^2 \) in which \( f \) has the matrix \( \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \). In the affine coordinate \( u \) of \( \mathbb{R}P^1 \) corresponding to this basis, the fixed points are 0 and \( \infty \) and the projective map \( \hat{f} \) is the scaling transformation \( u \mapsto \frac{\lambda_1}{\lambda_2} u \). In other affine coordinates, \( \hat{f} \) looks like the projected image of a scaling transformation.

2. \( f \) has two linearly independent eigenvectors to the same eigenvalue. Then \( f \) is a multiple of the identity and \( \hat{f} \) is the identity.

3. \( f \) has a double eigenvalue but only one linearly independent eigenvector, so \( \hat{f} \) has only one fixed point. Then there is a basis in which the matrix of \( f \) has the Jordan normal form \( \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} = \lambda \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \). In the affine coordinate \( u \) corresponding to this basis, the fixed point is \( \infty \) and \( \hat{f} \) is the translation \( u \mapsto u + \frac{1}{\lambda} \). In another choice of affine coordinate, \( \hat{f} \) looks like a projected image of a translation.

4. \( f \) has a pair of complex conjugate eigenvalues. Then there is a basis of \( \mathbb{R}^2 \) in which the matrix of \( f \) has the form \( \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \sqrt{a^2 + b^2} \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix} \) for some \( \varphi \), and we may forget about the factor \( \sqrt{a^2 + b^2} \). So \( \hat{f} \) comes from a linear map which looks in some basis of \( \mathbb{R}^2 \) like a rotation.
A proof of Pappus' theorem using central projections

Let $X = AC \cap C'A'$. We have to show $X = B''$. Consider the sequence of central projections

$$AC' \xrightarrow{A'} I \xrightarrow{B'} BC' \xrightarrow{C''} AC'',$$

where the centers of the projections are written above the arrows. They map

$$A \mapsto A \mapsto R \mapsto A \\
Q \mapsto B \mapsto B \mapsto Q \\
B'' \mapsto C \mapsto A'' \mapsto X \\
C'' \mapsto P \mapsto C' \mapsto C''.$$

So the composition is a projective transformation $AC'' \rightarrow AC'$ which fixes the three points $A, Q, C'$, so it must be the identity. It also maps $B''$ to $X$, so $B'' = X$. □

The cross ratio

The cross ratio of four distinct points $P_i = [\frac{x_i}{y_i}] \in \mathbb{RP}^1$ $(i = 1, \ldots, 4)$ is the number

$$\text{cr}(P_1, P_2, P_3, P_4) := \frac{\det(\frac{x_1}{y_1}, \frac{x_2}{y_2}) \det(\frac{x_3}{y_3}, \frac{x_4}{y_4})}{\det(\frac{x_1}{y_1}, \frac{x_3}{y_3}) \det(\frac{x_2}{y_2}, \frac{x_4}{y_4})} = \frac{(x_1y_2 - x_2y_1)(x_3y_4 - x_4y_3)}{(x_2y_3 - x_3y_2)(x_4y_1 - x_1y_4)}.$$

It is well defined because the expressions on the right hand side do not depend on the choice of representative vectors $(x_i, y_i)$ but only on the points $P_i$.

To derive an expression for the cross ratio in terms of the affine coordinate $u = \frac{x}{y}$, we assume at first that no $y_i$ is 0 so that no $u_i$ is infinite:

$$\text{cr}(P_1, P_2, P_3, P_4) = \frac{y_1y_2(x_2^{-1} - x_3^{-1})y_3y_4(x_3^{-1} - x_4^{-1})}{y_2y_3(x_2^{-1} - x_3^{-1})y_4y_1(x_4^{-1} - x_3^{-1})} = \frac{(u_1 - u_2)(u_3 - u_4)}{(u_2 - u_3)(u_4 - u_1)} := \text{cr}(u_1, u_2, u_3, u_4).$$

Even if one of the $u_i$ is infinity, one gets correct results using this formula if one “cancels infinities”. For example, if $y_1 = 0$ so that $u_1 = \infty$, one has

$$\text{cr}(P_1, P_2, P_3, P_4) = \frac{x_1y_2(x_1^{-1} - x_4^{-1})}{x_2y_3(-x_1^{-1})} = \frac{x_1y_2y_3y_4(x_3^{-1} - x_4^{-1})}{y_2y_3(-x_1^{-1})} = \frac{-u_3 - u_4}{u_2 - u_3},$$

so the following calculation gives the correct result:

$$\text{cr}(\infty, u_2, u_3, u_4) = \frac{(\infty - u_2)(u_3 - u_4)}{(u_2 - u_3)(u_4 - \infty)} = \frac{-u_3 - u_4}{u_2 - u_3}.$$  

**Proposition.** (i) If $f : \mathbb{RP}^1 \rightarrow \mathbb{RP}^1$ is a projective transformation and $f(P_i) = Q_i$, then

$$\text{cr}(P_1, P_2, P_3, P_4) = \text{cr}(Q_1, Q_2, Q_3, Q_4).$$

(ii) If $v, w$ is any basis of $\mathbb{R}^2$ and $\tilde{x}, \tilde{y}$ are homogeneous coordinates for $P_i$ in this basis (that is, $P_i = [\tilde{x}, v + \tilde{y}, w]$) and $\tilde{u} = \frac{x}{y}$ is the corresponding affine coordinate, then one may just as well use these coordinates to compute the cross-ratio:

$$\text{cr}(P_1, P_2, P_3, P_4) = \frac{(\tilde{x}_1\tilde{y}_2 - \tilde{x}_2\tilde{y}_1)(\tilde{x}_3\tilde{y}_4 - \tilde{x}_4\tilde{y}_3)}{(\tilde{x}_2\tilde{y}_3 - \tilde{x}_3\tilde{y}_2)(\tilde{x}_4\tilde{y}_1 - \tilde{x}_1\tilde{y}_4)} = \frac{(\tilde{u}_1 - \tilde{u}_2)(\tilde{u}_3 - \tilde{u}_4)}{(\tilde{u}_2 - \tilde{u}_3)(\tilde{u}_4 - \tilde{u}_1)}.$$

All this works not only for the real projective line $\mathbb{RP}^1$ but also for the complex projective line $\mathbb{CP}^1$ and any other projective space $P(V)$ of a 2-dimensional vector space $V$ over any field.

If $v, w$ is a basis of $V$, then the cross ratio of four points $P_i$ with homogeneous coordinates $\tilde{x}, \tilde{y}$ in this basis is defined defined by the equation above, and this is independent of the choice of basis.
Proposition. The cross-ratio \( cr(P_1, P_2, P_3, P_4) \) is the affine coordinate of the image of \( P_1 \) under the projective transformation that maps \( P_2, P_3, P_4 \) to the points with affine coordinates 0, 1, \( \infty \).

Corollary. The cross ratio of four distinct points can take all values except 0, 1, \( \infty \).

Proposition. There exists a projective transformation that maps four distinct points \( P_1, P_2, P_3, P_4 \) of a line to four distinct points \( Q_1, Q_2, Q_3, Q_4 \) on the same or another line if and only if
\[
    cr(P_1, P_2, P_3, P_4) = cr(Q_1, Q_2, Q_3, Q_4).
\]

The cross ratio depends on the order of the points. How does it change if the points are permuted?

- The cross ratio does not change if I simultaneously interchange two of the points and the remaining two:
\[
    cr(u_1, u_2, u_3, u_4) = cr(u_2, u_1, u_4, u_3) = cr(u_3, u_4, u_1, u_2) = cr(u_4, u_3, u_2, u_1).
\]

This is easy to see from the equation for the cross ratio in terms of the \( u_i \).

- Of the 24 permutations of \( u_1, u_2, u_3, u_4 \), I need therefore only consider the six which fix \( u_1 \) and permute \( u_2, u_3, u_4 \).

- If \( i, j, k, l \) is a permutation of 1, 2, 3, 4, then
\[
    cr(u_1, u_2, u_3, u_4) = cr(v_1, v_2, v_3, v_4) \iff cr(u_i, u_j, u_k, u_l) = cr(v_i, v_j, v_k, v_l).
\]

Indeed, there is a projective transformation that maps \( u_1, u_2, u_3, u_4 \) to \( v_1, v_2, v_3, v_4 \) if and only if there is one that maps \( u_i, u_j, u_k, u_l \) to \( v_i, v_j, v_k, v_l \). (It’s the same map in both cases.)

- Hence if \( cr(u_1, u_2, u_3, u_4) = q = cr(q, 0, 1, \infty) \), then
\[
    cr(u_1, u_3, u_2, u_4) = cr(q, 1, 0, \infty) = \frac{(q - 1)(0 - \infty)}{(1 - 0)(\infty - q)} = 1 - q,
\]
\[
    cr(u_1, u_2, u_4, u_3) = cr(q, 0, \infty, 1) = \frac{(q - 0)(\infty - 1)}{(0 - \infty)(1 - q)} = \frac{q}{q - 1},
\]
\[
    cr(u_1, u_4, u_2, u_3) = cr(q, \infty, 1, 0) = \frac{(q - \infty)(1 - 0)}{(\infty - 1)(0 - q)} = \frac{1}{q},
\]
\[
    cr(u_1, u_3, u_4, u_2) = cr(q, 1, \infty, 0) = \frac{(q - 1)(\infty - 0)}{(1 - \infty)(0 - q)} = \frac{q - 1}{q} = 1 - \frac{1}{q},
\]
\[
    cr(u_1, u_4, u_2, u_3) = cr(q, \infty, 0, 1) = \frac{(q - \infty)(0 - 1)}{(\infty - 0)(1 - q)} = \frac{1}{1 - q}.
\]
Projective involutions of the real projective line

For any four points $A, B, C, D \in \mathbb{R}P^1$ there is a unique projective transformation $f : \mathbb{R}P^1 \rightarrow \mathbb{R}P^1$ with $f(A) = B$, $f(B) = A$, $f(C) = D$, $f(D) = C$, because $cr(A, B, C, D) = cr(B, A, D, C)$. The transformation $f$ is an involution, that is, $f \neq \text{identity}$ but $f \circ f = \text{identity}$.

A pair of points $\{A, B\} \subset \mathbb{R}P^1$ separates another pair $\{C, D\} \subset \mathbb{R}P^1$ if $C$ and $D$ are in different connected components of $\mathbb{R}P^1 \setminus \{A, B\}$.

$\{A, B\}$ separates $\{C, D\} \iff cr(A, C, B, D) < 0$.

The involution $f$ has no fixed points if $\{A, B\}$ separates $\{C, D\}$, otherwise it has two fixed points. If $f$ has two fixed points $P$ and $Q$, then for all $X \in \mathbb{R}P^1$, $cr(X, P, f(X), Q) = -1$.

For any two points $P, Q \in \mathbb{R}P^1$ there is a unique projective involution of $\mathbb{R}P^1$ that fixes $P$ and $Q$. If $A, B, P, Q$ are four points in $\mathbb{R}P^1$, then one says the pair $\{A, B\}$ separates the pair $\{P, Q\}$ harmonically, if $cr(A, P, B, Q) = -1$.

The complete quadrilateral

**Theorem** (on the complete quadrilateral). Let $A, B, C, D \in \mathbb{R}P^2$ be four points in general position, let $P = AB \cap CD$, $Q = AD \cap BC$, $\ell = PQ$, $X = \ell \cap BD$, $Y = \ell \cap AC$. Then the pair of points $\{P, Q\}$ on $\ell$ separates the pair $\{X, Y\}$ harmonically:

$$cr(P, X, Q, Y) = -1.$$ 

Here are two proofs for this theorem, one computational, one using a projective involution of the plane.

**Proof** (by computation). Since the points are in general position, there is a projective transformation of $\mathbb{R}P^2$ that maps $A, B, C, D$ to

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}.$$ 

If the theorem holds for the projected figure, it holds also for the original one. It is thus enough to verify the theorem for

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \quad D = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$ 

In this case, $\ell$ is the line $x_3 = 0$, and $P = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $Q = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$, $X = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $Y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Homogeneous coordinates on $\ell$ are obtained by dropping the $x_3$-coordinate (which is zero), so

$$cr(P, X, Q, Y) = \frac{\det(1, -1, 0) \det(0, 1, 1)}{\det(1, -1, 0) \det(1, 0, 1)} = -1.$$
**Proof** (using a projective involution of \( \mathbb{RP}^2 \)). Since \( A, B, C, D \) are in general position, there is a projective transformation of the plane \( \mathbb{RP}^2 \) that maps \( A \mapsto B, B \mapsto A, C \mapsto D, D \mapsto C \). It is an involution of \( \mathbb{RP}^2 \) which maps the lines \( AB \) and \( CD \) onto themselves. It maps the line \( AD \) to \( BC \) and vice versa. Hence, the points \( P \) and \( Q \) are fixed, and the line \( \ell \) is mapped to itself. Since the line \( AC \) is mapped onto \( BD \) and vice versa, \( X \) is mapped to \( Y \) and \( Y \) to \( X \). Thus, the restriction to \( \ell \) is an involution of \( \ell \) with fixed points \( P, Q \) and interchanging \( X, Y \). So \( \text{cr}(P, X, Q, Y) = -1 \).

*Projective involutions of the real projective plane*

Suppose \( f : \mathbb{RP}^2 \to \mathbb{RP}^2 \) is a projective involution of the real projective plane. Let \( A \in \mathbb{RP}^2 \) be a point which is not a fixed point, and \( A' = f(A) \). Then also \( A = f(A') \) and hence the line \( AA' \) is mapped to itself.

Let \( B \in \mathbb{RP}^2 \) be a point not on this line which is also not a fixed point, and \( B' = f(B) \). Then the line \( BB' \) is also mapped to itself, so \( P = AA' \cap BB' \) is a fixed point of \( f \).

The restriction of \( f \) to the line \( AA' \) is an involution of \( AA' \) with a fixed point \( P \), so it has another fixed point \( Q \), and this is the point such that \( \{P, Q\} \) separates \( \{A, A'\} \) harmonically. Equally, the restriction of \( f \) to the line \( BB' \) is an involution of \( BB' \) with fixed points \( P \) and \( R \) such that \( \{P, R\} \) separates \( \{B, B'\} \) harmonically. Now \( f \) fixes every point on the line \( \ell = QR \). (Can you see why?) Thus:

Any projective involution of \( \mathbb{RP}^2 \) has a whole line \( \ell \) of fixed points and another fixed point \( P \not\in \ell \).

Conversely, if \( \ell \) is a line in \( \mathbb{RP}^2 \) and \( P \) is a point not on \( \ell \), then there is a unique projective involution \( f \) that fixes \( P \) and every point on \( \ell \). This is the projective reflection on \( \ell \) and \( P \).

Indeed if \( X, Y \) are any two points on \( \ell \), and any representative vectors of \( P, X, Y \) are chosen as basis of \( \mathbb{R}^3 \), then the matrix of \( f \) in this basis must be \( \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \).

(What does this reflection look like in an affine chart in which \( \ell \) is the line at infinity? What does it look like if \( P \) is a point at infinity?)

---

**Happy holidays!**

artwork: Franz Pedit  
creative directors: Ilya & Lysander
The fundamental theorem of real projective geometry

We know that projective transformations map lines to lines. For real projective spaces they are in fact all transformations that map lines to lines:

**Fundamental theorem of real projective geometry.** If a bijective map \( \mathbb{R}^n \to \mathbb{R}^n \) maps lines to lines, then it is a projective transformation.

**Remarks.**

- Note that it is not necessary to assume that the map is continuous.
- A map \( \mathbb{R}^n \to \mathbb{R}^n \) maps lines to lines if and only if it maps \( k \)-planes to \( k \)-planes. (Why?)
- The corresponding statement for \( \mathbb{C}P^n \) is false. For example, the map \( \mathbb{C}P^n \to \mathbb{C}P^n \),

\[
\begin{bmatrix}
    z_1 \\
    \vdots \\
    z_{n+1}
\end{bmatrix}
\mapsto
\begin{bmatrix}
    \bar{z}_1 \\
    \vdots \\
    \bar{z}_{n+1}
\end{bmatrix}
\]

is bijective and maps lines to lines, but it is not a projective transformation. The fundamental theorem for general projective spaces \( P(V) \) of a vector space \( V \) over a field \( F \) says the following:

If \( f : P(V) \to P(V) \) is bijective and maps lines to lines, then \( f \) comes from an almost linear map \( \varphi : V \to V \). A map \( \varphi : V \to W \) between vector spaces over \( F \) is called **almost linear** if for all \( u, v \in V \),

\[
\varphi(u + v) = \varphi(u) + \varphi(v) \quad \text{and} \quad \varphi(\lambda v) = \alpha(\lambda)\varphi(v),
\]

where \( \alpha : F \to F \) is a field automorphism. (For example, complex conjugation is an automorphism of \( \mathbb{C} \). The field \( \mathbb{R} \) of real numbers has no automorphism except the identity.)

**Corollary.** A bijective map \( f : \mathbb{R}^n \to \mathbb{R}^n \) which maps lines to lines is an affine transformation \( f(x) = Ax + b \) for some \( A \in GL(n, \mathbb{R}) \), \( b \in \mathbb{R}^n \).

(Because any such map \( \mathbb{R}^n \to \mathbb{R}^n \) can be extended to a bijective map \( \mathbb{R}P^n \to \mathbb{R}P^n \) which maps lines to lines and the hyperplane at infinity to itself. (How?))

For simplicity, I will present a proof of the fundamental theorem only for the case \( n = 2 \) of the real projective plane. This already contains all the important ideas, so you can figure out for yourself how it works for \( n > 2 \). The proof depends on the following two lemmas.

**Lemma.** Suppose a map \( f : \mathbb{R}^2P \to \mathbb{R}^2P \) is bijective and maps lines to lines. Then if \( A, B, C, D \) are four points on a line in \( \mathbb{R}^2P \) and \( \text{cr}(A, B, C, D) = -1 \), then \( \text{cr}(f(A), f(B), f(C), f(D)) = -1 \) as well.

**Proof.** Use the theorem on the complete quadrilateral. \( \square \)

**Lemma.** Suppose a map \( f : \mathbb{R}P^1 \to \mathbb{R}P^1 \) of the line has the property that if \( A, B, C, D \in \mathbb{R}P^1 \) are four points with \( \text{cr}(A, B, C, D) = -1 \), then also \( \text{cr}(f(A), f(B), f(C), f(D)) = -1 \). Then \( f \) is a projective transformation.

**Proof.** We will show that if \( f \) also fixes 0, 1, and \( \infty \), it must be the identity. This implies the lemma: For general \( f \) let \( g \) be the projective transformation that maps \( f(0) \mapsto 0, f(1) \mapsto 1, f(\infty) \mapsto \infty \). Then the composition \( g \circ f \) satisfies the assumptions of the theorem and fixes 0,1,\( \infty \). If it is the identity, then \( f = g^{-1} \) is a projective transformation.

So assume in addition that \( f \) fixes 0, 1, \( \infty \). Then for all \( x, y \in \mathbb{R} \):  

1. \( f(x+y) = f(x) + f(y) \), because \( \text{cr}(x, x+y, y, \infty) = -1 \).
2. \( f(2x) = 2f(x) \), because \( \text{cr}(0, x, 2x, \infty) = -1 \).
3. \( f(x+y) = f(x) + f(y) \). This follows from (1) and (2).
4. \( f(-x) = -f(x) \) because \( 0 = f(0) = f(x + (-x)) = f(x) + f(-x) \).
5. \( f(nx) = nf(x) \) for \( n \in \mathbb{Z} \). This follows from (3) and (4).
6. \( f(qx) = qf(x) \) for \( q \in \mathbb{Q} \). This follows from (5).
7. \( f(q) = q \) for \( q \in \mathbb{Q} \) because \( f(q) = f(q - 1) = qf(1) = q \cdot 1 \).
8. \( f(x^2) = (f(x))^2 \). This follows from (4) and \( \text{cr}(x, 1, x, x^2) \).
9. \( x > 0 \Rightarrow f(x) > 0 \). This follows from (8) because the a real number is positive if and only if it is the square of a real number.
10. \( f \) is increasing on \( \mathbb{R} \). This follows from (3,4,9) because  

\[
0 < x - y \implies 0 < f(x - y) = f(x) - f(y).
\]

Finally: An increasing function on \( \mathbb{R} \) which fixes the rationals is the identity. (Why?) \( \square \)
**Proof** (of the fundamental theorem, $n = 2$). We will show that if $f$ also fixes the four points

$$P_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad P_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

it must be the identity. This implies the theorem: For general $f$ let $g : \mathbb{RP}^2 \to \mathbb{RP}^2$ be the projective transformation that maps $f(P_1)$ to $P_1$. Then the composition $g \circ f$ is bijective, maps lines to lines and fixes the points $P_i$. If it is the identity, then $f = g^{-1}$ is a projective transformation.

So assume that $f : \mathbb{RP}^2 \to \mathbb{RP}^2$ is bijective, maps lines to lines and fixes $P_1, P_2, P_3, P_4$. Let $X \in \mathbb{RP}^2$ be any point not on the line $P_1 P_2$ (which we consider as the line at infinity). We will show that $f(X) = X$.

Let $\ell_1 = P_3 P_4$ and $\ell_2 = P_3 P_4$. Since $f$ fixes these points, it maps $\ell_1$ to $\ell_1$ and $\ell_2$ to $\ell_2$. By the lemmas, the restrictions $f|_{\ell} : \ell \to \ell$ are projective transformations. But $f|_{\ell_1}$ fixes $P_1, P_2,$ and $E_1 = P_2 P_4 \cap \ell_1$, so it is the identity. Equally, $f|_{\ell_2}$ fixes $P_3, P_4$, and $E_2 = P_1 P_4 \cap \ell_2$, so it is the identity. Hence $f$ fixes also $X_1 = P_2 X \cap \ell_1$ and $X_2 = P_1 X \cap \ell_2$. Since $f$ maps lines to lines, $X = X_1 P_2 \cap X_2 P_1$ implies $f(X) = f(X_1) f(P_2) \cap f(X_2) f(P_1) = X_1 P_2 \cap X_2 P_1 = X$.

We have shown that $f(X) = X$ for all $X$ not on $P_1 P_2$. But then it also fixes all points on $P_1 P_2$. (Why?) Hence, $f$ is the identity.

**Localized version of the fundamental theorem.** Let $U$ be a subset of $\mathbb{RP}^n$ that contains an open ball $B \subset \mathbb{R}^n \subset \mathbb{RP}^n$. Suppose an injective map $f : U \to \mathbb{RP}^n$ maps lines to lines in the following sense: If $\ell$ is a line in $\mathbb{RP}^n$ which intersects $U$, then there is a line $\ell'$ such that $f(\ell \cap U) = \ell' \cap f(U)$. Then $f$ is the restriction of a projective transformation of $\mathbb{RP}^n$ to $U$.

Again, for simplicity, I will present a proof for the case $n = 2$ only. This already contains all the important ideas so you can figure out for yourself how it works for $n > 2$.

**Proof.** (for $n = 2$) Define a map $\hat{f} : \mathbb{RP}^2 \to \mathbb{RP}^2$ as follows. For $X \in B$ let $\hat{f}(X) = f(X)$. If $X \not\in B$, let $\ell_1, \ell_2$ be two lines through $X$ that intersect $B$ and let $\ell' \cap f(X)$ be the intersection of the lines $\ell_1'$ and $\ell_2'$, the images of $\ell_1, \ell_2$ under $f$ (in the sense explained in the theorem). This point is well defined because it does not depend on the choice of $\ell_1$ and $\ell_2$. To see this, use Desargues’ theorem to show that if $\ell_1, \ell_2, \ell_3$ are three lines that intersect $B$ and all go through one point outside $B$, then their images under $f$ intersect in one point. You have to convince yourself that you always have enough room in the open ball to construct (the relevant part of) a Desargues figure. (See left figure below.)

We have defined $\hat{f}$ using only information about $f$ on $B$. In fact, $\hat{f}$ coincides with $f$ on $U$. (Why?) Further, $\hat{f}$ maps lines to lines. To see this, use (the inverse) Desargues’ theorem to show that $\hat{f}$ maps three points on a line to three points on a line. Again you have to convince yourself that you have enough room in $B$ to construct (the relevant part of) a Desargues figure. (See right figure below.) Finally, by the fundamental theorem (global version), $\hat{f}$ is a projective transformation.
Duality

In homogeneous coordinates \(x_1, x_2, x_3\), the equation for a line in a projective plane is

\[a_1 x_1 + a_2 x_2 + a_3 x_3 = 0,\]

where not all coefficients \(a_i\) are zero. The coefficients \(a_1, a_2, a_3\) can be seen as homogeneous coordinates for the line, because if we replace in the equation \(a_i\) by \(\lambda a_i\) for some \(\lambda \neq 0\) we get an equivalent equation for the same line. Thus, the set of lines in a projective plane is itself a projective plane, the dual plane. Points in the dual plane correspond to lines in the original plane. Moreover, if we consider in the above equation the \(x_i\) as fixed and the \(a_i\) as variables, we get an equation for a line in the dual plane. Points on this line correspond to points in the original plane that contain \([x]\). Thus, a the points on a line in the dual plane correspond to lines in the original plane through a point.

It makes sense to look at this phenomenon in a basis independent way and for arbitrary dimension. It boils down to the duality of vector spaces.

Let \(V\) be a finite dimensional vector space over a field \(F\).

The dual vector space \(V^*\) of \(V\) is the vector space of linear functions \(V \rightarrow F\) (linear forms on \(V\)). If \(v_1, \ldots, v_n\) is a basis of \(V\), the dual basis of \(V^*\) is \(\varphi_1, \ldots, \varphi_n\) with \(\varphi_i(v_j) = \delta_{ij}\). In particular \(\dim V = \dim V^*\). But there is no natural way to identify \(V^*\) with \(V\). ("Natural" means independent of any arbitrary choices. In this case: choice of a basis.)

There is, however, a natural identification of \(V\) with \(V^{**}\): A vector \(v \in V\) is identified with the linear form \(V^* \rightarrow F\), \(\varphi \mapsto \varphi(v)\). With this identification, \(V\) is also the dual vector space of \(V^*\).

Let \(f : V \rightarrow W\) be a linear map. The dual linear map \(f^* : W^* \rightarrow V^*\) is defined by \(f^*(\psi)(v) = \psi(f(v))\). Note that the dual map "goes in the opposite direction". If \(f\) is invertible, then \(f^{-1*} = f^{-1}\) is a map \(V^* \rightarrow W^*\).

If \(U \subseteq V\) is a linear subspace, the annihilator of \(U\) is the linear subspace

\[U^0 = \{\varphi \in V^* \mid \varphi(v) = 0 \text{ for all } v \in U\} \subseteq V^*\]

of linear forms that vanish on \(U\).

This provides a correspondence between subspaces of \(V\) with subspaces of \(V^*\).

The dimensions of \(U\) and \(U^0\) are related by

\[\dim U + \dim U^0 = \dim V.\]

Indeed, let \(v_1, \ldots, v_k\) be a basis for \(U\) and extend it to a basis \(v_1, \ldots, v_n\) of \(V\). Let \(\varphi_1, \ldots, \varphi_n\) be the dual basis of \(V^*\). Then (one sees easily that) \(\varphi_{k+1}, \ldots, \varphi_n\) is a basis of \(U^0\).

(In fact, the above dimension formula is just a coordinate free way of saying that each linearly independent homogeneous equation in the coordinates reduces the dimension of the solution space by 1.)

If \(U_1\) and \(U_2\) are subspaces of \(V\), then

\[(U_1 \cap U_2)^0 = U_1^0 + U_2^0 \quad \text{and} \quad (U_1 + U_2)^0 = U_1^0 \cap U_2^0.\]

(Can you see this?)

Now let \(P(V)\) be the \(n\)-dimensional projective space of an \((n + 1)\)-dimensional vector space \(V\). The dual projective space is \(P(V^*)\).

A point \([v] \in P(V)\) corresponds to the hyperplane \(P([v]^0) \subseteq P(V^*)\), and a point \([\varphi] \in P(V^*)\) corresponds to the hyperplane \(P([\varphi]^0) \subseteq P(V)\). Note that the points of the hyperplane \(P([\varphi]^0)\) correspond to the hyperplanes in \(P(V)\) that contain \([v]\).

In general, a \(k\)-plane \(P(U) \subseteq P(V)\) corresponds to the plane \(P(U^0) \subseteq P(V^*)\) of dimension

\[
\dim U^0 - 1 = \dim V - \dim U - 1 = (n + 1) - (k + 1) - 1 = n - k - 1.
\]

The points in \(P(U^0)\) correspond to the hyperplanes in \(P(V)\) that contain \(P(U)\).
Let us take another look at duality for projective planes. (Hyperplanes in a plane are lines.) To aid the imagination, let us focus on the real projective plane $\mathbb{R}P^2 = P(\mathbb{R}^3)$ and its dual plane $P(\mathbb{R}^3^*)$ which we denote by $\mathbb{R}P^{2*}$ (although everything holds in general).

So each point in $\mathbb{R}P^2$ corresponds to a line in $\mathbb{R}P^{2*}$ and vice versa. The points on a line in $\mathbb{R}P^2$ correspond to the lines through the corresponding point in $\mathbb{R}P^{2*}$. Lines through a point in $\mathbb{R}P^2$ correspond to the points on the corresponding line in $\mathbb{R}P^{2*}$.

Every theorem about $\mathbb{R}P^2$ can also be read as a theorem about $\mathbb{R}P^{2*}$. This leads to the following duality principle:

From every theorem that talks only about incidence relations between points and lines in a projective plane, one obtains another valid theorem by interchanging the words "point" and "line" (and the phrases "goes through" and "lies on").

For example, the theorem that is obtained from the Desargues theorem in this way (the dual Desargues theorem) turns out to be the converse of Desargues's theorem.

We had seen that the converse of Desargues is equivalent to Desargues, so Desargues's theorem turns out to be self-dual. The same is true for Pappus's theorem. (Check it out.)

Note that four lines through a point in $\mathbb{R}P^2$ correspond to four points on a line in $\mathbb{R}P^{2*}$. But for four points on a line we had defined the cross ratio. Via duality this gives us a definition for the cross ratio of four lines through a point.

**Proposition.** Let $l_1, l_2, l_3, l_4$ be four lines through a point $P$ in $\mathbb{R}P^2$. Let $l$ be a line not containing $P$ and let $P_1, P_2, P_3, P_4$ be the intersections of the four lines $l_i$ with $l$. Then

$$\text{cr}(l_1, l_2, l_3, l_4) = \text{cr}(P_1, P_2, P_3, P_4).$$

This proposition is an immediate consequence of the following one.

**Proposition.** Let $P$ be a point in $\mathbb{R}P^2$ and let $l^*$ be the corresponding line in $\mathbb{R}P^{2*}$, so that each point of $l^*$ corresponds to a line through $P$. Let $l$ be a line in $\mathbb{R}P^2$ that does not contain $P$. Then the map $l^* \rightarrow l$ that maps a point of $l^*$ to the intersection of the corresponding line with $l$ is a projective transformation.

**Proof.** Let $P = [v_1]$, and let $[v_2], [v_3]$ be two points on $l$. Then $v_1, v_2, v_3$ is a basis of $\mathbb{R}^3$. Let $\varphi_1, \varphi_2, \varphi_3$ be the dual basis of $\mathbb{R}^3^*$. The line $l^*$ is spanned by $[\varphi_2], [\varphi_3]$. Hence the points $[\varphi] \in l^*$ have representative vectors $\varphi = s\varphi_2 + t\varphi_3$, and $s, t$ are homogeneous coordinates on $l^*$. The line in $\mathbb{R}P^2$ corresponding to $[\varphi]$ intersects $l$ in a point $[v]$ such that $v = xv_2 + yv_3$ and

$$0 = \varphi(v) = (s\varphi_2 + t\varphi_3)(xv_2 + yv_3) = sx + ty.$$

This is the case for $x = t, y = -s$. So the map $l^* \rightarrow l$ in question comes from the linear map $s\varphi_2 + t\varphi_3 \mapsto tv_2 - sv_3$. □
Conic sections

The Euclidean point of view

The conic sections are ellipses (including circles), parabolas, hyperbolas, and the degenerate cases of a pair of lines, which may degenerate further to one “double” line, and a single point.

**ellipse**

\[
\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.
\]

\[r_1 + r_2 = \text{const.}\]

\[f = \sqrt{a^2 - b^2}\]

**hyperbola**

\[
\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.
\]

\[|r_1 - r_2| = \text{const.}\]

\[f = \sqrt{a^2 + b^2}\]

**parabola**

\[y = ax^2\]

\[r_1 = r_2\]

\[f = \frac{1}{4a}\]

**two intersecting lines**

\[
\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = 0
\]

**two parallel lines**

\[y^2 = a\]

**one “double” line**

\[y^2 = 0\]

What do these curves have in common?

1. They arise as intersection of a plane with a cone (or cylinder in the case of two parallel lines). Hence the name.

2. They are all described by quadratic equations in the two Euclidean coordinates. In fact:

**Theorem.** The set of solutions of any quadratic equation in two variables \(u, v\),

\[au^2 + 2buv + cv^2 + du + ev + f = 0\]

is empty or a conic section.

More precisely: There is a change of coordinates \((u, v) = A(x, y) + t\) with \(A \in O(2), t \in \mathbb{R}^2\) which reduces \((*)\) to one of the standard forms above. Do you still know how to prove this?
Optical properties of the conic sections

Here are two proofs for the case of an ellipse:

Proof #1 (more analytic). Let $\gamma : \mathbb{R} \to \mathbb{R}^2$ be a parameterization of an ellipse, like for example $\gamma(t) = (a \cos t, b \sin t)$. The direction of the tangent at a point $\gamma(t)$ is $\gamma'(t)$ and

$$
\cos(\alpha_1) = \frac{\langle \gamma', \gamma - F_1 \rangle}{\|\gamma\| \|\gamma - F_1\|}, \quad \cos(\alpha_2) = -\frac{\langle \gamma', \gamma - F_2 \rangle}{\|\gamma\| \|\gamma - F_2\|}.
$$

Now $\|\gamma - F_1\| + \|\gamma - F_2\| = \text{const.}$ implies

$$
0 = (\|\gamma - F_1\| + \|\gamma - F_2\|)' = \langle \gamma', \gamma - F_1 \rangle + \langle \gamma', \gamma - F_2 \rangle - (\gamma\gamma' \|\gamma - F_1\| + \gamma\gamma' \|\gamma - F_2\|) = \cos \alpha_1 - \cos \alpha_2.
$$

Proof #2 (more synthetic). Let $P$ be a point of the ellipse. Extend the line segment $F_2P$ a distance of $r_1$ beyond $P$. Call the new endpoint of the extended segment $F_2'$. Claim: The tangent of the ellipse at $P$ is the perpendicular bisector $\ell$ of $F_1F_2'$. Indeed, $P$ lies on $\ell$ because it has equal distance $r_1$ from $F_1$ and $F_2'$. Consider any other point $\tilde{P}$ on $\ell$ and let $\tilde{r}_1$ be its distance to both $F_1$ and $F_2'$ and let $\tilde{r}_2$ be its distance to $F_2$. Then $\tilde{r}_1 + \tilde{r}_2 > r_1 + r_2$ so $\tilde{P}$ does not lie on the ellipse. Hence, $\ell$ intersects the ellipse in precisely one point, $P$, which proves the claim. Now the equality of the angles follows easily.

The advantage of the second proof is that it suggests another theorem:

**Theorem.** Let $c$ be a circle with center $F_2$ and let $F_1$ be a point inside $c$. The locus of the centers of all circles that go through $F_1$ and touch $c$ is an ellipse with foci $F_1$ and $F_2$.\[42\]
The projective point of view
Five points determine a conic section
Pencils of conic sections
Pascal’s theorem

To be completed
See following handwritten notes by Alina Hinzmann.

*Thank you, Alina!
Inside and outside

The pole-polar relationship

The dual conic and Brianchon’s theorem

To be completed

See following handwritten notes by Alina Hinzmann.*

*Thank you, Alina!
The rational parameterization of conics

Steiner’s projective generation of conic sections

To be completed

See following handwritten notes by Alina Hinzmann.

*Thank you, Alina!*
Conic sections

Projective point of view

1. If \( q \) is a non-zero quadratic form on \( \mathbb{R}^2 \), then the set

\[
Q = \{ (x, y) \in \mathbb{R}P^2 \mid q(x, y) = 0 \}
\]

is called a conic section (or conic) in \( \mathbb{R}P^2 \).

2. In coordinates \( y_1, y_2, y_3 \) w.r.t. a suitable basis of \( \mathbb{R}^3 \),

\[ q(x) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 \]

with \( \lambda_1 = \pm 1, \lambda_2, \lambda_3 \) or 0.

3. If \( \lambda_1 > 0 \), then \( q \) and \( \lambda q \) define the same quadric.

According to the signature of \( q \) there are the following cases:

1. \( q \) is non-degenerate, i.e., all \( \lambda_i \neq 0 \).
2. \( q \) is definite, NLOA. \( q(x) = y_1^2 + y_2^2 + y_3^2 \)

\[ q(x) = 0 \implies x = 0 \]. So \( Q = \emptyset \).
3. \( q \) is indefinite, NLOA. \( q(x) = y_1^2 + y_2^2 - y_3^2 \)

Case of ellipses, hyperbolas and parabolas

2. \( q \) is degenerate

2.1. \( q(x) = y_1^2 + y_2^2 \)

\[ q(x) = 0 \implies y_1 = y_2 = 0 \]. \( Q \) is a point.

2.2. \( q(x) = y_1^2 - y_2^2 \)

\[ q(x) = 0 \implies y_1 = y_2 \] or \( y_1 = -y_2 \)

\( Q \) is the union of 2 lines.

2.3. \( q(x) = y_1^2 \)

\[ q(x) = 0 \implies y_1 = 0 \]

\( Q \) is one line. \( \rightarrow Q \) has rank 1

Every non-empty non-degenerate conic (1.2) can be mapped to any other such conic by a projective transformation.

From the projective point of view, there is no distinction between ellipses, hyperbolas and parabolas.

Whether such a conic looks like an ellipse, a hyperbola or a parabola in some affine image of \( \mathbb{R}P^2 \) depends on whether the line at infinity doesn't intersect the quadric or intersects it in two or one points.
In all cases except 1.1 (empty conic) and 2.1 (one point) the set \( Q \) defines the quadratic form \( q \) up to a scalar factor.

Let's exclude the one-point conic and the empty one from the further discussion.

A conic in \( \mathbb{RP}^2 \) is represented by a unique point in the projective space of the vector space of quadratic forms on \( \mathbb{R}^3 \).

The dimension of the vector space of quadratic forms is 6 because each quadratic form is represented by a symmetric \( 3 \times 3 \) matrix.

So the dimension of the projective space of conic sections is 5.

5 points determine a conic.

**Theorem.**

If \( P_1, P_2, P_3, P_4, P_5 \) are five points in \( \mathbb{RP}^2 \) then there is a conic containing all \( P_i \).

This conic is unique \( \iff \) no four of the points are on one line.

No three of the points are on one line \( \iff \) the conic is non-degenerate.

**Proof.**

1. Existence of a conic containing the \( P_i \).

   Let \( q(x) = x^T B x \)
   
   \[ \begin{align*}
   &= b_0 x_1^2 + 2 b_1 x_1 x_2 + 2 b_2 x_2^2 + 2 b_3 x_1 x_3 + b_4 x_3^2 + b_5 x_1 x_2 x_3 + b_6 x_2 x_3 + b_7 x_1 x_3 + b_8 x_3^2
   \end{align*} \]

   Now the condition that \( q(x) = 0 \) for a specific \( x \in \mathbb{R}^3 \) is a homogeneous linear equation in the 6 variables, bijective and the requirement that \( P_1, P_2, P_3, P_4, P_5 \in Q \) gives 5 equations so the space is at least 1-dimensional. (For uniqueness we must have that the 5 equations are linearly independent.)

2. Assume that no 4 points are on a line, but three of them are.

3. No three points are on a line.

   By part 1) there exists a conic through the 5 points. It cannot be degenerate because in that case there would be 3 on a line.

   Remains to show uniqueness.

   Suppose there would be two linearly independent quadratic forms \( q_1, q_2 \) such that \( q_i (P_i) = 0 \) and \( q_2 (P_i) = 0 \), \( i = 1, \ldots, 5 \).

   Then \( q_t = q_1 + t q_2 \) describes a 1-param. family of conics all of which contain all 5 points.

   But there is a value for \( t \) such that \( q_t \) is degenerate, \( \det q_t = \det (q_1 + t q_2) \) is a degree 2 polynomial in \( t \), so it has a 7th root.
Then we would have a degenerate conic containing 5 points no three of which are on one line, which is not possible.

(0) If four points are on one line, the conic is not unique.

**Pencils of conics (Kegelschnittwürfel)**

A pencil of conics is a line in the projective space of conics.

A pencil of conics is determined by two conics in it, \([q_1], [q_2]\). Then any conic in the pencil is of the form \([sq_1 + tq_2]\), \(s, t\) hom. coord. on the pencil.

If \(P_1, P_2, P_3, P_4\) are four points no three of which are on a line, then the conics through \(P_1, P_2, P_3, P_4\) form a pencil. This pencil contains precisely three degenerate conics.

Let \(l_{ij}\) be a linear form on \(\mathbb{R}^2\) such that \(l_{ij}(x) = 0\) is the equation for line \(P_i P_j\). Then the quadratic forms defining the degenerate conics are \(Q_2, Q_3, Q_4\).

Any conic in this pencil is described by a quadratic form \(Q = sQ_2 + tQ_3 + Q_4\) for some \(s, t\).

**Pascal's theorem**

**Theorem.**

Let \(A, B, C, D, E, F\) be six points on a non-degenerate conic. Then the intersection points of opposite sides of the hexagon \(ABCDEF\) lie on one line.
Proof.
Let \( q \) be the quadratic form describing the conic, and for two points \( x, y \) let \( l_{xy} \) denote the linear form describing the line \( xy \). Then

\[
q = \alpha_1 \ell_{AB} + \alpha_2 \ell_{BC} + \alpha_3 \ell_{DA} \quad \text{for some } \alpha_1, \alpha_2, \alpha_3
\]

\[
= \beta_1 \ell_{DE} + \beta_2 \ell_{FA} + \beta_3 \ell_{FLD}
\]

\[
= \beta_1 \ell_{DE} + \beta_2 \ell_{FA} - (\ell_{DE} + \ell_{FA}) = (\ell_{DE} - \ell_{FA}) \ell_{DA}
\]

Claim: \( ABnDE, CDnFA, BCnEF \) lie on the line \( t_2 \ell_{EF} - t_1 \ell_{BC} = 0 \).

On \( ABnDE \), \( \ell_{AB} = 0 \) and \( \ell_{DE} = 0 \), so \( t_2 \ell_{FA} - t_1 \ell_{BC} = 0 \) on \( ABnDE \).

Similar argument for \( CDnFA \).

Finally \( t_2 \ell_{EF} - t_1 \ell_{BC} \) is obviously zero on \( EFnBC \).

\( \square \)

\( \sim \) vgl. Pappus's theorem.

Inside & Outside

The inside and outside of a non-degenerate conic can be defined in a projectively invariant way.

The inside: the set of all points \( P \) such that any line through \( P \) intersects the conic.

The outside: the set of points \( P \) such that there is a line through \( P \) which does not intersect the conic.

Analytically, if the conic is defined by a quadratic form with signature \((+ + -)\) or \((- - +)\), then \( P = \{p\} \) is in the inside if \( q(P) < 0 \) and in the outside if \( q(P) > 0 \).

(Note: the signature is \((- - +)\), the inequality signs have to be reversed.)

The pole-polar relationship

Suppose \( b \) is a symmetric non-degenerate bilinear form on \( \mathbb{R}^3 \). This defines a map from the points of \( \mathbb{P}^2 \) to the lines of \( \mathbb{P}^2 \):

a point \([p]\) is mapped to the line of all \([x]\) with \( b(p, x) = 0 \).

(For arbitrary vector spaces \( V \), a symmetric non-degenerate bil. form defines a map \( V \rightarrow V^* : v \mapsto L(v, \cdot) \), and this ...
defines a projective transformation \( P(U) \rightarrow P(U^*) \).

The line corresponding to the point \([p]\) is called the **polar line** (or **polar**) of \([p]\), and \([p]\) is called the **pole** of this line.

If \( L \) is indefinite then the quadratic form \( q(U) = \langle U, U \rangle \)
defines a (non-empty) non-degenerate conic section. Conversely, a non-degenerate conic section determines its defining bilinear form up to scale, so it determines a pole-polar relationship between points and lines.

**Proposition.**

If \([p] \in Q\) then the polar of \([p]\) wrt. \( Q \) is the tangent through \([p]\).

**Proof.**

\([p] \in Q\) means that \( b(p,p) = 0 \). This means that \([p]\) (on its own) is contained in its polar.

Let \([q]\) be another point on the polar line. Then \( b(p,q) = 0 \). Now if \([q]\) were contained in \( Q \) then \( b(q,q) = 0 \) and hence \( b(sp+tg, sp+tg) = 0 \) for all \( s,t \). Then the whole line would be contained in \( Q \) which is not possible since \( Q \) is non-degenerate.

Alternatively, look for whole sections of the polar and \( Q \). Any point on line through \([p]\) at \( \lambda \) apart from \([p]\) is \( \lambda p + (1-\lambda)q \).

**Proposition.**

Suppose \( P_1, P_2 \) are on a line \( L \).

Then the polars \( l_1, l_2 \) of \( P_1, P_2 \) intersect in the pole \( P \) of \( L \).

**Proof.**

Let \( P = [p] \), \( P_i = [p_i] \).

We have to show that \( P \), the pole of \( L \), lies on the polars \( l_1, l_2 \).

So we have to show: \( b(p_1, p) = 0 \) and \( b(p_2, p) = 0 \).

But these are also the conditions for \( P_1 \) and \( P_2 \) to lie on the polar \( L \) of \( P \).

(So maybe one should have proved first:

\( P_1 \) lies on polar of \( P_2 \) \( \Rightarrow \) \( P_2 \) lies on polar of \( P_1 \))

**Construction of polars:**

- **Tangents are points for the touching points.**

- \( P \) in the outside.

- **Polar of \( P \).**

- \( L \), polar of \( P \).
**Theorem.**

The points $x, y, z$ form a polar triangle, this means:
Each of the three points is the pole of the line through the other two.

**Proof.** Since $A, B, C, D \in \mathcal{Q}$, the matrix $\mathcal{B}$ of the quadratic form defining $\mathcal{Q}$ must be diagonal.

Indeed:

(I). $0 = \begin{bmatrix} x \\ y \\ z \end{bmatrix}^T \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = b_{11} + b_{22} + b_{33} + 2b_{12} + 2b_{13} + 2b_{23}$

(II). $0 = \begin{bmatrix} -y \\ x \\ 0 \end{bmatrix}^T \mathcal{B} \begin{bmatrix} -y \\ x \\ 0 \end{bmatrix} = b_{11} + b_{22} + b_{33} - 2b_{12} - 2b_{13} + 2b_{23}$

(III). $0 = \begin{bmatrix} -z \\ 0 \\ x \end{bmatrix}^T \mathcal{B} \begin{bmatrix} -z \\ 0 \\ x \end{bmatrix} = b_{11} + b_{22} + b_{33} + 2b_{12} - 2b_{13} - 2b_{23}$

(IV). $0 = \begin{bmatrix} y \\ z \\ 0 \end{bmatrix}^T \mathcal{B} \begin{bmatrix} y \\ z \\ 0 \end{bmatrix} = b_{11} + b_{22} + b_{33} - 2b_{12} + 2b_{13} - 2b_{23}$

\[ (I)-(II) \Rightarrow b_{12} + b_{13} = 0 \]

\[ (I)-(III) \Rightarrow b_{12} + b_{23} = 0 \]

\[ \Rightarrow b_{12} = b_{23} = b_{13} = 0. \]

\[ \Rightarrow \mathcal{B} = \begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{pmatrix} \]

\[ b(x, y) = b \begin{bmatrix} \phi \\ \xi \end{bmatrix} = b \begin{bmatrix} \phi \\ \xi \end{bmatrix} \begin{bmatrix} \phi \\ \xi \end{bmatrix} = 0 \quad \text{etc.} \]

This gives a way to construct tangents to a circle from a point using only ruler. (finish)

---

\[ \text{choose } \mathcal{P} \]
\[ \text{draw any 2 lines through } \mathcal{P} \]
\[ \text{intersecting the circle} \]
\[ \text{compute quadrilateral} \]
\[ \text{touching pts of tangents} \]
The dual quadric & Brianchon's theorem

If \( Q \) is a non-degenerate quadric, then the set of tangents of \( Q \) form a quadric in the dual plane.

[Homework]

Pascal: Opposite sides of a hexagon inscribed in a conic intersect on a line.

Brianchon's theorem: In a hexagon which is circumscribed to a conic, the three diagonals connecting opposite points intersect in one point.

Rational parameterization of conics

\( Q \) a non-degenerate conic, \( P \) a point on \( Q \), \( l \) a line not containing \( P \).

Consider projection \( Q \to l \) from \( P \).

Example:

In homogeneous coordinates \( s = s_1/s_2 \), \( x = x_1/x_3 \), \( y = x_2/x_3 \), this map is

\[
\begin{pmatrix}
Sy_2 \\
S^2y_2
\end{pmatrix} =
\begin{pmatrix}
2s_1s_2 \\
S_1^2 - S_2^2 \\
S_1^2 + S_2^2
\end{pmatrix}.
\]

So the circle is parameterized over \( \mathbb{P}^2 \) by quadratic polynomials. That's why this is called the rational parameterization.
Back to the general case. Essentially, it does not matter which point we take as P and which line as \( l \):

1. Consider projecting to different lines.

2. Consider a different point.

Theorem: The projections \( Q \xrightarrow{\text{proj. from } P} l \) and \( Q \xrightarrow{\text{proj. from } P} l' \) differ by a proj. transformation \( l \rightarrow l' \) (proj. from \( P \)).

Proof.

Choose coordinates and line.

With this choice of coordinates, the matrix of the bilinear form defining \( Q \) has the form:

\[
\begin{bmatrix}
0 & d & 0 \\
d & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

So the equation for \( Q \) is

\[2dx_1x_2 - x_3^2 = 0.\]

Consider a point \( x \) on \( l \): \( x = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \)

Then the image under \( Q \xrightarrow{P_3} l \) is

\[x' = \begin{bmatrix} \frac{t_2^2}{2d} \\ \frac{s_1}{s} \\ t \end{bmatrix} = \begin{bmatrix} \frac{t_2^2}{2ds} \\ \frac{2ds}{2d} \\ t \end{bmatrix} \]
The image of $x'$ under $Q$ is $x'' = \begin{bmatrix} t^2 \\ t^2 \\ 2dst \end{bmatrix} = \begin{bmatrix} t \\ t \\ 2ds \end{bmatrix}$.

So $\mathbb{RP}^2 \rightarrow \mathbb{R}^2 \rightarrow Q \rightarrow \mathbb{R}^2 \rightarrow \mathbb{RP}^2$

$[t] \rightarrow \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} t \\ 2ds \end{bmatrix} \rightarrow \begin{bmatrix} t \\ 2ds \end{bmatrix} = \begin{bmatrix} (2d \cdot t) \cdot (t) \end{bmatrix}$

**Conclusion:** There is a correspondence between $Q$ and $\mathbb{RP}^2$ which is unique up to a proj. transformation of $\mathbb{RP}^2$.

**Steiner's projective generation of conic sections**

Let $P_1, P_2 \in \mathbb{RP}^2$. Consider them as lines in $P(\mathbb{R}^3, \ast)$. Let $\tau: P_1 \rightarrow P_2$ be a proj. transformation mapping each line through $P_1$ to a line through $P_2$.

**Theorem.** The set $Q = \{ \ln \tau(l) \mid l \in P_1 \}$ is a quadric in $\mathbb{RP}^2$ which is non-degenerate $\iff \tau$ does not map the line $P_1P_2$ to the line $P_1P_2$.

**Examples.**

- **Translation**

  $Q = \text{(line at infinity)} \cup P_1P_2$

- $Q = l \cup P_1P_2$

- **Conic section**
Proof. Choose coordinates so that $P_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $P_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

Let $[s \, t]$ be homogeneous coordinates of a line $sx_2 + tx_3 = 0$ through $P_2$, and let $[s' \, t']$ be hom. coordinates for a line $s'x_1 + t'x_3 = 0$ through $P_2$.

So $\Phi$ maps a line $sx_2 + tx_3 = 0$ to a line $(as + bt)x_1 + (cs + dt)x_3 = c$.

Line through $P_1, P_2$ is $x_2 = 0$, so it has coordinates $[s] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

This is mapped to line $P_2 P_3 : x_1 = 0$, or $[s'] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, so $c = 0$.

Similarly, line $P_1 P_4 : x_3 = 0$, $[s] = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ is mapped to $P_2 P_4 : x_1 - x_3 = 0$, $[s'] = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

\[ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} -1 \\ 1 \end{pmatrix} \]

\[ \begin{pmatrix} a-b & d \\ 0 & d \end{pmatrix} \]

Since $d \neq 0$ (otherwise matrix is singular), we may assume $d = 1$, and

\[ \Phi : [s \, t] \rightarrow \begin{bmatrix} \frac{s}{t} \n \end{bmatrix} \begin{pmatrix} a-b & d \\ 0 & 1 \end{bmatrix} = [s' \, t'] \]

So line $sx_2 + tx_3 = 0$ is mapped to the line $(b+1)s + bt)x_1 + tx_3 = c$.

Setting $x_3 = 1$ we get $x_2 = \frac{-s}{b}$, $x_1 = -\frac{t}{(b+1)s + bt}$.

Intersection point is

\[ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{-s}{(b+1)s + bt} \\ -\frac{t}{s} \\ 1 \end{bmatrix} \]

\[ \frac{(b+1)x_1 - x_2}{b(x_1 + x_3)} = \frac{(b+1)x_1 - x_2}{b(x_1 + x_3)} = \frac{b(b+1)x_1 - x_2 - x_2 x_3}{b(x_1 + x_3) - b x_1 x_2 - x_2 x_3} \]

\[ \frac{b(b+1)x_1 - x_2 - x_2 x_3}{b(x_1 + x_3) - b x_1 x_2 - x_2 x_3} \]
Quadrics

Quadrics are the generalization of conic sections to arbitrary dimension: They are the sets defined by one quadratic equation in the coordinates. Conic sections are the special case of quadrics in the plane.

The Euclidean point of view

A quadric in the Euclidean space $\mathbb{R}^n$ is defined by a quadratic equation

$$x^T A x + b^T x + c = 0,$$

with $A$ a symmetric $n \times n$ matrix, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

This can be brought to normal form by a Euclidean motion $x \mapsto Mx + v$, with $M \in O(n, \mathbb{R})$, $v \in \mathbb{R}^n$. In $\mathbb{R}^3$, the following cases can occur:

1. **ellipsoid**
   \[(\frac{x}{a})^2 + (\frac{y}{b})^2 + (\frac{z}{c})^2 = 1\]

2. **elliptic paraboloid**
   \[z = (\frac{x}{a})^2 + (\frac{y}{b})^2\]

3. **2-sheeted hyperboloid**
   \[(\frac{x}{a})^2 + (\frac{y}{b})^2 - (\frac{z}{c})^2 = -1\]

plus some degenerate cases:
- cones and cylinders over a conic
- two planes
- one “double” plane
- one line
- one point
- the empty set.

The projective point of view

If $q$ is quadratic form on a vector space $V$, $q \neq 0$, then

$$Q = \{ [v] \in P(V) \mid q(v) = 0 \}$$

is called a *quadric* in $P(V)$. Since $Q$ is defined by a quadratic polynomial in the homogeneous coordinates, the algebraic properties of the base field of scalars of $V$ play an important role. We will only consider $\mathbb{R}$ and comment on $\mathbb{C}$.

In $\mathbb{R}P^3$, there are only three non-degenerate cases depending on the signature of $q$:

1. **(0) (++++) or (------)**, the case of definite $q$ leading to an empty quadric $Q = \emptyset$. We exclude this case from now on.
2. **(1) (+ + --) or (+ -- +)**. In an affine image of $\mathbb{R}P^3$, $Q$ looks like an ellipsoid, an elliptic paraboloid, or a 2-sheeted hyperboloid, depending on whether the plane at infinity does not intersect, is tangent to, or intersects the quadric (without being tangent).
3. **(2) (+ + --)**. In an affine image of $\mathbb{R}P^3$, $Q$ looks like a 1-sheeted hyperboloid or a hyperbolic paraboloid, depending on whether the plane at infinity intersects (without being tangent) or is tangent to $Q$. (In this case, any plane meets $Q$.)

If $q$ is degenerate, let $U_0 = \ker q$, and let $U_1$ be a complementary subspace. Then $Q$ is the union of all lines joining a point in $P(U_0)$ with a point in the non-degenerate quadric in $P(U_1)$ defined by the restriction $q|_{U_1}$.

In $\mathbb{C}P^n$, there is up to projective transformations only one non-degenerate quadric. There are $n$ degenerate ones, depending on the rank of $q$ (which can be $1, \ldots, n$).

---

1. The three images in the first row are taken from Wikipedia.
Proposition. If $Q$ is a (non-empty) non-degenerate quadric in $\mathbb{RP}^n$, then its defining bilinear form is uniquely determined up to a scalar factor.

Proof. Suppose $q$ and $\tilde{q}$ determine the same quadric. Hence $q(v,v) = 0 \Leftrightarrow \tilde{q}(v,v) = 0$. We want to show $\tilde{q} = \lambda q$. Suppose $q$ has signature $(k, n+1-k)$, where $1 \leq k \leq n+1$, and let $e_1, \ldots, e_k, f_1, \ldots, f_{n+1-k}$ be an orthonormal basis for $q$ (with $q(e_i, e_i) = 1$ and $q(f_m, f_m) = -1$). We are done if we have shown that $\tilde{q}(e_i, e_j) = \lambda \delta_{ij}$, $\tilde{q}(f_m, f_l) = -\lambda \delta_{ml}$, and $\tilde{q}(e_i, f_m) = 0$. To see this, first note that for any $i = 1, \ldots, k$ and $m = 1, \ldots, n+1-k,$

$$q(e_i \pm f_m, e_i \pm f_m) = q(e_i, e_i) \pm 2q(e_i, f_m) + q(f_m, f_m) = 0,$$

so

$$0 = \tilde{q}(e_i \pm f_m, e_i \pm f_m) = \tilde{q}(e_i, e_i) \pm 2\tilde{q}(e_i, f_m) + \tilde{q}(f_m, f_m)$$

This implies $\tilde{q}(e_i, f_m) = 0$ and $\tilde{q}(e_i, e_i) = -\tilde{q}(f_m, f_m) = \lambda$ for some $\lambda \in \mathbb{R}$ independent of $i, m$. To see that $\tilde{q}(e_i, e_i) = 0$ for $i \neq j$ and $\tilde{q}(f_m, f_l) = 0$ for $m \neq l$, make a similar argument starting with $q(v,v) = 0$ for $v = (e_i \pm e_j) + (f_l \pm f_l)$ (here the signs may be chosen independently). \qed

Inside and outside

Any (non-empty) non-degenerate quadric divides $\mathbb{RP}^n$ into two connected components. If the signature is $(n, 1)$, then inside and outside can be defined in the same way as for conics: A point is inside iff every line through it intersects the quadric in two points.

Question: How can the two components be distinguished in general if the signature is $(k, m)$ with $k \neq m$?

If the signature is neutral $(k = m)$, the two components can be interchanged by a projective transformation.

Lines in a quadric

A line intersects a quadric in $\mathbb{RP}^n$ either not at all, in two points, in one point, or it lies entirely in the quadric. In the last two cases, the line is called a tangent. In $\mathbb{RP}^3$, the only non-degenerate quadrics that contain lines are the ones with neutral signature $(+ + - -)$.

Proposition. Let $Q$ be a quadric in $\mathbb{RP}^3$ with neutral signature. Through any point in $Q$ there are precisely two lines lying entirely in $Q$.

Proof. To see that there are no more than two lines through a point $[p] \in Q$ lying entirely in $Q$, show that any such line must lie in the plane $q(p, \cdot) = 0$, and note that the intersection of $Q$ with a plane is a conic section, so it cannot contain more that two lines.

To see that there are actually two such lines, we may assume (after a change of coordinates, if necessary) that $Q$ is the quadric $x_1^2 + x_2^2 - x_3^2 - x_4^2 = 0$. This equation is equivalent to $(x_1 + x_3)(x_1 - x_3) + (x_2 + x_4)(x_2 - x_4) = 0$, and, after changing to new coordinates

$$y_1 = x_1 + x_3, \quad y_2 = x_1 - x_3, \quad y_3 = - (x_2 + x_4), \quad y_4 = x_2 - x_4,$$

we have

$$y_1y_2 - y_3y_4 = 0.$$

Now the map

$$f : \mathbb{RP}^1 \times \mathbb{RP}^1 \to Q, \quad \left( \begin{bmatrix} s_1 \\ t_1 \end{bmatrix}, \begin{bmatrix} s_2 \\ t_2 \end{bmatrix} \right) \mapsto \left[ \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \right] = \begin{bmatrix} s_1s_2 \\ t_1t_2 \\ s_1t_2 \\ s_2t_1 \end{bmatrix}$$

is actually a bijection $\mathbb{RP}^1 \times \mathbb{RP}^1 \leftrightarrow Q$. Indeed, if $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \in Q$, then $\begin{bmatrix} s_1 \\ t_1 \end{bmatrix}$ is determined by $\frac{s_1}{t_1} = \frac{y_1}{y_4}$ or by $\frac{s_1}{t_1} = \frac{y_2}{y_3}$. (It can happen that one of the right hand sides is $\frac{y_1}{y_4}$, but not both. If neither is $\frac{y_2}{y_3}$, they are equal.) Similarly, $\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$ is determined by $\frac{s_1}{t_2} = \frac{y_1}{y_3}$ or by $\frac{s_1}{t_2} = \frac{y_2}{y_4}$. For any point $P = f(P_1, P_2) \in Q$, the images of the functions $f(P_1, \cdot) : \mathbb{RP}^1 \to Q$ and $f(\cdot, P_2) : \mathbb{RP}^1 \to Q$ are two lines through $P$ lying entirely in $Q$.

In fact this proof also shows:

- $Q$ contains two families of pairwise skew lines, and each line of the first family intersects each line of the second family.
- $\mathbb{RP}^1$ is homeomorphic to the circle $S^1$, $Q$ is homeomorphic to $S^1 \times S^1$, so it is topologically a torus.

57
Polarity

A non-degenerate symmetric bilinear form \( q \) on a vector space \( V \) defines a relation between the points and hyperplanes of \( \mathbb{P}(V) \): To each point \([v]\) \( \in \mathbb{P}(V) \) corresponds the polar hyperplane

\[
\{ [w] \in \mathbb{P}(V) \mid q(v, w) = 0 \},
\]

and to each hyperplane there is a corresponding point, its pole. Note that

\[
[x] \in \text{polar hyperplane of } [y] \iff [y] \in \text{polar hyperplane of } [x] \iff q(x, y) = 0.
\]

More generally, let \( U \subseteq V \) be a \((k+1)\)-dimensional linear subspace of \( V \), and let \( n + 1 = \dim V \).

The orthogonal subspace of \( U \) (with respect to \( q \)) is

\[
U^\perp = \{ w \in V \mid q(u, w) = 0 \text{ for all } u \in U \}.
\]

The dimension of \( U^\perp \) is \( \dim V - \dim U = n - k \), and \( U^{\perp\perp} = U \). The \( k \)-plane \( \mathbb{P}(U) \) and the \((n - k - 1)\)-plane \( \mathbb{P}(U)^\perp \) in \( \mathbb{P}(V) \) are called polar to each other. Polarity (with respect to \( q \)) is therefore a one-to-one relation between \( k \)-planes and \((n - k - 1)\)-planes in the \( n \)-dimensional projective space \( \mathbb{P}(V) \). In particular, if \( n = 3 \), polarity is a relation between points and planes and between lines and lines.

**Proposition.** Let \( Q \) be a (non-empty) non-degenerate quadric in \( \mathbb{RP}^n \) (\( \mathbb{CP}^n \)) defined by the symmetric bilinear form \( q \), and let \( X \in Q, Y \in \mathbb{RP}^n \) (\( \mathbb{CP}^n \)). Then

The line \( XY \) is tangent to \( Q \) \iff \( X \) is in the polar hyperplane of \( Y \).

**Proof.** Let \( X = [x], Y = [y] \). Then \( q(x, x) = 0 \) because \( X \in Q \). The line \( XY \) is tangent to \( Q \) either if it intersects \( Q \) in no other point but \( X \), or if it is contained entirely in \( Q \). The points on the line \( XY \) except \( X \) are parameterized by \([tx + y]\) with \( t \in \mathbb{R} \) (\( \mathbb{C} \)). Such a point lies in \( Q \) if

\[
0 = q(tx + y, tx + y) = t^2 q(x, x) + 2t q(x, y) + q(y, y) = 2 t q(x, y) + q(y, y).
\]

This equation for \( t \) has one solution if \( q(x, y) \neq 0 \), it has no solution if \( q(x, y) = 0 \) and \( q(y, y) \neq 0 \), and it is satisfied for all \( t \) if \( q(x, y) = q(y, y) = 0 \). So the line \( XY \) contains no other points of \( Q \) except \( X \) or lies entirely in \( Q \) precisely if \( q(x, y) = 0 \).

This provides a simple geometric interpretation of the polarity relationship between points and hyperplanes in the case when the polar hyperplane intersects \( Q \): The tangents from a point to the quadric touch the quadric in the points in which the quadric intersects the polar hyperplane.

If a quadric in \( \mathbb{RP}^3 \) is illuminated by a point light source outside the quadric (or by parallel light), the borderline between light and shadow on the quadric is a conic in the polar plane; and the shadow that the quadric throws on some other another plane is a projected image of this conic.

What about the polarity between lines in \( \mathbb{RP}^3 \)? If a point moves on a line, the polar planes rotate about a line, and these two lines are polar to each other.

**Projective transformations that map a quadric to itself**

Let \( Q \) be a (non-empty) non-degenerate quadric in \( \mathbb{RP}^n \) defined by the symmetric bilinear form \( q \) with signature \((k, n+1-k)\). If \( f : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1} \) is a linear map which is orthogonal with respect to \( q \) (that is, \( q(x, y) = q(f(x), f(y)) \) for all \( x, y \in \mathbb{R}^{n+1} \)), then the projective map \([x] \mapsto [f(x)]\) clearly maps \( Q \) to \( Q \). If the signature is not neutral (that is, if \( k \neq n+1-k \)), then these are all projective maps that map \( Q \) to \( Q \):

**Proposition.** If the signature is not neutral, then any projective transformation that maps \( Q \) to \( Q \) comes from a linear map which is orthogonal with respect to \( q \).

Hence, under the assumption of non-neutral signature, the group of projective transformations mapping \( Q \) to \( Q \) is \( PO(k, n+1-k) \), the projective orthogonal group for signature \((k, n+1-k)\).
We had defined the projective model of hyperbolic space to these form only a subgroup of index 2 within the group of all projective transformations mapping to itself, then it comes from an orthogonal transformation. (In the case of neutral signature, \( Q \) maps a non-degenerate quadric to itself and each connected component of the complement of \( Q \) contains \( k \) spacelike and \( n-1-k \) timelike vectors. This cannot be, because every orthogonal basis contains \( k \) spacelike and \( n-1-k \) timelike vectors)

To see that the non-neutral signature assumption is really necessary, consider the quadric \( Q \) defined by \( x_1^2 + x_2^2 - x_3^2 - x_4^2 = 0 \). The projective transformation \( \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \mapsto \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \) maps \( Q \) to \( Q \), but it does not come from an orthogonal map in \( O(2,2) \). Note that it also interchanges the two connected components of \( \mathbb{R}P^3 \setminus Q \). The following is true in general: If a projective transformation maps a non-degenerate quadric to itself and each connected component of the complement of \( Q \) to itself, then it comes from an orthogonal transformation. (In the case of neutral signature, these form only a subgroup of index 2 within the group of all projective transformations mapping \( Q \) to \( Q \)).

**The projective model of hyperbolic space**

We had defined \( n \)-dimensional hyperbolic space in the hyperboloid model as

\[ H^n = \{ x \in \mathbb{R}^{n+1} \mid \langle x, x \rangle = -1 \text{ and } x_{n+1} > 0 \} \]

where \( \langle x, y \rangle = x_1y_1 + \cdots + x_ny_n - x_{n+1}y_{n+1} \).

Equipped with the metric \( d \) defined by \( \cosh d(x,y) = -\langle x, y \rangle \).

In the projective model, \( n \)-dimensional hyperbolic space consists of the set \( H^n_{proj} \) of points inside a quadric \( Q \subset \mathbb{R}P^n \) defined by a symmetric bilinear form \( q \) of signature \((n,1)\),

\[ H^n_{proj} = \{ [x] \in \mathbb{R}P^n \mid q(x,x) < 0 \}, \]

equipped with the metric \( d_{proj} \) defined by

\[ \cosh d_{proj}([x],[y]) = \frac{|q(x,y)|}{\sqrt{q(x,x)q(y,y)}}. \]

This is indeed a metric space isometric to \( H^n \). First of all, if we use in \( \mathbb{R}P^n \) homogeneous coordinates with respect to an orthonormal basis for \( q \), then \( q = \langle \cdot, \cdot \rangle \), and the map

\[ H^n \longrightarrow H^n_{proj}, \quad x \mapsto [x] \]

is bijective and an isometry.

The projective model is actually the same as the Klein model (seen from the projective point of view, which was how Klein saw it in the first place): The points inside the unit circle—or unit sphere in higher dimension—are obviously the points inside a quadric of the right signature, and from the projective point of view we might as well take any other such quadric.

The group of isometries of \( H^n_{proj} \) is the projective orthogonal group \( PO(n,1) \) of projective transformations mapping \( Q \) to \( Q \) (and hence also the inside of \( Q \) to the inside of \( Q \)).

Via the polarity relation, points *outside* \( Q \) correspond to *hyperplanes* in \( H^n_{proj} \). Two hyperplanes intersect orthogonally if one contains the pole of the other. (Do you see why? What is the formula for the angle between two intersecting hyperplanes polar to two points \([v]\) and \([w]\) outside \( Q \)? What is the formula for their distance if they do not intersect?)

The following proposition is due to Arthur Cayley. (Actually, its history is usually told like this: Cayley discovered this way of defining a metric inside a quadric in terms of a cross ratio. Then Klein realized that this was a model for the non-Euclidean geometry discovered by Lobachevsky.)

**Proposition.** For two points \( A,B \in H^n_{proj} \), let \( X,Y \) be the points where the line \( AB \) intersects the quadric \( Q \), labelled such that \( X,B \) separates \( A,Y \). Then

\[ d_{proj}(A,B) = \frac{1}{2} \log \text{cr}(B,X,A,Y). \]

59
Proof. First convince yourself that \( cr(B, X, A, Y) > 1 \), so that the logarithm is positive (see Lectures 17 and 18). We will prove the proposition by showing that

\[
\cosh\left( \frac{1}{2} \log cr(B, X, A, Y) \right) = \frac{|q(a, b)|}{\sqrt{q(a, a)q(b, b)}}.
\]

Suppose \( A = [a] \) and \( B = [b] \), and introduce an affine parameter on the line \( AB \) by \( t \mapsto [a + tb] \). The points \( A \) and \( B \) correspond to the parameter values \( t = 0 \) and \( t = \infty \). The parameter values for \( X \) and \( Y \) are the roots \( t_{1,2} \) of the quadratic equation

\[
0 = q(a + tb, a + tb) = q(a, a) + 2q(a, b)t + q(b, b)t^2.
\]

On the one hand, \( cr(B, X, A, Y) = \frac{(\infty - t_1)(0 - t_2)}{(t_1 - 0)(t_2 - \infty)} = \frac{t_2}{t_1} \), so

\[
\cosh\left( \frac{1}{2} \log cr(B, X, A, Y) \right) = \frac{1}{2} \left( e^{\frac{1}{2} \log \frac{t_2}{t_1}} + e^{-\frac{1}{2} \log \frac{t_2}{t_1}} \right) = \frac{1}{2} \left( \sqrt{\frac{t_2}{t_1}} + \sqrt{\frac{t_1}{t_2}} \right).
\]

On the other hand, \( q(a, a) + 2q(a, b)t + q(b, b)t^2 = q(b, b)(t - t_1)(t - t_2) \) implies

\[
t_1 + t_2 = -\frac{2q(a, b)}{q(b, b)} \quad \text{and} \quad t_1t_2 = \frac{q(a, a)}{q(b, b)}.
\]

Since \( t_1 \) and \( t_2 \) have the same sign, and so do \( q(a, a) \) and \( q(b, b) \), we get

\[
\frac{1}{2} \left( \sqrt{\frac{t_2}{t_1}} + \sqrt{\frac{t_1}{t_2}} \right) = \frac{|t_1 + t_2|}{2\sqrt{t_1t_2}} = \frac{|q(a, b)|}{\sqrt{q(a, a)q(b, b)}}. \tag*{\square}
\]

Klein’s Erlangen program

For his inaugural lecture at the university of Erlangen in 1872, Felix Klein prepared a paper dealing with the question “What is geometry?”. In it, he breaks radically with the traditional point of view that geometry is the study of the “true” space around us. Instead, geometry should be viewed as

the study of invariants under a group of transformations.

Actually this was not meant as an abstract formal definition of the field of geometry. On the contrary, Klein saw a “joy in the pure form or shape” as the characteristic mark of a geometer, and he always emphasized the importance of space intuition. The point of the Erlangen program was to provide an organizing principle for the overabundant material that had accumulated in geometry, or rather, in the different geometries that had been discovered.

Let us see how some familiar geometries fit into this scheme. Every particular geometry deals with properties of figures in some space that remain invariant under some group of transformations of the space:

<table>
<thead>
<tr>
<th>geometry</th>
<th>space</th>
<th>transformation group</th>
<th>some invariant properties &amp; quantities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidean</td>
<td>( \mathbb{R}^n )</td>
<td>the group of Euclidean motions ( x \mapsto Ax + v ), ( A \in O(n) ), ( b \in \mathbb{R}^n )</td>
<td>distance of two points, angle of two lines</td>
</tr>
<tr>
<td>projective</td>
<td>( \mathbb{R}P^n )</td>
<td>the group of projective transformations, ( PGL(n + 1) )</td>
<td>the cross ratio, being a ( k )-plane, being a quadric, incidence relations</td>
</tr>
<tr>
<td>hyperbolic</td>
<td>( H_{proj}^n )</td>
<td>( PO(n, 1) )</td>
<td>hyperbolic distances and angles</td>
</tr>
<tr>
<td>spherical</td>
<td>( S^n )</td>
<td>( O(n + 1) )</td>
<td>spherical distances and angles</td>
</tr>
<tr>
<td>similarity</td>
<td>( \mathbb{R}^n )</td>
<td>the group of similarity transformations ( x \mapsto \lambda Ax + v ), ( \lambda \in \mathbb{R}_{&gt;0} ), ( A \in O(n) ), ( v \in \mathbb{R}^n )</td>
<td>angles, ratios of distances</td>
</tr>
<tr>
<td>affine</td>
<td>( \mathbb{R}^n )</td>
<td>the group of affine transformations ( x \mapsto Ax + v ), ( A \in GL(n) ), ( v \in \mathbb{R}^n )</td>
<td>parallelism, ratios of distances between points on a line</td>
</tr>
</tbody>
</table>

Actually, one does not distinguish between two geometries if there is a bijection between the spaces (or at least between subsets of them) which induces an isomorphism of the transformation
groups. Instead, one then speaks of different models for the same geometry. For example, we have considered different models for hyperbolic geometry: The points inside a quadric of signature \((n, 1)\) are in 1-to-1 correspondence with the points of the upper sheet of a 2-sheeted hyperboloid, and the projective transformations that fix the quadric correspond to the orthogonal transformations of \(\mathbb{R}^{n+1}\). Another model for elliptic geometry is \(\mathbb{RP}^n\) with group \(PO(n+1)\), the group of projective transformations that fix a definite symmetric bilinear form. The advantage of elliptic geometry over spherical geometry is that two lines intersect in one point, and there is a unique line through every two points.

Klein’s Erlangen program emphasizes the transformation group rather than the space on which it acts. If the transformation group of one geometry is a subgroup of the transformation group of another geometry, the first is called a subgeometry of the second. For example, the space \(\mathbb{R}^n\) of Euclidean geometry can be seen as a subset of \(\mathbb{RP}^n\)—the complement of a particular hyperplane \(x_n+1 = 0\) which is considered “at infinity”. The group of Euclidean motions then corresponds to the group of projective transformations \([A \, 0\, 0\, 0\, 0\, 1]\) with \(A \in O(n)\). Euclidean geometry is thus a subgeometry of projective geometry. The diagram illustrates the subgeometry relationship for some familiar geometries.

The invariants of one geometry are also invariants of any of its subgeometries (because the group is smaller). The same is true for theorems. Every theorem of projective geometry is also a theorem of Euclidean geometry. The converse is not true, but often one can see a Euclidean theorem as special case of a projective theorem. For example, the theorem on the inscribed angle over a chord of a circle is a special case of Steiner’s theorem on the projective generation of conics.

Klein’s group theoretical point of view on geometry illuminates the relationship between the different geometries. This is not only of theoretical interest, it is a great practical help in the day-to-day business of geometric research. When confronted with a geometric problem, it is usually an excellent idea to ask first: In which geometry should I treat this problem?

Let me illustrate this with a real-live example. A couple of months ago, Wolfgang Schief told me about the following striking theorem which he had discovered. It had somehow come up in his research on integrable systems and he had good reason to believe it was true.

**Theorem** (W. K. Schief). Consider three pairs \(a_1, a_2; b_1, b_2; c_1, c_2\) of lines in the plane. If the four intersection points \(a_1 \cap b_1\) lie on a circle and the four intersection points \(b_1 \cap c_1\) lie on a circle, then the four intersection points \(c_1 \cap a_2\) also lie on a circle.

How to prove this? It looks like a theorem of Euclidean geometry, or more precisely, of similarity geometry. But it can also be interpreted in terms of projective geometry and this leads to a surprisingly simple proof. Pairs of lines and circles are all conic sections. If we consider the Euclidean plane as the complement in \(\mathbb{RP}^2\) of the line \(x_3 = 0\), then the circles are characterized among the (non-empty) non-degenerate quadrics \(a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_{23}x_2x_3 + a_{33}x_3^2 = 0\) by the homogeneous linear equations in the coefficients \(a_{11} - a_{22} = 0\) and \(a_{12} = 0\). (A circle with center \((c_1, c_2)\) in \(\mathbb{R}^2\) and radius \(r\) has equation \(x_1^2 + x_2^2 - 2c_1x_1x_3 - 2c_2x_2x_3 + (c_1^2 + c_2^2 - r^2)x_3^2 = 0\).) Recall that a conic corresponds to a point in the 5-dimensional projective space of the vector space of quadratic forms. So a non-degenerate conic is a circle if it lies in a particular 3-plane in that projective space. Now the three pairs of lines correspond to points, and the three quadruples of intersection points correspond to pencils of non-degenerate conics, that is, to three lines connecting these points. If the 3-plane of circles intersects two of these lines, it also intersects the third. Thus the theorem is reduced to the fact that if a line in \(\mathbb{RP}^n\) intersects two sides of a triangle it also intersects the third.

61
Möbius geometry

Elementary model

Consider \( \mathbb{R}^n \) with the standard Euclidean scalar product \( (x, y) = \sum_1^n x_i y_i \).

Reflection in a hyperplane \( \{ x : (x - a, v) = 0 \} \) is the map

\[
x \mapsto x' = x - 2 \frac{(x - a, v)}{(v, v)} v.
\]

Reflection (or inversion) in a hypersphere with center \( c \) and radius \( r \) is the map

\[
x \mapsto x' = c + \frac{r^2}{\|x - c\|^2} (x - c).
\]

Note that \( x' \) lies on the same ray emanating from \( c \) and \( \|x - c\| \cdot \|x' - c\| = r^2 \).

Inversion in a sphere is an involution, except that the center \( c \) has no image and no preimage. We fix this by adding one extra point, \( \infty \), to \( \mathbb{R}^n \) and we declare it to be the image and preimage of \( c \). We also declare that reflections in hyperplanes map \( \infty \) to \( \infty \). Then both kinds of reflections are involutions on \( \mathbb{R}^n \cup \{ \infty \} \).

A Möbius transformation of \( \mathbb{R}^n \cup \{ \infty \} \) is a composition of reflections in hyperplanes and hyperspheres. The Möbius transformations form a group called the Möbius group denoted by \( \text{Möb}(n) \). A Möbius transformation is orientation reversing or preserving depending on whether it is the composition of an odd or even number of reflections. The subgroup of orientation preserving Möbius transformations is called the special Möbius group and denoted by \( \text{SMöb}(n) \) or \( \text{Möb}^+ (n) \).

The Möbius group contains all similarity transformations:

- A translation \( x \mapsto x + v \) is the composition of two reflections in parallel hyperplanes.
- An orthogonal transformation \( x \mapsto Ax \) with \( A \in \text{O}(n) \) is the composition of at most \( n \) reflections in hyperplanes through the origin. (Can you prove this?)
- A scaling transformation \( x \mapsto \lambda x \) with \( \lambda > 0 \) is the composition of a reflection in the unit sphere followed by a reflection in a sphere with center \( 0 \) and radius \( \sqrt{\lambda} \). (Check this.)

Proposition. A Möbius transformation maps any hyperplane or hypersphere to a hyperplane or hypersphere.

Proof. This is true for all similarity transformations. (These map hyperplanes to hyperplanes and hyperspheres to hyperspheres.) A reflection in a sphere with center \( c \) and radius \( r \) is the same as the similarity transformation \( x \mapsto \frac{1}{r}(x - c) \) mapping the sphere to the unit sphere, followed by inversion in the unit sphere, followed by the inverse similarity transformation \( x \mapsto rx + c \). (Check this.) So it remains to show that inversion in the unit sphere maps spheres and planes to spheres and planes. One could consider hyperspheres and hyperplanes separately but we will treat both cases simultaneously. Any hypersphere or hyperplane is determined by an equation of the form

\[
p\|x\|^2 - 2(v, x) + q = 0 \quad \text{with} \quad \|v\|^2 - pq > 0.
\]

If \( p = 0 \), the inequality implies \( v \neq 0 \) so the equation describes a hyperplane. If \( p \neq 0 \), it describes a hypersphere. Indeed, divide through by \( p \) to obtain

\[
0 = \|x\|^2 - 2\left(\frac{1}{p} v, x\right) + \frac{q}{p} = \|x - \frac{1}{p} v\|^2 - \frac{1}{p^2} \|v\|^2 + \frac{q}{p}.
\]

This is a sphere with center \( \frac{1}{p} v \) and radius \( \sqrt{\frac{1}{p^2} \|v\|^2 - \frac{q}{p}} \). (The assumed inequality ensures that the expression under the square root is positive.) Now for \( x' = \frac{1}{\|x\|^2} x \) one obtains

\[
p\|x'\|^2 - 2(v, x') + q = 0 \quad \iff \quad q\|x\|^2 - 2(v, x) + p = 0.
\]

So \( x' \) is contained in a particular hyperplane or hypersphere if and only if \( x \) is contained in some other hyperplane or hypersphere.
From now on we will consider hyperplanes as a special cases of hyperspheres that contain \( \infty \). So hypersphere will mean hyperspheres or hyperplane.

**Proposition.** Any bijective map \( f : \mathbb{R}^n \cup \{ \infty \} \to \mathbb{R}^n \cup \{ \infty \} \) which maps hyperspheres to hyperspheres is a M"obius transformation.

**Proof.** (i) Suppose \( f(\infty) = \infty \). Then \( f \) maps hyperplanes to hyperplanes. Then it also maps lines to lines, because a line is the intersection of \( n-1 \) hyperplanes. By the fundamental theorem of projective geometry (or rather the corollary of it, see Lecture 19), the restriction \( f|_{\mathbb{R}^n} \) is an affine transformation. Since it also maps spheres to spheres it must be a similarity.

(ii) Suppose \( f(\infty) = c \neq \infty \). Let \( g \) be the inversion in a sphere with center \( c \). Then \( g \circ f \) also maps hyperspheres to hyperspheres and also \( \infty \) to \( \infty \). By (i) it is a similarity transformation, so \( f = g \circ g \circ f \) is a M"obius transformation.

**Proposition.** The M"obius transformations are conformal.

**Proof.** Since the similarity transformations are conformal it remains only to show that inversion in the unit sphere is conformal. Let \( t \mapsto \gamma(t), t \mapsto \eta(t) \) be two parameterized curves intersecting in \( \gamma(t_0) = \eta(t_0) \). The intersection angle \( \alpha \) is determined by

\[
\cos \alpha = \frac{\langle \gamma'(t_0), \eta'(t_0) \rangle}{\| \gamma'(t_0) \| \| \eta'(t_0) \|}.
\]

Let \( \tilde{\gamma} = \frac{1}{(\gamma, \gamma)} \gamma, \tilde{\eta} = \frac{1}{(\eta, \eta)} \eta \), be the image curves after inversion in the unit sphere. One finds that

\[
\tilde{\gamma}' = \frac{1}{(\gamma, \gamma)^2} \left( \langle \gamma, \gamma \rangle \gamma' - 2 \langle \gamma, \gamma' \rangle \right),
\]

and similarly for \( \tilde{\eta}' \). From this one obtains

\[
\langle \tilde{\gamma}', \tilde{\gamma}' \rangle = \frac{1}{(\gamma, \gamma)^2} \langle \gamma', \gamma' \rangle, \quad \text{so} \quad \| \tilde{\gamma}' \| = \frac{1}{\| \gamma \|} \| \gamma' \|,
\]

and in the same way

\[
\| \tilde{\eta}' \| = \frac{1}{\| \eta \|} \| \eta' \|. \quad \text{Using} \quad \gamma(t_0) = \eta(t_0) = p \quad \text{one finds that}
\]

\[
\langle \tilde{\gamma}'(t_0), \tilde{\eta}'(t_0) \rangle = \frac{1}{\| p \|} \langle \gamma'(t_0), \eta'(t_0) \rangle
\]

and hence

\[
\frac{\langle \gamma'(t_0), \eta'(t_0) \rangle}{\| \gamma'(t_0) \| \| \eta'(t_0) \|} = \frac{\langle \tilde{\gamma}'(t_0), \tilde{\eta}'(t_0) \rangle}{\| \tilde{\gamma}'(t_0) \| \| \tilde{\eta}'(t_0) \|}
\]

Two-dimensional Möbius geometry

This case is special because we can identify \( \mathbb{R}^2 \) with \( \mathbb{C} \), and \( \mathbb{R}^2 \cup \{ \infty \} \) with the extended complex plane \( \mathbb{C} = \mathbb{C} \cup \{ \infty \} \), which is the same as \( \mathbb{C} \mathbb{P}^1 \), the complex projective line (see Lecture 13). The orientation preserving and reversing similarity transformations are \( z \mapsto az + b \) and \( z \mapsto a\bar{z} + b \) (\( a \neq 0 \)), reflection in the real line is \( z \mapsto \bar{z} \), and inversion in the unit circle \( |z| = 1 \) is the map \( z \mapsto \frac{\bar{z}}{|z|^2} = \frac{1}{z} \).

**Proposition.** The orientation preserving and reversing M"obius transformations of \( \mathbb{C} = \mathbb{C} \mathbb{P}^1 \) are precisely the maps of the form

\[
z \mapsto \frac{az + b}{cz + d} \quad \text{and} \quad z \mapsto \frac{a\bar{z} + b}{c\bar{z} + d} \quad \text{with} \quad \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \neq 0.
\]

**Proof.** First, these transformations form a group: The transformations of the first kind are the projective transformations of \( \mathbb{C} \mathbb{P}^1 \), and the transformations of the second kind are compositions of these with complex conjugation \( z \mapsto \bar{z} \). (Note that first performing a transformation of the first kind and then complex conjugation also leads to a transformation of the second kind.) This group contains the similarity transformations and inversion in the unit sphere, so it contains the M"obius group. On the other hand, it is not bigger than the M"obius group, because any of these transformations is a composition of reflections and similarity transformations: If \( c = 0 \), they are just similarity transformations. Otherwise, this follows from

\[
\frac{az + b}{cz + d} = \frac{a}{c} + \frac{bc - ad}{c(cz + d)}
\]

and the equation obtained by replacing \( z \) by \( \bar{z} \).
Two-dimensional Möbius geometry (continued)

We have identified the extended real plane \( \mathbb{R}^2 \cup \{\infty\} \) with the complex projective line \( \mathbb{CP}^1 = \hat{\mathbb{C}} = \mathbb{C}^2 \cup \{\infty\} \), and we have seen that the orientation preserving Möbius transformations are the complex projective transformations of \( \mathbb{CP}^1 \): \( SM\hat{\mathbb{Q}}(2) = \text{PGL}(2, \mathbb{C}) \). This has the following immediate consequences:

- The orientation preserving Möbius transformations of the plane preserve the complex cross ratio \( \text{cr}(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_2)(z_3 - z_4)}{(z_2 - z_3)(z_1 - z_4)} \). If \( f \) is an orientation reversing Möbius transformation, then \( \text{cr}(f(z_1), f(z_2), f(z_3), f(z_4)) = \text{cr}(z_1, z_2, z_3, z_4) \).

- For any three points \( z_1, z_2, z_3 \) and any three points \( w_1, w_2, w_3 \), there is a unique orientation preserving Möbius transformation \( f \in \text{PGL}(2, \mathbb{C}) \) with \( f(z_i) = w_i \). There is also a unique orientation reversing one mapping \( z_i \mapsto w_i \), namely \( f \) followed by an inversion in the circle through \( w_1, w_2, w_3 \).

- Four points \( z_1, z_2, z_3, z_4 \) lie on a circle if their cross ratio is real. Moreover, they are in that cyclic order on the circle if \( \text{cr}(z_1, z_2, z_3, z_4) < 0 \).

When we talked about the Poincaré disk and half-plane models of the hyperbolic plane, we did not say what the corresponding hyperbolic isometries were. We can do so now:

- The isometries of the hyperbolic plane in the Poincaré disk model are the Möbius transformations that map \( D^2 \to D^2 \). These are the maps
  \[
  z \mapsto \frac{az + b}{cz + d} \quad \text{and} \quad z \mapsto \frac{az + b}{cz + d} \quad \text{with} \quad a, b, c, d \in \mathbb{C} \quad \text{and} \quad |a|^2 - |b|^2 > 0.
  \]

- The isometries of the hyperbolic plane in the half-plane model are the Möbius transformations that map the upper half-plane \( \text{Im} z > 0 \) to the upper half-plane. These are the maps
  \[
  z \mapsto \frac{az + b}{cz + d} \quad \text{with} \quad ad - bc > 0 \quad \text{and} \quad z \mapsto \frac{az + b}{cz + d} \quad \text{with} \quad ad - bc < 0,
  \]
  and \( a, b, c, d \in \mathbb{R} \) in both cases.

- The Poincaré disk and half-plane models are related by the map \( D^2 \to \{ z \in \mathbb{C} \mid \text{Im} z > 0 \} \), \( z \mapsto \frac{z + 1}{1 - \overline{z}} \). Indeed, this maps \( 0 \mapsto i, 1 \mapsto 1, i \mapsto \infty, -1 \mapsto -1, -i \mapsto 0 \). So it maps the circle through \( 1, i, -1, -i \) (the unit circle) to the circle through \( 1, \infty, -1, 0 \) (the real line); and since it maps \( 0 \) to \( i \), it maps the inside of the unit circle to the upper half-plane.

The projective model of Möbius geometry

Stereographic projection maps \( \mathbb{R}^n \cup \{\infty\} \) to the \( n \)-dimensional sphere \( S^n \subset \mathbb{R}^{n+1} \subset \mathbb{RP}^{n+1} \). (The point \( \infty \) is mapped to the north pole.) Hyperspheres in \( \mathbb{R}^n \cup \{\infty\} \) are mapped to hyperspheres in \( S^n \), which are intersections of \( S^n \) with hyperplanes of \( \mathbb{RP}^{n+1} \). The Möbius transformations of \( \mathbb{R}^n \cup \{\infty\} \) are characterized by the property that they map hyperspheres to hyperspheres. Hence they correspond to those transformations of \( S^n \) that map intersections with hyperplanes to intersections with hyperplanes. The projective transformations of \( \mathbb{RP}^{n+1} \) that map \( S^n \to S^n \) have this property. Hence \( PO(n+1, 1) \subset \mathbb{M} \hat{\mathbb{Q}}(n) \). In fact, these are all Möbius transformations: \( PO(n+1, 1) = \mathbb{M} \hat{\mathbb{Q}}(n) \). To see this let us examine which transformations of \( S^n \) correspond to the scalings, translations and orthogonal transformation of \( \mathbb{R}^n \) and to inversion in the unit sphere.

First, \( S^n \subset \mathbb{RP}^{n+1} \) is
\[
S^n = \{ [x] \in \mathbb{RP}^{n+1} \mid x \in \mathbb{R}^{n+2}, \langle x, x \rangle = 0 \}, \quad \text{where} \quad \langle x, y \rangle = \sum_{i=1}^{n+1} x_i y_i - x_{n+2} y_{n+2}.
\]
So we can write \( \langle x, y \rangle = x^T E y \), where \( E = \begin{pmatrix} I_n & 0 \\ I_n & -1 \end{pmatrix} \) and \( I_n \) is the \( n \times n \) identity matrix. So
\[
\langle Ax, Ay \rangle = x^T A^T E A x \quad \text{and hence} \quad A \in O(n+1, 1) \iff A^T E A = E \iff A^{-1} = E A^T E.
\]
Stereographic projection \( \mathbb{R}^n \to S^n \) is
\[
u \mapsto \begin{pmatrix} 1 \\ u, u \end{pmatrix} \begin{pmatrix} 2u \\ (u, u) - 1 \end{pmatrix} \text{ in hom \ coods} \begin{bmatrix} 2u \\ (u, u) - 1 \\ (u, u) + 1 \end{bmatrix}
\]
where \( \langle \cdot, \cdot \rangle \) is the standard Euclidean scalar product of \( \mathbb{R}^n \).
Orthogonal transformations $u \mapsto Mu$, $M \in O(n)$. Stereographic projection maps

$$Mu \mapsto \left[ \frac{2Mu}{(Mu,Mu)-1} \right] = \left[ \begin{bmatrix} M & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2u \\ (u,u)+1 \end{bmatrix} \right] \quad \text{and} \quad \left[ \begin{bmatrix} M & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right] \in O(n+1,1).
$$

Scalings $u \mapsto \lambda u$. Stereographic projection maps

$$\lambda u \mapsto \left[ \frac{2u}{\lambda^2(u,u)-1} \right] = \left[ \begin{bmatrix} I_n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda(u,u)-\frac{1}{\lambda} \\ \lambda(u,u)+\frac{1}{\lambda} \end{bmatrix} \right] \left[ \begin{bmatrix} 2u \\ (u,u)+1 \end{bmatrix} \right].$$

To see that $S(\lambda) \in O(n+1,1)$, check that $S^{-1}(\lambda) = S\left(\frac{1}{\lambda}\right)$ and $S\left(\frac{1}{\lambda}\right) = EST(\lambda)E$.

Translations $u \mapsto u + v$. Stereographic projection maps

$$u + v \mapsto \left[ \frac{2u + 2v}{(u,u)+2(u,v)+(v,v)+1} \right] = \left[ \begin{bmatrix} I_n & -v \\ v^T & -\frac{1}{2}(v,v) \end{bmatrix} \begin{bmatrix} v \\ (u,u)+1 \end{bmatrix} \right] \left[ \begin{bmatrix} 2u \\ (u,u)+1 \end{bmatrix} \right].$$

To see that $T(v) \in O(n+1,1)$, check that $T^{-1}(v) = T(-v)$ and $T(-v) = ET^T(v)E$.

Inversion in unit sphere $u \mapsto \frac{1}{(u,u)} u$. Stereographic projection maps

$$\frac{1}{(u,u)} u \mapsto \left[ \frac{\frac{1}{(u,u)} u}{\frac{1}{(u,u)} u} \right] = \left[ \begin{bmatrix} I_n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{(u,u)} u \\ 1+(u,u) \end{bmatrix} \right] \left[ \begin{bmatrix} 2u \\ (u,u)+1 \end{bmatrix} \right], \quad \text{and} \quad \left[ \begin{bmatrix} I_n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right] \in O(n+1,1).
$$

Thus we have the correspondences:

<table>
<thead>
<tr>
<th>elementary model</th>
<th>projective model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{R}^n \cup {\infty} \leftrightarrow S^n \subset \mathbb{RP}^{n+1}$</td>
<td>$\text{M&quot;{o}bius}(n) \leftrightarrow \text{PO}(n+1,1)$</td>
</tr>
</tbody>
</table>

Hypersphere $\subset \mathbb{R}^n \cup \{\infty\} \leftrightarrow$ hyperplane $\subset \mathbb{RP}^{n+1}$ intersecting $S^n \leftrightarrow$ point outside $S^n$

Hyperspheres (including hyperplanes) in $\mathbb{R}^n \cup \{\infty\}$ correspond to points outside $S^n$, that is to points $[s]$ with $\langle s, s \rangle > 0$. Indeed we had seen that hyperspheres in $\mathbb{R}^n \cup \{\infty\}$ are determined by an equation of the form $p(u,u) - 2(u,v) + q = 0$ with $(v,v) = pq > 0$. For $x = \left(\begin{bmatrix} 2u \\ (u,u)+1 \end{bmatrix}\right)$, this equation is equivalent to $\langle x, s \rangle = 0$ with $s = \left(\begin{bmatrix} \frac{1}{2}(p+q) \\ -\frac{1}{2}(p+q) \end{bmatrix}\right)$. In particular, a hypersphere in $\mathbb{R}^n$ with center $c$ and radius $r$ corresponds to the point $[s]$ with $s = \lambda \left(\begin{bmatrix} \frac{1}{2}(1-c,c+r^2) \\ \frac{1}{2}(1-c,c+r^2) \end{bmatrix}\right)$ for some $\lambda \neq 0$. This can be used to show:

**Proposition.** The hyperspheres corresponding to two points $[s_1], [s_2]$ outside $S^2$ intersect at an angle $\theta$ determined by $\cos \theta = \frac{(s_1,s_2)}{\sqrt{(s_1,s_1)(s_2,s_2)}}$. (Since the right hand side is only determined up to sign, the angle $\theta$ is only defined up to $\theta \leftrightarrow \pi - \theta$.)

**Proof.** First check that $\langle s_1, s_2 \rangle = \lambda^2 r_1^2 r_2^2$ and

$$\frac{(s_1,s_2)}{\sqrt{(s_1,s_1)(s_2,s_2)}} = \pm \frac{r_1^2+r_2^2-(c_1-c_2,c_1-c_2)}{2r_1 r_2}.$$

So $\frac{(s_1,s_2)}{\sqrt{(s_1,s_1)(s_2,s_2)}} = \pm \frac{r_1^2+r_2^2-(c_1-c_2,c_1-c_2)}{2r_1 r_2}$. Now use $(c_1-c_2,c_1-c_2) = r_1^2 + r_2^2 - 2r_1 r_2 \cos \alpha$ (see Figure). \(\square\)

Möbius geometric pencils of circles

Lines in $\mathbb{RP}^3$ correspond to 1-parameter families of circles in $\mathbb{R}^2 \cup \{\infty\}$. The left figure shows two such families corresponding to a line intersecting $S^2$ (blue) and the polar line, which does not intersect $S^2$ (green). The right figure shows two families corresponding to two polar lines tangent to $S^2$. In both cases, the blue and green circles are orthogonal to each other.
Relationship between Möbius and other geometries

Möbius geometry deals with properties of figures in \( S^n \subset \mathbb{RP}^{n+1} \) that are invariant under the group \( PO(n+1,1) \) of projective transformations of \( \mathbb{RP}^{n+1} \) that map \( S^n \to S^n \). Thus, \( n \)-dimensional Möbius geometry is a subgeometry of \( (n+1) \)-dimensional projective geometry. The same group, \( PO(n+1,1) \), also maps \( B^{n+1} \) (the inside of \( S^n \)) to itself. This gives the Klein model of \( (n+1) \)-dimensional hyperbolic geometry. So \( n \)-dimensional Möbius geometry can be seen as the geometry of the points in the ideal boundary of \( (n+1) \)-dimensional hyperbolic space.

For a point \( P = [p] \in \mathbb{RP}^n \), let \( G_P \) be the subgroup of \( PO(n+1,1) \) consisting of all projective transformations that map \( P \mapsto P \) (in addition to mapping \( S^n \to S^n \)). These also map the polar plane of \( P \) to itself.

If \( P \) is outside \( S^n \), then the polar plane intersects \( B^{n+1} \), and the geometry of this intersection with the group \( G_P \) is \( n \)-dimensional hyperbolic geometry.

If \( P \) is the center of \( S^n \), then the polar plane is the plane at infinity, so \( G_P \) is the group of affine transformations mapping \( S^n \) to itself. This is the group of orthogonal transformations. So the space \( S^n \) with the group \( G_P \) is \( n \)-dimensional spherical geometry. If \( P \) is any other point inside \( S^n \), one obtains a Möbius geometrically equivalent model for \( n \)-dimensional spherical geometry.

If \( P \) is the north pole of \( S^n \), then \( G_P \) corresponds (via stereographic projection) to the Möbius transformations of \( \mathbb{R}^n \cup \{ \infty \} \) that fix \( \infty \). These are the similarity transformations. Thus, \( S^n \) with \( G_P \) is a model for \( n \)-dimensional similarity geometry. If \( P \) is any other point in \( S^n \), one obtains a Möbius geometrically equivalent model of similarity geometry.

If \( P \in S^n \), the group \( G_P \) consists of all projective transformations that come from orthogonal maps \( A \in O(n+1,1) \) with \( Ap = \lambda p \) for some \( \lambda \in \mathbb{R} \setminus \{0\} \). Because \( p \) is a lightlike vector \( \lambda \) is not always equal to \( \pm 1 \). (For example consider the orthogonal transformations that correspond to scalings in \( \mathbb{R}^n \cup \{ \infty \} \); see last lecture.) If instead of \( G_P \), one considers the (projectivized) group of all \( A \in O(n+1,1) \) with \( Ap = p \), then one obtains a model for \( n \)-dimensional Euclidean geometry.

The paraboloid model of Möbius geometry

Stereographic projection maps a point \( u \in \mathbb{R}^n \) to the point
\[
[2u_1 e_1 + \ldots + 2u_n e_n + (\|u\|^2 - 1)e_{n+1} + (\|u\|^2 + 1)e_{n+2}]
\]
in \( S^n \subset \mathbb{RP}^{n+1} \). Now let
\[
e_\infty = \frac{1}{2}(e_{n+1} + e_{n+2}), \quad e_0 = \frac{1}{2}(-e_{n+1} + e_{n+2}).
\]
The subscripts \( 0 \) and \( \infty \) are chosen because \([e_0]\) and \([e_\infty]\) are the “south pole” and “north pole” of \( S^n \), and these are the images of \( 0 \) and \( \infty \) under stereographic projection. Written in the basis \( B = (e_1, \ldots, e_n, e_\infty, e_0) \), the image of \( u \) is
\[
[u_1 e_1 + \ldots + u_n e_n + \|u\|^2 e_\infty + e_0].
\]
If we divide by the \( e_0 \)-coordinate to dehomogenize, this maps \( u \) to the point \((u, \|u\|^2)\) on the paraboloid \( u_{n+1} = \|u\|^2 \). So in the new coordinates, stereographic projection becomes vertical projection to a paraboloid.
For the new basis vectors, the Lorentz scalar product is \((e_0, e_0) = (e_\infty, e_\infty) = 0, (e_0, e_\infty) = -\frac{1}{2}\). So the matrix of the Lorentz scalar product in the new basis \(B\) is

\[
\begin{pmatrix}
1 & \cdots & 0 & -\frac{1}{2} \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 1 & -\frac{1}{2} \\
-\frac{1}{2} & \cdots & -\frac{1}{2} & 0
\end{pmatrix}.
\]

The new basis \(B\) is not orthonormal. The polar plane of \([e_\infty]\) is the set of points \([y_1 e_1 + \ldots + y_n e_n + y_\infty e_\infty + y_0 e_0]\) with \(y_0 = 0\). The paraboloid model obtained by using \(y_0\) to de-homogenize is related to the sphere model by a projective transformation that maps the tangent plane of \(S^n\) in the north pole to the plane at infinity and, consequently, the sphere to a paraboloid.

Because the paraboloid model is just a different projective image of the sphere model, a sphere in \(\mathbb{R}^n\) is related to the sphere model by a projective transformation that maps the tangent plane of \(S^n\) in the north pole to the plane at infinity and, consequently, the sphere to a paraboloid.

Let \((y_1, \ldots, y_n, y_\infty, y_0) = \lambda(u_1, \ldots, u_n, \|u\|^2, 1)\) be homogeneous coordinates of the image point on the paraboloid. In terms of these coordinates, the sphere equation is linear:

\[-2c_1 y_1 - \ldots - 2c_n y_n + y_\infty + (\|c\|^2 - r^2)y_0 = 0.\]

Let \(y_0 > 0\) be the center and radius of a sphere in \(\mathbb{R}^n\). A point \(u \in \mathbb{R}^n\) belongs to the sphere if it satisfies the equation

\[\|u\|^2 - 2\langle c, u \rangle + \|c\|^2 - r^2 = 0.\]

It is convenient to divide by \(-2\) to obtain

\[c_1 y_1 + \ldots + c_n y_n - \frac{1}{2} y_\infty - \frac{1}{2}(\|c\|^2 - r^2)y_0 = 0,\]

because this can be written

\[
\left(\sum_{i=1}^n c_i e_i + (\|c\|^2 - r^2)e_\infty + e_0, \sum_{i=1}^n y_i e_i + y_\infty e_\infty + y_0 e_0\right) = 0.
\]

So in homogeneous coordinates with respect to the basis \(e_1, \ldots, e_n, e_0, e_\infty\), a sphere with center \(c\) and radius \(r\) corresponds to the point

\[[c_1, \ldots, c_n, \|c\|^2 - r^2, 1].\]

Note that (after de-homogenization), the point corresponding to the circle with center \(c\) and radius \(r\) is at distance \(r^2\) directly below the point of the paraboloid corresponding to the center \(c\).

In the same way, the hyperplane in \(\mathbb{R}^n\) with equation \(\langle v, u \rangle - b = 0\) corresponds to the point \([v_1, \ldots, v_n, b, 0]\). Thus, in the new coordinates, hyperspheres and hyperplanes correspond to points \([s_1, \ldots, s_n, s_\infty, s_0]\) with \(s_1^2 + \ldots + s_n^2 - s_\infty s_0 > 0\). If \(s_0 = 0\), the point corresponds to a hyperplane in \(\mathbb{R}^n\) with normal vector \((s_1, \ldots, s_n)\). If \(s_0 \neq 0\), it corresponds to a sphere with center \(c\) and radius \(r\) determined by

\[
c = \frac{1}{s_0} (s_1, \ldots, s_n), \quad r^2 = \frac{1}{s_0} (s_1^2 + \ldots + s_n^2 - s_\infty s_0).
\]
Lie geometry

Suppose we distinguish between differently oriented spheres. An unoriented sphere corresponds to two oriented spheres which consist of the same points but differ in their orientation. The different orientations can be visualized by drawing arrows pointing inwards or outwards as shown in the figure. Let us define a signed radius for oriented spheres by saying that the signed radius is just the radius if the arrows point outward, and minus the radius if the arrows point inward. We define coordinates for oriented spheres by appending the signed radius to the M"obius geometric coordinates for spheres. Thus, in the basis of the paraboloid model, an oriented sphere with center \( c \) and signed radius \( r \) has homogeneous coordinates

\[
[ c_1, \ldots, c_n, ||c||^2 - r^2, 1, r ].
\]

These homogeneous coordinates are not independent anymore, because the first \( n + 2 \) already determine \( r^2 \). A point \([y_1, \ldots, y_n, y_0, y_\infty, y_{n+3}]\) corresponds to an oriented sphere only if

\[
y_1^2 + \ldots + y_n^2 - y_0 y_\infty - y_{n+3}^2 = 0.
\]

This is the equation of a quadric with signature \((n + 1, 2)\) in \(\mathbb{R}P^{n+2}\) called the Lie quadric. We have thus established a correspondence between the oriented spheres in \(\mathbb{R}^n \cup \{\infty\}\) and the points in the Lie quadric.

An oriented sphere in \(\mathbb{R}^n \cup \{\infty\}\) with center \(c\) and signed radius \(r\) corresponds to the point 

\[
[y_1, \ldots, y_n, y_0, y_\infty, y_{n+3}] = [c_1, \ldots, c_n, ||c||^2 - r^2, 1, r]
\]

in the Lie quadric.

While Lie geometry distinguishes between differently oriented spheres, it does not distinguish between spheres and points. A point is just a sphere with radius \(r = 0\), corresponding to a point in the Lie quadric with \(y_{n+3} = 0\). Consequently (and reassuringly) points do not come with two different orientations.

A oriented hyperplane in \(\mathbb{R}^n\) with unit normal vector \(n\) and equation

\[(n, u) - d = 0 \tag{2}\]

can be seen as the limit of a sphere with radius \(r\) and center \((d - r)n\) as \(r \to \infty\). For the corresponding points in the Lie quadric one has

\[
[(d - r)n, (d - r)^2 - r^2, 1, r] \to \infty \lim [(\frac{d}{r} - 1)n, \frac{d^2}{r} - 2d, \frac{1}{r}, 1] \to [n, -2d, 0, 1].
\]

Therefore, the oriented hyperplane with unit normal vector \(n\) and equation (2) correspond to the point

\[
[n, -2d, 0, 1]
\]

in the Lie quadric.

**Question:** Which oriented sphere in \(\mathbb{R}^n\) does a point 

\[
[y] = [y_1, \ldots, y_n, y_\infty, y_0, y_{n+3}]
\]

in the Lie quadric (1) represent? The answer depends on whether one or both of the coordinates \(y_0\) and \(y_{n+3}\) are zero. We consider the possible cases separately:

- \(y_0 \neq 0\) and \(y_{n+3} \neq 0\): The point \([y]\) represents an oriented sphere with center \(c\) and signed radius \(r\), where

\[
c = \frac{1}{y_0} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad r = \frac{y_{n+3}}{y_0}.
\]
• \(y_0 = 0, y_{n+3} \neq 0\): The point \([y]\) represents an oriented plane with unit normal vector \(n\) and equation (2), where
\[
n = -\frac{1}{y_{n+3}} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad d = -\frac{y_\infty}{2y_{n+3}}.
\]

• \(y_0 \neq 0, y_{n+3} = 0\): The point \([y]\) represents the point
\[
\frac{1}{y_0} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n.
\]

• \(y_0 = y_{n+3} = 0\): In this case, equation (1) implies \(y_1 = \ldots = y_n = 0\), and hence \([y] = [e_\infty]\). The point \([y]\) represents the point \(\infty \in \mathbb{R}^{\infty}\).

A Lie transformation is a transformation of the set of points and oriented spheres in \(\mathbb{R}^n\) that corresponds to a projective transformation of \(\mathbb{RP}^{n+2}\) that maps the Lie quadric to itself. Hence the group of Lie transformations is \(PO(n+1, 2)\).

A Lie transformation does in general not map points to points. Indeed, points in \(\mathbb{R}^n\) correspond to the points \([y]\) in the Lie quadric with \(y_{n+3} = 0\). This condition is in general not preserved by a projective transformation that fixes the Lie quadric. So Lie transformations are very different from the geometric transformation groups we have considered up to now. They act on the space of points and oriented spheres. They do not act on a space of points. In short (and in italics):

\textit{Lie transformations are not point transformations.}

How can these transformations be described geometrically? Let us examine what it means for the corresponding oriented spheres if two points in the Lie quadric are polar to each other.

The bilinear form corresponding to the quadratic form (1) is
\[
\langle y, \tilde{y} \rangle_L = y_1 \tilde{y}_1 + \ldots + y_n \tilde{y}_n - \frac{1}{2} y_\infty \tilde{y}_\infty - \frac{1}{2} y_0 \tilde{y}_0 - y_{n+3} \tilde{y}_{n+3}.
\]

(Its signature is \((n+1, 2)\).) Two circles with centers \(c, \tilde{c}\) and radii \(r, \tilde{r}\), respectively, correspond to the points
\[
[y] = \left[ c, \|c\|^2 - r^2, 1, r \right] \quad \text{and} \quad [\tilde{y}] = \left[ \tilde{c}, \|\tilde{c}\|^2 - \tilde{r}^2, 1, \tilde{r} \right]
\]
in the Lie quadric. These are polar to each other with respect to the Lie quadric if
\[
0 = \langle y, \tilde{y} \rangle = (c, \tilde{c}) - \frac{1}{2} (\|c\|^2 - r^2) - \frac{1}{2} (\|\tilde{c}\|^2 - \tilde{r}^2) - r\tilde{r}
\]
\[
= -\frac{1}{2} (\|c - \tilde{c}\|^2 - (r - \tilde{r})^2).
\]

This is equivalent to the condition
\[
\|c - \tilde{c}\| = |r - \tilde{r}|,
\]
which is fulfilled precisely if the spheres touch and the orientations of the spheres agree in the point of contact (this means that the arrows point in the same direction). Indeed, if the signs of the radii are different (as in the top figure) then the condition for oriented contact is \(\|c - \tilde{c}\| = |r| + |\tilde{r}|\); if the signs are equal (as in the bottom figure) then the condition is \(\|c - \tilde{c}\| = ||r| - |\tilde{r}|\). Thus:

\textit{Two points in the Lie quadric are polar to each other if the corresponding oriented spheres are in oriented contact.}

You may want to convince yourself that this is true also for the case when one or both spheres are in fact hyperplanes or points. For a point and a sphere, oriented contact means that the point is contained in the sphere. An immediate consequence of this is:
A Lie transformation maps spheres in oriented contact to spheres in oriented contact.

In fact, the following is true (although I will not prove this):

Any transformation of the space of spheres which maps spheres in oriented contact to spheres in oriented contact is a Lie transformation.

So Lie geometry is the geometry in the space of oriented spheres that studies invariants under the group of transformations that preserve oriented contact.