Geometry

Karl-Hermann Neeb

January 22, 2024

Contents

Introduction

Geometry has a long tradition in Erlangen. One of the most significant milestones in our present understanding of geometric structures was Felix Klein's "Erlanger Programm"[1] Its main point is to define geometric properties in terms of a corresponding group of transformations. This leads to an identification of geometries with the groups of transformations preserving the corresponding geometric properties. This correspondence is best understood in terms of some examples of geometries on the same underlying set $\mathbb{R}^n$.

Linear geometry: The linear geometry on $\mathbb{R}^n$ is specified by the group $\mathrm{GL}(\mathbb{R}^n)$ of linear automorphisms of $\mathbb{R}^n$, which can be identified with the group $\mathrm{GL}_n(\mathbb{R})$ of invertible $(n \times n)$-matrices. In this context geometric properties are: "being 0" (for an element of $\mathbb{R}^n$), "linear (in-)dependence" (for subsets of $\mathbb{R}^n$), "being collinear" (for subsets of $\mathbb{R}^n$).

Affine geometry: The affine geometry on $\mathbb{R}^n$ is specified by the group of all invertible affine maps $\phi(x) = Ax+b$ for $A \in \mathrm{GL}_n(\mathbb{R})$, $b \in \mathbb{R}^n$. The main difference to the linear geometry is that there is no origin specified, i.e., "being 0" is not an affine geometric property. Typical affine geometric properties are: "parallelity of two affine lines", "the dimension of the affine subspace spanned by a subset", "proportions", i.e., "the ratio of the lengths of two segments of a line". Since the affine group is larger than the linear group, all affine properties are in particular linear.

Euclidean geometry: The euclidean geometry on $\mathbb{R}^n$ is specified by the group of all isometries, i.e., length preserving transformations, with respect to the euclidean metric

$$d(x,y) = \Big(\sum_{j=1}^n (x_j - y_j)^2\Big)^{\frac{1}{2}}.$$

These are the affine maps of the form $\phi(x) = Ax + b$ for $A \in \mathrm{O}_n(\mathbb{R})$, $b \in \mathbb{R}^n$, also called *congruence transformations*. Typical metric properties are: "the distance $d(x,y)$ of two points", resp., "the length of a line segment", "the angle between two non-zero vectors", "the surface area of a triangle".

In addition to the three types of geometries mentioned above, we shall study

• **Spherical geometry**, i.e., the metric geometry of the unit sphere $\mathbb{S}(E)$ in a euclidean vector space E with respect to its natural metric. The corresponding group consists of the restrictions of elements of the orthogonal group $\mathrm{O}(E)$. For $E = \mathbb{R}^{n+1}$ we obtain in particular the sphere $\mathbb{S}_n$ and the group $\mathrm{O}_{n+1}(\mathbb{R})$.

• **Projective geometry** on the projective space $\mathbb{P}(V)$ of one-dimensional subspaces of a vector space. The corresponding group is the projective linear group $\mathrm{PGL}(V) \cong \mathrm{GL}(V)/\mathbb{K}^\times$, consisting of the *projectivities*, i.e., those maps $\mathbb{P}(V) \to \mathbb{P}(V)$ induced by invertible linear maps $\phi \in \mathrm{GL}(V)$ by $\overline{\phi}(\mathbb{K}v) := \mathbb{K}\phi(v)$. Over the real field we shall also discuss metric projective geometry on $\mathbb{P}(\mathbb{R}^n)$, for which the corresponding group is the projective orthogonal group $\mathrm{PSO}_n(\mathbb{R}) := \mathrm{O}_n(\mathbb{R})/\{\pm\mathbf{1}\}$.

[1] Christian Felix Klein (25 April 1849 to 22 June 1925) was a German mathematician, known for his work in group theory, complex analysis, non-euclidean geometry, and on the connections between geometry and group theory. His 1872 Erlangen Program, defining geometries in terms of their underlying symmetry groups, was a very influential synthesis of much of the mathematics of the day.

Notation and Terminology

We write
$\mathbb{R}_+ := [0, \infty[$ for the closed real half line, and
$\mathbb{N} = \{1, 2, 3, \ldots\}$ for the natural numbers.

1 Affine Geometry

We start our geometric journey with the kind of geometry that is closest to the concepts we know from Linear Algebra, namely vector spaces. Conceptually we want to deal with vector spaces where we "forget" the origin, i.e., we want a geometric structure for which all translations become isomorphisms. This leads naturally to the concept of an affine space which distinguishes between points and vectors, which is usually not done in Linear Algebra courses. In affine geometry vectors are arrows $\overrightarrow{ab}$ determined by a pair (a, b) of points and these vectors represent translations of our space moving a to b.

1.1 Affine spaces

Definition 1.1. (a) An *affine space (over the field $\mathbb{K}$)* is a triple $(A, V, +)$, where A is a set, V a $\mathbb{K}$-vector space and $+$ a map

$$+\colon A \times V \to A, \quad (a, \mathbf{x}) \mapsto a + \mathbf{x},$$

satisfying the following conditions:

(A1) $a + \mathbf{o} = a$ for all $a \in A$.

(A2) $a + (\mathbf{x} + \mathbf{y}) = (a + \mathbf{x}) + \mathbf{y}$ for $a \in A, \mathbf{x}, \mathbf{y} \in V$.

(A3) For $a, b \in A$, there exists a unique $\mathbf{x} \in V$ with $b = a + \mathbf{x}$. This element $\overrightarrow{ab} := \mathbf{x}$ is called the *translation (vector) from a to b*.

The vector space V is called the *translation space of* A. It is sometimes denoted $\overrightarrow{A}$. Elements of A are called *points* and elements of V are called *vectors*.

In this notation the axioms (A1-3) take the form

(A1) $\overrightarrow{aa} = \mathbf{o}$ for $a \in A$.

(A2) $\overrightarrow{ab} + \overrightarrow{bc} = \overrightarrow{ac}$ for $a, b, c \in A$.

(A3) $a + \overrightarrow{ab} = b$ and $\overrightarrow{ab}$ is uniquely determined by this relation for $a, b \in A$.

Example 1.2. (a) The prototypical example of an affine space is obtained as $(V, V, +)$, where V is a $\mathbb{K}$-vector space and $+$ is the addition map of V. Then (A1-3) are obviously satisfied with $\overrightarrow{ab} = b - a$.

(b) For $V = \mathbb{K}^n$ we so obtain the *n-dimensional affine (coordinate)-space*

$$\mathbb{A}_n := (\mathbb{K}^n, \mathbb{K}^n, +).$$

(c) $A = \emptyset$ is an affine space for every vector space V with respect to the empty map $A \times V \to A$.

1.2 Affine maps

Definition 1.3. (a) Let $(A, V, +)$ and $(B, W, +)$ be affine spaces. A map $\phi\colon A \to B$ is called *affine* if there exists a linear map $\phi_L\colon V \to W$ such that

$$\phi(a + v) = \phi(a) + \phi_L(v) \quad \text{for} \quad a \in A, v \in V.$$

Clearly, $\phi_L\colon V \to V$ is uniquely determined by ϕ.

(b) An affine map $\phi\colon A \to B$ is called an *affine isomorphism* if there exists an affine map $\psi\colon B \to A$ satisfying

$$\phi \circ \psi = \mathrm{id}_B \quad \text{and} \quad \psi \circ \phi = \mathrm{id}_A .$$

It is easy to see that this condition is equivalent to the requirement that ϕ is bijective because then the inverse map $\phi^{-1}\colon B \to A$ is automatically affine (Exercise 1.1). We write $\mathrm{Aff}(A, B)$ for the set of affine maps $A \to B$ and $\mathrm{Aff}(A)$ for the set of affine maps $A \to A$. For the group of affine *automorphisms of* A, i.e., isomorphisms $A \to A$, we write $\mathrm{Aut}(A)$.

Example 1.4. Translations

$$\tau_v\colon A \to A, \quad a \mapsto a + v$$

are affine maps with $(\tau_v)_L = \mathrm{id}_V$ because

$$\tau_v(a + w) = (a + w) + v = (a + v) + w = \tau_v(a) + w \quad \text{for} \quad a \in A, w \in V.$$

Example 1.5. For $A = (V, V, +)$ and $B = (W, W, +)$, the canonical affine spaces defined by the vector spaces V and W, a map $\phi\colon V \to W$ is affine if and only if

$$\phi(v) = \phi_L(v) + w_0$$

for some linear map $\phi_L\colon V \to W$ and an element $w_0 \in W$. In fact, that ϕ necessarily has this form follows immediately from the definition (for $a = 0$). Conversely, an easy calculation shows that all maps of this form are affine. We conclude that the affine maps between vector spaces are simply the sums of linear maps and constant ones.

For $V = \mathbb{K}^n$ and $W = \mathbb{K}^m$ the preceding observation shows that the affine maps $\phi\colon \mathbb{K}^n \to \mathbb{K}^m$ have the form

$$\phi(x) = Ax + b$$

for a matrix $A \in M_{m,n}(\mathbb{K})$ and $b \in \mathbb{K}^m$. For $n = m = 1$ we obtain in particular the form

$$\phi(x) = ax + b.$$

Remark 1.6. (Embedding affine spaces into linear ones) (a) Let V be a linear space. We identify the affine space $A := (V, V, +)$ with the affine subspace $A := V \times \{1\} \subseteq \widehat{V} := V \times \mathbb{K}$. We write $\mathrm{End}(\widehat{V})_A \subseteq \mathrm{End}(\widehat{V})$ for the monoid of those linear maps preserving the affine hyperplane A. Clearly, every element $\phi \in \mathrm{End}(\widehat{V})_A$ induces an affine map $\phi|_A$. If, conversely, $\phi\colon A \to A$ is an affine map, then it is of the form

$$\phi(v, 1) = (\psi v + b, 1), \quad v \in V, \psi \in \mathrm{End}(V), b \in V,$$

and

$$\widehat{\phi}\colon \widehat{V} \to \widehat{V}, \quad \widehat{\phi}(v, t) := (\psi v + tb, t)$$

is an element of $\operatorname{End}(\widehat{V})_A$ with $\widehat{\phi}|_A = \phi$. Since every $\phi \in \operatorname{End}(\widehat{V})_A$ is uniquely determined by its restriction to A, we thus obtain an isomorphism of monoids

$$\operatorname{End}(\widehat{V})_A \to \operatorname{Aff}(A), \quad \phi \mapsto \phi|_A.$$

Restricting to invertible affine maps, we obtain a group isomorphism

$$\operatorname{GL}(\widehat{V})_A := \operatorname{End}(\widehat{V})_A^\times \to \operatorname{Aut}(A), \quad \phi \mapsto \phi|_A.$$

(b) It is also instructive to take a look at the concrete matrix picture of $V = \mathbb{K}^n$. Then $\widehat{V} = \mathbb{K}^{n+1}$ and

$$\operatorname{End}(\mathbb{K}_{n+1})_A = \Big\{ \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix} : A \in M_n(\mathbb{K}), b \in \mathbb{K}^n \Big\}.$$

Remark 1.7. (a) Composition of affine maps are affine. In particular, for each affine plane A, the set $\operatorname{Aff}(A)$ of affine maps $A \to A$ is a monoid w.r.t. composition and the neutral element id_A.

(b) The map

$$\operatorname{Aff}(A) \to \operatorname{End}(V), \quad \phi \to \phi_L$$

is a morphism of monoids, i.e.,

$$(\operatorname{id}_A)_L = \operatorname{id}_V \quad \text{and} \quad (\phi \circ \psi)_L = \phi_L \circ \psi_L.$$

Example 1.8. (of affine maps) (a) For a point $a \in A$, the map

$$i_a \colon V \to A, \quad i_a(v) := a + v$$

is an affine isomorphism $(V, V, +) \to (A, V, +)$ with $\phi_L = \operatorname{id}_V$. In fact, it is bijective and, for $b = a + w$ and $v \in V$ we have

$$i_a(w + v) = a + (w + v) = (a + w) + v = i_a(w) + v.$$

We conclude that:
$\boxed{\text{Every affine space } (A, V, +) \text{ is isomorphic to an affine space of the form } (V, V, +).}$

(b) Fix points $a_0, b_0 \in A$. Then every linear map $\psi \colon V \to V$ defines a unique affine map $\phi \colon A \to A$ with $\phi(a_0) = b_0$ and $\phi_L = \psi$. This map is defined by

$$\phi \colon A \to A, \quad a_0 + v \mapsto b_0 + \psi(v).$$

To see that this map is affine, we first note that $\phi = i_{b_0} \circ \psi \circ i_{a_0}^{-1}$. Since all three factors are affine maps by (a) and Example 1.5, Remark 1.7 implies that ϕ is affine.

Proposition 1.9. (Structure of the group $\operatorname{Aut}(A)$) *Let $(A, V, +)$ be a non-empty affine space.*

(a) *The translations form an abelian subgroup $\tau_V \cong (V, +)$ of $\operatorname{Aut}(A)$. In particular, $\tau_v \tau_w = \tau_{v+w}$ for $v, w \in V$.*

(b) *The map $q \colon \operatorname{Aut}(A) \to \operatorname{GL}(V), \phi \mapsto \phi_L$ is a surjective homomorphism whose kernel $\ker q = \tau_V$ is the subgroup of translations.*

(c) *For every point $a \in A$, the stabilizer subgroup $\operatorname{Aut}(A)_a$ is mapped isomorphically by q to $\operatorname{GL}(V)$.*

(d) *For $\psi \in \mathrm{GL}(V)$, we write $\psi^a \in \mathrm{Aut}(A)_a$ for the unique affine map with $(\psi^a)_L = \psi$ and $\psi^a(a) = a$. Then the map*

$$\Gamma\colon V \times \mathrm{GL}(V) \to \mathrm{Aut}(A), \quad \Gamma(v,\psi) \mapsto \tau_v \circ \psi^a$$

is a bijection. It is an isomorphism of groups with respect to the group structure on $V \times \mathrm{GL}(V)$, given by

$$(v,\psi)(v',\psi') := (v + \psi(v'), \psi\psi'), \quad v, v' \in V, \psi, \psi' \in \mathrm{GL}(V).$$

Proof. (a) is an immediate consequence of the definitions.

(b) We have already observed in Remark 1.7 that

$$q(\phi \circ \psi) = (\phi \circ \psi)_L = \phi_L \circ \psi_L = q(\phi) \circ q(\psi),$$

so that q is a group homomorphism. It is surjective by Example 1.8(b). To determine its kernel, we assume that $\phi_L = \mathrm{id}_V$. We pick an element $a \in A$. For $w \in A$ we now find

$$\phi(w) = \phi(a + \overrightarrow{aw}) = \phi(a) + \overrightarrow{aw} = (a + \overrightarrow{a\phi(a)}) + \overrightarrow{aw} = a + \overrightarrow{aw} + \overrightarrow{a\phi(a)} = w + \overrightarrow{a\phi(a)}.$$

This shows that ϕ is the translation τ_v with $v := \overrightarrow{a\phi(a)}$. We conclude that $\ker q = \tau_V$ is the subgroup of translations. In particular, the kernel intersects the subgroup $\mathrm{Aut}(A)_a$ trivially.

(c) If a translation τ_v fixes a, then $v = \overrightarrow{a\tau_v(a)} = \overrightarrow{aa} = 0$ implies that $\tau_v = \mathrm{id}_A$. Therefore $\ker q \cap \mathrm{Aut}(A)_a = \{\mathrm{id}_A\}$, so that the restriction of q to $\mathrm{Aut}(A)_a$ is injective.

Example 1.8(b) implies in particular that, for every $\psi \in \mathrm{GL}(V)$, there exists an affine map $\phi\colon A \to A$ with $\phi_L = \psi$ that fixes a. Therefore $q|_{\mathrm{Aut}(A)_a}$ is also surjective.

(d) To see that Γ is surjective, we start with $\phi \in \mathrm{Aut}(A)$ and put $v := \overrightarrow{a\phi(a)}$. Then $\tau_{-v}\phi \in \mathrm{Aut}(A)_a$ is of the form ψ^a for some $\psi \in \mathrm{GL}(V)$, and therefore $\phi = \tau_v \circ \psi^a = \Gamma(v,\psi)$.

If $\phi = \Gamma(v,\psi) = \tau_v \circ \psi^a$, then $\psi = \phi_L$ and $v = \overrightarrow{a\phi(a)}$, so that Γ is also injective.

For $v \in V$ and $\psi \in \mathrm{GL}(V)$ we have

$$\psi^a \circ \tau_v = \tau_{\psi(v)} \circ \psi^a$$

because both sides have the same linear part and map a to the same point $a + \psi(v)$. For the composition of two affine maps, this leads to the formula

$$(\tau_{v_1} \circ \psi_1^a) \circ (\tau_{v_2} \circ \psi_2^a) = \tau_{v_1 + \psi_1(v_2)} \circ (\psi_1\psi_2)^a. \qquad \square$$

1.3 Semidirect products of groups

In the language of group theory, Proposition 1.9 asserts that $\mathrm{Aut}(A)$ is a semidirect product $V \rtimes_\alpha \mathrm{GL}(V)$ of the groups V and $\mathrm{GL}(V)$ with respect to the action of $\mathrm{GL}(V)$ on V given by $\alpha_\psi(v) := \psi(v)$. To understand what this means, we now take a closer look at the concept of a semidirect product of groups. We start with the formal definition of a semidirect product.

Definition 1.10. Let G and N be groups and $\alpha\colon G \to \mathrm{Aut}(N), g \mapsto \alpha_g$ be a homomorphism. Then $N \times G$ is a group with respect to the multiplication

$$(n,g)(n',g') := (n\alpha_g(n'), gg') \quad \text{and} \quad (n,g)^{-1} = (\alpha_g^{-1}(n^{-1}), g^{-1}).$$

It is called the *semidirect product of N and G with respect to α*. It is denoted $N \rtimes_\alpha G$.

The *direct product* of G and N corresponds to the trivial homomorphism with $\alpha_G = \{\mathrm{id}_N\}$. It is denoted $N \times G$, and the group structure is given by

$$(n, g)(n', g') := (nn', gg').$$

Remark 1.11. If $\widehat{G} := N \rtimes_\alpha G$ is a semidirect product, then

$$\pi\colon \widehat{G} \to G, \quad (n, g) \mapsto g,$$

$$\sigma\colon G \to \widehat{G}, \quad g \mapsto (\mathbf{1}, g) \quad \text{and} \quad \iota\colon N \to \widehat{G}, \quad n \mapsto (n, \mathbf{1})$$

are group homomorphisms with $\pi \circ \sigma = \mathrm{id}_G$. In particular, ι is an isomorphism of N onto the normal subgroup $\ker \pi$ of $\widehat{G}$ and σ is an isomorphism of G onto a subgroup of $\widehat{G}$. These two subgroups have the property that the multiplication map

$$N \times G \to \widehat{G}, \quad (n, g) \mapsto \iota(n)\sigma(g) = (n, \mathbf{1})(\mathbf{1}, g) = (n, g)$$

is bijective. The main point in the concept of a semidirect product is that it describes the group structure in these product coordinates.

The following proposition provides a criterion that permits us to recognize semidirect product groups in a similar way as for the affine group (see also Exercise 1.2).

Proposition 1.12. *Let $\widehat{G}$ be a group, $N \trianglelefteq \widehat{G}$ be a normal subgroup and $G \subseteq \widehat{G}$ be a subgroup such that the multiplication map*

$$N \times G \to \widehat{G}, \quad (n, g) \mapsto ng$$

is bijective. Then

$$\alpha\colon G \to \mathrm{Aut}(N), \quad \alpha_g(n) := gng^{-1}$$

is a homomorphism and

$$\Phi\colon N \rtimes_\alpha G \to \widehat{G}, \quad (n, g) \mapsto ng$$

is an isomorphism of groups.

Proof. Since N is a normal subgroup of $\widehat{G}$, it is invariant under conjugation with elements of G. From that it easily follows that $\alpha_g \in \mathrm{Aut}(N)$, and it is clear that α is a homomorphism of groups. We can thus construct the semidirect product group $N \rtimes_\alpha G$. A direct calculation shows that Φ is a group homomorphism (Exercise). $\square$

We have already seen that the group $\mathrm{Aut}(A)$ of affine automorphisms of an affine space is isomorphic to a semidirect product $V \rtimes \mathrm{GL}(V)$ for $V = \overrightarrow{A}$. Here are some examples of a more group theoretic flavor.

Example 1.13. (a) For $n = p + q$, we consider the group

$$\widehat{G} := \left\{ \begin{pmatrix} a & 0 \\ c & d \end{pmatrix} \in M_n(\mathbb{K})\colon a \in \mathrm{GL}_p(\mathbb{K}), d \in \mathrm{GL}_q(\mathbb{K}), c \in M_{q,p}(\mathbb{K}) \right\}.$$

Then

$$G := \left\{ g = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} : a \in \mathrm{GL}_p(\mathbb{K}), d \in \mathrm{GL}_q(\mathbb{K}) \right\} \cong \mathrm{GL}_p(\mathbb{K}) \times \mathrm{GL}_q(\mathbb{K})$$

is a subgroup, and

$$N := \left\{ g = \begin{pmatrix} \mathbf{1} & 0 \\ c & \mathbf{1} \end{pmatrix} : c \in M_{q,p}(\mathbb{K}) \right\} \cong (M_{q,p}(\mathbb{K}), +)$$

is a normal subgroup because it is the kernel of the homomorphism

$$\widehat{G} \to G, \quad \begin{pmatrix} a & 0 \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}.$$

In this case the multiplication map

$$N \times G \to \widehat{G}, \quad \left(\begin{pmatrix} \mathbf{1} & 0 \\ c & \mathbf{1} \end{pmatrix}, \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \right) \mapsto \begin{pmatrix} \mathbf{1} & 0 \\ c & \mathbf{1} \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} = \begin{pmatrix} a & 0 \\ ca & d \end{pmatrix}$$

is bijective. We thus obtain

$$\widehat{G} \cong N \rtimes_\alpha G \cong M_{q,p}(\mathbb{K}) \rtimes_\alpha (\mathrm{GL}_p(\mathbb{K}) \times \mathrm{GL}_q(\mathbb{K})), \quad \text{where} \quad \alpha(a,d)(c) := dca^{-1}.$$

(b) The group

$$\widehat{G} := \{\phi(x) = ax + b \colon a \in \mathbb{K}^\times, b \in \mathbb{K}\}$$

of all affine automorphism of the affine line $\mathbb{A}_1 = (\mathbb{K}, \mathbb{K}, +)$ has the subgroup

$$G := \{\phi(x) = ax \colon a \in \mathbb{K}^\times\} \cong (\mathbb{K}^\times, \cdot)$$

of all linear automorphisms of $\mathbb{K}$ and the normal subgroup

$$N := \{\phi(x) = x + b \colon b \in \mathbb{K}\} \cong (\mathbb{K}, +)$$

of all translations. We have already seen that $\widehat{G} \cong N \rtimes_\alpha G \cong \mathbb{K} \rtimes_\alpha \mathbb{K}^\times$ with $\alpha_a(b) = ab$.

(c) The cyclic group $C_4 := \{z \in \mathbb{C}^\times \colon z^4 = 1\}$ has a surjective homomorphism

$$q \colon C_4 \to C_2, \quad z \mapsto z^2$$

with kernel $N := \ker q = \{\pm 1\}$. However, there is no complementary subgroup G with $\widehat{G} \cong N \rtimes_\alpha G$ because every element $g \in C_4 \setminus N$ has order 4, hence generates the whole group C_4. Put more formally, the surjective homomorphism $q \colon C_4 \to C_2$ does not split in the sense that there exists a homomorphism $\sigma \colon C_2 \to C_4$ with $q \circ \sigma = \mathrm{id}_{C_2}$.

1.4 Affine coordinates

Definition 1.14. Let $(A, V, +)$ be an affine space. Note that A is allowed to be empty. If this is not the case, we call

$$\dim A := \dim V \in \mathbb{N}_0 \cup \{\infty\}$$

the *dimension of* A. An *affine line* is an affine space of dimension 1 and an *affine plane* is an affine space of dimension 2.

Example 1.15. If $(A, V, +)$ is an affine space of dimension 0, then $V = \{0\}$ and (A1/3) implies that A consists of a single point: Points are affine spaces of dimension 0.

In vector spaces we can form linear combinations of elements. In an affine space, there is a similar concept, but here the sum of all scalars must be one to obtain a well-defined expression.

Definition 1.16. (Affine combinations) Let $(A, V, +)$ be an affine space, $a_1, \ldots, a_n \in A$ and $t_1, \ldots, t_n \in \mathbb{K}$ with $\sum_{j=1}^n t_j = 1$. We want to define the corresponding *affine combination* by

$$t_1 a_1 + \cdots + t_n a_n := a_0 + \sum_{j=1}^n t_j \overrightarrow{a_0 a_j},$$

where $a_0 \in A$ is an arbitrary element. To see that this makes sense, we first have to show that the right hand side does not depend on the choice of the point a_0. In fact, for any other point $a_0' \in A$ we have

$$\begin{aligned} a_0' + \sum_{j=1}^n t_j \overrightarrow{a_0' a_j} &= a_0 + \overrightarrow{a_0 a_0'} + \sum_{j=1}^n t_j \overrightarrow{a_0' a_j} = a_0 + \sum_{j=1}^n t_j \overrightarrow{a_0 a_0'} + \sum_{j=1}^n t_j \overrightarrow{a_0' a_j} \\ &= a_0 + \sum_{j=1}^n t_j (\overrightarrow{a_0 a_0'} + \overrightarrow{a_0' a_j}) = a_0 + \sum_{j=1}^n t_j \overrightarrow{a_0 a_j}. \end{aligned}$$

Therefore our definition of affine combinations makes sense.

Remark 1.17. In the context of the preceding definition, we find for $a \in A$, $t_1, \ldots, t_n \in \mathbb{K}$ with $\sum_{j=1}^n t_j = 1$ and $v_i \in V$ the relation

$$\sum_{j=1}^n t_i (a + v_j) = a + \sum_{j=1}^n t_i v_j \tag{1}$$

by applying the definition of affine combinations with $a_0 = a$.

Lemma 1.18. *Affine maps are compatible with affine combinations: If $\phi \colon A \to B$ is affine, $a_1, \ldots, a_n \in A$ and $t_1, \ldots, t_n \in \mathbb{K}$ with $\sum_{j=1}^n t_j = 1$, then*

$$\phi\Big(\sum_{j=1}^n t_j a_j\Big) = \sum_{j=1}^n t_j \phi(a_j). \tag{2}$$

If, conversely, (2) *is always satisfied, then ϕ is affine.*

Proof. Fix $a_0 \in A$. If ϕ is affine, then $\phi(a_0 + v) = \phi(a_0) + \phi_L(v)$ for $v \in \overrightarrow{A}$. Writing $a_j = a_0 + v_j$, we obtain in particular

$$\begin{aligned} \phi\Big(\sum_{j=1}^n t_j a_j\Big) &= \phi\Big(a_0 + \sum_{j=1}^n t_j v_j\Big) = \phi(a_0) + \phi_L\Big(\sum_{j=1}^n t_j v_j\Big) = \phi(a_0) + \sum_{j=1}^n t_j \phi_L(v_j) \\ &\overset{(1)}{=} \sum_{j=1}^n t_j (\phi(a_0) + \phi_L(v_j)) = \sum_{j=1}^n t_j \phi(a_j). \end{aligned}$$

Now we assume that (2) is satisfied for the map $\phi \colon A \to B$ and that $a_0 \in A$. We consider the map

$$\psi \colon \overrightarrow{A} \to \overrightarrow{B}, \quad \psi(v) := \overrightarrow{\phi(a_0)\phi(a_0 + v)} \quad \text{with} \quad \phi(a_0 + v) = \phi(a_0) + \psi(v).$$

Then $\psi(0) = 0$, and for $v, w \in \overrightarrow{A}$, $\lambda, \mu \in \mathbb{K}$, we obtain

$$\begin{aligned}\phi(a_0) + \psi(\lambda v + \mu w) = \phi(a_0 + \lambda v + \mu w) &\overset{(1)}{=} \phi((1 - \lambda - \mu)a_0 + \lambda(a_0 + v) + \mu(a_0 + w)) \\ &\overset{(2)}{=} (1 - \lambda - \mu)\phi(a_0) + \lambda\phi(a_0 + v) + \mu\phi(a_0 + w) \\ &= (1 - \lambda - \mu)\phi(a_0) + \lambda(\phi(a_0) + \psi(v)) + \mu(\phi(a_0) + \psi(w)) \\ &= \phi(a_0) + \lambda\psi(v) + \mu\psi(w).\end{aligned}$$

This shows that

$$\psi(\lambda v + \mu w) = \lambda\psi(v) + \mu\psi(w),$$

so that ψ is linear, and hence ϕ is affine. □

Definition 1.19. (Affine basis/affine independence) Let $(A, V, +)$ be an affine space. We call an $(n+1)$-tuple $(a_0, \ldots, a_n)$ of points in A an *affine basis*, resp., *affinely independent*, if the vectors

$$\mathbf{b}_j := \overrightarrow{a_0 a_j}, \quad j = 1, \ldots, n$$

form a basis of the vector space V, resp., are linearly independent.

We now turn to two kinds of coordinates for elements of an affine space.

Definition 1.20. (Affine coordinates) Let $(A, V, +)$ be an affine space of dimension n and $(a_0, \ldots, a_n)$ an affine basis of A.

(a) Then there exists for every $a \in V$ uniquely determined numbers $x_1, \ldots, x_n$ in $\mathbb{K}$ with

$$a = a_0 + x_1 \overrightarrow{a_0 a_1} + \cdots + x_n \overrightarrow{a_0 a_n}.$$

We call $(x_1, \ldots, x_n)$ the *affine coordinates of a with respect to* $(a_0, \ldots, a_n)$.

The point a_0 is called the *origin*, its coordinates are $(0, \ldots, 0)$. The points a_j, $j = 1, \ldots, n$ have the coordinates $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ (1 in the jth place).

(b) For $x_0 := 1 - \sum_{j=1}^n x_j$, we obtain the relation

$$a = a_0 + x_1 \overrightarrow{a_0 a_1} + \cdots + x_n \overrightarrow{a_0 a_n} = x_0 a_0 + x_1 a_1 + \cdots + x_n a_n.$$

The $(n+1)$-tuple $(x_0, \ldots, x_n) \in \mathbb{K}^{n+1}$ is called the *barycentric coordinates of a with respect to the affine basis.*

Although the barycentric coordinates have the unpleasant feature of satisfying the equation $\sum_{j=0}^n x_j = 1$, they are often better adapted to the context of affine spaces than the affine coordinates $(x_1, \ldots, x_n)$ are.

Remark 1.21. (Changing coordinates) Suppose that $(a_0, \ldots, a_n)$ and $(a_0', \ldots, a_n')$ are affine bases of A. We put $\mathbf{b}_j := \overrightarrow{a_0 a_j}$ and $\mathbf{b}_j' := \overrightarrow{a_0' a_j'}$. Then there exist numbers x_j^0 and s_{ij} with

$$a_0' = a_0 + \sum_{j=1}^n x_j^0 \mathbf{b}_j \quad \text{and} \quad \mathbf{b}_j' := \sum_{i=1}^n s_{ij}\mathbf{b}_i.$$

We now obtain

$$a_0 + \sum_{j=1}^n x_j \mathbf{b}_j = a = a_0' + \sum_{j=1}^n x_j' \mathbf{b}_j' = a_0 + \sum_{j=1}^n x_j^0 \mathbf{b}_j + \sum_{i,j=1}^n s_{ij} x_j' \mathbf{b}_i = a_0 + \sum_{j=1}^n x_j^0 \mathbf{b}_j + \sum_{i,j=1}^n s_{ji} x_i' \mathbf{b}_j$$

and therefore

$$\mathbf{x} = \mathbf{x}^0 + S\mathbf{x}' \quad \text{for} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{K}^n \quad \text{and} \quad S = (s_{ij}) \in \mathrm{GL}_n(\mathbb{K}).$$

This means that changing the affine basis leads to the following transformation of the coordinates

$$\mathbf{x} = \mathbf{x}_0 + S\mathbf{x}' \quad \text{and} \quad \mathbf{x}' = S^{-1}\mathbf{x} - S^{-1}\mathbf{x}_0.$$

Note that these transformations define affine maps on $\mathbb{K}^n$.

We conclude this subsection with two important fact that we shall repeatedly use later on.

Lemma 1.22. *If A is a finite-dimensional affine space and $(a_0, \ldots, a_k)$ is affinely independent, then there exist $a_{k+1}, \ldots, a_n$ such that $(a_0, \ldots, a_n)$ is an affine basis of A.*

Proof. Extend the linearly independent system $\mathbf{b}_j := \overrightarrow{a_0 a_j}$, $j = 1, \ldots, k$ of $\overrightarrow{A}$ to a basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ and put $a_j := a_0 + \mathbf{b}_j$ for $j = k+1, \ldots, n$. □

Lemma 1.23. *Let A and B be affine spaces. If $(a_0, \ldots, a_n)$ is an affine basis of A and $b_0, \ldots, b_n \in B$ arbitrary elements, then there exists a unique affine map $\phi\colon A \to B$ with $\phi(a_j) = b_j$ for $j = 0, \ldots, n$.*

Proof. Put $\psi(a_0 + v) := b_0 + \psi(v)$, where $\psi\colon \overrightarrow{A} \to \overrightarrow{B}$ is the unique linear map with $\psi(\overrightarrow{a_0 a_j}) = \overrightarrow{b_0 b_j}$ for $j = 1, \ldots, n$ (Existence follows from Linear Algebra). Then $\phi(a_j) = b_j$, and that ϕ is uniquely determined by this property follows from Lemma 1.18 and the existence of affine coordinates (Definition 1.20). □

1.5 Affine subspaces

Definition 1.24. Let $(A, V, +)$ be an affine space. A subset $B \subseteq A$ is called an *affine subspace* if, for $b, b', b'' \in B$, we have

$$b + \mathbb{K} \cdot \overrightarrow{b'b''} \subseteq B.$$

The following lemma links our concept of an affine subspace to the well-known form of an affine subspace in the context of Linear Algebra.

Lemma 1.25. *If $B \subseteq A$ is a non-empty affine subspace, then*

$$U_B := \{u \in V : B + u \subseteq B\}$$

is a linear subspace of V, and, for every $b \in B$, we have

$$B = b + U_B.$$

Conversely, for every point $b \in A$ and every linear subspace $U \subseteq V$, the subset $B := b + U$ is an affine subspace with translation space $U = U_B$.

Proof. Clearly, $0 \in U$. For $u, u' \in U_B$, we have

$$B + (u + u') = (B + u) + u' \subseteq B + u' \subseteq B,$$

so that $U_B + U_B \subseteq U_B$. If $b \in B$, $u \in U_B$ and $b' := b + u$, then $u = \overrightarrow{bb'}$ and therefore $\mathbb{K}u \subseteq U_B$. This implies that $\mathbb{K}U_B = U_B$, i.e., that U_B is a linear subspace of V.

Now fix $b \in B$. We have seen above that every $u \in U_B$ is of the form $u = \overrightarrow{bb'}$ for some $b' \in B$. This implies that $B \subseteq b + U_B \subseteq B$, and hence that $B = b + U_B$.

Finally, we consider a subset $B \subseteq A$ of the form $B = b + U$ for some $b \in A$ and a linear subspace $U \subseteq V$. Then $\overrightarrow{b'b''} \in U$ for $b', b'' \in B$ follows from $\overrightarrow{b'b''} = \overrightarrow{bb''} - \overrightarrow{bb'}$. This implies that B is an affine subspace, and it is clear that $U_B = U$. $\square$

As a consequence of the preceding lemma, we immediately obtain:

Lemma 1.26. *Let $(A, V, +)$ be an affine space. Every affine subspace $B \subseteq A$ is an affine space with translation group U_B with respect to the restriction of $+$ to $B \times U_B$.*

Definition 1.27. Two non-empty affine subspaces $B_1, B_2 \subseteq A$ are called *parallel* if $U_{B_1} = U_{B_2}$. This means that there exist $a_1, a_2 \in A$ such that $U := U_{B_1} = U_{B_2}$ satisfies

$$B_1 = a_1 + U \quad \text{and} \quad B_2 = a_2 + U.$$

Lemma 1.28. *Parallel affine subspaces are either disjoint or equal.*

Proof. If $a \in (b + U) \cap (c + U)$, then Lemma 1.25 implies that $b + U = a + U = c + U$. $\square$

Example 1.29. (a) Points $B = \{a\}$ are affine subspaces of dimension 0.

(b) For $a \neq b \in A$, the subset $a + \mathbb{K}\overrightarrow{ab}$ is called the *affine line $\overline{ab}$ through a and b*. It is an affine subspace of dimension 1.

Using affine combinations, we obtain for $\lambda \in \mathbb{K}$ the relation

$$a + \lambda\overrightarrow{ab} = a + (1 - \lambda)\overrightarrow{aa} + \lambda\overrightarrow{ab} = (1 - \lambda)a + \lambda b.$$

For $\lambda = 0$ we obtain the point a, and for $\lambda = 1$ the point b. Therefore the affine line generated by a and b is the set of points

$$\overline{ab} = \{\lambda a + \mu b \colon \lambda, \mu \in \mathbb{K}, \lambda + \mu = 1\}.$$

Remark 1.30. After these preparations, we can rephrase the definition of an affine subspace $B \subseteq A$ as follows: For $b, b', b'' \in B$ with $b' \neq b''$ and $v := \overrightarrow{b'b''}$, the line $b + \mathbb{K}v$ through b which is parallel to the line $\overline{b'b''}$ is contained in B.

Proposition 1.31. *Let $(A, V, +)$ be an affine space and $B \subseteq A$. Then the following are equivalent:*

(i) *B is an affine subspace.*

(ii) *For $b_1, \ldots, b_n \in B$, all affine combinations of these elements belong to B.*

(iii) *(for $|\mathbb{K}| > 2$) With two elements $b_1, b_2 \in B$, the line $\overline{b_1b_2}$ is also contained in B.*

Proof. Since (i)-(iii) are trivially satisfied for $B = \emptyset$, we may assume that B is non-empty.

(i) $\Rightarrow$ (ii): We write $B = b_0 + U_B$. For $b_1, \ldots, b_n \in B$ and $t_1, \ldots, t_n \in \mathbb{K}$ with $\sum_{j=1}^n t_j = 1$, we then have

$$\sum_{j=1}^n t_j b_j = b_0 + \sum_{j=1}^n t_j \overrightarrow{b_0 b_j} \in b_0 + U = B.$$

(ii) $\Rightarrow$ (i): Let $b, b', b'' \in B$ and $\lambda \in \mathbb{K}$. Then

$$b + \lambda \overrightarrow{b'b''} = b + 1 \cdot \overrightarrow{bb} - \lambda \overrightarrow{bb'} + \lambda \overrightarrow{bb''} = 1 \cdot b - \lambda b' + \lambda b'' \in B.$$

(ii) $\Rightarrow$ (iii) is trivial because $\overline{b_1 b_2}$ consists of affine combinations of b_1 and b_2 (Example 1.29(b)).

(iii) $\Rightarrow$ (i): Suppose that $\mathbb{K}$ has at least three elements, i.e., there exists an element $\mu \in \mathbb{K} \setminus \{0, 1\}$. Let $b, b', b'' \in B$ and $\lambda \in \mathbb{K}$. Then

$$b + \lambda \overrightarrow{b'b''} = b + \lambda \overrightarrow{bb''} - \lambda \overrightarrow{bb'} = \mu \underbrace{\left(b + \frac{\lambda}{\mu} \overrightarrow{bb''}\right)}_{\in B} + (1 - \mu) \underbrace{\left(b - \frac{\lambda}{1 - \mu} \overrightarrow{bb'}\right)}_{\in B} \in B. \qquad \square$$

1.6 Affine geometric properties of tuples

In this subsection we discuss transitivity properties of the affine group $\mathrm{Aut}(A)$ on the finite-dimensional affine space A. We shall see that it is 2-transitive in the sense that it acts transitively of the set of all pairs $(a, b) \in A^2$ with $a \neq b$. This implies that pairs (a, b) of different points in an affine space have no affine geometric property. This changes when we consider triples. The affine group does not act transitively on all non-degenerate triples because it preserves the collinearity relation. On the set of collinear triples we shall see that the classification of the $\mathrm{Aut}(A)$-orbits on this space leads to the notion of "proportions", resp., ratios, as affine geometric properties. Further, the fact that $\mathrm{Aut}(A)$ acts transitively on the set of affinely independent triples, implies that triangles have no affine geometric properties.

Definition 1.32. Let X be a set and G be a group. A map $\sigma\colon G \times X \to X, (g.x) \mapsto \sigma_g(x) = g.x$ is called an *action of* G *on* X if the following two conditions are satisfied:

(GA1) $\sigma_{\mathbf{1}} = \mathrm{id}_X$, and

(GA2) $\sigma_{gh} = \sigma_g \circ \sigma_h$ for $g, h \in G$.

Note that this implies that the map $G \to S_X, g \mapsto \sigma_g$ is a group homomorphism from G into the group S_X of all permutations (=bijections) of X. Conversely, we obtain from any such homomorphism $\sigma\colon G \to S_X$ an action $g.x := \sigma_g(x)$.

Examples 1.33. (a) For any subgroup $G \subseteq S_X$, the map $\sigma_g(x) := g(x)$ defines an action of G on X.

More concrete examples are $G = \mathrm{GL}(V)$ for a vector space $X = V$ and $G = \mathrm{Aut}(A)$ for an affine space $X = A$.

(b) For any group G, we obtain an action of G on $X = G$ by $\sigma_g(x) := gxg^{-1}$. It is called the *conjugation action*.

(c) For $G = \mathrm{GL}_n(\mathbb{K})$ and $X = \mathbb{K}^n$ we have a natural action given by $\sigma_g(x) = gx$ (matrix multiplication).

Definition 1.34. Let $\sigma\colon G\times X\to X$ be an action of G on X. For $x\in X$, we call the subset

$$\mathcal{O}_x := G.x := \{g.x\colon g\in G\} = \{\sigma_g(x)\colon g\in G\}$$

the *orbit of* x.

The action is said to be *transitive* if there is only one orbit. Since the orbits form a partition of X (Exercise 1.11), this is equivalent to the requirement that, for $x,y\in X$, there exists a $g\in G$ with $g.x=y$

Definition 1.35. Let $\sigma\colon G\times X\to X, (g,x)\mapsto\sigma_g(x)$ be an action of the group G on the set X. Then we obtain for every $n\in\mathbb{N}$ an action

$$\sigma^n\colon G\times X^n\to X^n,\quad \sigma^n_g(x_1,\ldots,x_n)=(\sigma_g(x_1),\ldots,\sigma_g(x_n)).$$

Let $X^n_\times\subseteq X^n$ be the set of all n-tuples $(x_1,\ldots,x_n)$ for which $x_1,\ldots,x_n$ are pairwise different. We say that G acts *n-transitively* on X if the action σ^n on $X^n_\times$ is transitive. It acts *sharply n-transitively* if, in addition, the stabilizer subgroup of any element in $X^n_\times$ is trivial.

Remark 1.36. (a) If $|X|\geq 2$, then the action σ^2 on the set X^2 of pairs preserves the diagonal

$$\Delta_X := \{(x,x)\colon x\in X\},$$

so that it cannot act transitively on X^2. The partition $X^2=\Delta_X\dot\cup X^2_\times$ is G-invariant.

(b) The set X^3 contains many "partial diagonals":

$$\Delta_X^{1,2} := \{(x,x,y)\colon x\neq y\in X\},\quad \Delta_X^{1,3} := \{(x,y,x)\colon x\neq y\in X\},$$

and

$$\Delta_X^{2,3} := \{(y,x,x)\colon x\neq y\in X\},$$

and all these subsets are invariant under the G-action defined by σ^3. This leads to the partition

$$X^3=\Delta_X\dot\cup\Delta_X^{1,2}\dot\cup\Delta_X^{1,3}\dot\cup\Delta_X^{2,3}\dot\cup X^3_\times.$$

Suppose that G is 3-transitive on X, i.e., transitive on $X^3_\times$, and that this set is non-empty, i.e., $|X|\geq 3$. Then G acts transitively on all the sets $\Delta_X^{i,j}$ (Exercise), so that we obtain 5 orbits in X^3.

Lemma 1.37. *If the G-action on X is n-transitive and $|X|\geq n$, then it is k-transitive for every $k\leq n$.*

Proof. For $(x_1,\ldots,x_k),(y_1,\ldots,y_k)$ in $X^k_\times$, we find points $x_{k+1},\ldots,x_n\in X$ such that $\mathbf{x}:=(x_1,\ldots,x_n)\in X^n_\times$ and, likewise, $y_{k+1},\ldots,y_n\in X$ with $\mathbf{y}:=(y_1,\ldots,y_n)\in X^n_\times$. Let $g\in G$ with $\sigma^n_g(\mathbf{x})=\mathbf{y}$. Then $\sigma_g(x_j)=y_j$ for $j=1,\ldots,k$. $\square$

After these generalities, we now turn to the action of the affine group $\mathrm{Aut}(A)$ on a finite-dimensional affine space A.

Proposition 1.38. *Let A be an n-dimensional affine space. Then, for every $k\leq n+1$, the group $\mathrm{Aut}(A)$ acts transitively on the subset A^k_{indep} of affinely independent k-tuples in A^k.*

For $k=n+1$, we thus obtain a sharply transitive action on the set A^{n+1}_{indep} of all affine bases, i.e., for two affine bases $(a_0,\ldots,a_n)$ and $(a_0',\ldots,a_n')$, there exists a unique $\phi\in\mathrm{Aut}(A)$ with $\phi(a_j)=a_j'$ for $j=0,\ldots,n$.

Proof. Let $(a_0, \dots, a_{k-1})$ and $(a'_0, \dots, a'_{k-1})$ be affinely independent. Let $\mathbf{b}_j := \overrightarrow{a_0 a_j}$ and $\mathbf{b}'_j := \overrightarrow{a'_0 a'_j}$ be the corresponding vectors in V. Then $\mathbf{b}_1, \dots, \mathbf{b}_{k-1}$ are linearly independent in V, hence can be enlarged to a linear basis $b_1, \dots, b_n$, and then $a_j := a_0 + \mathbf{b}_j$, $j = k, \dots, n$, leads to an affine basis $(a_0, \dots, a_n)$. We likewise obtain an affine basis $(a'_0, \dots, a'_n)$. If $\psi \in \mathrm{GL}(V)$ is a linear isomorphism with $\psi(\mathbf{b}_j) = \mathbf{b}'_j$ for $j = 1, \dots, n$, then

$$\phi(a_0 + v) := a'_0 + \psi(v)$$

is an affine automorphism of A mapping each a_j, $j = 0, \dots, n$, to a'_j. This proves for each k the transitivity of $\mathrm{Aut}(A)$ on the set of affinely independent k-tuples.

To see that this action is sharply transitive for $k = n+1$, we note that the uniqueness of $\phi \in \mathrm{Aut}(A)$ mapping the affine basis $(a_0, \dots, a_n)$ to the affine basis $(a'_0, \dots, a'_n)$ follows from Lemma 1.23. $\square$

Since a pair $(a, b) \in A^2$ is affinely independent if and only if $a \neq b$, we have $A^2_\times = A^2_{\mathrm{indep}}$. For an affine line, this is the set of affine bases. Hence the preceding proposition implies:

Corollary 1.39. *Let A be a finite-dimensional affine space. Then the action of* $\mathrm{Aut}(A)$ *on A is 2-transitive. It is sharply transitive for* $\dim A = 1$*, i.e., if A is an affine line.*

Proposition 1.40. *Let A be a finite-dimensional affine space and $B_1, B_2 \subseteq A$ affine subspaces. Then every affine isomorphism $\psi\colon B_1 \to B_2$ extends to an affine isomorphism $\phi \in$* $\mathrm{Aut}(A)$*. In particular, for $k \in \mathbb{N}$, the group* $\mathrm{Aut}(A)$ *acts transitively on the set of all k-dimensional affine subspaces of A.*

Proof. Clearly, the second part follows from the first because affine subspaces of the same dimension k are isomorphic to $\mathbb{A}_k$ (Example 1.8(a)), hence in particular isomorphic.

So let $\psi\colon B_1 \to B_2$ be an affine isomorphism of affine k-dimensional subspaces of A. Let $(a_0, \dots, a_k)$ be an affine basis of B_1 and enlarge it to an affine basis $(a_0, \dots, a_n)$ of A (Lemma 1.22). We likewise enlarge the affine basis $(a'_0, \dots, a'_k) := (\psi(a_0), \dots, \psi(a_k))$ of B_2 to an affine basis $(a'_0, \dots, a'_n)$ of A. Then there exists a unique affine isomorphism $\phi \in \mathrm{Aut}(A)$ with $\phi(a_j) = a'_j$ for $j = 0, \dots, n$ (Proposition 1.38). This implies in particular that $\phi(B_1) = B_2$ with $\phi|_{B_1} = \psi$. $\square$

Definition 1.41. We call a subset $E \subseteq A$ *collinear* if $|E| \leq 1$ or E is contained in an affine line.

Proposition 1.38 implies in particular that affinely independent tuples carry no affine geometric information, they are all conjugate under the affine group $\mathrm{Aut}(A)$. For pairs (a, b), being different is the same as being affinely independent, so that the first non-trivial situation arises for triples (a_0, a_1, a_2) that are collinear. Since we know already the $\mathrm{Aut}(A)$-orbits in A^2 (Corollary 1.39), it suffices to consider triples (a_0, a_1, a_2) with $a_0 \neq a_1$. Let $A^3_r \subseteq A^3$ denote the set of all these triples. To any triple $(a_0, a_1, a_2) \in A^3_r$, we associate its *ratio*:

$$r(a_0, a_1, a_2) := t \in \mathbb{K} \quad \text{where} \quad a_2 = (1-t)a_0 + t a_1.$$

Proposition 1.42. (The ratio as an affine geometric property) *Two triples (a_0, a_1, a_2), $(a'_0, a'_1, a'_2) \in A^3_r$ are contained in the same* $\mathrm{Aut}(A)$*-orbit if and only if they have the same ratio.*

Proof. The function $r\colon A_r^3 \to \mathbb{K}$ is invariant under the action of $\mathrm{Aut}(A)$ because $a_2 = (1-t)a_0 + ta_1$ implies $\phi(a_2) = (1-t)\phi(a_0) + t\phi(a_1)$. If, conversely, the triples $(a_0, a_1, a_2), (a_0', a_1', a_2') \in A_r^3$ have the same ratio t, then there exists a $\phi \in \mathrm{Aut}(A)$ with $\phi(a_0) = a_0'$ and $\phi(a_1) = a_1'$ (Corollary 1.39), and this implies that

$$\phi(a_2) = \phi((1-t)a_0 + ta_1) = (1-t)\phi(a_0) + t\phi(a_1) = (1-t)a_0' + ta_1' = a_2'.$$

Therefore $\mathrm{Aut}(A)$ acts transitively on the subsets of triples with the same ratio. □

For $(a_0, a_1, a_2) \in A_r^3$, the ratio $t = r(a_0, a_1, a_2)$ can be computed by

$$\overrightarrow{a_0 a_2} = t\overrightarrow{a_0 a_1},$$

so it can be interpreted as a scale factor corresponding to the *relative length* of the vector $\overrightarrow{a_0 a_2}$ with respect to the gauge $\overrightarrow{a_0 a_1}$. In this sense the "ratio" of two collinear line segments is an affine geometric concept.

Exercises for Section 1

Exercise 1.1. Show that every bijective affine map $\phi\colon A \to B$ between affine spaces A and B is an isomorphism, i.e., there exists an affine map $\psi\colon B \to A$ with $\phi \circ \psi = \mathrm{id}_B$ and $\psi \circ \phi = \mathrm{id}_A$.

Exercise 1.2. Let N and G be groups and $\alpha\colon G \to \mathrm{Aut}(N), g \mapsto \alpha_g$ be a group homomorphism. Then we obtain on the set $G \times N$ a group structure by

$$(g, n)(g', n') := (gg', \alpha_{g'}^{-1}(n)n'),$$

and this group is denoted $G \ltimes_\alpha N$. Show also that

$$\Phi\colon N \rtimes_\alpha G \to G \ltimes_\alpha N, \quad (n, g) \mapsto (g, \alpha_g^{-1}(n))$$

is an isomorphism of groups.

Exercise 1.3. Show that

$$G := \left\{ \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_{n+1}(\mathbb{K})\colon A \in \mathrm{GL}_n(\mathbb{K}), b \in \mathbb{K}^n \right\}$$

is a subgroup of $\mathrm{GL}_{n+1}(\mathbb{K})$ and that G is isomorphic to the automorphism group $\mathrm{Aut}(\mathbb{A}_n)$ of the n-dimensional affine space $\mathbb{A}_n := (\mathbb{K}^n, \mathbb{K}^n, +)$ over $\mathbb{K}$.

Exercise 1.4. (Affine group action) Let G be a group and $A := (V, V, +)$ the affine space underlying the vector space V. An *affine action of* G *on* V is a group action $\sigma\colon G \times V \to V, (g, v) \mapsto g.v$ for which all maps $\sigma_g(v) := g.v$ are affine. [2]

(a) Show that affine actions of G on A are in one-to-one correspondence with homomorphisms $\gamma\colon G \to \mathrm{Aut}(A)$, the automorphism group of the affine space.

[2] Recall that, for a set X and a group G, a map $\sigma\colon G \times X \to X, (g, x) \mapsto g.x = \sigma_g(x)$ is called a *group action* if the map $g \mapsto \sigma_g$ defines a homomorphism of G into the group S_X of all permutations (=bijections) of X.

(b) Writing
$$\gamma(g)(v) = \pi(g)v + \beta(g), \quad \pi(g) \in \mathrm{GL}(V), \beta(g) \in V,$$
show that γ is a group homomorphism if and only if $\pi\colon G \to \mathrm{GL}(V)$ is a group homomorphism and β satisfies the *cocycle condition*
$$\beta(gh) = \beta(g) + \pi(g)\beta(h) \quad \text{for all} \quad g, h \in G.$$

(c) Show that the affine action $g.v = \gamma(g)v = \pi(g)v + \beta(g)$ has a fixed point if and only if there exists an element $v_0 \in V$ with
$$\beta(g) = v_0 - \pi(g)v_0 \quad \text{for all} \quad g \in G.$$

Exercise 1.5. Let $\mathbb{K} = \{0, 1\}$ be the 2-element field and $(A, V, +)$ an affine space over $\mathbb{K}$.

(a) Show that every subset $B \subseteq A$ has the property that $a, b \in B$ implies $\overline{ab} \subseteq B$.

(b) Which affine spaces V over $\mathbb{K}$ contain subsets that are not affine subspaces?

Exercise 1.6. Let $\phi\colon A \to B$ be an affine map. Show that:

(a) If $C \subseteq A$ is an affine subspace, then $\phi(C) \subseteq B$ is an affine subspace.

(b) If $C \subseteq B$ is an affine subspace, then $\phi^{-1}(C) \subseteq A$ is an affine subspace.

Exercise 1.7. Let A be an n-dimensional affine space and B be an m-dimensional affine space. We fix affine bases $a_0, \ldots, a_n$ in A and $b_0, \ldots, b_m$ in B. Show that, for every affine map $\phi\colon A \to B$, there exists a uniquely determined matrix $S = (s_{ij}) \in M_{m,n}(\mathbb{K})$ with
$$\phi\Big(\sum_{i=0}^{n} x_i a_i\Big) = \sum_{j=0}^{n}\sum_{i=0}^{n} s_{ji} x_i b_j.$$
Show further that this matrix satisfies
$$\sum_j s_{ji} = 1 \quad \text{for} \quad i = 1, \ldots, n,$$
and that every matrix satisfying this condition corresponds to an affine map $\phi\colon A \to B$.

Exercise 1.8. Let A and B be non-empty affine spaces and $\phi\colon A \to B$ be a map. Show that ϕ is affine if and only if its graph
$$\Gamma(\phi) := \{(a, \phi(a))\colon a \in A\} \subseteq A \times B$$
is an affine subspace with respect to the affine space structure on $A \times B$ with $\overrightarrow{A \times B} := \vec{A} \times \vec{B}$ and
$$(a, b) + (v, w) := (a + v, b + w), \quad a \in A, b \in B, v \in \vec{A}, w \in \vec{B}.$$

Exercise 1.9. Let A be an affine space. Show that:

(a) For every family $(B_j)_{j \in J}$ of affine subspaces of A the intersection $B := \bigcap_{j \in J} B_j$ is an affine subspace.

(b) Every subset $M \subseteq A$ is contained in a minimal affine subspace $\mathrm{span}_a(M)$. It coincides with the intersection of all affine subspaces of A containing M and the set of all affine combinations of elements of M.

Exercise 1.10. Let A and B be affine spaces and $\phi\colon A \to B$ be a map. Suppose that

$$\phi\Big(\sum_{j=1}^n t_j a_j\Big) = \sum_{j=1}^n t_j \phi(a_j) \tag{3}$$

holds for $a_1, \ldots, a_n \in A$ and $t_1, \ldots, t_n \in \mathbb{K}$ with $\sum_j t_j = 1$. Show that this implies that ϕ is affine. If, in addition, the field $\mathbb{K}$ has at least 3 elements, then it suffices that (3) holds for $n = 2$. Hint: Exercise 1.8.

Exercise 1.11. Let $\mu\colon G \times M \to M$ be an action of the group G on the set M.

(1) The orbits of G in M form a partition of M into pairwise disjoint subsets.

(2) For each $p \in M$ the set $G_p := \{g \in G\colon gp = p\}$ is a subgroup of G and the map

$$\sigma_p\colon G/G_p \to M, \quad gG_p \mapsto gp$$

is a well defined injective map which is equivariant, i.e., $\sigma_p(gxG_p) = g\sigma_p(xG_p)$ for all $g, x \in G$.

(3) The action of G on M is transitive if and only if there exists an equivariant bijection $\sigma\colon G/H \to M$ for some subgroup $H \subseteq G$.

Exercise 1.12. Recall the definition of an affinely independent tuple $(a_0, \ldots, a_k)$ from Definition 1.19. Show that this is equivalent to the condition that no a_j is contained in the affine subspace generated by $a_0, \ldots, a_{j-1}, a_{j+1}, \ldots, a_k$.

Exercise 1.13. Let $\mathbb{K}$ be a finite field with q elements and A an n-dimensional affine space over $\mathbb{K}$. Calculate the following numbers:

(a) the number $|A|$ of elements of A.

(b) the number $|A^k_{\mathrm{indep}}|$ of affinely independent k-tuples in A.

(c) the number $|\operatorname{Aut}(A)|$ of elements of the affine group.

Exercise 1.14. Suppose that the action $\sigma\colon G \times X \to X$ of the group G on the non-empty X is n-transitive and $|X| \geq n$. Show that X^n contains only finitely many G-orbits with respect to σ^n.

2 Euclidean Geometry

Euclidean geometry is the oldest branch of geometry. Its development started in the period between 350 and 200 BC. Up to that time people where dealing with mathematical facts as a collection of formulas and rules how to deal with numbers and geometric objects. A proper understanding of mathematics as a science with a clear (axiomatic) foundation from which mathematical truths can be derived by logical deduction emerged with Euclid's axiomatic approach to geometry in the plane. In the language used below, Euclid was dealing with the euclidean plane, i.e., a 2-dimensional affine euclidean space. Here we follow a more direct approach based on the concept of a euclidean vector space as a metric space. This permits us to define affine euclidean spaces and we recover the euclidean planes as the 2-dimensional objects in this theory. Restricting to the n-dimensional euclidean space $\mathbb{E}_n = \mathbb{R}^n$, this corresponds to the cartesian approach to euclidean geometry where points correspond to coordinate tuples. To keep the formulation of the results as "geometric" as possible, we try to avoid coordinates as much as possible because they are not invariant under the corresponding group, hence have no geometric meaning.

2.1 The euclidean metric

Euclidean metrics are defined in terms of a scalar product. We therefore recall briefly some concepts that are well-known from Linear Algebra.

Definition 2.1. A *euclidean vector space* E is a real vector space endowed with a positive definite symmetric bilinear form $(x, y) \mapsto \langle x, y\rangle$, called the scalar product.

Proposition 2.2. (Cauchy–Schwarz inequality) *In a euclidean vector space the Cauchy–Schwarz inequality holds:*

$$\langle x, y\rangle \leq \langle x, x\rangle^{\frac{1}{2}} \langle y, y\rangle^{\frac{1}{2}} \quad \text{for} \quad x, y \in E. \tag{CS}$$

If $0 \neq y$*, then equality holds in (CS) if and only if* $x \in \mathbb{R}_+ y$*.*

Proof. Fix $x, y \in E$. For $y = 0$, the CS inequality holds trivially, so that we may assume that $y \neq 0$. We consider the function

$$f\colon \mathbb{R} \to \mathbb{R}_+ = [0, \infty[, \quad f(\lambda) := \langle x - \lambda y, x - \lambda y\rangle.$$

Expanding the right hand side, we find

$$f(\lambda) = \langle x, x\rangle - 2\lambda\langle x, y\rangle + \lambda^2\langle y, y\rangle.$$

As $\langle y, y\rangle > 0$, this quadratic function has a unique minimal value at $\lambda_0 := \frac{\langle x,y\rangle}{\langle y,y\rangle}$:

$$0 \leq f(\lambda_0) = \langle x, x\rangle - 2\frac{\langle x, y\rangle^2}{\langle y, y\rangle} + \frac{\langle x, y\rangle^2}{\langle y, y\rangle} = \langle x, x\rangle - \frac{\langle x, y\rangle^2}{\langle y, y\rangle}.$$

Now the assertion follows by multiplication with $\langle y, y\rangle$.

If $x = \mu y$ for some $\mu \geq 0$, then we obtain the equality

$$\langle x, y\rangle = \mu\langle y, y\rangle = \langle x, x\rangle^{\frac{1}{2}} \langle y, y\rangle^{\frac{1}{2}}.$$

Suppose, conversely, that $y \neq 0$ and that

$$\langle x, y\rangle = \langle x, x\rangle^{\frac{1}{2}} \langle y, y\rangle^{\frac{1}{2}}.$$

Then $f(\lambda_0) = 0$ implies that $x - \lambda_0 y = 0$, i.e., $x = \lambda_0 y$, and since $\langle x, y\rangle \geq 0$, we obtain $\lambda_0 \geq 0$. $\square$

Lemma 2.3. *In a euclidean vector space E,*

$$\|x\| := \sqrt{\langle x, x\rangle}$$

defines a norm and

$$d(x,y) := \|x-y\| = \sqrt{\langle x-y, x-y\rangle}$$

defines a metric.

Proof. Clearly, $\|x\| \geq 0$, and $\|x\| = 0$ implies $x = 0$. Further, $\|\lambda x\| = |\lambda|\|x\|$ is immediate from the definition. From the Cauchy–Schwarz inequality, we further derive

$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y\rangle = \langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.\end{aligned}$$

This shows that $\|\cdot\|$ is a norm. Now it is clear that $d(x,y) := \|x-y\|$ is a metric. □

Definition 2.4. (Affine euclidean space) An affine space $\mathbb{E}$ whose translation group is a euclidean space E is called an *affine euclidean space*. Then we obtain a metric on $\mathbb{E}$ by

$$d(a,b) := \|\overrightarrow{ab}\|.$$

An affine basis $(a_0, a_1, \ldots, a_n)$ in $\mathbb{E}$ for which the vectors $\overrightarrow{a_0a_1}, \cdots, \overrightarrow{a_0a_n}$ form an orthonormal basis of E is called a *euclidean affine basis* or a *euclidean frame*.

Remark 2.5. (a) All translations $\tau_v(a) := a + v$ on an affine euclidean space are isometries:

$$d(a+v, b+v) = d(a+v, a+v+\overrightarrow{ab}) = \|\overrightarrow{ab}\| = d(a,b).$$

(b) Let p be a point in the euclidean affine space $\mathbb{E}$. Then the bijection

$$\eta_p \colon E \to \mathbb{E}, \quad v \mapsto p + v$$

is an affine isometry. Therefore E and $\mathbb{E}$ are isomorphic as affine euclidean spaces.

The following proposition shows that any affine euclidean space is metrically isomorphic to $\mathbb{R}^n$, endowed with the canonical metric.

Proposition 2.6. *If $\mathbb{E}$ is an affine euclidean space and $(a_0, a_1, \ldots, a_n)$ a euclidean frame in $\mathbb{E}$, then the corresponding coordinate map*

$$\Gamma \colon \mathbb{R}^n \to \mathbb{E}, \quad \Gamma(x) := a_0 + \sum_{j=1}^{n} x_j \overrightarrow{a_0a_j}$$

is an affine isometry with $\Gamma(0) = a_0$ and $\Gamma(\mathbf{e}_j) = a_j$ for $j = 1, \ldots, n$.

Lemma 2.7. (Triangle equality) *Three points $x, y, z \in \mathbb{E}$ satisfy the triangle equality*

$$d(x,y) + d(y,z) = d(x,z) \tag{4}$$

if and only if

$$y \in [x,z] := \{tx + (1-t)z \colon 0 \leq t \leq 1\},$$

i.e., y is contained in the line segment joining x and z.

Proof. If $y = tx + (1-t)z$ for some $t \in [0,1]$, then

$$d(x,y)+d(y,z) = \|(t-1)x+(1-t)z\|+\|tz-tx\| = (1-t)\|x-z\|+t\|x-z\| = \|x-z\| = d(x,z).$$

Suppose, conversely, that x, y, z satisfy the triangle equality (4). If $x = y$, then $y \in [x,z]$. We may therefore assume that $v := \overrightarrow{xy} \neq 0$. For $w := \overrightarrow{yz}$ we now obtain the equality

$$\|v\| + \|w\| = \|v + w\|.$$

Taking squares, we arrive at

$$\|v\|^2 + \|w\|^2 + 2\|v\|\|w\| = \|v\|^2 + \|w\|^2 + 2\langle v, w\rangle,$$

which leads to $\langle v, w\rangle = \|v\|\|w\|$. Now the second part of Proposition 2.2 on the Cauchy–Schwarz inequality implies that $w = \lambda v$ for some $\lambda \geq 0$. Hence $z = x + v + w = x + (1+\lambda)v$ and $y = x + v$ imply that $y \in [x,z]$. $\square$

Definition 2.8. In a euclidean vector space we define the *angle* between two non-zero vectors x, y by

$$\measuredangle(x,y) = \arccos\Big(\frac{\langle x,y\rangle}{\|x\| \cdot \|y\|}\Big),$$

where arccos is the inverse function of $\cos|_{[0,\pi]}$. According to the Cauchy–Schwarz inequality, the right hand side takes values in $[-1,1]$, so that $\measuredangle(x,y) \in [0,\pi]$ is defined. We say that x and y are *orthogonal* if $\measuredangle(x,y) = \frac{\pi}{2}$, i.e., if $\langle x, y\rangle = 0$.

The second part of Proposition 2.2 on the Cauchy–Schwarz inequality implies that $\measuredangle(x,y) \in \{0,\pi\}$ is equivalent to $x \in \mathbb{R}y$. If $x \in \mathbb{R}_+ y$ we have $\measuredangle(x,y) = 0$ and for $x \in -\mathbb{R}_+ y$ we have $\measuredangle(x,y) = \pi$.

In the axiomatic development of euclidean geometry one first develops the concept of the length of a line segment and further the concept of a right angle. In this context Pythagoras' Theorem is an important result derived from the axioms. Here it is a mere observation that follows from the description of affine euclidean spaces in terms of scalar products.

Lemma 2.9. (Pythagoras' Theorem) *If x, y, z are three points in an affine euclidean space and $\overrightarrow{xy} \perp \overrightarrow{yz}$, then*

$$d(x,z)^2 = d(x,y)^2 + d(y,z)^2.$$

Proof. For $v := \overrightarrow{xy}$ and $w := \overrightarrow{yz}$, this follows immediately by

$$d(x,z)^2 = \|v+w\|^2 = \langle v+w, v+w\rangle = \|v\|^2 + \|w\|^2 = d(x,y)^2 + d(y,z)^2. \qquad \square$$

2.2 The euclidean motion group $\mathrm{Mot}(\mathbb{E})$

In this subsection we show that all isometries of an affine euclidean space are affine maps. This means in particular that the group $\mathrm{Mot}(\mathbb{E}) := \mathrm{Isom}(\mathbb{E}, d)$ of all bijective isometries of $\mathbb{E}$, the *euclidean motion group*, is a subgroup of the affine group $\mathrm{Aut}(\mathbb{E})$. In the sense of the Erlangen Program, we thus consider euclidean geometry as a refinement of real affine geometry. Here the idea is that the subgroup $\mathrm{Mot}(\mathbb{E})$ of the affine group has more invariants, which leads to more euclidean geometric properties than affine ones.

Definition 2.10. For two points x, y in a metric space (X, d), a point $m \in X$ is called a *midpoint* if

$$d(x, m) = d(y, m) = \frac{1}{2} d(x, y).$$

Lemma 2.11. *In a euclidean affine space $\mathbb{E}$, two points $x, y \in \mathbb{E}$ have a unique midpoint. It is given by the arithmetic mean*

$$m := x + \tfrac{1}{2}\overrightarrow{xy} = y - \tfrac{1}{2}\overrightarrow{xy}.$$

Proof. Clearly

$$d(x, m) = \tfrac{1}{2}\|\overrightarrow{xy}\| = \tfrac{1}{2} d(x, y)$$

and likewise $d(y, m) = \frac{1}{2} d(x, y)$. Therefore m is a midpoint of x and y.

To prove uniqueness of midpoints, we may w.l.o.g. assume that $x \neq y$. Every midpoint z of x and y satisfies

$$d(x, y) = d(x, z) + d(z, y),$$

so that Lemma 2.7 implies the existence of some $t \in [0, 1]$ with $z = (1 - t)x + ty = x + t\overrightarrow{xy}$. Then

$$\frac{1}{2} d(x, y) = d(z, x) = t\|y - x\| = t d(x, y)$$

leads to $t = \frac{1}{2}$. □

The following theorem is one of a series of results asserting that, for several geometric structures endowed with a metric, any isometry automatically preserves the other structure. For affine euclidean spaces, this means that any isometry is automatically affine. This is also true for general normed spaces, but in this case it is slightly harder to prove (cf. Exercise 2.13).

Theorem 2.12. *Every isometry of an affine euclidean space $(\mathbb{E}, d)$ is affine. For bijective isometries we obtain in particular $\mathrm{Mot}(\mathbb{E}) \subseteq \mathrm{Aut}(\mathbb{E})$.*

Proof. Since $\mathbb{E}$ is metrically isomorphic to the euclidean vector space E (Remark 2.5(b)), we may w.l.o.g. assume that $\mathbb{E} = E$ is a euclidean vector space. Let $\phi \colon E \to E$ be isometric. Composing ϕ with the translation $\tau_{-\phi(0)}$, we may further assume that $\phi(0) = 0$. We now show that ϕ is linear.

For $x, y \in E$ with midpoint $m = \frac{1}{2}(x + y)$ we have

$$d(\phi(x), \phi(m)) = d(x, m) = \tfrac{1}{2} d(x, y) = \tfrac{1}{2} d(\phi(x), \phi(y)),$$

and likewise

$$d(\phi(y), \phi(m)) = d(y, m) = \tfrac{1}{2} d(\phi(x), \phi(y)).$$

Therefore $\phi(m)$ is the unique midpoint of $\phi(x)$ and $\phi(y)$ (Lemma 2.11), i.e.,

$$\phi(\tfrac{1}{2}(x + y)) = \tfrac{1}{2}(\phi(x) + \phi(y)), \quad x, y \in E. \tag{5}$$

For $y = 0$, we obtain in particular

$$\phi(\tfrac{1}{2}x) = \tfrac{1}{2}\phi(x) \quad \text{for} \quad x \in E. \tag{6}$$

Iterating this relation leads to

$$\phi\Big(\frac{1}{2^n} x\Big) = \frac{1}{2^n}\phi(x) \quad \text{for} \quad x \in E, n \in \mathbb{N}.$$

Combining (6) with (5) implies

$$\phi(x+y) = \phi(x) + \phi(y) \quad \text{for} \quad x, y \in E.$$

In particular, we get $\phi(mx) = m\phi(x)$ for $m \in \mathbb{Z}$. This leads to

$$\phi\Big(\frac{m}{2^n}x\Big) = \frac{m}{2^n}\phi(x) \quad \text{for} \quad x \in E, m \in \mathbb{Z}, n \in \mathbb{N}.$$

Since isometries are in particular continuous and the dyadic numbers $\frac{m}{2^n}$ form a dense subset of $\mathbb{R}$, we obtain

$$\phi(\lambda x) = \lambda\phi(x) \quad \text{for} \quad \lambda \in \mathbb{R}, x \in E.$$

This shows that ϕ is linear. □

Remark 2.13. If $\dim \mathbb{E} < \infty$, then every isometry $\phi\colon \mathbb{E} \to \mathbb{E}$ is surjective because the linear map $\phi_L\colon E \to E$ is injective, hence also surjective.

If $\mathbb{E}$ is infinite-dimensional, then there are isometries which are not surjective. A typical example is the right shift S on the space

$$E := \mathbb{R}^{(\mathbb{N})} = \{(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}\colon (\exists N \in \mathbb{N})(\forall n > N) a_n = 0\}$$

of all finite sequence, endowed with the canonical scalar product $\langle a, b\rangle := \sum_{n=1}^{\infty} a_n b_n$. It is defined by

$$S(a_1, a_2, \ldots) := (0, a_1, a_2, \ldots).$$

This map is a linear isometry which is not surjective.

Definition 2.14. (a) For a euclidean vector space E, we write

$$\mathrm{O}(E) := \{\psi \in \mathrm{GL}(E)\colon (\forall v, w \in E)\, \langle\psi(v), \psi(w)\rangle = \langle v, w\rangle\}$$

for the *orthogonal group of E*. In view of Exercise 2.14, we have

$$\mathrm{O}(E) := \{\psi \in \mathrm{GL}(E)\colon (\forall v \in E)\, \|\psi(v)\| = \|v\|\},$$

so that $\mathrm{O}(E)$ is the group of invertible linear isometries of E.

(b) If $E = \mathbb{R}^n$, then we identify linear endomorphisms with $(n\times n)$-matrices. Accordingly, we define the *orthogonal group*

$$\begin{aligned}\mathrm{O}_n(\mathbb{R}) &:= \{A \in M_n(\mathbb{R})\colon (\forall v, w \in \mathbb{R}^n)\, v^\top A^\top A w = v^\top w\}\\ &= \{A \in M_n(\mathbb{R})\colon A^\top A = \mathbf{1}\} = \{A \in \mathrm{GL}_n(\mathbb{R})\colon A^\top = A^{-1}\}.\end{aligned}$$

Observe that $\mathrm{O}(\mathbb{R}^n) \cong \mathrm{O}_n(\mathbb{R})$.

Example 2.15. (a) For the n-dimensional affine euclidean space $\mathbb{E}_n = \mathbb{R}^n$, Theorem 2.12 implies that every isometry is of the form

$$\phi(x) = Ax + b,$$

where the matrix $A \in \mathrm{GL}_n(\mathbb{R})$ corresponds to a length preserving linear map, i.e., $A \in \mathrm{O}_n(\mathbb{R})$ is an *orthogonal matrix*. We thus obtain

$$\mathrm{Mot}(\mathbb{R}^n, d) \cong \mathbb{R}^n \rtimes_\alpha \mathrm{O}_n(\mathbb{R}) \quad \text{with} \quad \alpha_g(x) = gx.$$

(b) If $\mathbb{E} = (E, E, +)$ is the affine space underlying a euclidean vector space E, then Theorem 2.12 implies that every surjective isometry is of the form

$$\phi(x) = \psi(x) + b, \quad \text{where} \quad \psi \in \mathrm{O}(E), b \in E.$$

Example 2.16. For $n = 1$ and $\mathbb{E} = \mathbb{R}$, we have $\mathrm{O}_1(\mathbb{R}) = \{\pm\mathbf{1}\}$, so that every isometry is either of the form $\phi(x) = x + b$ (a translation) or $\phi(x) = -x + b$, which is a reflection in the point $\frac{b}{2}$.

Example 2.17. For $n = 2$, we have

$$\mathrm{O}_2(\mathbb{R}) = \mathrm{O}_2(\mathbb{R})_+ \dot\cup \mathrm{O}_2(\mathbb{R})_-, \quad \text{where} \quad \mathrm{O}_2(\mathbb{R})_\pm := \{Q \in \mathrm{O}_2(\mathbb{R})\colon \det Q = \pm 1\}.$$

From Linear Algebra we know that, if $\det Q = 1$, then Q describes a rotation of the plane around the origin by some angle $\theta \in [0, 2\pi[$:

$$Q = D(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

If $\det Q = -1$, then

$$Q = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = D(\theta) \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

This describes a reflection in the line ℓ intersecting the x_1-axis in the angle $\frac{\theta}{2}$ (Exercise 2.5).

Theorem 2.18. (Geometric classification of planar isometries) *The isometries of the plane are of the form $\phi(x) = Qx + b$ for some $Q \in \mathrm{O}_2(\mathbb{R})$. The following, mutually exclusive, cases occur:*

(a) (Translations) *If $Q = \mathbf{1}$, then ϕ is a translation.*

(b) (Rotations) *If $Q \neq \mathbf{1}$ and $\det Q = 1$, then ϕ has a unique fixed point and ϕ is a rotation around this fixed point.*

(c) (Reflections) *If $\det Q = -1$ and ϕ has a fixed point, then ϕ is a reflection in a line.*

(d) (Glide reflections) *If $\det Q = -1$ and ϕ has no fixed point, then ϕ is a glide reflection in a line.*

Proof. (b) We first observe that the fixed point equation $\phi(x) = x$ has a unique solution x_0 because the matrix $\mathbf{1} - Q$ is invertible (it suffices to verify that its kernel is trivial because non-trivial rotations have no non-zero fixed vectors). Then $\phi(x_0 + v) = x_0 + Qv$ shows that ϕ is a rotation around x_0.

(c) If $\det Q = -1$ and ϕ has a fixed point x_0, then $\phi(x_0 + v) = x_0 + Qv$ is a reflection in the line $x_0 + \mathbb{R}v_1$, where v_1 is a non-zero fixed vector of Q.

(d) Let $v_\pm$ be an orthonormal basis of Q-eigenvectors for the eigenvalues ± 1. Then

$$\phi(x) = \phi(x_+ v_+ + x_- v_-) = x_+ v_+ - x_- v_- + b = (x_+ + b_+)v_+ + (b_- - x_-)v_-.$$

From this formula we see that $\phi(x)_- = x_-$ is equivalent to $x_- = \frac{b_-}{2}$ and, in general, the x_--component is reflected in $\frac{b_-}{2}$. The non-existence of fixed points therefore implies that $b_+ \neq 0$. Then $2x_- = b_-$ specifies a line ℓ which is ϕ-invariant, and on this line ϕ acts by translation by $b_+ v_+$. In particular, the condition that ϕ has no fixed point is equivalent to $b_+ \neq 0$. In the above coordinates, it is clear that ϕ is a glide reflection with axis ℓ. $\square$

2.3 Reflections

We now turn to an important fact about isometries of finite-dimensional euclidean spaces, namely that they are finite products of orthogonal reflections in hyperplanes.

Definition 2.19. Let $\mathbb{E}$ be an affine euclidean space and $\vec{n} \in E$ be a unit vector. For $a \in \mathbb{E}$ we call the subset

$$H = a + \vec{n}^{\perp}$$

an *affine hyperplane with (unit) normal vector* $\vec{n}$.[3]

Then we have an orthogonal decomposition $E = \overrightarrow{H} \oplus \overrightarrow{H}^{\perp}$, where $\overrightarrow{H}^{\perp} = \mathbb{R}\vec{n}$ is one-dimensional (Exercise!). We now define the *orthogonal reflection in* H by

$$r_H : \mathbb{E} \to \mathbb{E}, \quad r_H(a + w) := a - w \quad \text{ for } \quad a \in H, w \in \overrightarrow{H}^{\perp}.$$

To see that r_H is well-defined, we first observe that every element $p \in \mathbb{E}$ can be written in a unique way as $p = a + w$ as above. In fact, if $a + w = a' + w'$ with $a, a' \in H$ and $w, w' \in \overrightarrow{H}^{\perp}$, then $a + w = a + \overrightarrow{aa'} + w'$ leads to $w - w' = \overrightarrow{aa'} \in \overrightarrow{H} \cap \overrightarrow{H}^{\perp} = \{0\}$, so that $w = w'$ and $a = a'$. This implies that r_H is well-defined.

Fixing $a_0 \in H$, we have for $v \in \overrightarrow{H}$ and $w \in \overrightarrow{H}^{\perp}$ the relation

$$r_H(a_0 + v + w) = a_0 + v - w,$$

so that r_H is affine and the associated linear map is the orthogonal reflection

$$(r_H)_L(x) = x - 2\langle x, \vec{n}\rangle \vec{n}.$$

In this sense we can also write

$$r_H(p) = p - 2\langle \overrightarrow{ap}, \vec{n}\rangle \vec{n} \quad \text{ for any } \quad a \in H. \tag{7}$$

Example 2.20. (a) Let $a, b \in \mathbb{E}$ be two different points and m be their midpoint. Then $H := m + \overrightarrow{ab}^{\perp}$ is an affine hyperplane in $\mathbb{E}$ containing m. The reflection r_H then satisfies

$$r_H(b) = r_H\Big(m + \tfrac{1}{2}\overrightarrow{ab}\Big) = m - \tfrac{1}{2}\overrightarrow{ab} = a,$$

so that it exchanges a and b.

(b) (Linear case) If $a \neq b \in E$ (the corresponding euclidean vector space) and $\|a\| = \|b\|$, then $m = \frac{1}{2}(a + b) \in (b - a)^{\perp}$ implies that H is a linear subspace of E, so that r_H is a linear reflection.

Lemma 2.21. *Orthogonal reflections in hyperplanes are isometries.*

Proof. Let $H \subseteq \mathbb{E}$ be an affine hyperplane with unit normal vector $\vec{n}$ and r_H be the corresponding orthogonal reflection. We have already seen above that, for any $a_0 \in H$, $v \in \overrightarrow{H}$ and $w \in \overrightarrow{H}^{\perp}$ we have

$$r_H(a_0 + v + w) = a_0 + v - w.$$

[3] If E is complete, i.e., a real Hilbert space, one can show that all closed affine hyperplanes in $\mathbb{E}$ are of this form. Since all linear subspaces of finite-dimensional euclidean spaces are closed, this covers in particular the finite-dimensional case. However, we can ignore this difficulty by simply working with hyperplanes with a normal vector.

Since the map $\eta_{a_0}\colon E \to \mathbb{E}, v \mapsto a_0 + v$ is an affine isometry, it remains to show that the linear reflection defined by

$$r(x) := x - 2\langle x, \vec{n}\rangle \vec{n}$$

on E is isometric. This follows from

$$\begin{aligned}\langle r(x), r(y)\rangle &= \langle x - 2\langle x, \vec{n}\rangle \vec{n}, y - 2\langle y, \vec{n}\rangle \vec{n}\rangle \\ &= \langle x, y\rangle - 2\langle x, \vec{n}\rangle\langle \vec{n}, y\rangle - 2\langle y, \vec{n}\rangle\langle x, \vec{n}\rangle + 4\langle x, \vec{n}\rangle\langle y, \vec{n}\rangle = \langle x, y\rangle. \qquad \square\end{aligned}$$

Lemma 2.22. *Let E be a euclidean vector space and $(b_1, \ldots, b_k), (c_1, \ldots, c_k)$ be two orthonormal systems of the same length k. Then there exists a product ϕ of at most k linear reflections with $\phi(b_j) = c_j$ for $j = 1, \ldots, k$.*

Proof. We argue by induction on k. If $k = 0$, then we take $\phi := \mathrm{id}_E$.

If $b_1 = c_1$, then we put $\phi_1 := \mathrm{id}_E$, and otherwise, we write $\phi_1 := r_H$ for the reflection in the hyperplane $(b_1 - c_1)^\perp$. It is linear by Example 2.20. In both cases $\phi_1(b_1) = c_1$.

Now $b_j' := \phi_1(b_j)$, $j = 2, \ldots, k$, is an orthonormal system in the hyperplane $c_1^\perp \subseteq E$, and the induction hypothesis implies the existence of a product ϕ_2 of at most $k - 1$ reflections r_j in hyperplanes of $c_1^\perp$ such that $\phi_2(b_j') = c_j$ for $j = 2, \ldots, k$. Since each hyperplane in $c_1^\perp$ corresponds to a hyperplane in E containing c_1, we see that $\phi := \phi_2 \circ \phi_1$ is a product of at most k reflections, $\phi(b_1) = \phi_2(\phi_1(b_1)) = \phi_2(c_1) = c_1$, and, for $j > 1$, $\phi(b_j) = \phi_2(\phi_1(b_j)) = \phi_2(b_j') = c_j$. $\square$

Proposition 2.23. *For every euclidean vector space E, the orthogonal group $\mathrm{O}(E)$ acts transitively on the following sets:*

(i) *All spheres $\mathbb{S}_r(E) := \{v \in E \colon \|v\| = r\}$, $r \geq 0$.*

(ii) *The set $F_k \subseteq E^k$ of orthonormal k-tuples, $k \in \mathbb{N}$.*

(iii) *The set Gr_k of k-dimensional linear subspaces of E, $k \in \mathbb{N}$.*

Proof. (i) It suffices to consider the case $r = 1$ which follows from (ii).

(ii) Since reflections are isometries, this follows from Lemma 2.22.

(iii) Let $F_1, F_2 \subseteq E$ be two k-dimensional linear subspaces. Let $(b_1, \ldots, b_k)$ be an orthonormal basis of F_1 and $(c_1, \ldots, c_k)$ be an orthonormal basis of F_2. Then (ii) implies the existence of $\phi \in \mathrm{O}(E)$ with $\phi(b_j) = c_j$ for $j = 1, \ldots, k$, and this clearly implies $\phi(F_1) = F_2$. $\square$

Up to now, we did not assume that $\mathbb{E}$ is finite-dimensional, but the following theorem becomes false for infinite-dimensional spaces (Exercise 2.15).

Theorem 2.24. *If $\mathbb{E}$ is an n-dimensional affine euclidean space, then every element of $\mathrm{Mot}(\mathbb{E})$ can be written as a product of at most $n + 1$ reflections.*

Proof. Let $\phi \in \mathrm{Mot}(\mathbb{E})$. We fix a euclidean frame $(a_0, \ldots, a_n)$ of $\mathbb{E}$. If $\phi(a_0) \neq a_0$, then we write r_0 for an orthogonal reflection exchanging a_0 and $\phi(a_0)$ (Example 2.20). Then $\phi_0 := r_0 \circ \phi$ is an isometry and, in addition, it fixes a_0. Using the affine isometric isomorphism $\eta_{a_0}\colon E \to \mathbb{E}, v \mapsto a_0 + v$ and the fact that ϕ is affine, we see that it suffices to show that every linear isometry ψ of the n-dimensional euclidean vector space E is a product of n reflections.

Let $b_1, \ldots, b_n$ be an orthonormal basis of E. Then $c_j := \psi(b_j)$ is a second orthonormal basis. In view of Lemma 2.22, there exists a product γ of at most n reflections, such that $\gamma(b_j) = c_j$ for $j = 1, \ldots, n$. Since γ and ψ are determined by their values on a basis of E, it follows that ψ is a product of at most n reflections. $\square$

2.4 Geodesics

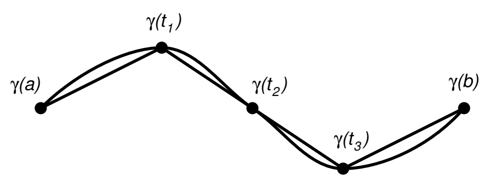
In this subsection we first introduce geodesics in general metric spaces. This requires in particular the concept of the length of a curve in a metric space. A geodesic is then defined as a length minimizing curve between two points whose velocity (defined suitably) is constant. We then show that geodesics in euclidean spaces are simply straight lines parametrized in an affine way.

Definition 2.25. (Curve length in metric spaces) Let (X, d) be a metric space and $\gamma\colon [a, b] \to X$ be a continuous path. We want to define the length of γ, whenever it makes sense. The basic idea is that, for $a \le t \le s \le b$, the length of $\gamma|_{[t,s]}$ should at least be $d(\gamma(t), \gamma(s))$. We call a tuple

$$\mathbf{t} = (t_0, \ldots, t_n) \quad \text{with} \quad t_0 = a < t_1 < \ldots < t_n = b$$

a *subdivision of the interval* $[a, b]$. For each such subdivision, we obtain the number

$$L_{\mathbf{t}}(\gamma) := \sum_{j=0}^{n-1} d(\gamma(t_j), \gamma(t_{j+1})).$$



For any refinement $\mathbf{t}'$ of $\mathbf{t}$ (i.e., $\mathbf{t}'$ is obtained by inserting more subdivision points), we then have

$$L_{\mathbf{t}'}(\gamma) \ge L_{\mathbf{t}}(\gamma)$$

by the triangle inequality. Now we say that γ is *rectifiable* (Germ.: rektifizierbar) if its *length*

$$L(\gamma) := \sup_{\mathbf{t}} L_{\mathbf{t}}(\gamma) \in [0, \infty]$$

is finite. By definition, this implies that

$$d(\gamma(a), \gamma(b)) \le L(\gamma). \tag{8}$$

Lemma 2.26. *If $\gamma\colon [a, b] \to X$ is a continuous curve in the metric space (X, d) and $a < c < b$. Then γ is rectifiable if and only if this is the case for its restrictions to $[a, c]$ and $[c, b]$, and in this case we have*

$$L(\gamma) = L(\gamma_{[a,c]}) + L(\gamma_{[c,b]}).$$

Proof. Since every subdivision of $[a, b]$ has a refinement containing c, it suffices to consider such subdivisions $\mathbf{t}$. Write $\mathbf{t}_1$ and $\mathbf{t}_2$ for the corresponding subdivisions of $[a, c]$ and $[c, a]$. Then

$$L_{\mathbf{t}}(\gamma) = L_{\mathbf{t}_1}(\gamma|_{[a,c]}) + L_{\mathbf{t}_2}(\gamma|_{[c,b]}).$$

Passing to the supremum over all such subdivisions, we obtain

$$L(\gamma) = L(\gamma_{[a,c]}) + L(\gamma_{[c,b]}).$$

In particular $L(\gamma)$ is finite if and only if this is the case for $L(\gamma_{[a,c]})$ and $L(\gamma_{[c,b]})$. $\square$

Definition 2.27. A curve $\gamma\colon [a,b] \to \mathbb{E}_n$ ($\mathbb{R}^n$, considered as an affine space) is said to be *differentiable in* $t \in [a,b]$ if its *velocity vector*

$$\gamma'(t) = \lim_{h\to 0} \frac{1}{h}\overrightarrow{\gamma(t)\gamma(t+h)}$$

exists. It must be interpreted as an element of the translation group $E = \mathbb{R}^n$. Accordingly, we define the *velocity of* γ *in* t as $\|\gamma'(t)\|$.

For differentiable curves in $\mathbb{E}$, the length can be computed as an integral. This is very helpful for concrete calculations.

Lemma 2.28. *If* $\gamma\colon [a,b] \to \mathbb{E}_n$ *is continuously differentiable, then it is rectifiable and its length can be computed by*

$$L(\gamma) := \int_a^b \|\gamma'(t)\|\, dt.$$

Proof. For $a \le t \le s \le b$, the Fundamental Theorem of Calculus yields

$$\overrightarrow{\gamma(t)\gamma(s)} = \int_t^s \gamma'(\tau)\, d\tau,$$

and this implies

$$d(\gamma(t),\gamma(s)) = \|\overrightarrow{\gamma(t)\gamma(s)}\| = \Big\|\int_t^s \gamma'(\tau)\, d\tau\Big\| \le \int_t^s \|\gamma'(\tau)\|\, d\tau.$$

For every subdivision

$$\mathbf{t} = (t_0,\ldots,t_n) \quad \text{with} \quad t_0 = a < t_1 < \ldots < t_n = b$$

this leads to

$$L_{\mathbf{t}}(\gamma) = \sum_{j=0}^{n-1} d(\gamma(t_j),\gamma(t_{j+1})) \le \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \|\gamma'(\tau)\|\, d\tau = \int_a^b \|\gamma'(\tau)\|\, d\tau,$$

and hence to

$$L(\gamma) = \sup_{\mathbf{t}} L_{\mathbf{t}}(\gamma) \le \int_a^b \|\gamma'(\tau)\|\, d\tau. \tag{9}$$

To prove equality, we consider the function

$$f\colon [a,b] \to \mathbb{R}, \quad f(t) := L(\gamma|_{[a,t]}) \le \int_a^t \|\gamma'(\tau)\|\, d\tau.$$

From Lemma 2.26 and (9) we derive

$$\frac{1}{h}\|\overrightarrow{\gamma(t)\gamma(t+h)}\| \le \frac{1}{h}L(\gamma|_{[t,t+h]}) = \frac{1}{h}(f(t+h) - f(t)) \le \frac{1}{h}\int_t^{t+h} \|\gamma'(\tau)\|\, dt.$$

By passing to the limit $h \to 0$ we see that f is differentiable in t with $f'(\tau) = \|\gamma'(\tau)\|$, and since this function is continuous, we arrive at the formula

$$L(\gamma) = f(b) = \int_a^b f'(\tau)\, d\tau = \int_a^b \|\gamma'(\tau)\|\, d\tau. \qquad \square$$

Lemma 2.29. *If the curve $\gamma\colon [a,b] \to X$ in the metric space (X,d) satisfies $L(\gamma) = d(\gamma(a),\gamma(b))$, then*

$$L(\gamma_{[t,s]}) = d(\gamma(t),\gamma(s)) \quad \textit{for} \quad a \leq t \leq s \leq b.$$

Proof. From

$$\begin{aligned} d(\gamma(a),\gamma(b)) = L(\gamma) &\overset{2.26}{=} L(\gamma|_{[a,t]}) + L(\gamma|_{[t,s]}) + L(\gamma|_{[s,b]}) \\ &\geq d(\gamma(a),\gamma(t)) + d(\gamma(t),\gamma(s)) + d(\gamma(s),\gamma(b)) \geq d(\gamma(a),\gamma(b)), \end{aligned}$$

we derive that, for each of the 3 summands, the corresponding inequality is an equality. In particular, $L(\gamma_{[t,s]}) = d(\gamma(t),\gamma(s))$. □

Definition 2.30. Let (X,d) be a metric space. A curve $\gamma\colon [a,b] \to X$ is called a *geodesic of speed c* if there exists a constant $c \geq 0$ such that

$$L(\gamma) = d(\gamma(a),\gamma(b)) \quad \text{and} \quad d(\gamma(a),\gamma(t)) = c(t-a) \quad \text{for} \quad a \leq t \leq b.$$

This property implies in particular that γ is a "shortest curve" from $\gamma(a)$ to $\gamma(b)$. In addition, it imposes a condition on the parametrization of the curve.

Examples 2.31. (of curves) (a) The simplest curves are parametrized lines

$$\gamma : \mathbb{R} \to \mathbb{E}_n, t \mapsto p + tv,$$

for $p \in \mathbb{E}_n$, $v \in E$ fixed. Then $\gamma'(t) = v$ is constant and the arclength is

$$L(\gamma) = |b-a| \cdot \|v\| = d(p, p+bv).$$

These curves are geodesics in $\mathbb{E}_n$.

(b) The curve $\gamma : [0,2\pi] \to \mathbb{E}_2 = \mathbb{R}^2, t \mapsto (r\cos t, r\sin t)$ parametrizes a circle of radius r with center 0. Its velocity vector is

$$\gamma'(t) = (-r\sin t, r\cos t) \in \mathbb{R}^2,$$

so that the velocity is $\|\gamma'(t)\| = r$ and $L(\gamma) = 2\pi r$ is the length of a circle of radius r.

(c) A screw line in the euclidean space $\mathbb{E}_3 \cong \mathbb{R}^3$ is given by the curve

$$\gamma : \mathbb{R} \to \mathbb{R}^3,\ \gamma(t) = (\cos t, \sin t, t).$$

Then

$$\gamma'(t) = (-\sin t, \cos t, 1),$$

and the velocity of γ is constant $\|\gamma'(t)\| = \sqrt{2}$.

Theorem 2.32. (Line segments as shortest curves) *If $\gamma\colon [a,b] \to \mathbb{E}_n$ is a continuous curve with $L(\gamma) = d(\gamma(a),\gamma(b))$, then*

$$\gamma([a,b]) = [\gamma(a),\gamma(b)] = \{t\gamma(s) + (1-t)\gamma(b)\colon 0 \leq t \leq 1\}$$

is the line segment from $\gamma(a)$ to $\gamma(b)$. If $v := \overrightarrow{\gamma(a)\gamma(b)} \neq 0$, it can be written as

$$\gamma(t) = \gamma(a) + f(t)v,$$

where $f\colon [a,b] \to [0,1]$ is surjective and monotone. Conversely, any such curve satisfies $L(\gamma) = d(\gamma(a),\gamma(b))$.

Proof. For every $t \in [a,b]$, we derive from Lemma 2.29 that

$$d(\gamma(a),\gamma(t)) = L(\gamma|_{[a,t]}) \quad \text{and} \quad d(\gamma(t),\gamma(b)) = L(\gamma|_{[t,b]}).$$

We thus obtain

$$d(\gamma(a),\gamma(b)) = L(\gamma) = L(\gamma|_{[a,t]}) + L(\gamma|_{[t,b]}) = d(\gamma(a),\gamma(t)) + d(\gamma(t),\gamma(b)).$$

Now Lemma 2.7 implies that

$$\gamma([a,b]) \subseteq [\gamma(a),\gamma(b)].$$

If $v \neq 0$, i.e., γ is not constant, then we can write $\gamma(t) = \gamma(a) + f(t)v$ with $0 \leq f(t) \leq 1$. For $a \leq s \leq t \leq b$ we likewise obtain from Lemma 2.29 that $\gamma(s) \in [\gamma(a),\gamma(t)]$, so that $f(s) \leq f(t)$, i.e., f is monotone.

If, conversely, $\gamma(t) = \gamma(a) + f(t)v$ with a monotone function f, then we obtain for every partition $\mathbf{t}$ of $[a,b]$ the relation

$$L_{\mathbf{t}}(\gamma) = \sum_{j=0}^{n-1} d(\gamma(t_j),\gamma(t_{j+1})) = \sum_{j=0}^{n-1}(f(t_{j+1}) - f(t_j))\|v\| = (f(b) - f(a))\|v\| = d(\gamma(a),\gamma(b)).$$

This implies that $L(\gamma) = d(\gamma(a),\gamma(b))$. □

Corollary 2.33. *A curve $\gamma\colon [a,b] \to \mathbb{E}$ is a geodesic if and only if it is affine, i.e.,*

$$\gamma(t) = \gamma(a) + \frac{t-a}{b-a}\overrightarrow{\gamma(a)\gamma(b)}.$$

Here we call γ affine because it is the restriction of an affine map to the interval $[a,b]$.

Proof. We may w.l.o.g. assume that $\gamma(a) \neq \gamma(b)$. If $\gamma\colon [a,b] \to \mathbb{E}$ is a geodesic of speed c, then we write it as $\gamma(t) = \gamma(a) + f(t)v$ with $v := \overrightarrow{\gamma(a)\gamma(b)}$ (Theorem 2.32). Then

$$c(t-a) = d(\gamma(a),\gamma(t)) = f(t)\|v\|$$

implies that $f(t) = \frac{c}{\|v\|}(t-a)$ is affine, and hence also γ is affine. □

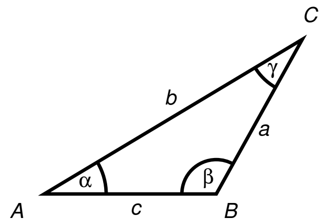
2.5 Triangles

In the following we study triangles in an affine euclidean space $\mathbb{E}$. Any such triangle is specified by the triple (A,B,C) of vertices, and we assume these to be pairwise different. We now put

$$\begin{aligned}
\vec{a} &:= \overrightarrow{BC}, a := d(B,C) = \|\vec{a}\|, \\
\vec{b} &:= \overrightarrow{CA}, b := d(C,A) = \|\vec{b}\|, \\
\vec{c} &:= \overrightarrow{AB}, c := d(A,B) = \|\vec{c}\|.
\end{aligned}$$

Then a,b,c are the lengths of the sides opposite to the vertices A,B,C, respectively. For the corresponding angles we write

$$\alpha := \measuredangle(\overrightarrow{AB},\overrightarrow{AC}), \quad \beta := \measuredangle(\overrightarrow{BA},\overrightarrow{BC}), \quad \gamma := \measuredangle(\overrightarrow{CB},\overrightarrow{CA}).$$



In this notation, we now derive the fundamental relations between these objects.

Proposition 2.34. (Cosine Rule of euclidean geometry)

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma).$$

Proof. This follows from $\vec{c} = -\vec{a} - \vec{b}$ and

$$c^2 = \langle \vec{c}, \vec{c} \rangle = \langle \vec{a} + \vec{b}, \vec{a} + \vec{b} \rangle = b^2 + a^2 + 2\langle \vec{a}, \vec{b} \rangle = b^2 + a^2 - 2ab\cos(\gamma). \qquad \square$$

Remark 2.35. (a) Note that the Cosine Rule generalizes Pythagoras' Theorem to arbitrary triangles.

(b) The Cosine Rule shows in particular, that the length c of the side $\overline{AB}$ is determined by the lengths of the other two sides and the angle γ.

Proposition 2.36. (Sine Rule of euclidean geometry)

$$\frac{a}{b} = \frac{\sin\alpha}{\sin\beta}.$$

Proof. According to the Cosine Rule we have

$$-2bc\cos(\alpha) = a^2 - (b^2 + c^2)$$

and thus

$$4b^2c^2\cos^2(\alpha) = (-a^2 + b^2 + c^2)^2.$$

We likewise obtain

$$4a^2c^2\cos^2(\beta) = (a^2 - b^2 + c^2)^2.$$

This leads to

$$\begin{aligned}\frac{\sin^2(\alpha)}{\sin^2(\beta)} &= \frac{4a^2b^2c^2(1-\cos^2(\alpha))}{4a^2b^2c^2(1-\cos^2(\beta))} = \frac{4a^2b^2c^2 - a^2(-a^2+b^2+c^2)^2}{4a^2b^2c^2 - b^2(a^2-b^2+c^2)^2} \\ &= \frac{a^2}{b^2} \cdot \frac{4b^2c^2 - (a^4+b^4+c^4-2a^2b^2-2a^2c^2+2b^2c^2)}{4a^2c^2 - (a^4+b^4+c^4-2a^2b^2+2a^2c^2-2b^2c^2)} = \frac{a^2}{b^2}. \qquad \square\end{aligned}$$

Proposition 2.37. (Sum of angles in a euclidean triangle)

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

Proof. Since any triangle is contained in an affine plane, we may w.l.o.g. assume that $\mathbb{E} = \mathbb{E}_2$ is the two-dimensional euclidean plane. Let $\phi_A \in \mathrm{Mot}(\mathbb{E})$ denote the unique motion mapping A to B and $\vec{c}$ to a positive multiple of $\vec{a}$. Then

$$\phi_A(A + x) = B + D(\pi - \beta)x.$$

We likewise define

$$\phi_B(B + x) = C + D(\pi - \gamma)x \quad \text{and} \quad \phi_C(C + x) = A + D(\pi - \alpha)x.$$

Then the composition $\phi := \phi_C \circ \phi_B \circ \phi_A$ is an affine map fixing A, and its linear part is given by the matrix

$$D := D(\pi - \alpha)D(\pi - \gamma)D(\pi - \beta) = D(3\pi - \alpha - \beta - \gamma)$$

(cf. Exercise 2.4). Since $D\vec{c}$ is a positive multiple of $\vec{c}$, we actually have $D = \mathbf{1}$, so that

$$3\pi - (\alpha + \beta + \gamma) \in 2\pi\mathbb{Z}.$$

Since the left hand side is positive and smaller than 3π, it must be equal to 2π. This completes the proof. □

2.6 Euclidean geometric properties

Since the euclidean motion group $\mathrm{Mot}(\mathbb{E})$ contains all translations, it acts transitively on $\mathbb{E}$. Therefore all points look alike and points are not distinguished by any geometric property. For pairs of points (a_0, a_1) in $\mathbb{E}$, their distance $d(a_0, a_1)$, i.e., the length of the line segment joining a_0 and a_1, is clearly a geometric property, i.e., invariant under the euclidean group. More generally, we can ask for geometric properties of finite or infinite configurations, i.e., subsets $S \subseteq \mathbb{E}$.

The following proposition implies that this is the only geometric property of a pair of points.

Proposition 2.38. *If $\mathbb{E}$ is a euclidean affine space, then the two pairs $(a_0, a_1), (a_0', a_1') \in \mathbb{E}^2$ lie in the same orbit of $\mathrm{Mot}(\mathbb{E})$ if and only if*

$$d(a_0, a_1) = d(a_0', a_1').$$

Proof. Clearly, $d(a_0, a_1) = d(a_0', a_1')$ holds if the two pairs lie in the same orbit. Suppose, conversely, that this condition is satisfied. Since $\mathrm{Mot}(\mathbb{E})$ contains all translations, it acts transitively on $\mathbb{E}$, and we may thus assume that $a_0' = a_0$. The stabilizer group $\mathrm{Mot}(\mathbb{E})_{a_0}$ is isomorphic to the linear isometry group $\mathrm{O}(E)$, so that we have to recall that $\mathrm{O}(E)$ acts transitively on all spheres in E (Proposition 2.23).

Alternative argument: Suppose that $a_0 = a_0', a_1 \neq a_1'$, and consider the hyperplane

$$H := \frac{1}{2}(a_1 + a_1') + \overrightarrow{a_1 a_1'}^{\perp}$$

orthogonal to the line segment from a_1 to a_1' and containing the midpoint. Then $d(a_0, a_1) = d(a_0, a_1')$ implies that $a_0 \in H$ (apply Example 2.20(b) with $a_0 = 0$). Therefore the reflection r_H fixes a_0 and exchanges a_1 and a_1'. □

The natural next step is to ask for the orbits of $\mathrm{Mot}(\mathbb{E})$ in the space $\mathbb{E}^3$ of triples which we discuss next:

Proposition 2.39. (Congruence of triangles) *If $\mathbb{E}$ is a euclidean affine space, then the triples $(x,y,z),(x',y',z') \in \mathbb{E}^3$ lie in the same orbit of $\mathrm{Mot}(\mathbb{E})$ if and only if*

$$d(x,y) = d(x',y'), \quad d(x,z) = d(x',z') \quad \textit{and} \quad d(y,z) = d(y',z').$$

Proof. Clearly, $d(x,y)$, $d(x,z)$ and $d(y,z)$ are functions on triples (x,y,z) that are invariant under the action of $\mathrm{Mot}(\mathbb{E})$. Suppose that (x,y,z), (x',y',z') are two triples with

$$d(x,y) = d(x',y'), \quad d(x,z) = d(x',z') \quad \text{and} \quad d(y,z) = d(y',z'),$$

then the preceding proposition implies the existence of some $\phi \in \mathrm{Mot}(\mathbb{E})$ with $\phi(x) = x'$ and $\phi(y) = y'$. If $x = y$, then $x' = y'$ and Proposition 2.38 even provides an element ϕ that also satisfies $\phi(z) = z'$.

We may therefore assume that $x = x' \neq y = y'$. Using the affine isometry $\eta_x \colon E \to \mathbb{E}$, we may w.l.o.g. assume that $\mathbb{E} = E$ and $x = 0$. Then $\mathrm{Mot}(E)_0 \cong \mathrm{O}(E)$ and $\mathrm{O}(E)_y \cong \mathrm{O}(y^\perp)$ is the orthogonal group of the hyperplane $y^\perp \subseteq E$.

Let $a := d(x,z) = \|z\|$, $b := d(y,z)$ and $c := d(x,y) = \|y\|$. We then have $a \leq \|y\| + b$ by the triangle inequality. Writing $z = \lambda y + z_1$ with $z_1 \bot y$ and $z' = \lambda' y + z_1'$ with $z_1' \bot y$,

$$a^2 = \|z\|^2 = \lambda^2 c^2 + \|z_1\|^2 \quad \text{and} \quad b^2 = \|z - y\|^2 = (\lambda - 1)^2 c^2 + \|z_1\|^2. \tag{10}$$

Therefore

$$b^2 - a^2 = (1 - 2\lambda)c^2 = (1 - 2\lambda')c^2$$

implies $\lambda = \lambda'$, so that $\|z_1\| = \|z_2\|$. Therefore the set of all $z' \in E$ with the fixed distance a from x and b from y is of the form

$$\lambda y + \{z' \in y^\perp \colon \|z'\| = \|z\|\},$$

i.e., a sphere in the affine euclidean space $y + y^\perp$. Since $\mathrm{O}(y^\perp)$ acts transitively on this sphere (Proposition 2.23), there exists a $\phi \in \mathrm{Mot}(\mathbb{E})$ fixing x and y and mapping z to z'.

Alternative argument: Suppose that $x = x'$, $y = y'$ and $z \neq z'$. We consider the hyperplane

$$H := \frac{1}{2}(z + z') + \overrightarrow{zz'}^\perp.$$

Then $d(x,z) = d(x,z')$ implies $x \in H$ (apply Example 2.20(b) with $x = 0$). We likewise obtain $y \in H$. Therefore the reflection r_H fixes x and y and exchanges z and z'. $\square$

Remark 2.40. The preceding proposition provides a classification of all triangles up to congruence, i.e., elements of the motion group. It says that two triangles with vertices (x,y,z) and (x',y',z') are congruent if and only if corresponding sides have the same length.

If $a := d(x,y)$, $b := d(y,z)$ and $c := d(x,z)$, then the triple (a,b,c) determines the congruence class of the triangle with sides of length a, b and c. According to the triangle inequality, we have

$$c \leq a + b, \quad a \leq b + c \quad \text{and} \quad b \leq c + a. \tag{11}$$

If, conversely, these 3 inequalities are satisfied for $(a,b,c) \in \mathbb{R}^3_+$, then (10) has a solution z (cf. Exercise 2.9). Therefore the orbits of $\mathrm{Mot}(\mathbb{E})$ in $\mathbb{E}^3$ are classified by the triples (a,b,c) satisfying (11).

2.7 Orientation

If E is a finite-dimensional euclidean vector space, then we define the *special orthogonal group*

$$\mathrm{SO}(E) := \{g \in \mathrm{O}(E)\colon \det g > 0\} := \{g \in \mathrm{O}(E)\colon \det g = 1\}.$$

From $\det(g) \in \{\pm 1\}$ for $g \in \mathrm{O}(E)$, it follows that $\mathrm{SO}(E)$ is a subgroup of index 2. The corresponding cosets are

$$\mathrm{SO}(E) \quad \text{and} \quad \mathrm{O}(E)_- := \{g \in \mathrm{O}(E)\colon \det g = -1\}.$$

For $E = \mathbb{R}^n$, we write

$$\mathrm{SO}_n(\mathbb{R}) := \mathrm{SO}(\mathbb{R}^n) = \{g \in \mathrm{O}_n(\mathbb{R})\colon \det g = 1\}.$$

Example 2.41. For $n = 2$ and $E = \mathbb{R}^2$, we have

$$\mathrm{SO}_2(\mathbb{R}) = \{D(\theta)\colon \theta \in \mathbb{R}\}.$$

i.e., the special orthogonal group consists of rotations.

Definition 2.42. We call a linear basis $b_1, \ldots, b_n$ of $\mathbb{R}^n$ *positively oriented* if

$$\det(b_1, \ldots, b_n) := \det(b_{ij})_{i,j=1,\ldots,n} > 0$$

and *negatively oriented* otherwise.

Remark 2.43. The action of $\mathrm{SO}_n(\mathbb{R})$ preserves the orientation of a basis because $g \in \mathrm{SO}_n(\mathbb{R})$ implies

$$\det(gb_1, \ldots, gb_n) = \det g \det(b_1, \ldots, b_n) = \det(b_1, \ldots, b_n).$$

For $n = 2$, the orientation leads to an additional invariant for the action of $\mathrm{SO}_2(\mathbb{R})$ on pairs of vectors.

Remark 2.44. Specializing Proposition 2.39 to the case where $z = z' = 0$ (0 is a vertex of the triangle), it follows that two pairs (x, y) and (x', y') are conjugate under the orthogonal group $\mathrm{O}_2(E) \cong \mathrm{Mot}(E)_0$ if and only if

$$\|x\| = \|x'\|, \quad \|y\| = \|y'\| \quad \text{and} \quad \|x - y\| = \|x' - y'\|.$$

Since

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y\rangle = \|x\|^2 + \|y\|^2 - 2\cos(\measuredangle(x, y))\|x\|\|y\|,$$

it follows that two pairs (x, y) and (x', y') are conjugate under the orthogonal group $\mathrm{O}_2(E)$ if and only if

$$\|x\| = \|x'\|, \quad \|y\| = \|y'\| \quad \text{and} \quad \measuredangle(x, y) = \measuredangle(x', y').$$

In terms of the corresponding triangles $\Delta(x, y, z)$, this means that they are determined, up to congruence, by the lengths $d(x, z)$ and $d(y, z)$ of the sides $\overline{xz}$ and $\overline{yz}$ and the (non-oriented) angle in z.

Example 2.45. Under the action of the subgroup

$$\mathrm{SMot}(\mathbb{R}^2) := \mathbb{R}^2 \rtimes \mathrm{SO}_2(\mathbb{R})$$

of orientation preserving isometries of the plane, the "orientation" of a triangle is also preserved. This is reflected by the fact that two pairs (x, y) and (x', y') of non-zero vectors in the plane $\mathbb{R}^2$ are conjugate under $\mathrm{SO}_2(\mathbb{R})$ if and only if

$$\|x\| = \|x'\|, \ \|y\| = \|y'\|, \quad \measuredangle(x,y) = \measuredangle(x',y'), \quad \text{and} \quad \mathrm{sgn}(\det(x,y)) = \mathrm{sgn}(\det(x',y')).$$

Accordingly, two planar triangles $\Delta(x, y, z)$ and $\Delta(x', y', z')$ are conjugate under $\mathrm{SMot}(\mathbb{R}^2)$ if and only if they are conjugate under $\mathrm{Mot}(\mathbb{R}^2)$ and, in addition,

$$\mathrm{sgn}(\det(\overrightarrow{zx}, \overrightarrow{zy})) = \mathrm{sgn}(\det(\overrightarrow{z'x'}, \overrightarrow{z'y'})).$$

In $\mathbb{R}^3$ the orientation is defined for triples of vectors via their determinant. Therefore it does not make sense to say that two triangles in space are equally oriented. We now make this more precise in terms of congruence with respect to $\mathrm{Mot}(\mathbb{R}^3)$ and $\mathrm{SMot}(\mathbb{R}^3)$.

Proposition 2.46. *If $n > 2$, then two triples (x, y, z) and (x', y', z') in $\mathbb{R}^n$ are congruent under $\mathrm{Mot}(\mathbb{R}^n)$ if and only if they are congruent under the smaller group $\mathrm{SMot}(\mathbb{R}^n)$.*

Proof. Every triangle $\Delta(x, y, z)$ in $\mathbb{R}^n$ is conjugate under $\mathrm{Mot}(\mathbb{R}^n)$ to a triangle in $\mathbb{R}^2$. The assumption $n > 2$ implies the existence of some $g_0 \in \mathrm{O}_n(\mathbb{R})$ with $\det(g_0) = -1$ and $g_0|_{\mathbb{R}^2} = \mathrm{id}_{\mathbb{R}^2}$. A typical example is the orthogonal reflection in the hyperplane $\mathbb{R}^{n-1} \cong \mathbb{R}^{n-1} \times \{0\} \subseteq \mathbb{R}^n$. Hence every triangle can be moved into $\mathbb{R}^2$ by an element of $\mathrm{SMot}(\mathbb{R}^n)$. If two triangles $\Delta(x, y, z)$ and $\Delta(x', y', z')$ in $\mathbb{R}^2$ are congruent under $\mathrm{Mot}(\mathbb{R}^n)$, composition with g_0 implies that they are also congruent under $\mathrm{SMot}(\mathbb{R}^n)$. □

2.8 Embeddings of metric spaces in euclidean spaces

In this section we discuss the problem of embedding a metric space (X, d) into some affine euclidean space $\mathbb{E}$, i.e., we are looking for a map $j\colon X \to \mathbb{E}$ with $d_{\mathbb{E}}(j(x), j(y)) = d_X(x, y)$ for $x, y \in X$. This problem has been studied for about 80 years starting with the work of Blumenthal, Menger, Schoenberg and Wilson. It has many interesting connections to other branches of mathematics such a representation theory, geometric group theory and probability theory, but first of all it is about the special metric properties of subsets of euclidean spaces.

To warm up, let us discuss some simple cases.

Examples 2.47. (a) If $X = \{x, y\}$ consists of at most 2 points, then there exists an isometric embedding $j\colon X \to \mathbb{R}$. Simply put $j(x) := 0$ and $j(y) := d(x, y)$.

(b) If $X = \{x, y, z\}$ consists of at most 3 points, then there exists an isometric embedding $j\colon X \to \mathbb{R}^2$. We put $j(x) := (0, 0)$, $j(y) := (d(x, y), 0)$. The triangle inequalities

$$d(x,y) \leq d(x,z) + d(z,y), \quad d(y,z) \leq d(y,x) + d(x,z), \quad d(z,x) \leq d(z,y) + d(y,x)$$

hold in X, so that there exists a point $p \in \mathbb{R}^2$ with $d(p, j(x)) = d(x, z)$ and $d(p, j(y)) = d(y, z)$ (Exercise 2.9). Then $j(z) := p$ yields an isometric embedding $j\colon X \to \mathbb{R}^2$.

(c) If there exists an isometric embedding $j\colon X \to \mathbb{E}$, then midpoints in X have to be unique because j maps midpoints to midpoints. We can thus construct metric spaces with four points which have no embedding into euclidean space.

Let $X := \{\pm e_1, \pm e_2\} \subseteq \mathbb{R}^2$, endowed with the metric

$$d(a, b) := \max(|a_1 - b_1|, |a_2 - b_2|).$$

Then

$$d(e_1, -e_1) = 2 \quad \text{and} \quad d(\pm e_1, \pm e_2) = 1.$$

Therefore $\pm e_2$ are two different midpoints for $\pm e_1$.

We have already seen that, for $|X| \geq 4$, an isometric embedding $j: X \to \mathbb{E}$ need not exist. Therefore it is a non-trivial problem to decide when a given metric space admits an isometric embedding into some euclidean space. The euclidean metric, when restricted to subsets with more than three elements of some euclidean space, has a property that is not shared by all metric spaces, and we have to see if we can formulate this property as explicitly as possible. The appropriate concept is that of a positive definite kernel.

Definition 2.48. Let X be a set and $K: X \times X \to \mathbb{R}$ be a symmetric function, i.e.,

$$K(x, y) = K(y, x) \quad \text{for} \quad x, y \in X.$$

We say that K is *positive definite* if, for $(x_1, c_1), \ldots, (x_n, c_n)$ in $X \times \mathbb{R}$, we have

$$\sum_{j,k=1}^{n} c_j c_k K(x_j, x_k) \geq 0.$$

This means that all the finite symmetric matrices $(K(x_j, x_k))_{1 \leq j,k \leq n}$ are positive semidefinite.[4]

The main point of the concept of a positive definite kernel is that it is an abstraction of the Gram matrix $(\langle v_i, v_j \rangle)_{i,j=1,\ldots,n}$ of vectors $v_1, \ldots, v_n$ in a euclidean space.

To verify positive definiteness of a kernel, one needs criteria for positive semidefiniteness of real symmetric matrices, such as the following.

Proposition 2.49. *A real symmetric matrix $A \in M_n(\mathbb{R})$ is positive semidefinite if and only if all its minors are non-negative. Recall that a minor is the determinant of a matrix $A_F := (a_{ij})_{i,j \in F}$, where $F \subseteq \{1, \ldots, n\}$ is a subset.*

Proof. If A is positive semidefinite, then all the matrices A_F are positive semidefinite. Hence all eigenvalues of A_F are non-negative, and thus $\det(A_F) \geq 0$.

Suppose, conversely, that $\det(A_F) \geq 0$ holds for all subsets $F \subseteq \{1, \ldots, n\}$. Proceeding by induction on n, we may assume that all matrices A_F, for $|F| < n$, are positive semidefinite. To see that A is positive semidefinite, we assume the contrary, i.e., that A has a unit eigenvector v with eigenvalue $\lambda < 0$. If A has only one eigenvalue ≤ 0, then all other eigenvalues of A are positive, and this leads to the contradiction $\det(A) < 0$. Hence there exists a unit eigenvector u orthogonal to v with eigenvalue $\mu \leq 0$. Since u is non-zero, some component u_i is non-zero. Hence there exists an $s \in \mathbb{R}$ so that the vector $w = v + su$ satisfies $w_i = 0$. If A' is the matrix obtained from A by removing the ith column and row and w' is obtained by removing the ith coordinate of w, then

$$(w')^\top A' w' = w^\top A w = (v + su)^\top (\lambda v + s\mu u) = \lambda + s^2 \mu < 0,$$

contradicting that A' is positive semidefinite. $\square$

[4]Actually this observation shows that it would be more natural to call these kernels positive semidefinite, but we follow the traditional terminology. However, this leads to the ugly situation, that a square matrix $A = (a_{ij})_{1 \leq i,j \leq n}$ is positive semidefinite if and only if the kernel defined by $K(i, j) := a_{ij}$ is positive definite.

Corollary 2.50. *A symmetric kernel $K\colon X \times X \to \mathbb{R}$ is positive definite if and only if, for every finite subset $\{x_1, \ldots, x_n\} \subseteq X$, the determinant of the matrix $(K(x_i, x_j))_{i,j=1,\ldots,n}$ is non-negative.*

Example 2.51. A symmetric matrix $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \in M_2(\mathbb{R})$ is positive semidefinite if and only if

$$a \geq 0, d \geq 0 \quad \text{and} \quad \det A = ad - b^2 \geq 0.$$

Proposition 2.52. (Characterization of positive definite kernels) *Let $K\colon X \times X \to \mathbb{R}$ be a symmetric kernel.*

(a) *The kernel K is positive definite if and only if there exists a euclidean vector space E and a map $\gamma\colon X \to E$ with $K(x, y) = \langle \gamma(x), \gamma(y) \rangle$ for $x, y \in X$.*

(b) *If $\gamma'\colon X \to E'$ is another map into a euclidean space E' with $K(x, y) = \langle \gamma'(x), \gamma'(y) \rangle$ for $x, y \in X$ and $\gamma(X)$ spans E, then there exists a unique linear isometry $\alpha\colon E \to E'$ with $\gamma' = \alpha \circ \gamma$.*

(c) *If $\gamma'\colon X \to E$ is another map with $K(x, y) = \langle \gamma'(x), \gamma'(y) \rangle$ for $x, y \in X$, then there exists an $\alpha \in \mathrm{O}(E)$ with $\gamma' = \alpha \circ \gamma$ if*

   (i) *$\gamma(X)$ and $\gamma'(X)$ span E, or*

   (ii) *$\gamma(X)$ spans a finite-dimensional subspace.*

Proof. (a) Suppose first that $\gamma\colon X \to E$ is a map into a euclidean vector space and $K(x, y) := \langle \gamma(x), \gamma(y) \rangle$. For $x_1, \ldots, x_n \in X$ and $c_1, \ldots, c_n \in \mathbb{R}$, we then have

$$\sum_{j,k=1}^{n} c_j c_k K(x_j, x_k) = \sum_{j,k=1}^{n} c_j c_k \langle \gamma(x_j), \gamma(x_k) \rangle = \| \sum_{j=1}^{n} c_j \gamma(x_j) \|^2 \geq 0.$$

Therefore K is positive definite.

Suppose, conversely, that K is positive definite. In the space $\mathbb{R}^X$ of real-valued functions on X, we consider the subspace E spanned by the functions $K_x(y) := K(x, y)$. We want to put

$$\Big\langle \sum_j c_j K_{x_j}, \sum_k d_k K_{x_k} \Big\rangle := \sum_{j,k} c_j d_k K(x_k, x_j), \tag{12}$$

so that we have to show that this is well-defined.

So let $f = \sum_j c_j K_{x_j}$ and $h = \sum_k d_k K_{x_k} \in E$. Then we obtain for the right hand side

$$\sum_{j,k} c_j d_k K(x_k, x_j) = \sum_{j,k} c_j d_k K_{x_j}(x_k) = \sum_k d_k f(x_k). \tag{13}$$

This expression does not depend on the representation of f as a linear combination of the K_{x_j}. Similarly, we see that the right hand side does not depend on the representation of h as a linear combination of the K_{x_k}. Therefore

$$\langle f, h \rangle := \sum_{j,k} c_j d_k K(x_k, x_j)$$

is well-defined. Since K is positive definite, we thus obtain a positive semidefinite symmetric bilinear form on E. From (13) we obtain for $h = K_x$ the relation

$$\langle f, K_x \rangle = f(x) \quad \text{for} \quad x \in X, f \in E.$$

If $\langle f, f \rangle = 0$, then the Cauchy–Schwarz inequality yields

$$|f(x)|^2 = |\langle f, K_x \rangle|^2 \leq \|K_x\|^2 \langle f, f \rangle = 0,$$

so that $f = 0$. Therefore E is a euclidean space. Now the map $\gamma \colon X \to E, \gamma(x) = K_x$ satisfies $K(x, y) = \langle \gamma(x), \gamma(y) \rangle$ for $x, y \in X$.

(b) Let $c_1, \ldots, c_n \in \mathbb{R}$ and $x_1, \ldots, x_n \in X$. Then

$$\Big\| \sum_{j=1}^n c_j \gamma'(x_j) \Big\|^2 = \sum_{j,k=1}^n c_j c_k \langle \gamma'(x_j), \gamma'(x_k) \rangle = \sum_{j,k=1}^n c_j c_k K(x_k, x_j) = \Big\| \sum_{j=1}^n c_j \gamma(x_j) \Big\|^2.$$

We therefore have a well-defined linear map

$$\alpha : E \to E', \quad \phi\Big(\sum_{i=1}^n c_i \gamma(x_i) \Big) := \sum_{i=1}^n c_i \gamma'(x_i).$$

As the preceding calculation shows, α is isometric. By definition, it satisfies $\alpha \circ \gamma = \gamma'$.

(c) (i) If $\gamma(X)$ and $\gamma'(X)$ both span E, the linear isometry α from (b) is bijective, hence an element of $\mathrm{O}(E)$.

(ii) If $\gamma(X)$ spans a finite-dimensional subspace, then (b) provides a bijective linear isometry

$$\alpha_0 \colon F := \operatorname{span} \gamma(X)) \to G := \operatorname{span}(\gamma'(X)).$$

In particular, $\dim F = \dim G$. In view of Proposition 2.23(iii), there exists a $\phi \in \mathrm{O}(E)$ with $\phi(F) = G$, which further implies that $\phi(F^\perp) = G^\perp$. Hence

$$\alpha(x + y) := \alpha_0(x) + \phi(y) \quad \text{for} \quad x \in F, y \in F^\perp$$

is a linear isometry whose range is $G + G^\perp = E$. Therefore $\alpha \in \mathrm{O}(E)$ and

$$\alpha \circ \gamma = \alpha_0 \circ \gamma = \gamma'. \qquad \square$$

Theorem 2.53. (Embedding Theorem for metric spaces) (a) *For a metric space (X, d), the following are equivalent:*

(i) *There exists an isometric embedding of X into an affine euclidean space $\mathbb{E}$.*

(ii) *For every $x_0 \in X$, the kernel $K(x, y) := d(x, x_0)^2 + d(y, x_0)^2 - d(x, y)^2$ is positive definite.*

(iii) *For some $x_0 \in X$, the kernel $K(x, y) := d(x, x_0)^2 + d(y, x_0)^2 - d(x, y)^2$ is positive definite.*

(b) *If $\gamma \colon X \to \mathbb{E}$ and $\gamma' \colon X \to \mathbb{E}'$ are two isometric embeddings and $\gamma(X)$ is not contained in a proper affine subspace of $\mathbb{E}$, then there exists a unique affine isometry $\alpha \colon \mathbb{E} \to \mathbb{E}'$ with $\gamma' = \alpha \circ \gamma$.*

(c) *If $\gamma, \gamma' \colon X \to \mathbb{E}$ are two isometric embeddings and either $\gamma(X)$ and $\gamma'(X)$ are not contained in a proper affine subspace of $\mathbb{E}$ or $\gamma(X)$ is contained in a finite-dimensional affine subspace, then there exists an $\alpha \in \mathrm{Mot}(\mathbb{E})$ with $\gamma' = \alpha \circ \gamma$.*

Proof. (a) (i) ⇒ (ii): Let $\gamma\colon X \to E$ be an isometric embedding into a euclidean vector space E and $x_0 \in X$. For $\eta(x) := \gamma(x) - \gamma(x_0)$ we then obtain

$$\begin{aligned} K(x,y) &= d(x,x_0)^2 + d(y,x_0)^2 - d(x,y)^2 \\ &= \|\gamma(x) - \gamma(x_0)\|^2 + \|\gamma(y) - \gamma(x_0)\|^2 - \|\gamma(y) - \gamma(x)\|^2 \\ &= \|\eta(x)\|^2 + \|\eta(y)\|^2 - \|\eta(x) - \eta(y)\|^2 = 2\langle \eta(x), \eta(y)\rangle. \end{aligned}$$

Therefore K is positive definite by Proposition 2.52.

(ii) ⇒ (iii) is trivial.

(iii) ⇒ (i): Suppose that the kernel $K(x,y) := d(x,x_0)^2 + d(y,x_0)^2 - d(x,y)^2$ is positive definite. Then there exists a map $\eta\colon X \to E$ into a euclidean space E with

$$K(x,y) = 2\langle \eta(x), \eta(y)\rangle \quad \text{for} \quad x, y \in X$$

(Proposition 2.52). From $K(x_0,x_0) = 0$ it then follows that $\eta(x_0) = 0$. We thus obtain

$$2\|\eta(x)\|^2 = K(x,x) = 2d(x,x_0)^2,$$

and therefore

$$\|\eta(x)-\eta(y)\|^2 = \|\eta(x)\|^2 + \|\eta(y)\|^2 - 2\langle \eta(x), \eta(y)\rangle = d(x,x_0)^2 + d(y,x_0)^2 - K(x,y) = d(x,y)^2.$$

This shows that η is isometric.

(b) Fix a point $x_0 \in X$. As we have seen in (a) above, $\eta := \gamma - \gamma(x_0)$ satisfies

$$K(x,y) := 2\langle \eta(x), \eta(y)\rangle = d(x,x_0)^2 + d(y,x_0)^2 - d(x,y)^2.$$

Likewise, $\eta' := \gamma' - \gamma'(x_0)$ satisfies

$$2\langle \eta'(x), \eta'(y)\rangle = d(x,x_0)^2 + d(y,x_0)^2 - d(x,y)^2 = K(x,y).$$

Therefore Proposition 2.52(b) implies the existence of a unique linear isometry $\beta\colon E \to E$ with $\beta \circ \eta = \eta'$. Now

$$\beta \circ \gamma = \beta \circ \eta + \beta(\gamma(x_0)) = \eta' + \beta(\gamma(x_0)) = \gamma' - \gamma'(x_0) + \beta(\gamma(x_0)).$$

Hence the isometry

$$\alpha := \beta + \gamma'(x_0) - \beta(\gamma(x_0))$$

satisfies $\alpha \circ \gamma = \gamma'$. Since α is an affine map and $\gamma(X)$ generates the affine space E, it is uniquely determined by its values on $\gamma(X)$.

(c) follows by combining the proof of (b) with Proposition 2.52(c). □

Remark 2.54. If $|X| = n$ and $\gamma\colon X \to \mathbb{E}$ is an isometric embedding generating the affine space $\mathbb{E}$, then $\dim \mathbb{E} \leq n - 1$.

More precisely, the construction of the embedding from the kernel

$$K(x,y) = d(x,x_0)^2 + d(y,x_0)^2 - d(x,y)^2$$

implies that $\dim \mathbb{E} = \operatorname{rank}(K)$, where K is considered as an $(n \times n)$-matrix (Exercise).

The preceding theorem also solves the so-called *congruence problem* for euclidean spaces: We call two subsets $X, Y \subseteq \mathbb{E}$ *isometric* if there exists a surjective isometry $\gamma\colon X \to Y$. We say that X and Y are *congruent* if there exists a motion $\phi \in \mathrm{Mot}(\mathbb{E})$ with $\phi(X) = Y$. Clearly, congruent subsets are isometric, and if X and Y generate $\mathbb{E}$ as an affine space, the converse is also true:

Theorem 2.55. (Congruence Theorem for euclidean spaces) *Let X and Y be two isometric subsets X, Y of a euclidean affine space $\mathbb{E}$. If X and Y both generate $\mathbb{E}$ as an affine space or X is contained in a finite-dimensional affine subspace, then they are congruent.*

Proof. Let $\gamma\colon X \to Y$ be a surjective isometry. Then Theorem 2.53(c) implies the existence of a $\phi \in \mathrm{Mot}(\mathbb{E})$ with $\phi|_X = \gamma$. In particular, we then have $\phi(X) = \gamma(X) = Y$. $\square$

Remark 2.56. Let $d\colon X \times X \to \mathbb{R}$ be a symmetric non-negative function on the 3-element set $X = \{a, b, c\}$. We put $x_0 := a$ and consider the kernel

$$K(x, y) := d(x, x_0)^2 + d(y, x_0)^2 - d(x, y)^2.$$

Then

$$\begin{aligned} A &:= \begin{pmatrix} K(a,a) & K(a,b) & K(a,c) \\ K(b,a) & K(b,b) & K(b,c) \\ K(c,a) & K(c,b) & K(c,c) \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2d(a,b)^2 & d(a,b)^2 + d(a,c)^2 - d(b,c)^2 \\ 0 & d(a,b)^2 + d(a,c)^2 - d(b,c)^2 & 2d(a,c)^2 \end{pmatrix} \end{aligned}$$

is positive semidefinite if and only if the lower right (2×2)-minor is non-negative (cf. Proposition 2.49, Example 2.51):

$$4d(a,b)^2 d(a,c)^2 \geq \big(d(a,b)^2 + d(a,c)^2 - d(b,c)^2\big)^2$$

This relation is equivalent to the two inequalities

$$2d(a,b)d(a,c) \geq \pm\big(d(a,b)^2 + d(a,c)^2 - d(b,c)^2\big).$$

The $+$-inequality is equivalent to

$$2d(a,b)d(a,c) \geq d(a,b)^2 + d(a,c)^2 - d(b,c)^2,$$

which is equivalent to

$$d(b,c)^2 \geq (d(a,b) - d(a,c))^2,$$

i.e.,

$$d(b,c) \geq \pm(d(b,a) - d(a,c)),$$

i.e.,

$$d(a,b) \leq d(b,c) + d(a,c) \quad \text{ and } \quad d(a,c) \leq d(a,b) + d(b,c).$$

The $-$-inequality is equivalent to

$$2d(a,b)d(a,c) \geq d(b,c)^2 - d(a,b)^2 - d(a,c)^2,$$

which is equivalent to

$$d(b,c)^2 \leq (d(a,b) + d(a,c))^2,$$

i.e.,

$$d(b,c) \leq d(a,b) + d(a,c).$$

We conclude that the positive definiteness of the kernel K is equivalent to the triangle inequality for (X,d).

If, conversely, (X,d) is a metric space, then Theorem 2.53 leads to an isometric embedding of X into a euclidean space, actually into $\mathbb{R}^2$.

Example 2.57. On $\mathbb{R}^2$ we consider the sup-metric

$$d(x,y) := \max(|x_1 - y_1|, |x_2 - y_2|) = \|x - y\|_\infty.$$

Then $X := \{0, e_1, e_1 \pm e_2\}$ is a four-element metric space. We have

$$d(0,e_1) = d(0, e_1 \pm e_2) = 1, d(e_1, e_1 \pm e_2) = 1, d(e_1 - e_2, e_1 + e_2) = 2.$$

For $x_0 := 0$, $x_1 := e_1$, $x_2 := e_1 + e_2$ and $x_3 := e_1 - e_2$, we thus obtain the kernel

$$K(x,y) = d(x,x_0)^2 + d(y,x_0)^2 - d(x,y)^2$$

with the matrix

$$(K(x_i,x_j))_{1 \leq i,j \leq 4} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 1 & 2 & -2 \\ 0 & 1 & -2 & 2 \end{pmatrix}.$$

This matrix has a negative minor

$$\det \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & -2 \\ 1 & -2 & 2 \end{pmatrix} = 8 - 2 - 2 - (2 + 8 + 2) = -8 < 0,$$

so that K is not positive definite. This means that (X,d) has no isometric embedding into a euclidean space $\mathbb{R}^3$.

Remark 2.58. We have above that four-point subsets of a metric space (X,d) are not necessarily isometric to a subset of euclidean space. Accordingly one says that a metric space (X,d) has the *n-point property* if every n-element subset $F \subseteq X$ is isometric to a subset of euclidean space. That every metric space has the 3-point property corresponds to the triangle inequality (cf. Example 2.47 and Remark 2.56). Likewise the inequalities corresponding to the 4-point property are called *tetrahedral inequalities*. These are inequalities of the form $\det A \geq 0$, where $A = (K(x_i,x_j))_{1 \leq i,j \leq 3}$ and

$$K(x,y) = d(x,x_0)^2 + d(y,x_0)^2 - d(x,y)^2.$$

As a consequence of the Embedding Theorem 2.53, a metric space has the n-point property for every $n \in \mathbb{N}$ if and only if it can be embedded in some euclidean space. This follows from the fact that a kernel is positive definite if and only if all its restrictions to finite subsets are positive definite.

It can be shown for a rather large class of metric spaces that the 4-point property implies the n-point property for every $n \in \mathbb{N}$ ([Wi32]). For an interesting survey on the 4-point property we refer to [Bl75].

Definition 2.59. (The parallelogram law in metric spaces) Let $\mathbb{E}$ be an affine euclidean space and $E = \overrightarrow{\mathbb{E}}$ be the corresponding euclidean vector space. For $v, w \in E$, we then have the relation

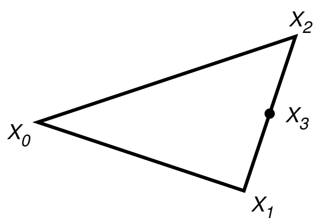
$$\|v - w\|^2 + \|v + w\|^2 = 2\|v\|^2 + 2\|w\|^2, \tag{EPL}$$

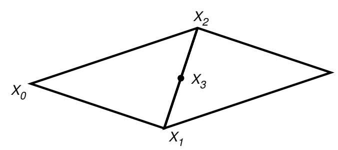
which is also called the *euclidean parallelogram law.*

For any triangle $\Delta(x_0, x_1, x_2)$ and the midpoint x_3 of x_1 and x_2 (with $v = \overrightarrow{x_0x_1}$ and $w = \overrightarrow{x_0x_2}$) we thus obtain the relation

$$d(x_1, x_2)^2 + 4d(x_0, x_3)^2 = 2d(x_0, x_1)^2 + 2d(x_0, x_2)^2 \tag{PL}$$

which is called the *(metric) parallelogram law* because it relates the sides of the parallelogram with vertices $(x_0, x_1, x_2, x_0 + \overrightarrow{x_0x_1} + \overrightarrow{x_0x_2})$ to the lengths of the diagonals.





Example 2.60. (The parallelogram law) Let (X, d) be a metric space and

$$S := \{x_0, x_1, x_2, x_3\} \subseteq X$$

such that x_3 is a midpoint of x_1 and x_2, i.e.,

$$d(x_1, x_3) = d(x_2, x_3) = \tfrac{1}{2} d(x_1, x_2). \tag{14}$$

We want to see when S has an embedding into a euclidean space. If this is the case for all subsets S as above, then (X, d) is said to have the *weak* 4*-point property* because it is a condition referring only to special point configurations (cf. [Bl75]).

We consider the kernel

$$K(x, y) = d(x, x_3)^2 + d(y, x_3)^2 - d(x, y)^2.$$

For $h := d(x_1, x_3)$ and $d_j := d(x_0, x_j)$, $j = 1, 2, 3$, we obtain for K the matrix

$$\begin{pmatrix} 2d_3^2 & d_3^2 + h^2 - d_1^2 & d_3^2 + h^2 - d_2^2 & 0 \\ d_3^2 + h^2 - d_1^2 & 2h^2 & -2h^2 & 0 \\ d_3^2 + h^2 - d_2^2 & -2h^2 & 2h^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since the triangle equality holds in X, all 2×2-minors are non-negative. Therefore K is positive definite if and only if

$$\begin{aligned} &\det \begin{pmatrix} 2d_3^2 & d_3^2 + h^2 - d_1^2 & d_3^2 + h^2 - d_2^2 \\ d_3^2 + h^2 - d_1^2 & 2h^2 & -2h^2 \\ d_3^2 + h^2 - d_2^2 & -2h^2 & 2h^2 \end{pmatrix} \\ &= -2h^2(2(d_3^2 + h^2 - d_1^2)(d_3^2 + h^2 - d_2^2) + (d_3^2 + h^2 - d_2^2)^2 + (d_3^2 + h^2 - d_1^2)^2) \\ &= -2h^2\big((d_3^2 + h^2 - d_2^2) + (d_3^2 + h^2 - d_1^2)\big)^2 \\ &= -2h^2\big(2(d_3^2 + h^2) - (d_2^2 + d_1^2)\big)^2 \leq 0. \end{aligned}$$

This expression is non-negative if and only if it vanishes, which means that

$$2(d_3^2 + h^2) = d_2^2 + d_1^2,$$

after multiplication by 2,

$$4d(x_0, x_3)^2 + d(x_1, x_2)^2 = 2d(x_0, x_1)^2 + 2d(x_0, x_2)^2. \tag{15}$$

For points x_0, x_1, x_2 in a euclidean affine space $\mathbb{E}$ and the midpoint x_3 of x_1 and x_2, this is the parallelogram law. Therefore (15) is necessary for the embeddability of (X, d) into some euclidean space. Conversely, the Embedding Theorem 2.53 also shows that it is sufficient for the subset $S \subseteq X$. In this sense the weak 4-point property is an abstract form of the parallelogram law in the metric space (X, d).

A closely related problem is to determine when a normed spaces $(X, \|\cdot\|)$ is euclidean, i.e., there exists a scalar product $\langle\cdot,\cdot\rangle$ such that $\|v\| = \sqrt{\langle v, v\rangle}$ holds for $v \in X$. Clearly, the parallelogram law

$$\|v - w\|^2 + \|v + w\|^2 = 2\|v\|^2 + 2\|w\|^2$$

is a necessary condition. It has been shown by Jordan and von Neumann that this condition, which corresponds to the weak 4-point property in the corresponding metric space (X, d), is also sufficient ([JvN35]). An earlier result in this context is the observation by M. Fréchet [Fr35] that a necessary and sufficient condition is that the kernel

$$K(v, w) := \|v\|^2 + \|w\|^2 - \|v - w\|^2$$

is positive definite. By Theorem 2.53, Fréchet's condition implies the existence of an isometric embedding $\gamma\colon X \to E$ into a euclidean space E with $\gamma(0) = 0$, so that it encodes the n-point property of (X, d) for every n. The main point of the Jordan–von Neumann Theorem is that it reduces all that to the weak 4-point property, which is much easier to verify.

Theorem 2.61. (Jordan–von Neumann) *Let $(X, \|\cdot\|)$ be a normed space satisfying the parallelogram law*

$$\|v - w\|^2 + \|v + w\|^2 = 2\|v\|^2 + 2\|w\|^2 \quad \text{for} \quad v, w \in X.$$

Then X is euclidean, i.e., there exists a symmetric bilinear form $\langle\cdot,\cdot\rangle$ on X with $\|v\|^2 = \langle v, v\rangle$ for $v \in X$.

Proof. (cf. [Gl66]) For $Q(x) := \frac{1}{4}\|x\|^2$, we consider the map

$$\gamma\colon X \times X \to \mathbb{R}, \quad \gamma(x, y) := Q(x + y) - Q(x - y).$$

The parallelogram law asserts that

$$Q(x + y) + Q(x - y) = 2Q(x) + 2Q(y) \quad \text{for} \quad x, y \in X. \tag{16}$$

We clearly have $Q(0) = 0$, so that we get for $x = 0$ the relation $Q(-y) = Q(y)$. This implies that γ is symmetric. For $x = y$, we further get $\gamma(x, x) = Q(2x) = 4Q(x) = \|x\|^2$.

Next we show that γ is additive in the first argument. For $x, y, z \in X$, we obtain

$$\begin{aligned} 2\gamma(x, z) + 2\gamma(y, z) &= 2Q(x + z) + 2Q(y + z) - 2Q(x - z) - 2Q(y - z) \\ &= Q(x + y + 2z) + Q(x - y) - Q(x + y - 2z) - Q(x - y) \\ &= \gamma(x + y, 2z), \end{aligned}$$

so that

$$2\gamma(x,z) + 2\gamma(y,z) = \gamma(x+y, 2z). \tag{17}$$

For $y = 0$ we obtain with $\gamma(x,0) = 0 = \gamma(0,y)$,

$$2\gamma(x,z) = \gamma(x, 2z),$$

so that (17) yields

$$\gamma(x,z) + \gamma(y,z) = \gamma(x+y, z).$$

For $x = y$ we obtain by induction $\gamma(nx, z) = n\gamma(x,z)$ for $n \in \mathbb{N}$. Replacing x by $\frac{x}{n}$, leads to $\gamma(\frac{x}{n}, z) = \frac{1}{n}\gamma(x,z)$ for $n \in \mathbb{N}$. Therefore the map $\gamma(\cdot, z)$ is $\mathbb{Q}$-linear, and since it is continuous, it is $\mathbb{R}$-linear. Finally we recall that γ is symmetric, hence bilinear. Therefore γ is a scalar product on X for which $\|\cdot\|$ is the corresponding norm. $\square$

We conclude this section with a result on the compatibility of embeddings into euclidean spaces with isometries of (X,d).

Theorem 2.62. *Let (X,d) be a metric space and $\gamma\colon X \to \mathbb{E}$ be an isometric embedding into an affine euclidean space whose image is not contained in a proper affine subspace of $\mathbb{E}$. Then, for every isometry $g \in \mathrm{Isom}(X,d)$, there exists a unique $\alpha_g \in \mathrm{Mot}(\mathbb{E})$ with*

$$\alpha_g \circ \gamma = \gamma \circ g. \tag{18}$$

The following assertions hold:

(i) $\alpha\colon \mathrm{Isom}(X,d) \to \mathrm{Mot}(\mathbb{E})$ *is a group homomorphism, so that we obtain an affine isometric action of* $\mathrm{Isom}(X,d)$ *on* $\mathbb{E}$.

(ii) $\gamma\colon X \to \mathbb{E}$ *is* $\mathrm{Isom}(X,d)$*-equivariant, i.e.,* (18) *holds.*

Proof. Since $\gamma \circ g\colon X \to \mathbb{E}$ is also isometric, Theorem 2.55 implies for each $g \in \mathrm{Isom}(X,d)$ the existence of a unique isometry $\alpha_g \in \mathrm{Mot}(\mathbb{E})$ with $\alpha_g \circ \gamma = \gamma \circ g$.

(i) For $g, h \in \mathrm{Isom}(X,d)$ we have

$$\alpha_g\alpha_h \circ \gamma = \alpha_g \circ \gamma \circ h = \gamma \circ gh = \alpha_{gh} \circ \gamma,$$

and since $\gamma(X)$ generates the affine space $\mathbb{E}$, we obtain $\alpha_g \circ \alpha_h = \alpha_{gh}$.

(ii) is a consequence of the definition of α_g. $\square$

Remark 2.63. The preceding theorem can be used to obtain a new proof for the fact that every isometry of an affine euclidean space $\mathbb{E}$ is affine (cf. Theorem 2.12). In fact, let $g\colon \mathbb{E} \to \mathbb{E}$ be an isometry. We apply Theorem 2.62 with $X = \mathbb{E}$ and $\gamma = \mathrm{id}_{\mathbb{E}}$. This leads to $g = \alpha_g$, so that g is affine.

Example 2.64. For $X := \mathbb{R}$, we call a metric d *translation invariant* if all the translations $\tau_x(y) := x + y$ are isometries. Any such metric is of the form $d(x,y) = f(|x-y|)$ for a function $f\colon \mathbb{R}_+ \to \mathbb{R}_+$. The functions f for which we thus obtain a metric that permits an isometric embedding $\gamma\colon \mathbb{R} \to \mathbb{E}$ into euclidean spaces have been determined by Schoenberg and von Neumann ([SvN41]). Then the embedding γ is called a *screw line* or a *helix*.

Typical examples of helices in $\mathbb{R}^3$ are given by

$$\xi\colon \mathbb{R} \to \mathbb{R}^3, \quad \xi(x) = (a\cos x, a\sin x, bx).$$

Then

$$d(x,y) := \|\xi(x) - \xi(y)\| = \sqrt{4a^2 \sin^2\left(\frac{x-y}{2}\right) + b^2(x-y)^2}$$

is a translation invariant metric on $\mathbb{R}$ and ξ is an isometric embedding $(\mathbb{R}, d) \hookrightarrow \mathbb{R}^3$.

The corresponding affine isometric action α of the group $G = (\mathbb{R}, +)$ on $\mathbb{R}^3$ whose existence is ensured by Theorem 2.62 is given by

$$\alpha_t(x,y,z) := (\cos t \cdot x - \sin t \cdot y, \sin t \cdot x + \cos t \cdot y, z + bt),$$

so that $\xi(x) = \alpha_x(a, 0, 0)$ is an orbit of this action.

Exercises for Section 2

Exercise 2.1. Find examples of points x, y in metric spaces for which:

(i) No midpoint exists.

(ii) More than one midpoint exists.

Exercise 2.2. Consider the euclidean space

$$E := \mathbb{R}^{(\mathbb{N})} = \{(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \colon (\exists N \in \mathbb{N})(\forall n > N) a_n = 0\}$$

of all finite sequence, endowed with the canonical scalar product $\langle a, b\rangle := \sum_{n=1}^{\infty} a_n b_n$. Show that

$$H := \left\{a = (a_n) \in E \colon \sum_n a_n = 0\right\}$$

is a hyperplane with $H^\perp = \{0\}$.

Exercise 2.3. Let E be a euclidean space and $0 < k_1 < \ldots < k_N$ natural numbers. We consider the corresponding set of flags

$$\mathcal{F} := \left\{(F_1, \ldots, F_N) \in \prod_{j=1}^{n} \mathrm{Gr}_{k_j}(E) \colon F_1 \subseteq \ldots \subseteq F_N, \dim F_j = k_j\right\},$$

where the $F_j \subseteq E$ are linear subspaces of dimension k_j. Show that the orthogonal group $\mathrm{O}(E)$ acts transitively on $\mathcal{F}$. Hint: Proposition 2.23.

Exercise 2.4. Show that, for $\alpha, \beta \in \mathbb{R}$, we have

$$D(\alpha)D(\beta) = D(\alpha + \beta) \quad \text{for the rotation matrices} \quad D(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Exercise 2.5. Show that the linear endomorphism $\phi \colon \mathbb{R}^2 \to \mathbb{R}^2, x \mapsto Qx$, defined by the matrix

$$Q = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = D(\theta) \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

is a reflection in the line ℓ intersecting the x_1-axis in the angle $\frac{\theta}{2}$.

Exercise 2.6. Let $M \in \mathbb{R}^2$ and $\theta \in \mathbb{R}$. Find a formula for the rotation ϕ of the euclidean plane $\mathbb{R}^2$ with fixed point M and angle θ.

Exercise 2.7. Show that, for a closed convex polygon in the plane, with vertices $P_1, \ldots, P_N$ and the corresponding angles

$$\alpha_1 := \measuredangle(\overrightarrow{P_1P_N}, \overrightarrow{P_1P_2}), \quad \alpha_j := \measuredangle(\overrightarrow{P_jP_{j-1}}, \overrightarrow{P_jP_{j+1}}) \quad \text{for} \quad j > 1,$$

we have

$$\alpha_1 + \cdots + \alpha_N = (N-2)\pi.$$

Hint: Generalize the proof of Proposition 2.37.

Exercise 2.8. Let $a, b, c \in \mathbb{R}$ with $(a, b) \neq (0, 0)$ and consider the affine line

$$H := \{(x, y) \in \mathbb{R}^2 \colon ax + by = c\}.$$

Find a formula for the orthogonal reflection $r \colon \mathbb{R}^2 \to \mathbb{R}^2$ in terms of a, b, c.

Exercise 2.9. Let $\mathbb{E}$ be an affine euclidean space. For $m \in \mathbb{E}$, we consider the sphere

$$\mathbb{S}(m, r) := \{P \in \mathbb{E} \colon d(P, m) = r\}$$

of radius r and center m.

(i) Show that, for two such spheres $\mathbb{S}(m, r)$ and $\mathbb{S}(m', r')$, the intersection is either empty or a sphere in an affine subspace of $\mathbb{E}$. Determine the center and the radius of this sphere. Hint: If this is a little too abstract, discuss the case $\mathbb{E} = \mathbb{R}^2$ first. In this case we are dealing with intersections of two circles.

(ii) $\mathbb{S}(m, r)$ and $\mathbb{S}(m', r')$ intersect if and only if the following "triangle inequalities" hold:

$$d(m, m') \leq r + r', \quad r' \leq d(m, m') + r, \quad r \leq d(m, m') + r'.$$

Exercise 2.10. Let $\mathbb{E}$ be an affine euclidean space and $r_{H_1}, r_{H_2} \in \mathrm{Mot}(\mathbb{E})$ be two orthogonal reflections in parallel hyperplanes H_1 and H_2. Show that their composition $r_{H_1} \circ r_{H_2}$ is a translation in the direction of a vector orthogonal to both hyperplanes.

Exercise 2.11. Let $\mathbb{E}$ be an affine euclidean space and $r_{H_1}, r_{H_2} \in \mathrm{Mot}(\mathbb{E})$ be two orthogonal reflections in two hyperplanes H_1 and H_2 which are not parallel. Give a geometric description of their composition $r_{H_1} \circ r_{H_2}$. Hint: Start with the case $\dim \mathbb{E} = 2$ and reduce the general case to the 2-dimensional case.

Exercise 2.12. (Metric characterization of midpoints) Let $(X, \|\cdot\|)$ be a normed space and $x, y \in X$ distinct points. Let

$$M_0 := \{z \in X : \|z - x\| = \|z - y\| = \tfrac{1}{2}\|x - y\|\} \quad \text{and} \quad m := \frac{x + y}{2}.$$

For a subset $A \subseteq X$ we define its *diameter*

$$\delta(A) := \sup\{\|a - b\| : a, b \in A\}.$$

Show that:

(1) For $z \in M_0$, we have $\|z - m\| \leq \frac{1}{2}\delta(M_0) \leq \frac{1}{2}\|x - y\|$.

(2) For $n \in \mathbb{N}$ we define inductively:

$$M_n := \{p \in M_{n-1} : (\forall z \in M_{n-1})\ \|z - p\| \leq \tfrac{1}{2}\delta(M_{n-1})\}.$$

Then we have for each $n \in \mathbb{N}$:

(a) M_n is a convex set.

(b) M_n is invariant under the point reflection $r_m(a) := 2m - a$ in m.

(c) $m \in M_n$.

(d) $\delta(M_n) \leq \frac{1}{2}\delta(M_{n-1})$.

(3) $\bigcap_{n\in\mathbb{N}} M_n = \{m\}$.

Exercise 2.13. (Isometries of normed spaces are affine maps) Let $(X, \|\cdot\|)$ be a normed space endowed with the metric $d(x, y) := \|x - y\|$. Show that each isometry $\phi\colon (X, d) \to (X, d)$ is an affine map by using the following steps:

(1) It suffices to assume that $\phi(0) = 0$ and to show that this implies that ϕ is a linear map.

(2) $\phi(\frac{x+y}{2}) = \frac{1}{2}(\phi(x) + \phi(y))$ for $x, y \in X$. Hint: Exercise 2.12.

(3) ϕ is continuous.

(4) $\phi(\lambda x) = \lambda\phi(x)$ for $\lambda \in 2^{\mathbb{Z}} \subseteq \mathbb{R}$.

(5) $\phi(x + y) = \phi(x) + \phi(y)$ for $x, y \in X$.

(6) $\phi(\lambda x) = \lambda\phi(x)$ for $\lambda \in \mathbb{R}$.

Exercise 2.14. Let $\beta\colon V \times V \to V$ be a symmetric bilinear form on the vector space V and

$$q\colon V \to V, \quad v \mapsto \beta(v, v)$$

the corresponding quadratic form. Then, for $\phi \in \mathrm{End}(V)$, the following are equivalent:

(1) $(\forall v \in V)\ q(\phi(v)) = q(v)$.

(2) $(\forall v, w \in V)\ \beta(\phi(v), \phi(w)) = \beta(v, w)$.

Hint: Use the polarization identity $\beta(v, w) = \frac{1}{4}\big(q(v + w) - q(v - w)\big)$.

Exercise 2.15. Let E be a euclidean vector space and $\phi \in \mathrm{O}(E)$ a product of n reflections. Show that

$$\mathrm{rank}(\phi - \mathbf{1}) \leq n.$$

Conclude that, if V is infinite-dimensional, then not every element of $\mathrm{O}(E)$ is a finite product of reflections.

Exercise 2.16. Consider in $\mathbb{R}^2$ the curve

$$\gamma\colon [0, 2\pi] \to \mathbb{R}^2, \quad \gamma(t) = (\cos t, \sin t).$$

Calculate its length and explain how the definition of $L(\gamma)$ as a limit of the numbers of the form $L_{\mathfrak{t}}(\gamma)$ can be used to derive a practical method to calculate π.

3 Spherical Geometry

After the euclidean affine geometry in the preceding section, we now turn to spherical geometry, i.e., the metric geometry of the unit sphere in a euclidean vector space. Here the 2-sphere $\mathbb{S}_2 \subseteq \mathbb{R}^3$ is of particular interest because it corresponds quite well to the geometry on the surface of the earth.

We start by defining the metric on the unit sphere $\mathbb{S}(E)$ in a euclidean vector space E. Then we show that its group of isometries can be identified with the orthogonal group $\mathrm{O}(E)$ of linear isometries of E. We then turn to geodesics and triangles and conclude this section with a brief discussion of the link with the Riemannian approach to geometry where the distance is derived from the lengths of curves.

3.1 The metric on the euclidean sphere

We consider the *n-dimensional unit sphere*

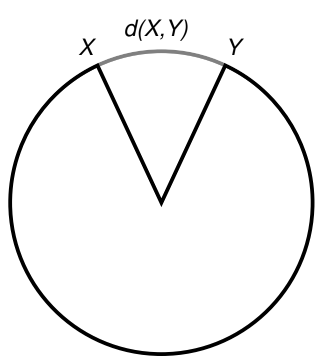
$$\mathbb{S}_n := \{x \in \mathbb{R}^{n+1} \colon \|x\| = 1\} = \{x \in \mathbb{R}^{n+1} \colon x_1^2 + \cdots + x_{n+1}^2 = 1\}$$

and more generally, the unit sphere

$$\mathbb{S}(E) := \{x \in E \colon \|x\| = 1\}$$

in a euclidean vector space. For $x, y \in \mathbb{S}(E)$, the Cauchy–Schwarz inequality implies that $|\langle x, y\rangle| \leq 1$, so that we may define a function

$$d \colon \mathbb{S}(E) \times \mathbb{S}(E) \to \mathbb{R}_+, \quad d(x,y) := \measuredangle(x,y) = \arccos(\langle x, y\rangle) \in [0, \pi].$$



Lemma 3.1. *$(\mathbb{S}(E), d)$ is a metric space.*

Proof. Clearly, $d(x,y) = d(y,x)$ and $d(x,x) \geq 0$. If $d(x,y) = 0$, then $\langle x, y\rangle = 1$, so that the second part of Proposition 2.2 on the Cauchy–Schwarz inequality implies that $x = y$.

It therefore remains to verify the triangle inequality. So let $x, y, z \in \mathbb{S}(E)$. To verify that

$$d(x,z) \leq d(x,y) + d(y,z),$$

we may w.l.o.g. assume that $d(x,y) + d(y,z) \leq \pi$; otherwise there is nothing to show. Then it suffices to verify the relation

$$\begin{aligned}\langle x, z\rangle &= \cos(d(x,z)) \geq \cos(d(x,y) + d(y,z)) \\ &= \cos(d(x,y))\cos(d(y,z)) - \sin(d(x,y))\sin(d(y,z)) \\ &= \langle x,y\rangle\langle y,z\rangle - \sqrt{1-\langle x,y\rangle^2}\sqrt{1-\langle y,z\rangle^2}.\end{aligned}$$

To verify this relation, we write $x = x_0 + x_1$, $z = z_0 + z_1$ with $x_0, z_0 \in \mathbb{R}y$ and $x_1, z_1 \perp y$, i.e.,

$$x_0 := \langle x,y\rangle y, \quad x_1 := x - x_0, \quad z_0 := \langle z,y\rangle y, \quad z_1 := z - z_0.$$

Then Pythagoras' Theorem (Lemma 2.9) leads to

$$1 = \|x\|^2 = \|x_0\|^2 + \|x_1\|^2 = \langle x,y\rangle^2 + \|x_1\|^2 \quad \text{and} \quad 1 = \|z\|^2 = \langle z,y\rangle^2 + \|z_1\|^2.$$

We thus obtain with the CS inequality

$$\langle x,y\rangle\langle y,z\rangle - \sqrt{1-\langle x,y\rangle^2}\sqrt{1-\langle y,z\rangle^2} = \langle x_0, z_0\rangle - \|x_1\|\|z_1\| \leq \langle x_0, z_0\rangle + \langle x_1, z_1\rangle = \langle x, z\rangle.$$

This proves the triangle inequality. □

Remark 3.2. For any metric space (X,d), it is of some interest to analyze for which triples (x,y,z) we have the *triangle equality*:

$$d(x,z) = d(x,y) + d(y,z). \tag{19}$$

If this is the case for $x,y,z \in \mathbb{S}(E)$, then the preceding proof implies that

$$\langle x_1, z_1\rangle = -\|x_1\|\|z_1\|.$$

This means that either $x_1 = 0$ or that $z_1 \in \mathbb{R}x_1$ with $\langle z_1, x_1\rangle \leq 0$ (cf. the second part of Proposition 2.2 on the Cauchy–Schwarz inequality).

In the first case $x \in \mathbb{R}y$ implies $x \in \{\pm y\}$. The case $x = y$ is trivial, and if $x = -y$, then the assumption $d(x,y) + d(y,z) \leq \pi$ leads to $y = z$. Likewise $z = -y$ leads to $x = y$. This is the case where all three points lie in a one-dimensional linear subspace of E.

In the other case ($x_1 \neq 0$), so that x, y and z generate a 2-dimensional plane F in which y and $\frac{1}{\|x_1\|}x_1$ form an orthonormal basis and $x,y,z \in \mathbb{S}(F) \cong \mathbb{S}_1$. In this case the triangle equality holds if and only if $z_1 \in -\mathbb{R}_+ x_1$, i.e., $\langle z_1, x_1\rangle \leq 0$, and $d(x,y) + d(y,z) \leq \pi$ (Exercise: Visualize this condition on the unit circle!).

Remark 3.3. If $\iota\colon E_1 \to E_2$ is an isometric embedding of euclidean vector spaces, then the corresponding map

$$\iota|_{\mathbb{S}(E_1)}\colon \mathbb{S}(E_1) \to \mathbb{S}(E_2)$$

also is an isometry.

3.2 Geodesics on spheres

Proposition 3.4. (Uniqueness of midpoints on spheres) *For two points $x, y \in \mathbb{S}(E)$ with $d(x,y) < \pi$, there exists a unique midpoint m, i.e., $d(x,m) = d(y,m) = \frac{1}{2}d(x,y)$. It lies in the plane generated by x and y.*

Proof. We may w.l.o.g. assume that $x \neq y$, which implies that both are linearly independent because $x \neq -y$ follows from $d(x,y) < \pi$. Clearly $F := \mathbb{R}x + \mathbb{R}y$ is 2-dimensional because $y \notin \mathbb{R}x$, and therefore $\mathbb{S}(F) \cong \mathbb{S}_1$ contains a midpoint z of x and y. It can be obtained as one of the two points in $\mathbb{S}(F)$ that are fixed by the linear reflection in the line $(x-y)^\perp$ which exchanges x and y (Example 2.20).

Any midpoint m of x and y satisfies the triangle equality $d(x,m)+d(m,y) = d(x,y) < \pi$, so that Remark 3.2 implies that x, y, m lie in a linear subspace of dimension 2, so that $m \in F$. Now an elementary consideration in $\mathbb{S}_1$ shows that $m = z$. □

Remark 3.5. (Maximal geodesics) Suppose that $v_1, v_2 \in E$ are orthogonal unit vectors. We consider the curve

$$\gamma\colon [0,\pi] \to \mathbb{S}(E), \quad \gamma(t) := \cos t \cdot v_1 + \sin t \cdot v_2$$

from $\gamma(0) = v_1$ to $\gamma(\pi) = -v_1$. For $0 \leq t \leq s \leq \pi$ we then have

$$d(\gamma(t),\gamma(s)) = \arccos(\langle\gamma(t),\gamma(s)\rangle) = \arccos(\cos t\cos s+\sin t\sin s) = \arccos(\cos(s-t)) = s-t.$$

Therefore γ is a geodesic of unit speed. Since the maximal distance between two points in $\mathbb{S}(E)$ is π, the geodesic γ cannot be extended to a geodesic curve defined on any larger interval.

Definition 3.6. If $F \subseteq E$ is a 2-dimensional linear subspace, then $F \cap \mathbb{S}(E) = \mathbb{S}(F)$ is called a *great circle* in the sphere $\mathbb{S}(E)$. In Remark 3.5 we have seen that any half of a great circle leads to a geodesic in the metric space $\mathbb{S}(E)$, and this is also true for any subsegment of length smaller than π.

Clearly, for $\dim E \geq 2$, there are many geodesics connecting two opposite points $\pm x$ in the sphere $\mathbb{S}(E)$. However, if two points do not lie on the same line, they can be connected by a unique geodesic:

Theorem 3.7. (Uniqueness of geodesics) *For any two points $x, y \in \mathbb{S}(E)$ with $d(x,y) < \pi$, there is a unique geodesic*

$$\gamma\colon [0,1] \to \mathbb{S}(E) \quad \textit{with} \quad \gamma(0) = x, \gamma(1) = y.$$

Proof. The existence of a geodesic follows from the observation that $F := \mathbb{R}x + \mathbb{R}y$ is a plane for which the great circle $\mathbb{S}(F)$ contains x and y (Recall that $x \neq y$ and $d(x,y) < \pi$ imply that x and y are linearly independent). The smaller piece of $\mathbb{S}(F)$ obtained by cutting in the two points x and y now provides a geodesic segment γ as required.

To see that this is unique, let $\eta\colon [0,1] \to \mathbb{S}(E)$ be any geodesic from x to y. Then its speed is $c = d(x,y)$. Any $z := \eta(t)$ satisfies

$$d(x,z) = d(\eta(0),\eta(t)) = td(x,y) \quad \text{and} \quad d(z,y) = d(\eta(t),\eta(1)) = (1-t)d(x,y),$$

so that

$$d(x,z) + d(z,y) = d(x,y).$$

Now Remark 3.2 implies that $\eta([0,1]) \subseteq F$. The uniqueness of midpoints further leads to

$$\eta(\tfrac{1}{2}) = \gamma(\tfrac{1}{2}),$$

and we inductively obtain $\eta(t) = \gamma(t)$ for all dyadic numbers $t = \frac{n}{2^k}$, $0 \leq n \leq 2^k$, $k \in \mathbb{N}$. Since both curves η and γ are continuous, it follows that $\eta = \gamma$. □

3.3 The group of isometries

In this subsection we show that every isometry of the sphere $\mathbb{S}(E)$ is obtained by the restriction of a unique orthogonal linear map. This shows that the symmetry group $\mathrm{Isom}(\mathbb{S}(E), d)$ of spherical geometry is the orthogonal group $\mathrm{O}(E)$ of the euclidean space E.

Theorem 3.8. (Isometries are orthogonal maps) *For $\phi \in \mathrm{O}(E)$, the restriction to $\mathbb{S}(E)$ defines a surjective isometry ϕ_S of $\mathbb{S}(E)$ and all surjective isometries of $\mathbb{S}(E)$ are of this form.*

Proof. It is obvious that every $\phi \in \mathrm{O}(E)$ of E preserves the sphere

$$\mathbb{S}(E) = \{v \in E \colon d(v, 0) = 1\}$$

because it fixes 0. Since ϕ preserves scalar products, it induces a surjective isometry ϕ_S of $\mathbb{S}(E)$.

Suppose, conversely, that $\psi \colon \mathbb{S}(E) \to \mathbb{S}(E)$ is a surjective isometry. For $x, y \in \mathbb{S}(E)$ we then have the relation

$$\langle \phi(x), \phi(y) \rangle = \cos(d(\phi(x), \phi(y))) = \cos(d(x, y)) = \langle x, y \rangle,$$

so that Proposition 2.52 implies the existence of a linear isometry $\alpha \colon E \to E$ with $\alpha|_{\mathbb{S}(E)} = \phi$. The surjectivity of ϕ now implies that the linear map α is also surjective, hence an element of $\mathrm{O}(E)$. $\square$

Corollary 3.9. *For every euclidean vector space E, we have $\mathrm{Isom}(\mathbb{S}(E), d) \cong \mathrm{O}(E)$.*

The preceding theorem shows that $\mathrm{O}(E)$ is the symmetry group of spherical geometry. Proposition 2.23 asserts in particular that $\mathrm{O}(E)$ acts transitively on $\mathbb{S}(E)$. For its action on the set $\mathbb{S}(E)^2$ of pairs by $g.(x, y) := (gx, gy)$, the distance function $d \colon \mathbb{S}(E) \times \mathbb{S}(E) \to \mathbb{R}$ is invariant. We actually have the following converse, which means that the distance functions actually separates the $\mathrm{O}(E)$-orbits in $\mathbb{S}(E)$.

Theorem 3.10. *The action of $\mathrm{O}(E)$ on $\mathbb{S}(E)$ is 2-point transitive, i.e., if $x, y, x', y' \in \mathbb{S}(E)$ satisfy*

$$d(x, y) = d(x', y'),$$

then there exists a $\phi \in \mathrm{O}(E)$ with $\phi(x) = x'$ and $\phi(y) = y'$.

Proof. Using Proposition 2.23, we find a $\phi \in \mathrm{O}(E)$ with $\phi(x) = x'$, so that we may assume that $x = x'$. For every $r \in [0, \pi]$, the set

$$B_r(x) = \{v \in \mathbb{S}(E) \colon d(x, v) = r\}$$

is of the form

$$\{v \in \mathbb{S}(E) \colon \langle v, x \rangle = \cos r\} = \big\{v \in \mathbb{S}(E) \colon v = (\cos r)x + w, w \in x^\perp, \|w\| = \sin r\big\}.$$

It may therefore be identified with the sphere of radius $\sin r$ in the hyperplane $x^\perp$. Since $\mathrm{O}(x^\perp) = \mathrm{O}(E)_x$ acts transitively on this sphere (Proposition 2.23), there exists a $\phi \in \mathrm{O}(E)$ with $\phi(x) = x$ and $\phi(y) = y'$. $\square$

Remark 3.11. For a metric space (X, d), the condition that its isometry group $\mathrm{Isom}(X, d)$ acts 2-point transitively in the sense that, for $x, y, x', y' \in X$

$$d(x, y) = d(x', y') \quad \Rightarrow \quad \exists \phi \in \mathrm{Isom}(X, d) \;\; \phi(x) = x', \;\; \phi(y) = y'$$

is quite strong. Metric spaces with this property are called *2-point homogeneous*. We know already that affine euclidean spaces (Proposition 2.38) and euclidean spheres are 2-point homogeneous.

For $x = y$, we see in particular that, for any 2-point homogeneous space $\mathrm{Isom}(X, d)$ acts transitively on (X, d), so that (X, d) is *homogeneous*. For a homogeneous metric space, the 2-point homogeneity is equivalent to the transitivity of the stabilizer groups $\mathrm{Isom}(X, d)_x$ of every $x \in X$ on the spheres

$$S_r(x) := \{y \in X \colon d(x, y) = r\}$$

(Exercise!). Therefore it can be interpreted as an abstract version of *isotropy*, i.e., viewed from each point $x \in X$, the space X looks the same in the "direction" of any other $y \in X$.

One can show that there are very view types of *2-point homogeneous spaces*. The basic examples are the affine spaces, spheres, projective (Section 4) and hyperbolic spaces (Section **??**) over the fields $\mathbb{R}$, $\mathbb{C}$ and the skew-field $\mathbb{H}$ of quaternions. There is also an exceptional space corresponding to the "projective plane" over the 8-dimensional algebra $\mathbb{O}$ of octonions (cf. [Wa51, Wa52] for details and classification results).

Remark 3.12. (The Banach–Mazur problem) We have seen above that, for a euclidean space E, the group $\mathrm{O}(E)$ of linear isometries of E acts transitively on the unit sphere $\mathbb{S}(E)$. Simple examples, such as the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ on $\mathbb{R}^2$, show that, for a normed space $(X, \|\cdot\|)$, the group

$$\mathrm{O}(X) := \{\phi \in \mathrm{GL}(X) \colon (\forall v \in X)\, \|\phi(x)\| = \|x\|\}$$

need not act transitively on the unit sphere $\mathbb{S}(X) = \{v \in X \colon \|v\| = 1\}$. In particular, for $\|\cdot\|_1$ and $\|\cdot\|_\infty$ on $\mathbb{R}^2$, the group $\mathrm{O}(X)$ has 8 elements, it is the symmetry group of a square.

Therefore it is a natural question to ask if there are normed spaces for which $\mathrm{O}(X)$ acts transitively on $\mathbb{S}(X)$ but which are not euclidean. If $\dim X < \infty$, the answer to this question is 'no', but the proof requires some more advanced tools. There exist examples of inseparable Banach spaces X which are not euclidean, but for which $\mathrm{O}(X)$ acts transitively on $\mathbb{S}(X)$. For separable Banach spaces, it is still an open problem, whether such examples exist or not (cf. [Ra02] for recent progress and more references concerning this problem).

Sketch of the proof for $\dim X < \infty$: We may assume that $X = \mathbb{R}^n$. Then $\mathrm{O}(X) \subseteq \mathrm{GL}_n(\mathbb{R})$ can be viewed as a group of matrices. It is easily seen to be closed, and since all norms on $\mathbb{R}^n$ are equivalent, it is also bounded, hence compact. For compact groups the existence of an invariant probability measure on the group implies the existence of an invariant scalar product $\langle\cdot,\cdot\rangle$, so that there exists an $\mathrm{O}(X)$-invariant euclidean norm $\|\cdot\|'$ on X. Pick $x_0 \in \mathbb{S}(X)$ and let $c := \|x_0\|'$. Then $\|\phi(x_0)\|' = \|x_0\|' = c$ for every $\phi \in \mathrm{O}(X)$, and since $\mathrm{O}(X)$ acts transitively on $\mathbb{S}(X)$, it follows that $\|\cdot\|' = c\|\cdot\|$. This shows that the norm $\|\cdot\|$ is euclidean and corresponds to the scalar product $\frac{1}{c}\langle\cdot,\cdot\rangle$.

3.4 Congruence in the sphere

As in euclidean spaces, we call two subsets $X, Y \subseteq \mathbb{S}(E)$ *isometric* if there exists a surjective isometry $\gamma \colon X \to Y$. We say that X and Y are *congruent* if there exists a motion

$\phi \in \mathrm{Iso}(\mathbb{S}(E))$ with $\phi(X) = Y$. Clearly, congruent subsets are isometric, and if X and Y generate E as a linear space, the converse is also true. The following theorem is a substantial generalization of Theorem 3.10.

Theorem 3.13. (Congruence Theorem for spheres) *If two isometric subsets X, Y of $\mathbb{S}(E)$ generate E linearly or X is contained in a finite-dimensional subspace, then they are congruent.*

Proof. Let $\gamma\colon X \to Y$ be a surjective isometry. Then the definition of the metric on $\mathbb{S}(E)$ implies that

$$\langle x, x' \rangle = \langle \gamma(x), \gamma(x') \rangle \quad \text{for} \quad x, x' \in X.$$

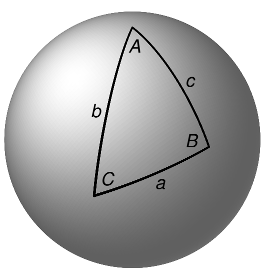
Therefore Proposition 2.52(c) implies the existence of $\phi \in \mathrm{O}(E)$ with $\phi|_X = \gamma$. $\square$

3.5 Spherical Triangles

In this section we study triangles on the euclidean sphere $\mathbb{S}_2$.

Definition 3.14. A *spherical triangle* $\Delta(ABC)$ is given by three points $A, B, C \in \mathbb{S}_2$ connected by geodesic segments (pieces of great circles) a from B to C, b from C to A, and c from A to B. [5]

We write $\angle A$, $\angle B$ and $\angle C$ for the *vertex angles* between the two sides of the triangle in A, B and C, respectively.



Theorem 3.15. *In a spherical triangle with vertices A, B, C in $\mathbb{S}_2$, we have*

$$\angle A + \angle B + \angle C = \pi + \mathrm{area}(\Delta(ABC)).$$

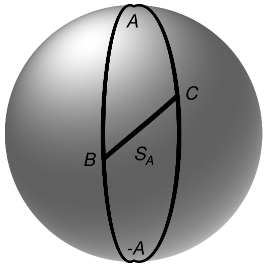
Proof. Let $S_A \subseteq \mathbb{S}_2$ be the sector bounded by the two geodesic arcs $\bar{b}$ and $\bar{c}$ from A to $-A$, obtained by extending b and c. Then

$$\mathrm{area}(S_A) = \frac{\angle A}{2\pi}\,\mathrm{area}(\mathbb{S}_2) = 2\angle A.$$

We likewise defines S_B and S_C and obtain

$$\mathrm{area}(S_B) = 2\angle B \quad \text{and} \quad \mathrm{area}(S_C) = 2\angle C.$$

[5] Recall that the geodesic segment connecting two points P, Q is uniquely determined if $d(P,Q) < \pi$ (Theorem 3.7). Otherwise $P = -Q$ and every great circle through P defines two geodesic segments from P to Q. To specify a triangle for which one side has length π, we therefore have to specify the corresponding geodesic segment explicitly. It is not determined by the vertices.



We write C_{AB} for the great circle containing A and B, and H_C for the hemisphere bounded by C_{AB} and containing C. It is the union of 4 triangles intersecting only in mutual boundaries:

$$H_C = \Delta(ABC) \cup \Delta((-A)BC) \cup \Delta((-A)(-B)C) \cup \Delta(A(-B)C).$$

We also have

$$S_A = \Delta(ABC) \cup \Delta((-A)BC), \quad S_B = \Delta(ABC) \cup \Delta(A(-B)C), \quad S_C = \Delta(ABC) \cup \Delta(AB(-C)).$$

Since the map $f\colon \mathbb{S}_2 \to \mathbb{S}_2, f(x) := -x$ is an isometry preserving area, we have

$$\operatorname{area}\big(\Delta((-A)(-B)C)\big) = \operatorname{area}\big(\Delta(AB(-C))\big).$$

We thus obtain

$$\begin{aligned} 2(\angle A + \angle B + \angle C) &= \operatorname{area}(S_A) + \operatorname{area}(S_B) + \operatorname{area}(S_C) = \operatorname{area}(H_C) + 2\operatorname{area}(\Delta(ABC)) \\ &= 2\pi + 2\operatorname{area}(\Delta(ABC)). \end{aligned}$$ □

The preceding result asserts in particular that the sum of the angles in a spherical triangle is larger than π. This is interpreted as *positive curvature* of the sphere. In Section **??** below, we shall see that, in hyperbolic space, the sum of angles in a geodesic triangle is smaller than π, and this is interpreted as *negative curvature*.

3.6 Euclidean versus spherical metric

The unit sphere $\mathbb{S}(E)$ in a euclidean vector space E carries two natural metrics. One is the metric d_E obtained by simply restricting the metric $\|x - y\|$ of E, and the other is the metric $d_S(x, y) = \arccos(\langle x, y\rangle)$ that we introduced above. In this short subsection we briefly discuss the difference between these two metrics. From Theorem 3.8 we know that both metrics have the same isometry group $\mathrm{O}(E)$, and thus define the same geometry in the sense of the Erlangen Program. Our main point is that both metrics lead to the same lengths of curves and since geodesics exist for d_S, this makes the metric d_S preferable from the point of view of metric spaces.

Lemma 3.16. *For a rectifiable curve $\gamma\colon [a, b] \to \mathbb{S}_n$, the length with respect to the metric d_S on $\mathbb{S}_n$ and the euclidean metric d_E coincide.*

Proof. For $x, y \in \mathbb{S}_n$, the spherical distance

$$d_S(x, y) = \arccos(\langle x, y\rangle)$$

is larger than the euclidean distance

$$d_E(x,y) = \|x-y\| = \sqrt{\|x-y\|^2} = \sqrt{2(1-\langle x,y\rangle)}.$$

In fact, for $0 \leq t \leq 2$, we have

$$1 - \frac{t^2}{2} \leq \cos t,$$

and for $t := \sqrt{2(1-\langle x,y\rangle)} = d_E(x,y)$ this leads to $\langle x,y\rangle \leq \cos t$, and further to $d_E(x,y) \leq d_S(x,y)$ (Exercise: Give a geometric reason for this fact!). This already shows that the length $L_E(\gamma)$ with respect to the euclidean metric and the length $L_S(\gamma)$ with respect to the spherical metric satisfy $L_E(\gamma) \leq L_S(\gamma)$.

Next we observe that the function

$$f(t) := \sum_{n=0}^{\infty} (-1)^n \frac{t^n}{(2n)!}$$

is analytic on $\mathbb{R}$ with $f(0) = 1$ and $f'(0) = -\frac{1}{2}$. From $f(t^2) = \cos t$ it follows that, in a neighborhood of 1, we have

$$\arccos(s)^2 = f^{-1}(s),$$

and that this function is differentiable with derivative -2 in 1. For $s = \langle x,y\rangle$, we therefore have

$$\lim_{d_S(x,y)\to 0} \frac{d_S(x,y)^2}{d_E(x,y)^2} = \lim_{s\to 1} \frac{\arccos(s)^2}{2(1-s)} = \lim_{s\to 1} \frac{f^{-1}(s)}{2(1-s)} = -\tfrac{1}{2}(f^{-1})'(0) = 1. \tag{20}$$

Hence we find, for any $C > 1$, a $\delta_C > 0$ such that

$$d_S(x,y) \leq \delta_C \quad \Rightarrow \quad d_S(x,y) \leq C d_E(x,y).$$

Let $\varepsilon > 0$ and $\mathbf{t} = (t_0, \ldots, t_N)$ be a subdivision of $[a,b]$ such that

$$\sum_{j=0}^{N-1} d_S(\gamma(t_j), \gamma(t_{j+1})) = L_{S,\mathbf{t}}(\gamma) > L_S(\gamma) - \varepsilon.$$

Since γ is uniformly continuous, we may further assume that $d_S(\gamma(t_j), \gamma(t_{j+1})) \leq \delta_C$ for every j. This leads to

$$L_S(\gamma) \leq \varepsilon + \sum_{j=0}^{N-1} d_S(\gamma(t_j), \gamma(t_{j+1})) \leq \varepsilon + C \sum_{j=0}^{N-1} d_E(\gamma(t_j), \gamma(t_{j+1})) \leq \varepsilon + C L_E(\gamma).$$

Since $\varepsilon > 0$ and $C > 1$ are arbitrary, we obtain $L_S(\gamma) \leq L_E(\gamma)$. □

An important consequence of the preceding lemma is that the metric on the sphere is the "length metric" of $\mathbb{S}_n$, considered as a subset of $\mathbb{R}^{n+1}$. We only have to recall that, for the geodesics on $\mathbb{S}_n$ defined by great circles, the euclidean length coincides with the spherical length.

Proposition 3.17. *For $x, y \in \mathbb{S}_n$, we have*

$$d_S(x,y) = \inf\{L_E(\gamma) \colon \gamma \in C^1([0,1], \mathbb{S}_n), \gamma(0) = x, \gamma(1) = y\}.$$

Remark 3.18. The angle between two unit vectors x and y in $E = \mathbb{R}^{n+1}$ can also be expressed quite directly in terms of the metric:

$$\cos(\measuredangle(x,y)) = 1 - \tfrac{1}{2} d_E(x,y)^2$$

because

$$d_E(x,y)^2 = \|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x,y\rangle = 2(1 - \langle x,y\rangle).$$

If, more generally, $\gamma, \eta\colon [0,1] \to E$ are curves with

$$p := \gamma(0) = \eta(0) \quad \text{and} \quad \gamma'(0) = x, \quad \eta'(0) = y, \quad \|x\| = \|y\| = 1,$$

then

$$\begin{aligned}\frac{1}{t^2}\|\gamma(t)-\eta(t)\|^2 &= \frac{1}{t^2}\|\gamma(t)-p\|^2 + \frac{1}{t^2}\|\eta(t)-p\|^2 - 2\frac{1}{t^2}\langle\gamma(t)-p, \eta(t)-p\rangle \\ &\to \|x\|^2 + \|y\|^2 - \langle x,y\rangle = 2(1-\cos\measuredangle(x,y)),\end{aligned}$$

which leads to

$$\cos(\measuredangle(x,y)) = 1 - \lim_{t\to 0} \frac{d_E(\gamma(t),\eta(t))^2}{2t^2}.$$

If $\gamma([0,1]), \eta([0,1]) \subseteq \mathbb{S}(E)$, then (20) further leads to

$$\cos(\measuredangle(x,y)) = 1 - \lim_{t\to 0} \frac{d_S(\gamma(t),\eta(t))^2}{2t^2}. \tag{21}$$

We shall use this relation in the following proof.

Theorem 3.19. *There is no isometry $f\colon \Delta ABC \to \mathbb{E}$ from a spherical triangle with positive area to a euclidean space.*

Proof. Any such isometry f maps geodesics to geodesics. Therefore it maps the sides of ΔABC into straight line segments (Corollary 2.33). Applying the same argument to geodesic segments connecting C with points on the side from A to B, it follows that $f(\Delta ABC)$ is contained in an affine plane, so that it is the planar triangle with vertices $f(A)$, $f(B)$ and $f(C)$.

From (21) we derive that $\angle A = \angle f(A)$ etc., which leads to

$$\pi = \angle f(A) + \angle f(B) + \angle f(C) = \angle A + \angle B + \angle C = \pi + \text{area}(\Delta ABC).$$

This contradicts our assumption that the area of ΔABC is positive. $\square$

Exercises for Section 3

Exercise 3.1. Let (X,d) be a complete metric space with the property that each pair of points $x_1, x_2 \in X$ has a midpoint. Show that there exists a continuous curve $\gamma\colon [0, d(x_1,x_2)] \to X$ with $\gamma(0) = x_1$, $\gamma(d(x_1,x_2)) = x_2$ and

$$d(\gamma(t), x_1) = t \quad \text{for all} \quad t \in [0, d(x_1,x_2)].$$

Such curves in a metric space are called *geodesics* because they are "shortest paths" from x_1 to x_2. Hint: Define γ first on dyadic subdivision points of $[0, d(x_1,x_2)]$.

4 Projective Geometry

The historical roots of Projective Geometry lie in Florence in the 15th century (Brunelleschi, Alberti). With the invention of central perspective in paintings one had to deal with maps from one plane to another where "points at infinity" had to be taken into account. The mathematical development of this theory started much later in France in the 19th century when the concepts of matrices, vector spaces and linear maps emerged.

Our approach to projective geometry is based on the definition of a projective space as the set $\mathbb{P}(V)$ of one-dimensional subspaces of a vector space V.

4.1 The projective space of a vector space

Definition 4.1. Let V be a vector space over the field $\mathbb{K}$. Then the set $\mathbb{P}(V)$ of all one-dimensional subspaces of V is called *the projective space of* V.

For $0 \neq v \in V$, we write $[v] := \mathbb{K}v$ for the one-dimensional subspace generated by v and $q \colon V \setminus \{0\} \to \mathbb{P}(V), v \mapsto \mathbb{K}v = [v]$ for the canonical projection.

Definition 4.2. (The projective group) (a) Let V and W be $\mathbb{K}$-vector spaces. Then any linear isomorphism $\phi \colon V \to W$ induces a map

$$\overline{\phi} \colon \mathbb{P}(V) \to \mathbb{P}(W), \quad [v] \mapsto \overline{\phi}([v]) = [\phi(v)].$$

Then $\overline{\phi}$ is called an *isomorphism of projective spaces* or a *projectivity* (Germ.: Projektivität).

(b) For $V = W$, the map $\overline{\phi}$ permutes the one-dimensional subspaces of V. Clearly, the elements of the form $g = \lambda\mathbf{1}$, $\lambda \in \mathbb{K}^\times$ preserve all one-dimensional subspaces. If, conversely, $g \in \mathrm{GL}(V)$ preserves all one-dimensional subspaces, then $g = \lambda\mathbf{1}$ for some $\lambda \in \mathbb{K}^\times$ (Exercise 4.1). Therefore we obtain an action of the *projective linear group*

$$\mathrm{PGL}(V) := \mathrm{GL}(V)/\mathbb{K}^\times\mathbf{1}$$

on $\mathbb{P}(V)$. We write $\overline{g} = \mathbb{K}^\times g$ for the image of $g \in \mathrm{GL}(V)$ in $\mathrm{PGL}(V)$, so that the group structure is given by

$$\overline{g}\overline{h} = \overline{gh} \quad \text{for} \quad g, h, \in \mathrm{GL}(V).$$

The elements of $\mathrm{PGL}(V)$, considered as maps $\mathbb{P}(V) \to \mathbb{P}(V)$, are called *projectivities* (Germ.: Projektivitäten).

The preceding arguments show that the action of $\mathrm{PGL}(V)$ on $\mathbb{P}(V)$ is *faithful* (or *effective*) in the sense that no element $\overline{g} \neq [\mathbf{1}]$ acts as the identity on $\mathbb{P}(V)$. In this sense $\mathrm{PGL}(V)$ is the *group of projective automorphisms* of $\mathbb{P}(V)$.[6]

Remark 4.3. If $g \in \mathrm{GL}(V)$, then $\overline{g}[v] = [v]$ in $\mathbb{P}(V)$ is equivalent to $v \in V$ being an eigenvector of ϕ.

Examples 4.4. (a) If $\dim V = 0$, then $\mathbb{P}(V) = \emptyset$, and if $\dim V = 1$, then $\mathbb{P}(V)$ is a single point.

(b) For $V = \mathbb{K}^{n+1}$, $\mathbb{P}_n(\mathbb{K}) := \mathbb{P}(\mathbb{K}^{n+1})$ is called the *n-dimensional projective space over* $\mathbb{K}$. For $n = 1$, it is called the *projective line*, and for $n = 2$ the *projective plane*. The corresponding projective linear group is also denoted $\mathrm{PGL}_n(\mathbb{K}) := \mathrm{GL}_n(\mathbb{K})/\mathbb{K}^\times\mathbf{1}$, where we identify $\mathrm{GL}_n(\mathbb{K})$ and $\mathrm{GL}(\mathbb{K}^n)$ in the usual way.

[6] In projective geometry, there is a certain ambiguity in what the corresponding "projective maps" should be. We refer to [Pa13] for some recent results and a discussion of this issue. Here we adopt the point of view of the Erlangen Program, where we "define" projective geometry by defining what the corresponding group is (Definition 4.2(b)).

Definition 4.5. (Homogeneous coordinates) For $V = \mathbb{K}^{n+1}$, there are natural ways to introduce coordinates on the n-dimensional projective space $\mathbb{P}_n(\mathbb{K}) = \mathbb{P}(\mathbb{K}^{n+1})$. Its points are represented in the form

$$[x_0 : x_1 : \ldots : x_n] := \mathbb{K}(x_0, \ldots, x_n), \quad x_j \in \mathbb{K}, (x_0, \ldots, x_n) \neq \mathbf{0}.$$

The tuple $(x_0, \ldots, x_n)$ is called the *homogeneous coordinates* of the corresponding line (=one-dimensional subspace). This representation is not unique because

$$[\lambda x_0 : \ldots : \lambda x_n] := [x_0 : \ldots : x_n] \quad \text{for} \quad \lambda \neq 0.$$

Remark 4.6. If $g = (g_{ij}) \in \mathrm{GL}_n(\mathbb{K})$ is an invertible matrix and $\mathbf{x}' = g\mathbf{x}$ for $\mathbf{x} = (x_1, \ldots, x_n)$, then

$$x_i' = \sum_{j=1}^{n} g_{ij} x_j$$

leads to

$$\overline{g}[x_1 : \ldots : x_n] = [x_1' : \ldots : x_n'] = \Big[\sum_{j=1}^{n} g_{1j} x_j : \ldots : \sum_{j=1}^{n} g_{nj} x_j\Big].$$

For $x_n = 1$, we obtain in particular

$$\overline{g}[x_1 : \ldots : x_{n-1} : 1] = \Big[\frac{\sum_{j=1}^{n-1} g_{1j} x_j + g_{1n}}{\sum_{j=1}^{n-1} g_{nj} x_j + g_{nn}} : \ldots : \frac{\sum_{j=1}^{n-1} g_{n-1,j} x_j + g_{n-1,n}}{\sum_{j=1}^{n-1} g_{nj} x_j + g_{nn}} : 1\Big],$$

provided $\sum_{j=1}^{n-1} g_{nj} x_j + g_{nn} \neq 0$. This is how projective maps can be expressed in terms of homogeneous coordinates. They correspond to fractional linear functions in each coordinate.

4.2 The projective line

We take a closer look at the projective line $\mathbb{P}_1(\mathbb{K}) := \mathbb{P}(\mathbb{K}^2)$ of all one-dimensional subspaces of $\mathbb{K}^2$. In terms of homogeneous coordinates, its elements are represented by pairs $[x_0 : x_1]$ with $(x_0, x_1) \neq (0, 0)$ and $[\lambda x_0 : \lambda x_1] = [x_0 : x_1]$ for $\lambda \neq 0$.

If $e_0 := (1, 0)$ and $e_1 := (0, 1)$ are the canonical basis vectors of $\mathbb{K}^2$, then

$$[e_0] = [1 : 0] \quad \text{and} \quad [e_1] = [0 : 1].$$

If $x_1 \neq 0$, then $[x_0 : x_1] = \big[\frac{x_0}{x_1} : 1\big]$, and if $x_1 = 0$, then $[x_0 : 0] = [1 : 0]$. Therefore the elements of the projective line can be represented as

$$\mathbb{P}_1(\mathbb{K}) = \{[x : 1] : x \in \mathbb{K}\} \cup \{[1 : 0]\}.$$

For the disjoint union $\mathbb{K}_\infty := \mathbb{K} \dot{\cup} \{\infty\}$, we therefore have a bijective map

$$\eta \colon \mathbb{K}_\infty \to \mathbb{P}_1(\mathbb{K}), \quad \eta(x) := \begin{cases} [x : 1] & \text{for } x \in \mathbb{K} \\ [1 : 0] & \text{for } x = \infty. \end{cases}$$

Therefore one can view $\mathbb{P}_1(\mathbb{K})$ as the field $\mathbb{K}$ with an additional point ∞. Here $[1 : 0]$ is called the *"point at infinity"*.

For a projective linear transformation $\overline{g}$, represented by the matrix

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{K}),$$

we have

$$\overline{g}[x : y] = [ax + by : cx + dy] = \begin{cases} \left[\frac{ax+by}{cx+dy} : 1\right] & \text{if } cx + dy \neq 0 \\ [1 : 0] & \text{otherwise.} \end{cases}$$

In view of $(c, d) \neq (0, 0)$ (the matrix g is invertible), the equation $cx + dy = 0$ determines a unique element $[x_\infty : y_\infty] \in \mathbb{P}_1(\mathbb{K})$ for which

$$\overline{g}[x_\infty : y_\infty] = [1 : 0].$$

This also follows from the fact that $\overline{g} \colon \mathbb{P}_1(\mathbb{K}) \to \mathbb{P}_1(\mathbb{K})$ is bijective.

In the η-coordinates, this corresponds to the *fractional linear transformation*

$$\phi \colon \mathbb{K}_\infty \to \mathbb{K}_\infty, \quad \phi(x) = \begin{cases} \frac{ax+b}{cx+d} & \text{for } cx + d \neq 0 \\ \infty & \text{for } cx + d = 0. \end{cases}$$

For $\mathbb{K} = \mathbb{C}$, these transformations are called *Möbius transformations*.

Example 4.7. To visualize the projective line for $\mathbb{K} = \mathbb{R}$ and $\mathbb{C}$, it is natural to use the stereographic projection.

In $\mathbb{S}_n \subseteq \mathbb{R}^{n+1}$, we call the unit vector $e_0 := (1, 0, \ldots, 0)$ the *north pole*. We then have the *stereographic projection map*

$$\phi \colon \mathbb{S}^n \setminus \{e_0\} \to \mathbb{R}^n, \quad (y_0, \mathbf{y}) \mapsto \frac{1}{1 - y_0}\mathbf{y}.$$

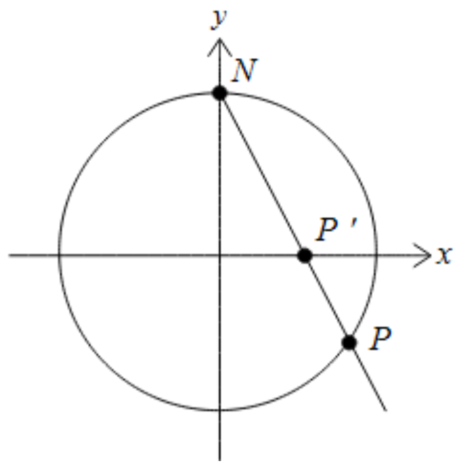
This map is bijective with inverse

$$\phi^{-1}(x) = \Big(\frac{\|x\|_2^2 - 1}{\|x\|_2^2 + 1}, \frac{2x}{\|x\|_2^2 + 1}\Big)$$

(Exercise).

As a point set, we can therefore consider $\mathbb{S}^n$ as the union of $\mathbb{R}^n$ and the north pole. For $n = 1$, we thus obtain a bijection

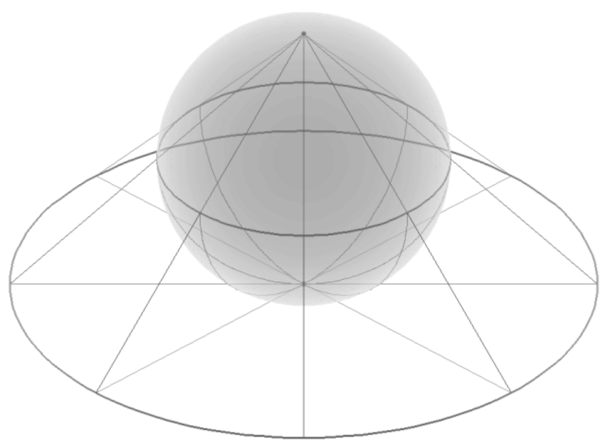
$$\psi \colon \mathbb{R}_\infty \cong \mathbb{P}_1(\mathbb{R}) \to \mathbb{S}_1, \quad \psi(x) = \begin{cases} \left(\frac{x^2-1}{x^2+1}, \frac{2x}{x^2+1}\right) & \text{for } x \in \mathbb{R} \\ e_0 & \text{for } x = \infty. \end{cases}$$



For $n = 1$ and $\mathbb{K} = \mathbb{C}$, we likewise obtain a bijection

$$\psi\colon \mathbb{C}_\infty \cong \mathbb{P}_1(\mathbb{C}) \to \mathbb{S}_2 \subseteq \mathbb{R}^3 = \mathbb{R} \times \mathbb{C}, \quad \psi(z) = \begin{cases} \left(\frac{|z|^2-1}{|z|^2+1}, \frac{2z}{1+|z|^2}\right) & \text{for } z \in \mathbb{C} \\ e_0 & \text{for } \ z = \infty. \end{cases}$$

For this reason, the space $\mathbb{P}_1(\mathbb{C})$ is also called the *Riemann sphere* (Germ.: Riemannsche Zahlenkugel).



4.3 Affine subspaces of projective spaces

In this section we turn to the close connection between affine spaces and projective spaces. We have already seen in Remark 1.6 that affine spaces embed naturally as hyperplanes in vector spaces and we shall see below that this even leads to embeddings into projective spaces. From this point of view, projective spaces appear as "completions" of affine spaces A to which we add certain "points at infinity" which can be represented by elements of $\mathbb{P}(\overrightarrow{A})$. Geometrically elements $[v] \in \mathbb{P}(\overrightarrow{A})$ correspond to families of parallel lines $p + [v]$, $p \in A$.

Definition 4.8. (Affine subspaces) Let V be a vector space and $A = a_0 + H \subseteq V$ an affine hyperplane not containing 0. Then we obtain a natural map

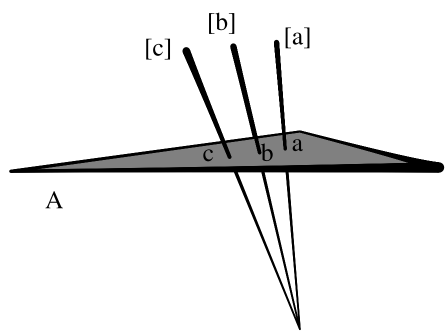
$$\eta_A := q|_A \colon A \to \mathbb{P}(V), \quad a \mapsto [a].$$

This map is injective because $[a] = [a']$ implies that $a' = \lambda a$ for some $\lambda \neq 0$. This is only possible for $\lambda = 1$ because $\mathbb{K}a \cap A = \{a\}$ follows from $V = H \oplus \mathbb{K}a_0$.

Writing $[A] := q(A)$ for the image of A in $\mathbb{P}(V)$, we may therefore identify A with a subset of $\mathbb{P}(V)$. It contains all lines intersecting A. From $A = a_0 + H$ we derive that a line $[v] = \mathbb{K}v$ does not intersect A if and only if $v \in H$. This means that

$$\mathbb{P}(V) = [A] \dot{\cup} \mathbb{P}(H) = [A] \dot{\cup} \mathbb{P}(\overrightarrow{A}).$$

We call $[A]$ an *affine subspace of* $\mathbb{P}(V)$.



Example 4.9. For the affine hyperplanes $A_j := \{x \in \mathbb{K}^{n+1} \colon x_j = 1\}$, $j = 0, \ldots, n$, we obtain the subsets

$$[A_j] = \{[x_0 : \ldots : x_{j-1} : 1 : x_{j+1} : \ldots : x_n] \colon x_i \in \mathbb{K}\} \subseteq \mathbb{P}_n(\mathbb{K}).$$

Remark 4.10. If $H \subseteq V$ is a linear hyperplane, then the complement $\mathbb{P}(V) \setminus \mathbb{P}(H)$ coincides with the affine subspaces $[A]$, where A is of the form $A = a_0 + H$, $a_0 \notin H$.

Proposition 4.11. *For two affine subspaces $[A], [B] \subseteq \mathbb{P}(V)$, the following assertions hold:*

(a) *For every affine map $\phi \colon A \to B$, there exists a unique linear map $\Phi \colon V \to V$ extending ϕ.*

(b) *ϕ is an affine isomorphism if and only if Φ is a linear isomorphism. Then $\overline{\Phi}[A] = [B]$.*

(c) *If $\Psi \in \mathrm{PGL}(V)$ satisfies $\Psi[A] = [B]$, then there exists a unique affine isomorphism $\phi \colon A \to B$ whose unique linear extension $\Phi \in \mathrm{GL}(V)$ satisfies $\Psi = \overline{\Phi}$.*

Proof. (a) Write $A = a_0 + H$ and $B = b_0 + H'$. Then $V = H \oplus \mathbb{K}a_0$ and

$$\Phi \colon V \to V, \quad \Phi(h + \lambda a_0) := \phi_L(h) + \lambda \phi(a_0)$$

is a linear map. It is uniquely determined by $\Phi|_H = \phi_L$ and $\Phi(a_0) = \phi(a_0)$.

(b) That ϕ is an affine isomorphism is equivalent to $\phi_L\colon H \to H'$ being a linear isomorphism, and then $\Phi\colon V \to V$ also is a linear isomorphism. If, conversely, Φ is a linear isomorphism, then its restriction $\phi_L = \Phi|_H\colon H \to H'$ also is a linear isomorphism and this implies that ϕ is an affine isomorphism.

(c) Write $\Psi = \overline{\Phi}$ for $\Phi \in \mathrm{GL}(V)$. Then $[B] = \Psi([A]) = [\Phi(A)]$ implies in particular that

$$\mathbb{P}(\Phi(H)) = \Psi(\mathbb{P}(H)) = \Psi([A]^c) = [\Phi(A)]^c = \mathbb{P}(H'),$$

so that $\Phi(H) = H'$. Therefore the affine subspace $\Phi(A) \subseteq V$ is parallel to B and does not contain 0 (Remark 4.10). Pick $\lambda \in \mathbb{K}^\times$ with $\lambda\Phi(a_0) \in B$. Then $\lambda\Phi(A) = B$ and $\lambda\Phi|_A\colon A \to B$ is an affine isomorphism. Now $\overline{\lambda\Phi} = \overline{\Phi} = \Psi$ completes the proof. $\square$

From the preceding proposition, we obtain in particular that projective isomorphisms preserving an affine subspace act as affine maps on this subspace:

Corollary 4.12. *If $\Psi \in \mathrm{PGL}(V)$ preserves the affine subspace $[A] \subseteq \mathbb{P}(V)$, then there exists a unique affine automorphism $\phi \in \mathrm{Aut}(A)$ with $\Psi = \overline{\Phi}$. This leads to an isomorphism of the group $\mathrm{Aut}(A)$ of affine automorphisms of A with the stabilizer of $[A]$ in $\mathrm{PGL}(V)$:*

$$\mathrm{Aut}(A) \cong \mathrm{PGL}(V)_{[A]} := \{\Psi \in \mathrm{PGL}(V)\colon \Psi([A]) = [A]\}.$$

Example 4.13. For the affine line $A = \{(x,1)\colon x \in \mathbb{K}\} \subseteq \mathbb{K}^2$ we have

$$[A] = \{[x:1]\colon x \in \mathbb{K}\} = \mathbb{P}_1(\mathbb{K}) \setminus \{[1:0]\}.$$

That a projective automorphism $\overline{g}([x:1]) = [ax+b : cx+d]$ preserves the affine subspace $[A]$ is equivalent to $\overline{g}([1:0]) = [1:0]$. In view of $\overline{g}([1:0]) = [a:c]$, this is equivalent to $c = 0$, which implies that $\overline{g}([x:1]) = [\frac{ax+b}{d} : 1]$, i.e., $\overline{g}$ corresponds to the affine map $x \mapsto d^{-1}(ax+b)$ on $\mathbb{K}$.

Proposition 4.14. *The projective group $\mathrm{PGL}_2(\mathbb{K})$ of the projective line $\mathbb{P}_1(\mathbb{K}) \cong \mathbb{K}_\infty$ is generated by the affine maps $ax+b$, $a \neq 0$, and the inversion $x \mapsto x^{-1}$.*

Proof. If $\overline{\phi} \in \mathrm{PGL}_2(\mathbb{K})$ fixes the point $\infty = [1:0]$, we have seen in Example 4.13 that $\overline{\phi}([x:1]) = [ax+b:1]$ for $a \in \mathbb{K}^\times$ and $b \in \mathbb{K}$, i.e., $\overline{\phi}$ is affine.

If this is not the case, then $\overline{\phi}(\infty) \neq \infty$, i.e., $\overline{\phi}(\infty) \in \mathbb{K}$. Then there exists a translation τ with $\tau \circ \overline{\phi}(\infty) = 0$, so that the inversion $\sigma(x) = x^{-1}$ leads to the map $\alpha := \sigma \circ \tau \circ \overline{\phi}$ preserving ∞. We have already seen that this implies that α is affine, and now $\overline{\phi} = \tau^{-1} \circ \sigma^{-1} \circ \alpha = \tau^{-1} \circ \sigma \circ \alpha$ implies that $\mathrm{PGL}_2(\mathbb{K})$ is generated by σ and affine maps. $\square$

Example 4.15. For the affine plane $A = \{(x_1, x_2, 1)\colon x_j \in \mathbb{K}\} \subseteq \mathbb{K}^3$, we have

$$\mathbb{P}_2(\mathbb{K}) \setminus [A] = \{[x_1 : x_2 : 0]\colon (x_1, x_2) \neq (0,0)\} \cong \mathbb{P}_1(\mathbb{K}),$$

so that $\mathbb{P}_2(\mathbb{K})$ arises from $[A]$ by adding a projective line at infinity.

Let $\phi \in \mathrm{GL}(\mathbb{K}^3)$ be given by multiplication with the matrix $\begin{pmatrix} A & b \\ c & d \end{pmatrix}$, where $A \in M_2(\mathbb{K})$, $b, c^\top \in \mathbb{K}^2$, and $d \in \mathbb{K}$. That the projectivity

$$\overline{\phi}([x : x_3]) = [Ax + bx_3 : cx + dx_3]$$

preserves the affine subspace $[A] = [\mathbb{K}^2 : 1]$ is equivalent to $\overline{\phi}([x:0]) = [x':0]$ for every non-zero $x \in \mathbb{K}^2$ and some $x' \in \mathbb{K}^2$. In view of $\overline{\phi}([x:0]) = [Ax : cx]$, this is equivalent to

$c = 0$, which implies that $\overline{\phi}([x : 1]) = [d^{-1}(Ax + b) : 1]$, i.e., $\overline{\phi}$ corresponds to the affine map $x \mapsto d^{-1}(Ax + b)$ on $\mathbb{K}^2$.

A general element $\overline{\phi} \in \mathrm{PGL}_3(\mathbb{K})$ induces on the subset $[\mathbb{K}^2 : 1]$ a map of the form

$$\overline{\phi}([x : 1]) = [(cx + d)^{-1}(Ax + b) : 1] \quad \text{for} \quad cx + d \neq 0.$$

The relation $cx + d = 0$ describes a line in the affine plane $\mathbb{K}^2$ which is mapped by $\overline{\phi}$ "to infinity", here represented by $\mathbb{P}_1(\mathbb{K})$. If $cx + d = 0$, then

$$\overline{\phi}([x : 1]) = [Ax + b : 0].$$

4.4 Projective maps between affine hyperplanes

We have already seen that any affine hyperplane $A \subseteq V$ not containing 0 leads to a subset $[A] \subseteq \mathbb{P}(V)$. If we have two such hyperplanes A_1 and A_2, any projectivity $\Psi \in \mathrm{PGL}(V)$ restricts to a map $[A_1] \cap \Psi^{-1}([A_2]) \to [A_2]$. Since the maps $q|_{A_j} : A_j \to [A_j]$ are bijective, this map can be transferred to a (not everywhere defined) map $A_1 \to A_2$. We thus obtain the concept of a projective map between affine hyperplanes. For the special case $\Psi = \mathrm{id}_{\mathrm{PGL}(V)}$ we obtain in particular central projections which play a crucial role in perspective drawing. Historically, these maps actually led to the development of projective geometry.

We now turn to the details. Let V be a vector space and $A_j = a_j + H_j \subseteq V$, $j = 1, 2$, be two affine hyperplanes not containing 0. We write $\lambda_j \in V^* = \mathrm{Hom}(V, \mathbb{R})$ for the uniquely determined linear functionals on V satisfying $A_j = \lambda_j^{-1}(1)$. With respect to the linear isomorphism $H_j \oplus \mathbb{R} \to V, (h, t) \mapsto h + ta_j$, we have $\lambda_j(h + ta_j) = t$ for $h \in H_j$. Note that $H_j = \ker \lambda_j$.

Definition 4.16. (Projective maps between affine subspaces) Let $\phi \in \mathrm{GL}(V)$. Then

$$\widetilde{\phi} \colon A_1 \setminus \phi^{-1}(H_2) \to A_2, \quad x \mapsto \frac{1}{\lambda_2(\phi(x))}\phi(x)$$

is called a *projective map from A_1 to A_2*.

Remark 4.17. Clearly, the map $\widetilde{\phi}$ satisfies vor $v \in A_1 \setminus \phi^{-1}(H_2)$ the relation

$$[\widetilde{\phi}(v)] = [\phi(v)] = \overline{\phi}[v],$$

so that it is simply a description of the projectivity in terms of "coordinates" provided by the affine subspaces $[A_1]$ and $[A_2] \subseteq \mathbb{P}(V)$.

Remark 4.18. (a) The projective map from A_1 to A_2 defined by $\phi \in \mathrm{GL}(V)$ is not defined on the subset $A_1 \setminus \phi^{-1}(H_2)$ because this is the set of all points $x \in A_1$, where $\lambda_2(\phi(x)) = 0$.

On the level of the projective space, this is due to the fact that, for any $x \in A_1$ with $\lambda_2(\phi(x)) = 0$, we have $\phi(x) \in H_2$, i.e., $\overline{\phi}(x) \in \mathbb{P}(H_2) = \mathbb{P}(V) \setminus [A_2]$.

(b) That the exceptional subset $\phi^{-1}(H_2) \cap A_1$ is empty is equivalent to $\phi(A_1) \cap H_2 = \emptyset$. This means that the affine function $\lambda_2 \colon \phi(A_1) \to \mathbb{K}$ has no zeros, and this happens only if it is constant. This in turn means that $\phi(A_1) = x + H_2$ for some $x \in V \setminus H_2$. We conclude that the exceptional set is empty if and only if the two hyperplanes A_2 and $\phi(A_1)$ are parallel. Clearly, this is equivalent to the two hyperplanes A_1 and $\phi^{-1}(A_2)$ being parallel.

(c) If the exceptional set is non-empty, then it is the zero set of an affine function $\lambda_2 \circ \phi \colon A_1 \to \mathbb{K}$, hence an affine hyperplane of A_1 and therefore an affine subspace of codimension 2 of V.

For $\dim V = 2$ this leads to exceptional points and, for $\dim V = 3$, to exceptional lines.

Lemma 4.19. *Suppose that V is finite-dimensional and $|\mathbb{K}| \geq 4$. Then the exceptional set $A_1 \cap \phi^{-1}(H_2)$ is empty if and only if $\widetilde{\phi}$ extends to an affine map $A_1 \to A_2$.*

Proof. We have already seen in Remark 4.18(b) that $A_1 \cap \phi^{-1}(H_2) = \emptyset$ is equivalent to $\phi(A_1)$ and A_2 being parallel, which means that $\phi(H_1) = H_2$. In view of $\mathbb{P}(H_j) = \mathbb{P}(V) \setminus [A_j]$, this is equivalent to

$$\overline{\phi}([A_1]) = [A_2].$$

(a) If $A_1 \cap \phi^{-1}(H_2) = \emptyset$, then the denominator $\lambda_2 \circ \phi \colon A_1 \to \mathbb{K}$ is constant non-zero, so that $\widetilde{\phi}$ is a multiple of $\phi|_{A_1}$ and therefore affine.

(b) If $A_1 \cap \phi^{-1}(H_2) \neq \emptyset$, then we pick $x_0 \in A_1 \cap \phi^{-1}(H_2)$ and $0 \neq v \in H_1 \setminus \phi^{-1}(H_2)$ (the existence follows from $\operatorname{codim}_{A_1}(\phi^{-1}(H_1) \cap A_1) \geq 1$). Since V is assumed to be finite-dimensional, we also find a linear functional $\mu \in V^*$ with $\mu(\phi(x_0)) \neq 0$. Then, for $t \neq 0$,

$$\mu(\widetilde{\phi}(x_0 + tv)) = \frac{\mu(\phi(x_0 + tv))}{\lambda_2(\phi(x_0 + tv))} = \frac{\mu(\phi(x_0)) + \mu(\phi(v))t}{t\lambda_2(\phi(v))}.$$

This function is of the form $a + bt^{-1}$ with $b \neq 0$. If it extends to an affine function $a't + b'$, then $a' \neq 0$ and $at + b = a't^2 + b't$ for all $t \neq 0$. Since $a' \neq 0$, this equation has at most two solutions, so that $|\mathbb{K}| \geq 4$ implies $a' = 0$. This leads to the contradiction that the function $a + bt^{-1}$ is constant. We conclude that $\widetilde{\phi}$ cannot be extended to an affine function if the exceptional set is non-zero. $\square$

Example 4.20. We return to the situation of Remark 4.6, where $g = (g_{ij}) \in \mathrm{GL}_n(\mathbb{K})$ is an invertible matrix. We consider the affine hyperplane

$$A_1 = A_2 = A := \{x \in \mathbb{K}^n \colon x_n = 1\}.$$

Then $\lambda(x) = x_n$ and $\phi(x) = gx$, so that

$$\widetilde{\phi}(x_1, \ldots, x_{n-1}, 1) = \frac{1}{\sum_{j=1}^{n-1} g_{nj}x_j + g_{nn}} \Big(\sum_{j=1}^{n-1} g_{1j}x_j + g_{1n}, \ldots, \sum_{j=1}^{n-1} g_{n,j}x_j + g_{n,n} \Big).$$

This map is affine if and only if the denominator is constant, which is equivalent to $g_{nj} = 0$ for $j = 1, \ldots, n-1$. This is turn means that $(gx)_n = 0$ for $x_n = 0$, i.e., that g preserves the subspace $\mathbb{K}^{n-1} \times \{0\} \subseteq \mathbb{K}^n$.

4.5 Central projections

In this subsection we turn to the special case of projective maps between hyperplanes that are induced by the identity on the projective space. Geometrically, these maps correspond to central projections of one affine hyperplane A_1 to another hyperplane A_2, where the center of projection C is neither contained in A_1 nor in A_2. These maps are important to understand the geometry of perspective drawings or projecting images (A_1 is the object plane and A_2 is the screen).

Example 4.21. (Central projections) (a) We consider the special case where $\phi = \mathrm{id}$. Then

$$\widetilde{\phi} \colon A_1 \setminus H_2 \to A_2, \quad \widetilde{\phi}(x) = \frac{1}{\lambda_2(x)} x.$$

This map has the following geometric interpretation. Let $C = 0 \in V$ be the origin. For a point $x \in A_1$, the condition that the line $\mathbb{K}x$ intersects A_2 is equivalent to $x \notin H_2$ (otherwise this line is parallel to A_2). If $\mathbb{K}x \cap A_2 \neq \emptyset$, then this intersection is a single point $x' = \mu x$ with $1 = \lambda_2(x') = \mu\lambda_2(x)$, i.e.,

$$x' = \frac{1}{\lambda_2(x)} x = \widetilde{\phi}(x).$$

Therefore the map $\widetilde{\phi}$ can be interpreted as a central projection of $A_1 \setminus H_2$ to the affine hyperplane with respect to the origin 0 as center of projection.

(b) If, more generally, $\mathbb{A}$ is an affine space with two affine hyperplanes $A_1, A_2 \subseteq \mathbb{A}$ and $C \in \mathbb{A}$ is a point not contained in $A_1 \cup A_2$, then we consider the vector space structure on $\mathbb{A}$ obtained by defining $0 := C$ to be the origin. Then the central projection of A_1 to A_2 can be described as under (a).

Example 4.22. We consider the real affine space $\mathbb{A}_3 = \mathbb{R}^3$. It contains the affine hyperplane

$$A_1 := \{x \in \mathbb{R}^3 \colon x_3 = 0\}.$$

We think of this hyperplane as an hyperplane that a painter whose eye is located in the point $C := (0, 0, 1)$ wants to draw on a (transparent) canvas which is represented by the affine plane

$$A_2 := \{x \in \mathbb{R}^3 \colon x_1 = 1\}.$$

This means that a point $x \in A_1$ corresponds to the point $x' \in A_2$ which is specified by the condition that

$$(C + \mathbb{R}(x - C)) \cap A_2 = \{x'\}.$$

Choosing C as the origin of a new coordinate system, we write $y = (y_1, y_2, y_3)$ for the new coordinates and find

$$y_1 = x_1, \quad y_2 = x_2 \quad \text{and} \quad y_3 := x_3 - 1.$$

Accordingly

$$A_1 = \{y \in \mathbb{A}_3 \colon y_3 = -1\} \quad \text{and} \quad A_2 = \{y \in \mathbb{A}_3 \colon y_1 = 1\}.$$

Therefore the projection from A_1 to A_2 is given by

$$\widetilde{\phi} \colon A_1 \setminus H_2 \to A_2, \quad \widetilde{\phi}(y) = \frac{1}{y_1}(y_1, y_2, -1) = \Big(1, \frac{y_2}{y_1}, -\frac{1}{y_1}\Big).$$

Transforming back to the old coordinates, we obtain

$$\widetilde{\phi} \colon A_1 \setminus H_2 \to A_2, \quad \widetilde{\phi}(x) = \frac{1}{x_1}(x_1, x_2, 0) + (0, 0, 1) = \Big(1, \frac{x_2}{x_1}, 1 - \frac{1}{x_1}\Big).$$

The exceptional set in A_1 corresponds to the line $x_1 = 0$. We also observe that

$$\widetilde{\phi}(A_1 \setminus H_2) = \{x \in A_2 \colon x_3 \neq 1\}.$$

Thinking of $\widetilde{\phi}(A_1 \setminus H_2)$ as a picture of the plane A_1, then the line $\{x \in A_2 \colon x_3 = 1\}$ is the "horizon", resp., the "line at infinity" of the plane A_1 on the "screen" A_2.

4.6 Projective subspaces

We now turn to the discussion of projective geometric properties, resp., the transitivity properties of the action of $\mathrm{PGL}(V)$ on $\mathbb{P}(V)$. We start with the action on the set of projective subspaces.

Definition 4.23. A *projective subspace* of the projective space $\mathbb{P}(V)$ is a subset of the form $\mathbb{P}(W)$, where $W \subseteq V$ is a linear subspace.

In an affine plane, lines intersect if and only if they are not parallel. This becomes simpler in projective planes:

Proposition 4.24. (a) *In a projective plane, any two different projective lines $L_j = \mathbb{P}(W_j) \subseteq \mathbb{P}(V)$, $j = 1, 2$, intersect in a unique point $[W_1 \cap W_2]$.*

(b) *In a 3-dimensional projective space, any two different projective planes $E_j = \mathbb{P}(W_j) \subseteq \mathbb{P}(V)$, $j = 1, 2$, intersect in a unique projective line $[W_1 \cap W_2]$.*

Proof. (a) By assumption $\dim V = 3$, and W_1, W_2 are two different two-dimensional subspaces. This implies that $W_1 + W_2 = V$. Therefore the Dimension Formula from Linear Algebra leads to

$$\dim(W_1 \cap W_2) = \dim W_1 + \dim W_2 - \dim(W_1 + W_2) = 2 + 2 - 3 = 1.$$

This means that $[W_1] \cap [W_2] = [W_1 \cap W_2]$ is a single point.

(b) By assumption $\dim V = 4$, and W_1, W_2 are two different three-dimensional subspaces. This implies that $W_1 + W_2 = V$. Therefore the Dimension Formula from Linear Algebra leads to

$$\dim(W_1 \cap W_2) = \dim W_1 + \dim W_2 - \dim(W_1 + W_2) = 3 + 3 - 4 = 2.$$

This means that $[W_1] \cap [W_2] = [W_1 \cap W_2]$ is a projective line. $\square$

Proposition 4.25. (Extension of projectivities) *Let V be a finite-dimensional vector space and $\mathbb{P}(V)$ be its projective space. If $W_1, W_2 \subseteq V$ are projective subspaces, then every projectivity $\overline{\psi}\colon \mathbb{P}(W_1) \to \mathbb{P}(W_2)$ extends to an element $\overline{\phi} \in \mathrm{PGL}(V)$.*

Proof. Let $\psi\colon W_1 \to W_2$ be a linear isomorphism. Let $(a_1, \ldots, a_k)$ be a linear basis of W_1 and enlarge it to a linear basis $(a_1, \ldots, a_n)$ of V. We likewise enlarge the linear basis $(a_1', \ldots, a_k') := (\psi(a_1), \ldots, \psi(a_k))$ of W_2 to a linear basis $(a_1', \ldots, a_n')$ of V. Then there exists a unique linear isomorphism $\phi \in \mathrm{GL}(V)$ with $\phi(a_j) = a_j'$ for $j = 1, \ldots, n$. This implies that ϕ extends ψ, and therefore $\overline{\phi}$ extends $\overline{\psi}$. $\square$

Corollary 4.26. *The group $\mathrm{PGL}(V)$ acts transitively on the set of all k-dimensional projective subspaces of $\mathbb{P}(V)$.*

4.7 Projective properties of tuples

In this subsection we finally turn to projective geometric properties of tuples, i.e., properties which are invariant under the action of the projective group $\mathrm{PGL}(V)$.

Definition 4.27. We call $([v_1], \ldots, [v_n])$ in $\mathbb{P}(V)$ *projectively independent* if the corresponding vectors $v_1, \ldots, v_n$ are linearly independent. If, in addition, they form a basis of V, we say that $([v_1], \ldots, [v_n])$ is a *projective basis.*

From Linear Algebra we recall:

Lemma 4.28. *Let V be an n-dimensional linear space. Then, for every $k \leq n$, the group $\mathrm{GL}(V)$ acts transitively on the subset V^k_{ind} of linearly independent k-tuples in V^k.*

For $k = n$, we thus obtain a sharply transitive action of the group $\mathrm{GL}(V)$ on the set V^n_{ind} of all linear bases, i.e., for two linear bases $(v_1, \ldots, v_n)$ and $(v_1', \ldots, v_n')$, there exists a unique $\phi \in \mathrm{GL}(V)$ with $\phi(v_j) = v_j'$ for $j = 1, \ldots, n$.

Proof. (Sketch) For $k = n$, the assertion follows from the fact that, for two linear bases $(v_1, \ldots, v_n)$ and $(v_1', \ldots, v_n')$, there exists a unique linear map $\phi \colon V \to V$ with $\phi(v_j) = v_j'$ for $j = 1, \ldots, n$. Then ϕ is injective and surjective, hence an element of $\mathrm{GL}(V)$.

The case $k < n$ follows from the preceding argument because every linearly independent k-tuple can be extended to a linear basis. $\square$

As an immediate consequence, we obtain:

Proposition 4.29. *Let V be an $n+1$-dimensional linear space and $\mathbb{P} := \mathbb{P}(V)$ be its projective space. Then, for every $k \leq n+1$, the group $\mathrm{PGL}(V)$ acts transitively on the subset $\mathbb{P}^k_{\mathrm{ind}}$ of projectively independent k-tuples in $\mathbb{P}^k$.*

Since a pair $(a, b) \in \mathbb{P}^2$ is projectively independent if and only if $a \neq b$, we have

$$\mathbb{P}^2_\times = \{(p, q) \in \mathbb{P} \colon p \neq q\} = \mathbb{P}^2_{\mathrm{ind}}.$$

For a projective line, this is the set of projective bases in $\mathbb{P}$. The preceding proposition implies:

Corollary 4.30. *Let V be a finite-dimensional vector space. Then the action of $\mathrm{PGL}(V)$ on $\mathbb{P}(V)$ is 2-transitive.*

Remark 4.31. The preceding corollary means that pairs of distinct points have no projective geometric property.

Definition 4.32. We call a subset E of the projective space $\mathbb{P}$ *collinear* if $|E| \leq 1$ or E is contained in a projective line.

Proposition 4.29 implies in particular that projectively independent tuples carry no projective geometric information. They are all conjugate under the projective group $\mathrm{PGL}(V)$. For pairs (a, b), being different is the same as being projectively independent, so that the first non-trivial situation arises for triples (a_0, a_1, a_2) that are collinear. Since we know already that $\mathrm{PGL}(V)$-acts 2-transitively on $\mathbb{P}$ (Corollary 4.30), it suffices to consider triples $(a_0, a_1, a_2) \in \mathbb{P}(V)^3_\times$ of mutually distinct points.

Theorem 4.33. (Fundamental Theorem of analytic projective geometry) *Let V be a finite-dimensional vector space. Then the action of $\mathrm{PGL}(V)$ on the subset of mutually distinct collinear triples (a_1, a_2, a_3) is transitive.*

Proof. First we show that the stabilizer group $\mathrm{PGL}(V)_{a_1, a_2}$ of a_1 and a_2 acts transitively on the set of collinear triples (a_1, a_2, a) with $a \neq a_1, a_2$. Write $a_j = [v_j]$ and observe that v_1 and v_2 are linearly independent with $v_3 \in W := \mathbb{K}v_1 + \mathbb{K}v_2$. Further $a_3 \neq a_1, a_2$ implies that $v_3 = \lambda v_1 + \mu v_2$ with $0 \neq \lambda, \mu$. Let $a_4 = [\lambda' v_1 + \mu' v_2] \in \mathbb{P}(W)$ be another element different from a_1 and a_2.

Then there exists a linear isomorphism $\psi \in \mathrm{GL}(W)$ with $\psi(v_1) = \lambda'\lambda^{-1}v_1$ and $\psi(v_2) = \mu'\mu^{-1}v_2$. Then $\psi(v_3) = \lambda'v_1 + \mu'v_2$, so that

$$\overline{\psi}(a_1) = a_1, \quad \overline{\psi}(a_2) = a_2 \quad \text{and} \quad \overline{\psi}(a_3) = a_4.$$

With Proposition 4.25 we now find $\overline{\phi} \in \mathrm{PGL}(V)$ fixing a_1 and a_2 and satisfying $\overline{\phi}(a_3) = a_4$. Therefore the stabilizer group $\mathrm{PGL}(V)_{a_1,a_2}$ acts transitively on the set of all collinear triples (a_1, a_2, a_3) with $a_3 \neq a_1, a_2$.

If (a_1', a_2', a_3') is a collinear mutually distinct triple, then Corollary 4.30 implies the existence of $\overline{\phi'} \in \mathrm{PGL}(V)$ with $\overline{\phi'}(a_1') = a_1$ and $\overline{\phi'}(a_2') = a_2$. Then $a_4 := \overline{\phi'}(a_3')$ is collinear with a_1 and a_2, so that the preceding argument implies the existence of $\overline{\phi} \in \mathrm{PGL}(V)_{a_1,a_2}$ with $\overline{\phi}(a_4) = a_3$. Then $\overline{\phi} \circ \overline{\phi'}$ maps a_j' to a_j for $j = 1, 2, 3$. $\square$

Corollary 4.34. *For a projective line $\mathbb{P}(V)$, the action of $\mathrm{PGL}(V)$ on $\mathbb{P}(V)^3_\times$ is sharply transitive. In particular, the group $\mathrm{PGL}_2(\mathbb{K})$ acts sharply 3-transitively on the projective line $\mathbb{P}_1(\mathbb{K})$.*

Proof. From Theorem 4.33 we know that $\mathrm{PGL}(V)$ acts transitively on the set $\mathbb{P}(V)^3_\times$ of mutually different triples because all triples are collinear. To see that this action is sharply transitive, let $a_j = [v_j]$, $j = 1, 2, 3$, be mutually distinct elements on $\mathbb{P}(V)$. If $\overline{\phi} \in \mathrm{PGL}(V)$ fixes these three lines, then v_1, v_2 and v_3 are eigenvectors of ϕ, i.e., $\phi(v_j) = \lambda_j v_j$ for $\lambda_j \in \mathbb{K}^\times$. Writing $v_3 = \alpha v_1 + \beta v_2$, we obtain

$$\lambda_3\alpha v_1 + \lambda_3\beta v_2 = \lambda_3 v_3 = \phi(v_3) = \alpha\lambda_1 v_1 + \beta\lambda_2 v_2,$$

which implies that $\lambda_1 = \lambda_3 = \lambda_2$. Therefore $\phi = \lambda_1\mathbf{1}$, and thus $\overline{\phi} = \mathrm{id}_{\mathbb{P}(V)}$. $\square$

Remark 4.35. With the insight of the preceding corollary, we return to the fractional linear transformations

$$\phi(x) = \frac{ax + b}{cx + d}, \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{K}) \quad \text{on} \quad \mathbb{K}_\infty \cong \mathbb{P}_1(\mathbb{K}).$$

We want to understand the fixed points of such maps. We have already observed in Remark 4.3 that, for $x = [v]$, $v \in \mathbb{K}^2 \setminus \{0\}$, the relation $\phi(x) = x$ is equivalent to v being an eigenvector of the matrix g.

- If ϕ has 3 fixed points, then Corollary 4.34 implies that $\phi = \mathrm{id}_{\mathbb{K}_\infty}$. One can also argue directly that the fact that g has three linearly independent eigenvectors implies that it is a multiple of the identity.

- That ϕ has exactly 2 fixed points is equivalent to g having two linearly independent eigenvectors, i.e., that g is diagonalizable. Assuming w.l.o.g. that e_1 and e_2 are eigenvectors, we obtain $g = \mathrm{diag}(\lambda_1, \lambda_2)$ and $\phi(x) = \frac{\lambda_1}{\lambda_2}x$ is linear. In terms of fractional linear transformations, the 2-transitivity of $\mathrm{PGL}_2(\mathbb{K})$ on $\mathbb{P}_1(\mathbb{K})$ thus implies that ϕ has two fixed points if and only if it is conjugate to a linear map.

- If ϕ has exactly 1 fixed point, then g has exactly one eigenvalue. Changing the basis, we may assume that $ge_1 = \lambda e_1$. Then $c = 0$ and the fact that there is only one eigenvalue means that $g = \begin{pmatrix} \lambda & b \\ 0 & \lambda \end{pmatrix}$. Since $\lambda^{-1}g$ induces the same fractional linear map, we may w.l.o.g. assume that $\lambda = 1$, i.e., $g = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$ for some $b \in \mathbb{K}$. Then $\phi(x) = x + b$ is a translation and ∞ is the only fixed point.

- The case where ϕ has no fixed point is the most complicated one and depends very much on the field $\mathbb{K}$. If $\mathbb{K}$ is algebraically closed, such as $\mathbb{K} = \mathbb{C}$, every matrix has an eigenvalue, so that ϕ always has at least one fixed point.

  However, for $\mathbb{K} = \mathbb{R}$, the matrix $g = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ with $g^2 = -\mathbf{1}$ leads to the map $\phi(x) = -x^{-1}$ which has no fixed point in $\mathbb{R}_\infty$.

  In general, the fixed point relation $\phi(x) = x$ is equivalent to the quadratic equation

$$ax + b = cx^2 + dx \quad \text{or} \quad cx^2 + (d-a)x - b = 0$$

  which need not have a solution in $\mathbb{K}$.

4.8 The cross ratio

In this subsection we introduce the cross ratio as a projective invariant of collinear quadruples in $\mathbb{P}(V)$. The cross ratio has manifold applications, some of which we shall explore below. In particular, it permits us to study quadruples of points in the complex projective line whose cross ratio is real, and this leads to the beautiful projective similarity between circles and lines in the Riemann sphere. Later, we shall see that the cross ratio can also be used to obtain a formula for the metric in hyperbolic space in a suitable projective model.

Definition 4.36. (Cross ratio of 4 collinear points) The next step of complexity is represented by collinear quadruples (a_1, a_2, a_3, a_4) for which a_1, a_2, a_3 are mutually distinct. We write $\mathbb{P}(V)^4_r$ for the set of these quadruples. Then there exist $v_j \in V$ with $a_j = [v_j]$, $j = 1, \ldots, 4$, and $v_3 = v_1 + v_2$. In addition, we may choose v_4 in such a way that $v_4 = tv_1 + v_2$ for some $t \in \mathbb{K}$ if $a_4 \neq a_1$, and $v_4 = v_1$ otherwise.

We show that the number t is independent of the choice of the vectors v_j. If $w_j \in V$ are other elements such that $a_j = [w_j]$, $j = 1, \ldots, 4$, $w_3 = w_1 + w_2$ and $w_4 = sw_1 + w_2$, then there exist $\alpha_j \in \mathbb{K}^\times$ with $w_j = \alpha_j v_j$. From

$$\alpha_3(v_1 + v_2) = \alpha_3 v_3 = w_3 = w_1 + w_2 = \alpha_1 v_1 + \alpha_2 v_2$$

we obtain $\alpha_1 = \alpha_3 = \alpha_2$, so that

$$\alpha_1(sv_1 + v_2) = sw_1 + w_2 = w_4 = \alpha_4(tv_1 + v_2)$$

further leads to $\alpha_4 = \alpha_1$, and hence to $t = s$. Therefore

$$c(a_0, a_1, a_2, a_3) := \begin{cases} t \in \mathbb{K} & \text{for } a_1 \neq a_4 \\ \infty \in \mathbb{K}_\infty & \text{for } a_1 = a_4 \end{cases}$$

does not depend on the choice of the representing vectors v_j. It is called the *cross ratio of* (a_1, a_2, a_3, a_4) (Germ.: Doppelverhältnis).

Remark 4.37. Let $a_1, a_2, a_3 \in \mathbb{P}(V)$ be mutually distinct and collinear and $L = \mathbb{P}(W)$ be the projective line generated by a_1, a_2 and a_3 (it is actually generated by any two of them). Then the definition of the cross ratio implies that the map

$$F \colon L = \mathbb{P}(W) \to \mathbb{K}_\infty \cong \mathbb{P}_1(\mathbb{K}), \quad F(a) = c(a_1, a_2, a_3, a)$$

is a projectivity whose restriction to the affine subspace $[\mathbb{K}v_1 + v_2]$ is affine:

$$F([av_1 + bv_2]) = F([\frac{a}{b}v_1 + v_2]) = \frac{a}{b} \cong [a : b]$$

and which satisfies

$$F(a_2) = 0, \quad F(a_3) = 1 \quad \text{and} \quad F(a_1) = \infty.$$

In view of Corollary 4.34, F is uniquely determined by this property.

Proposition 4.38. (The cross ratio as a projective geometric property) *Two collinear quadruples (a_1, a_2, a_3, a_4), $(a_1', a_2', a_3', a_4') \in \mathbb{P}(V)^4_r$ are contained in the same $\mathrm{PGL}(V)$-orbit if and only if they have the same cross ratio.*

Proof. It follows immediately from the definition that

$$c(\overline{\phi}(a_1), \overline{\phi}(a_2), \overline{\phi}(a_3), \overline{\phi}(a_4)) = c(a_1, a_2, a_3, a_4) \quad \text{for} \quad \phi \in \mathrm{GL}(V), (a_1, a_2, a_3, a_4) \in \mathbb{P}(V)^4_r.$$

In fact, write $a_j = [w_j]$ with $w_3 = w_1 + w_2$ and $w_4 = tw_1 + w_2$. For $w_j' := \phi(w_j)$ we then have $\overline{\phi}(a_j) = [\phi(w_j)]$ with

$$w_3' = w_1' + w_2' \quad \text{and } w_4' = tw_1' + w_2'.$$

Suppose, conversely, that (a_1, a_2, a_3, a_4), $(a_1', a_2', a_3', a_4') \in \mathbb{P}(V)^4_r$ have the same cross ratio $t \in \mathbb{K}$. The case $t = \infty$ is trivial. We first find a $\phi \in \mathrm{GL}(V)$ with $\overline{\phi}(a_j) = a_j'$ for $j = 1, 2, 3$ (Theorem 4.33). Choose $v_j \in V$ with $a_j = [v_j]$ and $v_3 = v_1 + v_2$. Then the vectors $v_j' := \phi(v_j)$ satisfy $v_3' := v_1' + v_2'$. From $v_4 = tv_1 + v_2$ we now derive that

$$a_4' = [tv_1' + v_2'] = [\phi(v_4)] = \overline{\phi}(a_4).$$

□

Corollary 4.39. *Let L_1 and L_2 be two projective lines, $a_0, a_1, a_2, a_3 \in L_1$ with the first three mutually distinct and $a_0', a_1', a_2', a_3' \in L_2$ with the first three mutually distinct. Then there exists a projectivity $\psi\colon L_1 \to L_2$ with $\psi(a_j) = a_j'$ for $j = 1, 2, 3, 4$ if and only if*

$$c(a_0, a_1, a_2, a_3) = c(a_0', a_1', a_2', a_3').$$

Proof. The necessity of the equality of the cross ratios follows as in the preceding proof immediately from the definition. To see that it is sufficient, we fix a projectivity $\phi\colon L_1 \to L_2$. With Proposition 4.38 we obtain a projectivity $\phi'\colon L_2 \to L_2$ with $\phi'(\phi(a_j)) = a_j'$ for $j = 1, 2, 3, 4$. Now $\phi' \circ \phi\colon L_1 \to L_2$ is the required projectivity. □

Example 4.40. We now calculate the cross ratio of four points $[z_j : 1]$, $j = 1, 2, 3, 4$, in the projective line $\mathbb{P}_1(\mathbb{K})$. We assume that z_1, z_2 and z_3 are pairwise different. Any representative v_j of $[z_j : 1]$ is of the form $v_j = (w_j z_j, w_j)$ with $w_j \neq 0$. Then $v_3 = v_1 + v_2$ corresponds to $w_3 = w_1 + w_2$ and $w_3 z_3 = w_1 z_1 + w_2 z_2$, i.e.,

$$z_3 = \frac{w_1 z_1 + w_2 z_2}{w_1 + w_2}.$$

Note that $w_1 + w_2 = w_3 \neq 0$. Accordingly $v_4 = tv_1 + v_2$ leads to $w_4 = tw_1 + w_2$ and $w_4 z_4 = tw_1 z_1 + w_2 z_2$, i.e.,

$$z_4 = \frac{tw_1 z_1 + w_2 z_2}{tw_1 + w_2}.$$

This leads to

$$\frac{z_4 - z_2}{z_4 - z_1} = \frac{\frac{tw_1 z_1 - tw_1 z_2}{tw_1 + w_2}}{\frac{-w_2 z_1 + w_2 z_2}{tw_1 + w_2}} = -\frac{tw_1}{w_2},$$

and for $t = 1$ we obtain in particular

$$\frac{z_3 - z_2}{z_3 - z_1} = -\frac{w_1}{w_2}.$$

This shows that

$$c([z_1 : 1], [z_2 : 1], [z_3 : 1], [z_4 : 1]) = t = \frac{z_4 - z_2}{z_4 - z_1} \frac{z_3 - z_1}{z_3 - z_2} = \frac{(z_4 - z_2)(z_3 - z_1)}{(z_4 - z_1)(z_3 - z_2)}. \tag{22}$$

For $z_4 = z_1$ this expression must be interpreted as ∞.

One may also define the function

$$C(z_1, z_2, z_3, z_4) := \frac{(z_4 - z_2)(z_3 - z_1)}{(z_4 - z_1)(z_3 - z_2)} = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)} \in \mathbb{K}_\infty$$

directly for $z_2 \neq z_3$. The invariance of the cross ratio on $\mathbb{P}_1(\mathbb{K})^4_r$ now implies the invariance of the function C under the action of the matrix group $\mathrm{GL}_2(\mathbb{K})$ by fractional linear transformations $z' = \frac{az+b}{cz+d}$ for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R})$, i.e.,

$$C(z_1', z_2', z_3', z_4') = C(z_1, z_2, z_3, z_4).$$

Note that the function

$$D(z_1, z_2, z_3) := \frac{z_3 - z_1}{z_3 - z_2}$$

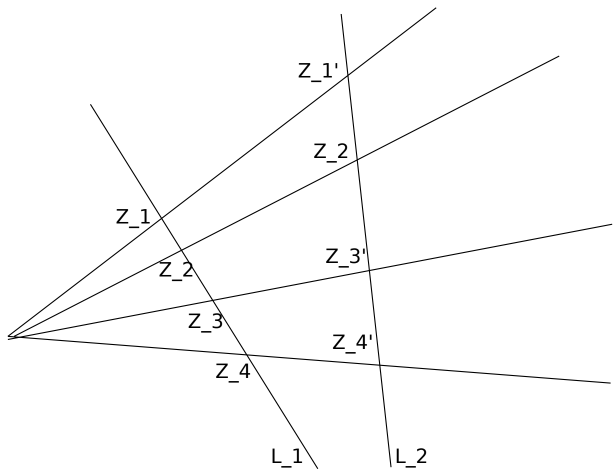
(the ratio of the collinear triple (z_1, z_2, z_3)) is invariant under all affine transformations $z' = az + b$, $a \neq 0$. The cross ratio has the form

$$C(z_1, z_2, z_3, z_4) = \frac{D(z_1, z_2, z_3)}{D(z_1, z_2, z_4)}.$$

Remark 4.41. (Geometric interpretation of the cross ratio) Let $\mathbb{E}$ be the euclidean affine plane, $C \in \mathbb{E}$ and L_j, $j = 1, 2$, two affine lines not containing C. We consider the central projection

$$\pi \colon L_1 \setminus (C + \overrightarrow{L_2}) \to L_2.$$

Let $Z_1, Z_2, Z_3, Z_4 \in L_1 \setminus (C + \overrightarrow{L_2})$ be 4 different points and Z_j' be their images under π.



Using $0 := C$ as the origin, we obtain a euclidean vector space E and $\mathbb{P}(E)$ is a projective line. We consider the expression

$$\widetilde{C}(Z_1, Z_2, Z_3, Z_4) := \frac{d(Z_1, Z_3)d(Z_2, Z_4)}{d(Z_2, Z_3)d(Z_1, Z_4)}.$$

Introducing coordinates on L_1, by an affine isomorphism $\alpha\colon \mathbb{R} \to L_1$, we obtain a projectivity

$$\overline{\alpha}\colon \mathbb{P}_1(\mathbb{R}) \to \mathbb{P}(L_1), \quad [x:1] \mapsto [\alpha(x)] \quad \text{for} \quad x \in \mathbb{R}.$$

Therefore (22) implies

$$\widetilde{C}(Z_1, Z_2, Z_3, Z_4) = |c([Z_1], [Z_2], [Z_3], [Z_4])|$$

is the absolute value of the cross ratio. The same argument shows that

$$\widetilde{C}(Z_1', Z_2', Z_3', Z_4') = |c([Z_1'], [Z_2'], [Z_3'], [Z_4'])|.$$

Now $[Z_j'] = [Z_j]$, $j = 1, 2, 3, 4$, leads to the relation

$$\widetilde{C}(Z_1, Z_2, Z_3, Z_4) = \widetilde{C}(Z_1', Z_2', Z_3', Z_4'),$$

which is a metric consequence of the invariance of the cross ratio under projective isomorphisms.

The following proposition is a converse to the invariance of the cross ratio under fractional linear transformations.

Proposition 4.42. *Every bijection $\phi\colon \mathbb{K}_\infty \to \mathbb{K}_\infty$ preserving the cross ratio*

$$C(z_1, z_2, z_3, z_4) := \frac{(z_4 - z_2)(z_3 - z_1)}{(z_4 - z_1)(z_3 - z_2)} \in \mathbb{K}_\infty$$

is a fractional linear transformation.

Proof. First we observe that

$$C(x,1,0,\infty) = \frac{-x}{-1} = x.$$

This leads to

$$\phi(x) = C(\phi(x),1,0,\infty) = C(x,\phi^{-1}(1),\phi^{-1}(0),\phi^{-1}(\infty)).$$

For $\phi(a) = 1$, $\phi(b) = 0$ and $\phi(c) = \infty$, this leads to

$$\phi(x) = \frac{(c-a)(b-x)}{(c-x)(b-a)}.$$

□

Remark 4.43. If $z_2, z_3, z_4 \in \mathbb{K}_\infty \cong \mathbb{P}_1(\mathbb{K})$ are three different points, then there exists a unique $\phi \in \mathrm{PGL}_2(\mathbb{K})$ with $\phi(z_2) = 1$, $\phi(z_3) = 0$ and $\phi(z_4) = \infty$. Now the argument from the preceding proof implies that

$$\phi(z) = C(z,\phi^{-1}(1),\phi^{-1}(0),\phi^{-1}(\infty)) = C(z,z_2,z_3,z_4).$$

This formula provides an interpretation of the cross ratio, as a function of its first argument, as a fractional linear transformation.

Definition 4.44. In the Riemann sphere $\mathbb{P}_1(\mathbb{C}) \cong \mathbb{C}_\infty$, we call a subset C a *generalized circle* if it either is the projective completion

$$[L] := L \cup \{\infty\}$$

of a real affine line $L = z_0 + \mathbb{R}w \subseteq \mathbb{C}$, $w \neq 0$, or a circle in $\mathbb{C}$.

Remark 4.45. Generalized circles, considered as subsets of $\mathbb{C}$ (resp. $\mathbb{C}_\infty$), are the solutions of an equation

$$a|z|^2 + cz + \overline{cz} + d = 0 \quad \text{with} \quad a, d \in \mathbb{R}, ad < |c|^2. \tag{23}$$

Here $a = 0$ corresponds to affine lines. For $a \neq 0$ we may divide by a, so that we may w.l.o.g. assume that $a = 1$. In this case, we obtain for $z = x + iy$ and $c = c_1 + ic_2$ the equation

$$0 = x^2 + y^2 + 2(c_1x - c_2y) + d = (x + c_1)^2 + (y - c_2)^2 + d - |c|^2$$

of a circle in $\mathbb{R}^2$ with radius $\sqrt{|c|^2 - d}$ and center $(-c_1, c_2)$. In particular, we obtain a proper circle if and only if $d < |c|^2$.

Proposition 4.46. (Generalized circles in the Riemann sphere) *Then the following assertions hold for generalized circles in $\mathbb{C}_\infty$:*

(a) *If C is a generalized circle, then the same holds for $\overline{g}(C)$, $g \in \mathrm{GL}_2(\mathbb{C})$.*

(b) *The generalized circles are precisely the sets of the form $\overline{g}(\mathbb{R}_\infty)$, where $g \in \mathrm{GL}_2(\mathbb{C})$.*

(c) *Four distinct points z_j, $j = 1, 2, 3, 4$, in $\mathbb{C}_\infty$ lie on a generalized circle if and only if their cross ratio $C(z_1, z_2, z_3, z_4)$ lies in $\mathbb{R}_\infty$.*

Proof. (a) In Proposition 4.14 we have seen that the group $\mathrm{PGL}_2(\mathbb{C})$ is generated by affine maps $g(z) = az + b$, $a \in \mathbb{C}^\times$, and the inversion $\sigma(z) = z^{-1}$. The affine maps $g(z) = az + b$ can also be written as $g(z) = |a|e^{i\theta}z + b$ for some $\theta \in \mathbb{R}$. These maps are linear and satisfy $|g(z) - g(w)| = L|z - w|$ for $L = |a|$. This shows that all these maps map generalized circles to generalized circles.

To see that σ maps generalized circles to generalized circles, we first recall that generalized circles are solutions of some equation

$$a|z|^2 + cz + \overline{cz} + d = 0 \quad \text{for} \quad a, d \in \mathbb{R}, ad < |c|^2.$$

Replacing z by z^{-1} in the above equation leads to

$$a|z|^{-2} + cz^{-1} + \overline{cz}^{-1} + d = 0,$$

and multiplying with $|z|^2$ further leads to

$$a + c\overline{z} + \overline{c}z + d|z|^2 = 0.$$

Since this is the same type of equation, σ maps generalized circles to generalized circles.

(b) From (a) we derive that $\overline{g}(\mathbb{R}_\infty)$ is a generalized circle for every $g \in \mathrm{GL}_2(\mathbb{C})$. Suppose, conversely, that $C \subseteq \mathbb{C}_\infty$ is a generalized circle. Pick three different points $z_1, z_2, z_3 \in C$. Then there exists a $g \in \mathrm{GL}_2(\mathbb{C})$ with $\overline{g}(z_1) = -1$, $\overline{g}(z_2) = 0$, $\overline{g}(z_3) = 1$. Then $\overline{g}(C)$ is a generalized circle containing the points $-1, 0, 1$, which implies that $\overline{g}(C) = \mathbb{R}_\infty$. This shows that $C = \overline{g}^{-1}(\mathbb{R}_\infty)$, which proves (b).

(c) If z_j, $j = 1, 2, 3, 4$, lie in a generalized circle C and $C = g(\mathbb{R}_\infty)$ according to (a), then the invariance of the cross ratio implies that $C(z_1, z_2, z_3, z_4) \in \mathbb{R}_\infty$.

If, conversely, $x := C(z_1, z_2, z_3, z_4) \in \mathbb{R}_\infty$, then $x = C(x, 1, 0, \infty)$ implies the existence of some $g \in \mathrm{PGL}_2(\mathbb{C})$ with $g(z_1) = x$, $g(z_2) = 1$, $g(z_3) = 0$ and $g(z_4) = \infty$ (Proposition 4.38). Hence all the z_j lies in the generalized circle $g^{-1}(\mathbb{R}_\infty)$. □

Examples 4.47. (a) (The Cayley transform) The map

$$\mathrm{Cay}\colon \mathbb{C}_\infty \to \mathbb{C}_\infty, \quad \mathrm{Cay}(z) := \frac{z - i}{z + i}$$

is called the *Cayley transform*. Is corresponds to the matrix

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

For $x \in \mathbb{R}$ we clearly have $|\mathrm{Cay}(x)| = 1$, so that the preceding proposition implies that

$$\mathrm{Cay}(\mathbb{R}_\infty) = \mathbb{S}_1 = \{z \in \mathbb{C} : |z| = 1\}.$$

(b) For the affine line $C := i + \mathbb{R}$ and $\sigma(z) = z^{-1}$, we obtain

$$\sigma(i) = -i, \quad \sigma(i+1) = (i+1)^{-1} = \tfrac{1}{2}(1 - i), \quad \sigma(i-1) = (i-1)^{-1} = -\tfrac{1}{2}(1 + i),$$

and since theese three points determine the circle

$$C' := \{z \in \mathbb{C}\colon |z + \tfrac{1}{2}i| = \tfrac{1}{2}\},$$

we obtain $\sigma(C \cup \{\infty\}) = C'$.

Proposition 4.48. *Every Möbuis transformation of the form* $\phi(z) = (az + b)(cz + d)^{-1}$ *for* $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})$ *satisfies*

$$\phi(\mathbb{R}_\infty) = \mathbb{R}_\infty \quad \textit{and} \quad \phi(\mathbb{C}_+) = \mathbb{C}_+,$$

where $\mathbb{C}_+ := \{z \in \mathbb{C}_+\colon \mathrm{Im}\, z > 0\}$ *is the upper half plane. Conversely, all Möbius transformations preserving* $\mathbb{C}_+$ *are of this form.*

Proof. The first assertion follows from

$$\begin{aligned}
2i \operatorname{Im} \phi(z) = \phi(z) - \overline{\phi(z)} &= \frac{(az+b)(c\overline{z}+d) - (a\overline{z}+b)(cz+d)}{|cz+d|^2} \\
&= \frac{ac|z|^2 + adz + bc\overline{z} + bd - ac|z|^2 - bcz - ad\overline{z} - bd}{|cz+d|^2} \\
&= \frac{\det(g)(z-\overline{z})}{|cz+d|^2} = \frac{z-\overline{z}}{|cz+d|^2} = 2i\frac{\operatorname{Im} z}{|cz+d|^2}
\end{aligned}$$

implies that $\operatorname{Im} z$ and $\operatorname{Im} \phi(z)$ have the same sign.

Suppose, conversely, that ϕ preserves the upper half plane $\mathbb{C}_+$. From $\phi(\mathbb{C}_+) = \mathbb{C}_+$ it also follows that ϕ preserves $\mathbb{R}_\infty$. Since it is uniquely determined by the image of $0, 1, 2$, we have $\phi(z) = \frac{az+b}{cz+d}$ for some $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R})$. If $\det g < 0$, then the preceding calculation shows that $g(\mathbb{C}_+) = -\mathbb{C}_+$, so that we must have $\det g > 0$. $\square$

Proposition 4.49. *For the open unit disc $\mathcal{D} := \{z \in \mathbb{C}_+ \colon |z| < 1\}$, we have*

$$\mathrm{SU}_{1,1}(\mathbb{C}) := \left\{ \begin{pmatrix} \alpha & \overline{\beta} \\ \beta & \overline{\alpha} \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C}) \colon |\alpha|^2 - |\beta|^2 = 1 \right\} = \{g \in \mathrm{SL}_2(\mathbb{C}) \colon \overline{g}(\mathcal{D}) = \mathcal{D}\}.$$

Proof. In the preceding proposition we have seen that

$$\mathrm{SL}_2(\mathbb{R}) = \{g \in \mathrm{SL}_2(\mathbb{C}) \colon \overline{g}(\mathbb{C}_+) = \mathbb{C}_+\}.$$

We also know from Example 4.47 that the Cayley transform

$$\mathrm{Cay}(z) := \overline{T}(z) = \frac{z-i}{z+i} \quad \text{for } T := \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$$

satisfies $\mathrm{Cay}(\mathbb{R}_\infty) = \mathbb{S}_1$. From $\mathrm{Cay}(i) = 0$, we thus obtain $\mathrm{Cay}(\mathbb{C}_+) = \mathcal{D}$. Therefore

$$G := \{g \in \mathrm{SL}_2(\mathbb{C}) \colon \overline{g}(\mathcal{D}) = \mathcal{D}\} = \{g \in \mathrm{SL}_2(\mathbb{C}) \colon \mathrm{Cay}^{-1} \overline{g} \, \mathrm{Cay}(\mathbb{C}_+) = \mathbb{C}_+\} = T \, \mathrm{SL}_2(\mathbb{R}) T^{-1}.$$

For $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the condition $T^{-1}gT \in \mathrm{SL}_2(\mathbb{R})$ is equivalent to

$$T^{-1}gT = \overline{T^{-1}gT} = \overline{T}^{-1}\overline{g}\overline{T}, \qquad \text{i.e.,} \qquad \overline{T}T^{-1}g(\overline{T}T^{-1})^{-1} = \overline{g}.$$

As

$$\overline{T}T^{-1} = \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \frac{1}{2i} \begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix} = \frac{1}{2i} \begin{pmatrix} 0 & 2i \\ 2i & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

we obtain the condition

$$\begin{pmatrix} \overline{a} & \overline{b} \\ \overline{c} & \overline{d} \end{pmatrix} = \overline{g} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} g \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} d & c \\ b & a \end{pmatrix},$$

which is equivalent to $d = \overline{a}$ and $c = \overline{b}$. This shows that $T \, \mathrm{SL}_2(\mathbb{R}) T^{-1} = \mathrm{SU}_{1,1}(\mathbb{C})$. $\square$

Remark 4.50. To understand why the group in the preceding proposition is called $\mathrm{SU}_{1,1}(\mathbb{C})$, we consider on $\mathbb{C}^2$ the hermitian form

$$h(z,w) := z_1\overline{w_1} - z_2\overline{w_2} = z^\top D\overline{w} \quad \text{for} \quad D := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The corresponding symmetry group is

$$\begin{aligned} \mathrm{U}(\mathbb{C}^2, h) &= \{g \in \mathrm{GL}_2(\mathbb{C})\colon (\forall z, w \in \mathbb{C}^2)\, h(gz, gw) = h(z,w)\} \\ &= \{g \in \mathrm{GL}_2(\mathbb{C})\colon g^\top D\overline{g} = D\} \\ &= \{g \in \mathrm{GL}_2(\mathbb{C})\colon g^* = Dg^{-1}D^{-1}\}. \end{aligned}$$

For $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C})$, we have

$$g^* = \begin{pmatrix} \overline{\alpha} & \overline{\gamma} \\ \overline{\beta} & \overline{\delta} \end{pmatrix} \quad \text{and} \quad g^{-1} = \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}, \quad \text{so that} \quad Dg^{-1}D^{-1} = \begin{pmatrix} \delta & \beta \\ \gamma & \alpha \end{pmatrix}.$$

Therefore $g \in \mathrm{U}(\mathbb{C}^2, h)$ is equivalent to $\delta = \overline{\alpha}$ and $\gamma = \overline{\beta}$. This leads to

$$\mathrm{SU}_{1,1}(\mathbb{C}) = \Big\{ \begin{pmatrix} \alpha & \overline{\beta} \\ \beta & \overline{\alpha} \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C})\colon |\alpha|^2 - |\beta|^2 = 1 \Big\} = \mathrm{SU}(\mathbb{C}^2, h) := \mathrm{SL}_2(\mathbb{C}) \cap \mathrm{U}(\mathbb{C}^2, h).$$

Remark 4.51. (More on cross ratio and generalized circles) For $\mathbb{K} = \mathbb{C}$, we consider the cross ratio on the projective line $\mathbb{C}_\infty \cong \mathbb{P}_1(\mathbb{C})$:

$$C(z_1, z_2, z_3, z_4) = \frac{(z_4 - z_2)(z_3 - z_1)}{(z_4 - z_1)(z_3 - z_2)}.$$

For $z_1, z_2, z_3 \in \mathbb{C}$, there exists a uniquely determined generalized circle C containing all three points (by elementary geometry). For a fourth point $z \in \mathbb{C}$, the condition

$$C(z, z_1, z_2, z_3) \in \mathbb{R}_\infty$$

is equivalent to z, z_1, z_2, z_3 lying in one generalized circle, so that it means that $z \in C$ (Proposition 4.46). We therefore obtain

$$C = \{z \in \mathbb{C}_\infty\colon C(z, z_1, z_2, z_3) \in \mathbb{R}_\infty\}.$$

4.9 A projective view on conic sections

In this subsection we briefly discuss conic section from the point of view of real projective geometry. We start by explaining how the zero set of a polynomial can be defined as a subset of projective space and then apply this to the equation $x^2 + y^2 = 1$ of the euclidean circle. We shall then derive the projective equivalence of regular conic sections.

Definition 4.52. (Projective zero sets) A polynomial $f \in \mathbb{K}[x_0, x_1, \ldots, x_n]$ is called *homogeneous of degree* d if it is a linear combinations of monomials $x^{\mathbf{m}} := x_0^{m_0} \cdots x_n^{m_n}$ with $\mathbf{m} = (m_0, \ldots, m_n)$ and $|\mathbf{m}| := m_0 + m_1 + \ldots + m_n = d$. This implies that $f(\lambda x) = \lambda^d f(x)$ for $x \in \mathbb{K}^{n+1}$. Therefore we can define the *projective zero set* of f:

$$Z(f) := \{[x] \in \mathbb{P}_n(\mathbb{K})\colon f(x) = 0\}.$$

This set is well-defined because the relation $f(x) = 0$ does not depend on the choice of a representative $x \in [x] = \mathbb{K}x$.

Examples 4.53. (a) If $f(x) = a_0x_0 + \cdots + a_nx_n$ is linear, then its projective zero set

$$Z(f) := \Big\{[x] \in \mathbb{P}_n(\mathbb{K})\colon \sum_{j=0}^n a_jx_j = 0\Big\}$$

is the projective space of the hyperplane $\ker f \subseteq \mathbb{K}^{n+1}$.

(b) If $f(x) = \sum_{i,j=0}^n a_{ij}x_ix_j$ with a symmetric matrix $A = (a_{ij}) \in \mathrm{Sym}_{n+1}(\mathbb{K})$, then

$$Z(f) := \{[x] \in \mathbb{P}_n(\mathbb{K})\colon f(x) = 0\}$$

is called a *projective quadric*.

(c) For $f(x_0, x_1) := x_0x_1$, we have

$$Z(f) := \{[x] \in \mathbb{P}_1(\mathbb{K})\colon x_0x_1 = 0\} = \{[1 : 0], [0 : 1]\}.$$

Remark 4.54. (Homogeneous completion of a polynomial) Let $f(x) = \sum a_{\mathbf{m}}x^{\mathbf{m}} \in \mathbb{K}[x_1, \ldots, x_n]$ be a not necessarily homogeneous polynomial. We can associate to f a homogeneous polynomial as follows. Let

$$N := \max\{|\mathbf{m}|\colon a_{\mathbf{m}} \neq 0\}.$$

Then

$$\widetilde{f}(x) := \sum a_{\mathbf{m}}x_0^{N-|\mathbf{m}|}x^{\mathbf{m}} \in \mathbb{K}[x_0, x_1, \ldots, x_n]$$

is a homogeneous polynomial. It has the property that,

$$\widetilde{f}(1, x_1, \ldots, x_n) = f(x_1, \ldots, x_n) \quad \text{for} \quad x \in \mathbb{K}^n.$$

Therefore the projective zero set $Z(\widetilde{f}) \subseteq \mathbb{P}_n(\mathbb{K})$ can be considered as a "projective completion" of the zero set of f in $\mathbb{K}^n$. In fact, it corresponds to the intersection of $Z(\widetilde{f})$ with the affine subspace

$$\mathbb{K}^n \cong \{[1 : x_1 : \cdots : x_n]\colon x \in \mathbb{K}^n\} \subseteq \mathbb{P}_n(\mathbb{K}).$$

Example 4.55. (a) For the polynomial $f(x) = x_1^3 + x_2^2 - x_3 + 1$, we obtain the homogeneous polynomial

$$\widetilde{f}(x) = x_1^3 + x_0x_2^2 - x_0^2x_3 + x_0^3.$$

(b) For the polynomial $f(x) = x_1^5 - x_3^2$, we get

$$\widetilde{f}(x) = x_1^5 - x_0^2x_3^2.$$

We consider the homogeneous polynomial

$$f(x_1, x_2, x_3) := x_1^2 + x_2^2 - x_3^2.$$

The corresponding subset

$$C := \{[x] \in \mathbb{P}_2(\mathbb{R})\colon f(x) = 0\}$$

is called the *standard conic* and

$$\widetilde{C} := \{x \in \mathbb{R}^3\colon f(x) = 0\}$$

is called the *standard cone*.

A *conic section* is a planar curve of the type $S := \widetilde{C} \cap H$, where $H \subseteq \mathbb{R}^3$ is an affine plane. If H is a linear subspace, i.e., $0 \in H$, then the conic is called *singular* and otherwise *regular*.

In the following we discuss regular conic sections. We write the hyperplane H as

$$H = \{x \in \mathbb{R}^3\colon \langle x, \mathbf{n}\rangle = c\} \quad \text{for} \quad c \in \mathbb{R}, 0 \neq \mathbf{n} \in \mathbb{R}^3.$$

Lemma 4.56. *We obtain the following types of conic sections $S := \widetilde{C} \cap H$:*

(1) $f(\mathbf{n}) < 0$, $c = 0$: $S = \{0\}$.

(2) $f(\mathbf{n}) < 0$, $c \neq 0$: *S is an ellipse. Typical case*

$$\mathbf{n} = e_3, \qquad S = \{(x_1, x_2, c) \colon x_1^2 + x_2^2 = c^2\}.$$

(3) $f(\mathbf{n}) = 0$, $c = 0$: $S = \mathbb{R}(n_1, n_2, -n_3)$ *is a line.*

(4) $f(\mathbf{n}) = 0$, $c \neq 0$: *S is a parabola. Typical case*

$$\mathbf{n} = e_2 + e_3, \qquad S = \{(x_1, x_2, c - x_2) \colon x_1^2 + 2cx_2 = c^2\}.$$

(5) $f(\mathbf{n}) > 0$, $c = 0$: *$S = \mathbb{R}x \cup \mathbb{R}y$ is a union of two lines, where $x, y \in \mathbf{n}^\perp$ are two linearly independent elements in S.*

(6) $f(\mathbf{n}) > 0$, $c \neq 0$: *S is a hyperbola. Typical case*

$$\mathbf{n} = e_2, \qquad S = \{(x_1, c, x_3) \colon x_1^2 - x_3^2 = -c^2\}.$$

Proof. Rotating around the x_3-axis, which leaves $\widetilde{C}$ invariant, we may w.l.o.g. assume that $n_1 = 0$.

(1), (2) From $0 > f(\mathbf{n}) = n_2^2 - n_3^2$ it follows that $0 \leq |n_2| < |n_3|$. We may w.l.o.g. assume $n_3 = 1$. We then obtain $x_3 = c - n_2 x_2$, which leads to

$$f(x) = x_1^2 + x_2^2 - (c - n_2 x_2)^2 = x_1^2 + (1 - n_2^2)x_2^2 + 2cn_2 x_2 - c^2.$$

As $n_2^2 < 1$, this describes an ellipse which degenerates to a point if $c = 0$. That we obtain a proper ellipse for $c \neq 0$ follows from $f(0) = -c^2 < 0$ and $\lim_{|x| \to \infty} f(x) = \infty$.

(3), (4) From $f(\mathbf{n}) = 0$ we derive $|n_2| = |n_3|$, so that we may assume $n_3 = 1$ and $x_3 = c - n_2 x_2$. Now

$$f(x) = x_1^2 + x_2^2 - (c - n_2 x_2)^2 = x_1^2 + (1 - n_2^2)x_2^2 + 2cn_2 x_2 - c^2 = x_1^2 + 2cn_2 x_2 - c^2.$$

For $c = 0$, $f(x) = 0$ implies $x_1 = 0$, so that $S = \mathbb{R}(0, 1, -n_2)$, and for $c \neq 0$, S is a parabola.

(5), (6) From $0 < f(\mathbf{n}) = n_2^2 - n_3^2$ it follows that $|n_2| > |n_3|$. We may w.l.o.g. assume $n_2 = 1$. We then obtain $x_2 = c - n_3 x_3$, which leads to

$$f(x) = x_1^2 + (c - n_3 x_3)^2 - x_3^2 = x_1^2 + (n_3^2 - 1)x_3^2 - 2cn_3 x_3 + c^2.$$

For $c \neq 0$, S is a hyperbola. For $c = 0$ it degenerates to a pair of lines: $x_1 = \pm\sqrt{1 - n_3^2} \cdot x_3$. $\square$

We now consider the subet $[C] \subseteq \mathbb{P}_2(\mathbb{R})$ as a curve in projective space. For $c \neq 0$, the affine hyperplane $H = \{x \in \mathbb{R}^3 \colon \langle x, \mathbf{n} \rangle = c\}$ does not contain zero, so that we obtain the affine subspace $[H] \subseteq \mathbb{P}_2(\mathbb{R})$. We can now ask for the type of the curve $[H] \cap [C]$, which corresponds under the embedding $H \cong [H] \subseteq \mathbb{P}_2(\mathbb{R})$ to the curve $S := H \cap \widetilde{C}$.

In the preceding lemma we have seen that S may be a parabola, a hyperbola or an ellipse, depending on the position of the affine hyperplane H.

We now change our perspective and fix the hyperplane H. For $g \in \mathrm{GL}_3(\mathbb{R})$ we then obtain conic sections

$$[H] \cap \overline{g}[C] = \overline{g}(\overline{g}^{-1}[H] \cap [C]),$$

corresponding to the curves $H \cap g\widetilde{C}$ which are of the same type as $g^{-1}H \cap \widetilde{C}$. Since every hyperplane $H' \subseteq \mathbb{R}^3$ not containing zero is of the form gH for some $g \in \mathrm{GL}_3(\mathbb{R})$ (Exercise!), we see that $\overline{g}[C] \cap [H]$ can be of all three types: parabola, hyperbola or ellipse. This proves the following theorem:

Theorem 4.57. (Projective equivalence of conic sections) *Projective transformations of an affine plane can turn every non-degenerate conic section, parabola, hyperbola or ellipse into any of these forms.*

Example 4.58. For $H = \{x \in \mathbb{R}^3 \colon x_3 = 1\}$ we obtain the circle

$$S = H \cap \widetilde{C} = \{(x_1, x_2, 1) \colon x_1^2 + x_2^2 = 1\}.$$

For

$$g := \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \in \mathrm{GL}_3(\mathbb{R})$$

the corresponding projective transformation on the hyperplane $H \cong \mathbb{R}^2$ is given by

$$\overline{g}[x_1 : x_2 : 1] = [x_1 + 1 : x_2 : x_1 - 1] = \Big[\frac{x_1+1}{x_1-1} : \frac{x_2}{x_1-1} : 1\Big].$$

Put

$$\phi(x_1, x_2) := \Big(\frac{x_1+1}{x_1-1}, \frac{x_2}{x_1-1}\Big).$$

Then ϕ is singular in the line $x_1 = 1$ that is tangent to the circle

$$S = \{(x_1, x_2) \colon x_1^2 + x_2^2 = 1\}.$$

Now

$$\phi^{-1}(S) = \{(x_1, x_2) \colon (x_1+1)^2 + x_2^2 = (x_1-1)^2, x_1 \neq 1\} = \{(x_1, x_2) \colon x_2^2 = -4x_1\}$$

is a parabola.

For

$$h := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \in \mathrm{GL}_3(\mathbb{R})$$

the corresponding projective transformation on $H \cong \mathbb{R}^2$ is given by

$$\overline{h}[x_1 : x_2 : 1] = [1 : x_2 : x_1] = \Big[\frac{1}{x_1} : \frac{x_2}{x_1} : 1\Big].$$

Put

$$\psi(x_1, x_2) := \Big(\frac{1}{x_1}, \frac{x_2}{x_1}\Big).$$

Then ψ is singular in the line $x_1 = 0$ intersecting the circle S. Now

$$\psi^{-1}(S) = \{(x_1, x_2) \colon 1 + x_2^2 = x_1^2, x_1 \neq 0\} = \{(x_1, x_2) \colon x_1^2 - x_2^2 = 1\}$$

is a hyperbola.

Exercises for Section 4

Exercise 4.1. Let V be a vector space and $\phi\colon V \to V$ be a linear map mapping every one-dimensional subspace into itself. Show that there exists a $\lambda \in \mathbb{K}$ with $\phi = \lambda\, \mathrm{id}_V$.

Exercise 4.2. (Transformation properties of the cross ratio) Let z_j, $j = 1, 2, 3, 4$, be four distinct points in $\mathbb{K}_\infty$. Show that their cross ratio transforms as follows under permutation of the entries:

$$C(z_2, z_1, z_3, z_4) = C(z_1, z_2, z_3, z_4)^{-1}, \quad C(z_1, z_3, z_2, z_4) = 1 - C(z_1, z_2, z_3, z_4)$$

$$C(z_1, z_2, z_4, z_3) = C(z_1, z_2, z_3, z_4)^{-1}.$$

Conclude from these relations that

$$C(z_2, z_3, z_4, z_1) = \frac{C(z_1, z_2, z_3, z_4)}{C(z_1, z_2, z_3, z_4) - 1}.$$

How many different elements of $\mathbb{K}_\infty$ do we get at most by permuting the arguments of the cross ratio?

Exercise 4.3. Recall the stereographic projection

$$\phi\colon \mathbb{S}^2 \setminus \{e_0\} \to \mathbb{R}^2, \quad (x_0, x_1, x_2) \mapsto \frac{1}{1 - x_0}(x_1, x_2)$$

(Example 4.7). Show that:

(a) ϕ is continuous with continuous inverse $\phi^{-1}(x) = \left(\frac{\|x\|_2^2 - 1}{\|x\|_2^2 + 1}, \frac{2x}{\|x\|_2^2 + 1}\right)$.

(b) If $C \subseteq \mathbb{S}^2$ is a circle containing e_0, then $\phi(C \setminus \{e_0\})$ is an affine line.

(c)* If $C \subseteq \mathbb{S}^2$ is a circle not containing e_0, then $\phi(C)$ is a circle.

Exercise 4.4. Show that the natural map

$$q\colon \mathrm{SL}_n(\mathbb{K}) \to \mathrm{PGL}_n(\mathbb{K}), \quad g \mapsto \mathbb{K}^\times g = \overline{g}$$

is surjective if and only if, for every $\lambda \in \mathbb{K}^\times$, there exists a $\mu \in \mathbb{K}$ with $\mu^n = \lambda$. Discuss the cases $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$; for which n is this condition satisfied?

References

[Ar97] Artmann, B., "Geometrie," Vorlesung an der TU Darmstadt, WS 96/97

[Ba01] Bär, C., "Elementare Differentialgeometrie," de Gruyter, 2001

[Ba64] Bargmann, V., *Note on Wigner's Theorem on Symmetry Operations*, Journal of Mathematical Physics **5:7** (1964), 862–868 (A self-contained clear proof of Wigner's Theorem mentioned in the introduction.)

[Bl75] Blumenthal, L. M., *Four-point properties and norm postulates*, in "The Geometry of Metric and Linear Spaces", Proc. Conf. Michigan State Univ., Lecture Notes in Math. **490**, Springer, Berlin, 1975; 1–13

[Bo41] Bochner, S., *Hilbert distances and positive definite functions*, Ann. of Math. **42** (1941), 647–656

[FF00] Faure, C.-A., and A. Frölicher, "Modern Projective Geometry," Kluwer Acad. Press, Dordrecht, 2000

[Fr35] Fréchet, M., *Sur la définition axiomatique d'une classe d'espaces vectoriels distanciés applicables vectoriellement sur l'espace de Hilbert*, Ann. Math. **36** (1935), 705–718

[Gl66] Gleason, A. M., *The definition of a quadratic form*, Amer. Math. Monthly **73** (1966), 1049–1056

[W89] Hsiang, W.-Y., *On the laws of trigonometries of two-point homogeneous spaces*, Ann. Global Anal. Geom. **7:1** (1989), 29–45

[JvN35] Jordan, P., and J. von Neumann, *On inner product in linear metric spaces*, Ann. of Math. **36:2** (1935), 719–723

[Me28] Menger, K., *Untersuchungen zur allgemeinen Metrik*, Math. Ann. **100:1** (1928), 618–627

[Pa13] Pankov, M., *Characterization of geometrical mappings of projective spaces in terms of the projective linear group*, Preprint, arXive:math.RT.1304.1626

[Ra02] Randrianantoanina, B., *A note on the Banach–Mazur problem*, Glasgow Math. Journal **44** (2002), 159–165

[Sch38] Schoenberg, I. J., *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. **44** (1938), 522–536

[Sch38b] Schoenberg, I. J., *Metric spaces and completely monotone functions*, Ann. of Math. (2) **39:4** (1938), 811–841

[Sch42] Schoenberg, I. J., *Positive definite functions on spheres*, Duke Math. J. **9** (1942), 96–108

[SvN41] Schoenberg, I. J., and J. von Neumann, *Fourier integrals and metric geometry*, Trans. Amer. Math. Soc. **50** (1941), 226–251

[PF80] Pfeffer, W. F., *Lorentz transformations of a Hilbert space*, Amer. J. Math. **103:4** (1980), 691–709

[Wa51] Wang, H.-C., *Two theorems on metric spaces*, Pac. J. Math. **1:3** (1951), 473–480

[Wa52] Wang, H.-C., *Two-point homogeneous spaces*, Ann. Math. **55:1** (1952), 177–191

[We11] Werner, D., "Funktionalanalysis," Springer-Verlag, 2011

[Wi32] Wilson, W. A., *A relation between metric and euclidean spaces*, Amer. J. Math. **54** (1932), 505–517