

Lecture Notes Geometry

Roland van der Veen

7-4-2020

1 Introduction

It is indeed very satisfying for a mathematician to define an affine space as being a set acted on by a vector space (and this is what I do here) but this formal approach, although elegant, must not hide the phenomenological aspect of elementary geometry, its own aesthetics: yes, Thales theorem expresses the fact that a projection is an affine mapping, no, you do not need to orient the plane before defining oriented angles. . . but this will prevent neither the Euler circle from being tangent to the incircle and excircles, nor the Simson lines from enveloping a three-cusped hypocycloid! -Michelle Audin.

In these lecture notes we aim to summarize some of the main points of the first chapters of the book *Geometry* by Michelle Audin: We will also treat some elementary notions of differential geometry after. All our vector spaces will be over \mathbb{R} for simplicity.

2 Affine space

A natural setting for discussing elementary geometry of lines and planes is affine space. We are mostly concerned with properties of parallelism and intersection, not yet with distances. The approach is firmly rooted in linear algebra as we will see affine space is just a vector space where one refuses to say where the origin is as geometrically there is no natural choice of origin.

Definition 2.1. (Audin 1.1)

A set \mathcal{E} is called an **affine space** if there exist a vector space E and a map $\Theta : \mathcal{E} \times \mathcal{E} \rightarrow E$. We use the notation $\Theta(A, B) = \overrightarrow{AB}$. The map should satisfy two conditions:

1. For all $A \in \mathcal{E}$ the map $B \mapsto \overrightarrow{AB}$ is a bijection from \mathcal{E} to E .
2. For all $A, B, C \in \mathcal{E}$ we have $\overrightarrow{AC} + \overrightarrow{CB} = \overrightarrow{AB}$.

We say \mathcal{E} is directed by vector space E or E underlies \mathcal{E} . Also sometimes the image of Θ_A is named \mathcal{E}_A : it is \mathcal{E} viewed as a vector space with the origin at A .

For two points $A, B \in \mathcal{E}$ we may express $\Theta_A \circ \Theta_B^{-1} : E \rightarrow E$ as follows: $\Theta_A \circ \Theta_B^{-1} v = \overrightarrow{AB} + v$. This is because $\Theta_B^{-1} v = C$ for a unique $C \in \mathcal{E}$ such that $v = \overrightarrow{BC}$. But then $\Theta_A C = \overrightarrow{AC}$ and the second property of affine space says $\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC}$. In other words $\Theta_A \circ \Theta_B^{-1} v = \overrightarrow{AC} = \overrightarrow{AB} + v$.

Going in the opposite direction we can promote any vector space V to an associated affine space \mathcal{V} as follows. Define $\mathcal{V} = V$ as a set and define $\Theta : \mathcal{V} \times \mathcal{V} \rightarrow V$ by $\Theta(A, B) = B - A$. This provides many examples of affine spaces but just like with vector spaces there is in some sense only one for each dimension. We call it affine n -space. To get even more concrete we could do the above procedure to $V = \mathbb{R}^n$. However being so concrete is not always the easiest approach to difficult geometric problems. Often it is not the coordinates that count but rather the relationships between them and these are observed more readily in an affine setting.

Definition 2.2. (Audin p.10)

A subset $\mathcal{F} \subset \mathcal{E}$ of an affine space \mathcal{E} is called an **affine subspace** if it is empty or for some $A \in \mathcal{F}$ we have $\Theta_A(\mathcal{F})$ is a linear subspace of E .

Definition 2.3. (Audin p.11)

For $S \subset \mathcal{E}$ the **subspace generated** by S is the intersection of all affine subspaces of \mathcal{E} containing S . Notation $\langle S \rangle$.

For points $A, B \in \mathcal{E}$ the line $\langle A, B \rangle$ is also described as the line AB . The line segment from A to B is the set $\{M \in \mathcal{E} | \overrightarrow{AM} = \lambda \overrightarrow{AB}, \lambda \in [0, 1]\}$.

Definition 2.4. (Audin p.12)

Points $A_0 \dots A_k \in \mathcal{E}$ are **affine independent** if $\langle A_0, \dots, A_k \rangle$ has dimension k . When $k = \dim \mathcal{E}$ we say (A_0, \dots, A_k) is an **affine frame** of \mathcal{E} .

Affine independence is closely related to linear independence as becomes clear when we choose one of the points as an origin. More precisely,

Lemma 2.1. The points $A_0 \dots A_k \in \mathcal{E}$ are affine independent if and only if the vectors $\overrightarrow{A_0 A_i}$ for $i = 1, \dots, k$ are linearly independent.

Proof. $\mathcal{F} = \langle A_0, \dots, A_k \rangle$, is an affine subspace so $\Theta_{A_0} : \mathcal{E} \rightarrow E$ sends \mathcal{F} to a linear subspace $F = \Theta_{A_0}(\mathcal{F})$. Affine independence means that $\dim F = k$. On the other hand, by definition of Θ_{A_0} the space F is spanned by the k vectors $\overrightarrow{A_0 A_i}$ for $i = 1, \dots, k$. Therefore $\dim F = k$ implies these are all linearly independent and vice versa. \square

In particular the above lemma asserts the relation between affine frames in \mathcal{E} and bases of E . This is of great importance for applying linear algebra to understand affine space. Of course the choice of origin A_0 in the above was arbitrary, we could just as well have chosen some other point A_j .

Definition 2.5. (Audin p.12)

Two affine subspaces are said to be **parallel** if they have the same direction (underlying vector space).

More precisely this means that affine subspaces $\mathcal{F}, \mathcal{G} \subset \mathcal{E}$ are parallel iff for every $A \in \mathcal{F}$ and $B \in \mathcal{G}$ we have $\Theta_A(\mathcal{F}) = \Theta_B(\mathcal{G})$ as linear subspaces of E .

Definition 2.6. (Audin p.14)

A map $\mathcal{E} \xrightarrow{\phi} \mathcal{F}$ between two affine spaces is called an **affine map** if for some $O \in \mathcal{E}$ there is a linear map $E \xrightarrow{f} F$ such that for all $M \in \mathcal{E}$ we have $\overrightarrow{\phi(O)\phi(M)} = f(\overrightarrow{OM})$. Sometimes we use the notation $f = \overrightarrow{\phi}$.

Sometimes it is useful to reformulate this condition in terms of the bijection $\Theta_O : \mathcal{E} \rightarrow E$ and $\Theta_{\phi(O)} : \mathcal{F} \rightarrow F$ as follows: $f \circ \Theta_O = \Theta_{\phi(O)} \circ \phi$. These compositions are conveniently visualized in a (commutative) diagram:

$$\begin{array}{ccc} \mathcal{E} & \xrightarrow{\phi} & \mathcal{F} \\ \downarrow \Theta_O & & \downarrow \Theta_{\phi(O)} \\ E & \xrightarrow{f} & F \end{array}$$

So after choosing an origin O and the compatible choice of origin $\phi(O)$ we find that ϕ is represented by the linear map f .

Examples of affine maps:

- 1) For $u \in E$ the translation $t_u : \mathcal{E} \rightarrow \mathcal{E}$ is an affine map defined by $t_u(A) = B$ with $\overrightarrow{AB} = u$.
- 2) For $O \in \mathcal{E}$ and $\lambda \in \mathbb{R}$ the (central) dilation with center O and ratio λ , denoted $h_{O,\lambda} : \mathcal{E} \rightarrow \mathcal{E}$, is the affine map defined by $\overrightarrow{Oh_{O,\lambda}(M)} = \lambda \overrightarrow{OM}$.
- 3) Another example is the projection $\pi_{\mathcal{F},L}$ onto affine subspace $\mathcal{F} \subset \mathcal{E}$ in the direction of linear subspace $L \subset E$. This makes sense only if $L + F = E$ and $L \cap F = 0$ where F is the direction of \mathcal{F} . In that case $\pi_{\mathcal{F},L}(A) = B$ where $B \in \mathcal{F}$ is the unique point such that $\overrightarrow{AB} \in L$. Point B is indeed unique because if C was another such point then $\overrightarrow{AB} - \overrightarrow{AC}$ would be in $F \cap L$. Point B exists because for any $O \in \mathcal{F}$ we can write $\overrightarrow{OA} = \ell + v$ for some $v \in F, \ell \in L$ and set B to be such that $\overrightarrow{OB} = v$. The corresponding linear map is the projection $\ell + v \xrightarrow{f} v$ for any $\ell \in L, v \in F$. Indeed $f(\overrightarrow{OA}) = \pi_{\mathcal{F},L}(O)\pi_{\mathcal{F},L}(A)$.

Lemma 2.2. (Audin 2.5 p.16)

Any affine map sends affine subspaces to affine subspaces.

Proof. Given affine subspace $\mathcal{G} \subset \mathcal{E}$ and affine map $\phi : \mathcal{E} \rightarrow \mathcal{F}$ as above we need to check that $\phi(\mathcal{G})$ is an affine subspace of \mathcal{F} . When $\mathcal{G} = \emptyset$ this is clear. Otherwise choose a point $O \in \mathcal{G}$, since it is an affine subspace $G = \Theta_O(\mathcal{G}) \subset E$ is a linear subspace. Now $\phi(O) \in \phi(\mathcal{G})$ so we need to check that $J = \Theta_{\phi(O)}(\phi(\mathcal{G})) \subset F$ is also a linear subspace. Looking at the diagram of maps above and writing $f = \overrightarrow{\phi}$ we have $\Theta_{\phi(O)} \circ \phi = f \circ \Theta_O$ so $J = f \circ \Theta_O(\mathcal{G}) = f(G)$. From linear algebra we know that the linear map f sends linear subspace $G \subset E$ to a linear subspace $f(G) \subset F$. \square

Lemma 2.3. (Audin 2.10 p.18)

An affine mapping is determined by the images of an affine frame. Any affine mapping of n -dimensional affine space that fixes $n + 1$ affine independent points is the identity.

Proof. Given affine frame A_0, \dots, A_n and $O = A_0$ Lemma 2.1 tells us that Θ_O sends A_1, \dots, A_n to a basis v_1, \dots, v_n of E . Using the same notation as above the linear map $f = \vec{\phi}$ completely determines ϕ . Also $f(v_i) = \Theta_{\phi(O)} \circ \phi \circ \Theta_O^{-1}(v_i) = \Theta_{\phi(O)}(\phi(A_i))$ shows that f in turn is completely determined by the images of the frame A_i under ϕ . Here we used that a linear map is determined by what it does on the basis v_i . \square

For four collinear points $A, B, C \neq D$ the ratio $\frac{\overrightarrow{AB}}{\overrightarrow{CD}}$ means the scalar λ such that $\overrightarrow{AB} = \lambda \overrightarrow{CD}$.

Theorem 2.1. (*Pappus, Audin Thm 3.4, p.25*)

In an affine plane imagine points A, B, C on line \mathcal{D} and A', B', C' three points on a second distinct line \mathcal{D}' . If AB' is parallel to $A'B$ and BC' parallel to CB' then AC' must be parallel to CA' .

Lemma 2.4. *Given two distinct lines $\mathcal{L}, \mathcal{L}'$ intersecting at O and points X, Y on \mathcal{L} and X', Y' on \mathcal{L}' . The dilation $h_{O, \lambda}$ sends both X to Y and X' to Y' if and only if XX' is parallel to YY' .*

Proof. Lines XX' and YY' are parallel if and only if $\overrightarrow{YY'} = \mu \overrightarrow{XX'}$ for some $\mu \in \mathbb{R}$. If $\overrightarrow{OY} = \lambda \overrightarrow{OX}$ and $\overrightarrow{OY'} = \lambda \overrightarrow{OX'}$ then

$$\overrightarrow{YY'} = \overrightarrow{OY'} - \overrightarrow{OY} = \lambda(\overrightarrow{OX'} - \overrightarrow{OX}) = \lambda \overrightarrow{XX'}$$

Conversely, suppose $\overrightarrow{OY} = \lambda \overrightarrow{OX}$ and $\overrightarrow{YY'} = \mu \overrightarrow{XX'}$ for some $\mu \in \mathbb{R}$. Then

$$\lambda \overrightarrow{OX'} = \lambda \overrightarrow{OX} + \lambda \overrightarrow{XX'} = \overrightarrow{OY} + \lambda \mu \overrightarrow{YY'} = \overrightarrow{OY'} - \overrightarrow{YY'} + \lambda \mu \overrightarrow{YY'}$$

Since the lines $\mathcal{L}, \mathcal{L}'$ are assumed distinct $\overrightarrow{YY'}$ is independent from the (dependent) vectors $\overrightarrow{OY'}, \overrightarrow{OX'}$ and the two terms on the right hand side must cancel out. \square

Lemma 2.5. *Given two distinct parallel lines $\mathcal{L}, \mathcal{L}'$ and points X, Y on \mathcal{L} and X', Y' on \mathcal{L}' . The translation $t_{\overrightarrow{XY}}$ sends both X to Y and X' to Y' if and only if XX' is parallel to YY' .*

Proof. If $\overrightarrow{X'Y'} = \overrightarrow{XY}$ then using $\overrightarrow{XY} = -\overrightarrow{YX}$ we get

$$\overrightarrow{YY'} = \overrightarrow{YX} + \overrightarrow{XX'} + \overrightarrow{X'Y'} = \overrightarrow{XX'}$$

Conversely, suppose $\overrightarrow{YY'} = \mu \overrightarrow{XX'}$ for some $\mu \in \mathbb{R}$. Then $\overrightarrow{X'Y'} = \overrightarrow{X'X} + \overrightarrow{XY} + \overrightarrow{YY'} = \overrightarrow{XY} + (-1 + \mu) \overrightarrow{XX'}$ implies that $\mu = 1$ since $\overrightarrow{X'Y'}$ and \overrightarrow{XY} are parallel vectors assumed independent of $\overrightarrow{XX'}$ because the lines $\mathcal{L}, \mathcal{L}'$ are assumed distinct. \square

Proof. (of Pappus)

First we will assume the lines $\mathcal{D}, \mathcal{D}'$ intersect in point O . For some λ, μ we have $h_{O, \lambda}A = B$ and $h_{O, \mu}B = C$. By Lemma 2.4 $h_{O, \lambda}B' = A'$ and $h_{O, \mu}C' = B'$. This means that the composition $h_{O, \lambda\mu} = h_{O, \lambda} \circ h_{O, \mu} = h_{O, \mu} \circ h_{O, \lambda}$ must map both A to C and C' to A' . According to Lemma 2.4 this means AC' is parallel to $C'A$.

When lines $\mathcal{D}, \mathcal{D}'$ in the plane do not intersect they must be parallel and we argue in the same way using translations: The translation $t_{\overrightarrow{AB}}$ sends A to B and the translation $t_{\overrightarrow{BC}}$ sends B to C . By Lemma 2.5 we also have $t_{\overrightarrow{AB}}B' = A'$ and $t_{\overrightarrow{BC}}C' = B'$. So $t_{\overrightarrow{AB+BC}}$ sends both A to C and C' to A' . According to Lemma 2.4 this means AC' is parallel to $C'A$. \square

Theorem 2.2. (*Desargues*)

Given distinct triangles A, B, C and A', B', C' in an affine plane such that AB is parallel to $A'B'$ and AC parallel to $A'C'$ and BC parallel to $B'C'$, the lines AA', BB' and CC' are either all parallel or intersect in a single point.

Proof. First consider the case where AA' and BB' intersect in point O . By Lemma 2.4 there is a $\lambda \in \mathbb{R}$ such that $h(O, \lambda)A = A'$ and $h_{O, \lambda}B = B'$. We aim to show that the point $C'' = h_{O, \lambda}C$ is equal to C' . Again Lemma 2.4 assures us that $A'C''$ and AC are parallel so C'' lies on $A'C'$. For the same reason C'' also lies on line $B'C'$. Together this means $C'' = C'$. Also CC' passes through O by definition of C'' .

Second, the case where AA' and BB' are parallel is approached analogously using Lemma 2.5. So the translation $t_{\overrightarrow{AA'}}$ sends A to A' and also B to B' . Define $C'' = t_{\overrightarrow{AA'}}C$ then we want to show $C'' = C'$. Lemma 2.5 implies that $A'C''$ is parallel to AC so C'' lies on $A'C'$. For the same reason C'' also lies on line $B'C'$. Together this means $C'' = C'$ and so CC' is parallel to AA' . \square

Exercises

- Can you give an example of an affine space that is not a vector space?
- Prove that for any points B, C of affine subspace $\mathcal{F} \subset \mathcal{E}$ we have $\Theta_B(\mathcal{F}) = \Theta_C(\mathcal{F})$.
- Prove that through any two points of an affine space passes a unique line.
- Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Define $\mathcal{F} = \{x \in \mathbb{R}^n : Ax = b\}$. Prove that \mathcal{F} is an affine subspace of \mathbb{R}^n . What is its direction? When is it empty? Express the dimension of \mathcal{F} in terms of the rank of A .
- Given an affine map $\phi : \mathcal{E} \rightarrow \mathcal{F}$ between affine spaces and \mathcal{M} an affine subspace of \mathcal{F} . Prove that $\phi^{-1}(\mathcal{M})$ is an affine subspace of \mathcal{E} .
- Prove that an affine map is completely determined by the image of an affine frame.
- Is Pappus theorem valid if the lines $\mathcal{D}, \mathcal{D}'$ are two distinct lines in three-dimensional affine space?
- The set of all invertible affine maps from \mathcal{E} to itself is denoted $GA(\mathcal{E})$.
 - Prove that the composition of two affine maps ϕ and ψ is again an affine map, and that the linear map associated to $\psi \circ \phi$ is the composition of the linear maps associated to ψ and ϕ , i.e. that $\overrightarrow{\psi \circ \phi} = \overrightarrow{\psi} \overrightarrow{\phi}$.
 - Let $\phi \in GA(\mathcal{E})$, and let $h(O, \lambda)$ be a central dilation. Compute $\phi \circ h(O, \lambda) \circ \phi^{-1}$.
 - Prove that two dilations with the same center commute.
 - Compute $h(B, \lambda') \circ h(A, \lambda)$.
 - Explain whether or not the set of all dilations is a subgroup of $GA(\mathcal{E})$.
 - Prove $\overrightarrow{AB} = -\overrightarrow{BA}$ for any $A, B \in \mathcal{E}$.
- In this exercise we aim to prove **Menelaus' theorem** in a couple of steps, see also Audin Exercise I.37. In an affine plane consider three distinct non-collinear points A, B, C and points A' on line BC , B' on CA and C' on AB all distinct from A, B, C . Points A', B', C' are collinear if and only if $\frac{\overrightarrow{A'B}}{A'C} \cdot \frac{\overrightarrow{B'C}}{B'A} \cdot \frac{\overrightarrow{C'A}}{C'B} = 1$.
 - Draw a picture.
 - Explain why $h_{C', \frac{\overrightarrow{C'A}}{C'B}}$ sends B to A and $h_{B', \frac{\overrightarrow{B'C}}{B'A}}$ sends A to C and $h_{A', \frac{\overrightarrow{A'B}}{A'C}}$ sends C to B .
 - The composition of the three maps above sends B to itself, why must it be of the form $h_{B, \lambda}$ for some $\lambda \in \mathbb{R}$?
 - The first two maps fix the line $C'B'$.
 - The third map fixes $C'B'$ if and only if A' is on that line.
 - Conclude that if the composition of the three is the identity then A', B', C' are collinear.
 - The point B is not on $A'C'$ because then $\overrightarrow{A'B}$ and $\overrightarrow{C'B}$ are proportional and hence the sides of the triangle BC and AB would not be independent.
 - Conversely if A', B', C' are collinear then show the composition of the three is the identity because it fixes both the line $C'B'$ and B (not on that line).
- Let V_1 and V_2 be linear subspaces of a vector space V . Under which conditions is $V_1 \cup V_2$ a linear subspace of V ? Prove your claim.
 - Let \mathcal{F}_1 and \mathcal{F}_2 be affine subspaces of an affine space \mathcal{E} , directed by E . Under which condition is $\mathcal{F}_1 \cup \mathcal{F}_2$ an affine subspace of \mathcal{E} ? Prove your claim.
- Let A be a point on \mathcal{D} be a line in an affine plane \mathcal{P} with underlying vector space P . For a one-dimensional linear subspace $L \subset P$ not equal to $\Theta_A(\mathcal{D})$ we consider the mapping $\pi : \mathcal{P} \rightarrow \mathcal{P}$ defined as follows. $\pi(M)$ is the point M' defined by $M' \in \mathcal{D}$ and $\overrightarrow{MM'} \in L$.
 - Why can there be only one point M' with properties $M' \in \mathcal{D}$ and $\overrightarrow{MM'} \in L$?
 - Prove that π is an affine map.
 - What is the linear map $\overrightarrow{\pi} : P \rightarrow P$ associated to π ?
 - Why is π not defined if L is in the same direction as \mathcal{D} ?
- All affine n -spaces are 'the same'.
 - Prove that any two vector spaces of dimension n are isomorphic (i.e.) there is a linear bijection from one to the other that has a linear inverse.
 - Suppose \mathcal{E} and \mathcal{F} are two n -dimensional affine spaces. Pick $B \in \mathcal{E}$ and $C \in \mathcal{F}$ and use Θ_B, Θ_C and the previous part to construct an affine bijection between \mathcal{E} and \mathcal{F} with affine inverse.

3 Euclidean space in general

By Euclidean space we mean affine space that also has a notion of dot-product (aka scalar or inner)-product $\langle \cdot, \cdot \rangle$. This allows us to discuss angles and distances as Euclid did but still using vector spaces as our foundation.

Definition 3.1. (Audin p.44)

A Euclidean vector space is a vector space with a choice of scalar product. A Euclidean affine space \mathcal{E} is an affine space directed by a Euclidean vector space. The distance between two points A, B is $d(A, B) = |\overrightarrow{AB}|$.

In a Euclidean vector space E the notion of perpendicular is very important. For any subset $S \subset E$ we define $S^\perp = \{x \in E : \forall s \in S, \langle s, x \rangle = 0\}$. When S is a subspace of E then $E = S \oplus S^\perp$. Recall that $A = B \oplus C$ means every element a of vector space A can uniquely be written as $a = b + c$ for some b in subspace $B \subset A$ and c in subspace $C \subset A$.

3.1 Euclidean isometries

The motions that preserve the geometry are fundamental so we will spend some time defining and studying them before moving on to more classical theorems in geometry.

Definition 3.2. (Audin p.44)

A linear isometry is a linear map between Euclidean vector spaces that preserves the scalar product. An affine isometry is an affine map between Euclidean affine spaces if it preserves the distance between any pair of points. The set of all linear isometries of E is $O(E)$. The set of all affine isometries of \mathcal{E} is $Isom(\mathcal{E})$.

Examples of isometries include

1. Translations are affine isometries.
2. Dilations $h_{O,\lambda}$ are affine isometries if and only if $\lambda = \pm 1$.
3. Orthogonal linear isometries are the following. For a linear subspace $F \subset E$ define $E \xrightarrow{s_F} E$ by $s_F(x+y) = x - y$ where $x \in F$ and $y \in F^\perp$.
4. Orthogonal affine isometries are defined similarly. If $\mathcal{F} \subset \mathcal{E}$ is an affine subspace then $\mathcal{E} \xrightarrow{\sigma_{\mathcal{F}}} \mathcal{E}$ is defined by choosing any $O \in \mathcal{F}$ and setting $F = \Theta_O(\mathcal{F})$ and $\overrightarrow{\sigma_{\mathcal{F}}(M)} = s_F(\overrightarrow{OM})$ for any $M \in \mathcal{E}$.
5. In case the subspace is a hyperplane (so the dimension is one less than the whole space) we say the orthogonal isometry is a reflection.

When $v \in E$ is a vector in Euclidean space E there is a simple formula for the reflection s_{v^\perp} in the hyperplane orthogonal to v . Namely $s_{v^\perp} w = w - 2 \frac{\langle v, w \rangle}{|v|^2} v$. This formula is correct since it fixes any $w \in v^\perp$ and sends v to $-v$. Notice that s_{v^\perp} is its own inverse.

Theorem 3.1. Suppose Euclidean vector space E is n -dimensional and $g \in O(E)$. If there is a k dimensional subspace F such that $\forall f \in F : g(f) = f$ then g is the composition of at most $n - k$ reflections.

Proof. We argue by induction on $n - k$. When $k = n$ the g must be the identity. Suppose g fixes a $k - 1$ dimensional subspace F and suppose $g(v) = w \neq v$ for some $v \in E$. We will show that reflection in u^\perp where $u = v - w$ fixes all vectors in F and sends w to v . Since g is an isometry $|v| = |w|$ we have $-2\langle u, w \rangle = |u|^2$ so $s_{u^\perp} w = w - 2 \frac{\langle u, w \rangle}{|u|^2} u = w + v - w = v$. Indeed for any $f \in F$ we have $\langle f, v \rangle = \langle g(f), g(v) \rangle = \langle f, w \rangle$ so $\langle f, u \rangle = 0$ meaning s_{u^\perp} fixes f .

We conclude that the composition $s_{u^\perp} \circ g$ fixes vectors in F and g sends v to w and s_{u^\perp} maps it back to v so the k -dimensional subspace spanned by both F and v . By the induction hypothesis $s_{u^\perp} \circ g$ is a composition of at most $n - k$ reflections so composing by s_{u^\perp} we see g is obtained from at most $n - k + 1$ reflections. \square

Lemma 3.1. (Audin exercise II.7)

Given two points $A, A' \in \mathcal{E}$ there exists an affine hyperplane $\mathcal{H} \subset \mathcal{E}$ such that $\sigma_{\mathcal{H}}(A) = A'$.

Proof. Define $\mathcal{H} = \{B \in \mathcal{E} : |\overrightarrow{AB}| = |\overrightarrow{A'B}|\}$. Then \mathcal{H} contains $O = t_{\overrightarrow{AA'}/2} A$ since $\overrightarrow{OA} = -\overrightarrow{AA'}/2$ and $\overrightarrow{OA'} = \overrightarrow{OA} + \overrightarrow{AA'} = \overrightarrow{AA'}/2$. Next \mathcal{H} is an affine hyperplane since $\Theta_O(\mathcal{H}) = \overrightarrow{AA'}^\perp$. To see why notice that for $B \in \mathcal{H}$ we have $|\overrightarrow{AB}| = |\overrightarrow{A'B}|$ so $|\overrightarrow{OA} - \overrightarrow{OB}|^2 = |\overrightarrow{OA'} - \overrightarrow{OB}|^2$ meaning $\langle \overrightarrow{OA}, \overrightarrow{OB} \rangle = \langle \overrightarrow{OA'}, \overrightarrow{OB} \rangle$ so $\overrightarrow{OB} \perp \overrightarrow{OA'} - \overrightarrow{OA} = \overrightarrow{AA'}$. Conversely if $\overrightarrow{OB} \perp \overrightarrow{AA'}$ then $\langle \overrightarrow{OB}, \overrightarrow{OA} \rangle = \langle \overrightarrow{OB}, \overrightarrow{OA'} \rangle$ so $|\overrightarrow{AB}| = |\overrightarrow{A'B}|$ meaning $B \in \mathcal{H}$.

Finally if $\sigma_{\mathcal{H}}(A) = A''$ then $\overrightarrow{OA''} = s_{\overrightarrow{AA'}^\perp}(\overrightarrow{OA}) = s_{(\overrightarrow{OA'} - \overrightarrow{OA})^\perp}(\overrightarrow{OA}) = \overrightarrow{OA'}$ by the calculation in the proof of Theorem 3.1. Therefore $A'' = A'$. \square

Theorem 3.2. (Audin thm 2.2, p.49)

Suppose \mathcal{E} is an n dimensional Euclidean affine space. Any element of $\text{Isom}(\mathcal{E})$ can be written as the composition of at most $n + 1$ reflections.

Proof. If the affine isometry ψ has a fixed point O then there is a linear isometry $L = \Theta_O \circ \psi \circ \Theta_O^{-1}$ which by Theorem 3.1 is the composition of at most n reflections through hyperplanes containing O . This means that the same is true for ψ because we may conjugate each of these reflections by Θ_O . More precisely if $L = s_{H_1} \circ \dots \circ s_{H_j}$ then $\psi = \Theta_O^{-1} \circ L \circ \Theta_O = \Theta_O^{-1} \circ s_{H_1} \circ \Theta_O \circ \Theta_O^{-1} \circ \dots \circ \Theta_O^{-1} \circ s_{H_j} \circ \Theta_O$ and each $\Theta_O^{-1} \circ s_{H_i} \circ \Theta_O$ is a reflection in an affine hyperplane through O .

In case ψ does not fix any point then choose $A \in \mathcal{E}$ and set $A' = \psi A$. Using Lemma 3.1 we see that $\sigma_{\mathcal{H}} \circ \psi$ fixes A so we conclude ψ is the composition of at most $n + 1$ reflections. \square

Linear reflections reverse orientation in the sense that they have determinant -1 . Since the determinant of the product is the product of the determinants this means that linear isometries have determinant ± 1 . The linear isometries with determinant 1 are known as the rigid motions. The same terminology is used for the affine isometries. We use the notation $\text{Isom}^+(\mathcal{E})$ for the rigid affine motions and $O^+(E)$ or $SO(E)$ for the linear rigid motions.

Choosing an orthonormal basis of Euclidean n -dimensional space E we may identify $O(E)$ with a subset $O(n) \subset M_n$ inside the set M_n of the $n \times n$ matrices. Recall the columns of the matrix of a linear map are the images of the base vectors and a linear isometry preserves the inner product. Therefore the columns of the matrix with respect to an orthonormal basis must again form an orthonormal basis. In other words $O(n) = \{A \in M_n : AA^t = I\}$, where A^t is the transpose of matrix A . Identifying M_n with \mathbb{R}^{n^2} we see that $A \mapsto AA^t$ is a continuous map and $O(n)$ is the inverse image of a point so $O(n)$ must be a closed subset of M_n with respect to this topology. Also $O(n)$ is bounded since all the columns must be unit vectors so all in all we concluded that $O(n)$ is a compact subset of M_n .

More precisely the determinant shows that $O(n)$ decomposes neatly into two equal parts, the matrices with determinant 1 called $O^+(n)$ and those with determinant -1 .

The two dimensional case is fundamental so we examine $O^+(2)$ in more detail. It will be shown to be identified with the unit circle U in the complex plane.

Lemma 3.2. (Audin prop 3.4, p. 53)

The group $O^+(2)$ is isomorphic and homeomorphic to the complex units $U \subset \mathbb{C}$.

Proof. Matrix $A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ belongs to $O(2)$ if and only if the columns form an orthonormal basis. So $a^2 + b^2 = 1$ and $c^2 + d^2 = 1$ and $ac + bd = 0$. This means that $c = -eb$ and $d = \epsilon a$ for some $\epsilon \in \{-1, 1\}$. In fact $\epsilon = \det A$. The map $\varphi : O^+(2) \rightarrow U$ given by $\varphi \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a + ib$ is a homomorphism of groups and it is also a continuous bijection with continuous inverse as the reader can check by explicit calculations, finishing the proof. \square

We thus find a surjective map $\mathbb{R} \rightarrow U \rightarrow O^+(2)$ sending $\theta \in \mathbb{R}$ to $e^{i\theta}$. The image in $O^+(2)$ is called the rotation with angle θ . In section 4 we will have more to say about angles and plane geometry.

3.2 Centroid, Orthocenter and Circumcenter

We start with generalizing triangles to any dimension. These are known as simplices.

Definition 3.3. Given affine independent points A_0, \dots, A_n in affine space \mathcal{E} , the n -simplex $[A_0, \dots, A_n]$ is defined as

$$[A_0, \dots, A_n] = \{B \in \mathcal{E} \mid \exists \beta_0, \dots, \beta_n \in [0, 1] : \sum_{i=0}^n \beta_i \overrightarrow{BA_i} = 0\}$$

A 0-simplex is a point, a 1-simplex is simply an interval and a 2-simplex is a triangle. 3-simplices are known as tetrahedra. The face of simplex $[A_0, \dots, A_n]$ opposite to A_i is the $(n-1)$ -simplex $[A_0, \dots, \hat{A}_i, \dots, A_n]$ where the hat means we removed A_i from the sequence. The faces of a 2-simplex are simply its sides viewed as segments.

The numbers β_i appearing in the definition of simplex are known as barycentric coordinates.

Definition 3.4. Numbers $\beta_0, \dots, \beta_n \in \mathbb{R}$ are **barycentric coordinates** of point B with respect to an affine frame A_0, \dots, A_n if $\sum_{i=0}^n \beta_i \overrightarrow{BA_i} = 0$.

Barycentric coordinates are not unique but they are unique when we add the condition that $\sum_{i=0}^n \beta_i = 1$. This condition can always be met because affine independence of the A_i implies $\sum_{i=0}^n \beta_i \neq 0$. The n -simplex can be defined as those points with barycentric coordinates in $[0, 1]^{n+1}$.

What is the middle of a simplex? There are several competing notions and we will survey a few of them: centroid, circumcenter and orthocenter. For 2-simplices (triangles) we will investigate how they relate to each other.

Looking at the definition of simplex we defined the **centroid** G of the simplex as follows. $G \in \mathcal{E}$ is the unique point such that $\sum_{i=0}^n \frac{1}{n+1} \overrightarrow{GA_i} = 0$.

The faces of the simplex also have centroids called G_i and the lines $G_i A_i$ all meet in G . To see why this is the case, notice that the segment $A_i G_i$ is precisely the set of points with barycentric coordinates $\beta_j = \begin{cases} \frac{1-t}{n} & j \neq i \\ t & j = i \end{cases}$ for $t \in [0, 1]$. Since G has barycentric coordinates $\beta_k = \frac{1}{n+1}$ this is the case $t = \frac{1}{n+1}$. In fact $n \overrightarrow{GG_i} = \overrightarrow{A_i G}$ (exercise).

A competing notion of the middle of a simplex is known as the **circumcenter** O . It is the center of the unique sphere that passes through all vertices of the simplex. First we should define the $(n-1)$ -**sphere** $S^{n-1}(c, r)$ with center $c \in \mathcal{E}$ and radius $r \geq 0$ in affine n -space \mathcal{E} to be $S^{n-1}(c, r) = \{A \in \mathcal{E} : |c\overrightarrow{A}| = r\}$. The following lemma shows that the circumcenter is uniquely defined by the simplex.

Lemma 3.3. *For an n -simplex $[A_0, \dots, A_n]$ in affine Euclidean n -space \mathcal{E} there exists a unique $(n-1)$ -sphere $S^{n-1}(O, r)$ such that $\forall i : A_i \in S^{n-1}(O, r)$.*

Proof. We argue by induction on the dimension n . The base case $n = 0$ is clear so now assume the statement is true in all dimensions $< n$. The face of our simplex opposite to A_n spans an affine hyperplane \mathcal{H}_n in which there is a unique $(n-1)$ -sphere with center O_n passes through all points A_i with $i < n$. By Lemma 3.1 there exists a unique hyperplane \mathcal{F} in \mathcal{E} equidistant from A_0 and A_n and also a line \mathcal{L} orthogonal to \mathcal{H}_n passing through O_n . Since A_n is affine independent from the other A_i there must be a (unique) intersection point $O = \mathcal{L} \cap \mathcal{F}$. We claim that $S^{n-1}(O, |\overrightarrow{OA_n}|)$ passes through all points A_i of the simplex. Indeed orthogonal projection on \mathcal{H}_n shows $|\overrightarrow{OA_0}| = |\overrightarrow{OA_i}|$ with $i < n$ and by construction $|\overrightarrow{OA_n}| = |\overrightarrow{OA_0}|$. \square

Finally the **altitudes** of a simplex are the affine lines \mathcal{H}_i through A_i orthogonal to the faces opposite A_i . Interestingly the altitudes do not necessarily meet in a single point when $n > 2$. For example the 3-simplex $[(1, 0, 0), (0, 0, 1), (1, 1, 0), (0, 0, 0)]$ has non-intersecting altitudes \mathcal{H}_2 and \mathcal{H}_1 as is easily seen by inscribing this tetrahedron inside a regular cube.

For $n = 2$ the altitudes do meet in a single point called the **orthocenter** H of the triangle $\mathcal{A} = [A_0, A_1, A_2]$. To see why this is the case consider the triangle $\mathcal{B} = [B_0, B_1, B_2]$ defined by $h_{G, -2}(A_i) = B_i$ where G is the centroid of \mathcal{A} . Notice that $A_i A_j$ is parallel to $B_i B_j$ by Lemma 2.4. Also when i, j, k are distinct A_i is the midpoint of the segment $B_j B_k$. This is because the midpoint of segment $A_j A_k$ is G_i and $\overrightarrow{B_j A_i} = \overrightarrow{GA_i} - \overrightarrow{GB_j} = -2\overrightarrow{GG_i} + 2\overrightarrow{GA_j} = 2\overrightarrow{G_i A_j}$ and likewise $\overrightarrow{B_k A_i} = 2\overrightarrow{G_i A_k}$. Finally the altitudes of \mathcal{A} are precisely the perpendicular bisectors of the sides of \mathcal{B} . The latter meet in the circumcenter of \mathcal{B} so this point is the orthocenter of \mathcal{A} .

Looking carefully at what we just proved we find the following theorem of Euler:

Theorem 3.3. *(Euler line)*

The circumcenter, centroid and orthocenter O, G, H of triangle $\mathcal{B} = [B_0, B_1, B_2]$ lie on a line and more precisely $-2\overrightarrow{GO} = \overrightarrow{GH}$. Also on this line is center N of the circle through the midpoints G_i of the sides of our triangle. Moreover, N is the midpoint of the segment HO .

Proof. As before we use two triangles \mathcal{A}, \mathcal{B} , where $\mathcal{B} = h_{G, -2}\mathcal{A}$. By construction they have the same centroid G and $A_i = G_i$. Moreover, we showed the circumcenter $O = O_{\mathcal{B}}$ of \mathcal{B} is the orthocenter $H_{\mathcal{A}}$ of \mathcal{A} . By definition this means the orthocenter $H = H_{\mathcal{B}}$ of \mathcal{B} is $h_{G, -2}H_{\mathcal{A}} = h_{G, -2}O$. We conclude that O, H, G lie on a line and moreover $\overrightarrow{GH} = h_{G, -2}\overrightarrow{GO} = -2\overrightarrow{GO}$.

Notice N is the circumcenter $O_{\mathcal{A}}$ of \mathcal{A} and so $h_{G, -2}O_{\mathcal{A}} = O_{\mathcal{B}} = O$. Therefore O, G, N, H lie on the same line and $\overrightarrow{GO} = h_{G, -2}\overrightarrow{GN} = -2\overrightarrow{GN}$. This means that $\overrightarrow{NO} = \overrightarrow{NG} + \overrightarrow{GO} = \frac{3}{2}\overrightarrow{GO} = 2\overrightarrow{GO} + \overrightarrow{GN} = \overrightarrow{HG} + \overrightarrow{GN} = \overrightarrow{HN}$. \square

By similar methods much more can be said about the circle through the midpoints of the sides of our triangle. It is the famous nine-point circle.

Theorem 3.4. *(Nine point circle)*

Given triangle $\mathcal{B} = [B_0, B_1, B_2]$, there is a circle that passes through the following nine special points: the midpoints G_i of the sides and intersections H_i of the altitude \mathcal{H}_i with the opposite side and the midpoints of the segments $B_i H$.

Proof. Keeping the previous notation we call our triangle \mathcal{B} and define the nine-point circle \mathcal{C} as the circle through the $G_i = A_i$. There must be a unique such circle by Lemma 3.3, also it has center N . We now need to show that \mathcal{C} contains both the H_i and the midpoints M_i of the segments $B_i H$.

First $h_{G, -\frac{1}{2}} \circ h_{H, 2} = h_{N, -1}$ because the left hand side has fixed point N by the previous theorem. Also looking at the left hand side we see that $h_{N, -1}(M_i) = A_i$. Since $h_{N, -1}$ is an isometry we have $|NM_i| = |NA_i|$ so the M_i are on \mathcal{C} .

Second H_i and G_i are both on line $B_j B_k$ for i, j, k distinct. To show that $|NH_i| = |NG_i|$ we argue as follows. The lines HH_i and OG_i are parallel because both perpendicular to $B_j B_k$. Since by the previous theorem N is the midpoint of segment HO , Lemma 2.4 shows NN_i is also perpendicular to $B_j B_k$ where N_i is the midpoint of segment $H_i G_i$. It follows by a reflection in line NN_i that $|NH_i| = |NG_i|$ as desired. \square

There are in fact many more interesting points on the nine-point circle. For example Feuerbach found there are precisely four circles that are tangent to all three sides of our triangle and the nine-point circle is tangent to all four of those!

Exercises

- Is it true that for any subset S in a Euclidean vector space E we have $(S^\perp)^\perp$?
- Imagine a line \mathcal{L} in affine Euclidean plane \mathcal{E} and a vector $u \in L$ in the direction (underlying vector subspace) of \mathcal{L} . The affine map $\gamma_{\mathcal{L}, u} : \mathcal{E} \rightarrow \mathcal{E}$ defined by $\gamma_{\mathcal{L}, u} = t_u \circ \sigma_{\mathcal{L}}$ is called a glide reflection.
 - Prove that the glide reflection $\gamma_{\mathcal{L}, u}$ is an affine isometry.
 - Why can $\gamma_{\mathcal{L}, u}$ not be written as the composition of two or less reflections? (Hint: fixed points)
 - Write $\gamma_{\mathcal{L}, u}$ explicitly as a composition of three reflections.
- Verify that $s_F \circ s_G$ is the identity. Give an example of linear subspaces $F, G \subset E$ such that $s_F \circ s_G \neq s_G \circ s_F$.
- If we choose a basis $O(E)$ can be identified with a subset of \mathbb{R}^{n^2} where $n = \dim(E)$. Show that if f is the composition of an even number of reflections then there is a continuous function $\gamma : [0, 1] \rightarrow O(E)$ with $\gamma(0) = id_E$ and $\gamma(1) = f$.
- Imagine a 2-simplex $[A_0, A_1, A_2]$ with centroid G and G_i the centroid of the face opposite to A_i . Using $\overrightarrow{GA_0} + \overrightarrow{GA_1} + \overrightarrow{GA_2} = 0 = \overrightarrow{G_0 A_1} + \overrightarrow{G_0 A_2}$, prove that $2\overrightarrow{GG_0} = \overrightarrow{A_0 G}$.
- Prove that for any nonzero $\lambda \in \mathbb{R}$ and $O \in \mathcal{E}$ the dilation $h_{O, \lambda}$ sends any n -simplex in \mathcal{E} to another n -simplex in \mathcal{E} .
- By a rotation around point A in affine Euclidean plane \mathcal{E} we mean any composition of two affine reflections in lines passing through A . For $A, B \in \mathcal{E}$ let ρ_A be a rotation around A and ρ_B a rotation around $B \neq A$.
 - What is the determinant of the linear map $\overrightarrow{\rho_A} : E \rightarrow E$ associated to ρ_A ?
 - Prove that $\overrightarrow{\rho_A \circ \rho_B} = \overrightarrow{\rho_A} \circ \overrightarrow{\rho_B}$.
 - What is the determinant of $\overrightarrow{\rho_A \circ \rho_B}$?
 - Is it true that $\rho_A \circ \rho_B$ must also be a rotation around some point $C \in \mathcal{E}$? Prove or provide a counterexample.

4 Plane Euclidean geometry

Throughout this section \mathcal{E} denotes a two-dimensional Euclidean affine space with underlying (2d) vector space E .

We begin our study of the plane with a more detailed discussion of the isometries, starting with E . From Theorem 3.1 we know that every linear isometry is either the identity, a reflection in a line through the origin or a composition of two such reflections. The latter are usually called rotations and they will play an important role in what follows.

Definition 4.1. A **rotation** is an element of $O(E)$ that can be written as the composition of two reflections.

In fact the set $O^+(E)$ of linear isometries of positive determinant coincides with the set of all rotations. It forms a commutative subgroup isomorphic to $O^+(2)$ or the unit circle in the complex plane, see Lemma 3.2. Notice that the identity is also a rotation according to this definition. As we mentioned before, reflections reverse orientation while rotations preserve it. Orientation is an important and often overlooked concept so we will provide a definition.

Definition 4.2. (*Orientation*)

Recall a basis for a vector space is an ordered sequence of independent vectors that span V . An **orientation** of a vector space V is a choice of an equivalence class of bases of V . Here two bases of V are said to be equivalent if the linear map sending one to the other has positive determinant.

For $V = \mathbb{R}^2$ the standard choice of orientation is to take the equivalence class of the standard basis (e_1, e_2) . It often appears as saying that the positive direction of rotation is counter-clockwise. In \mathbb{R}^3 orientation is often referred to as the 'right hand rule' but we will deal with space later.

As promised reflections do reverse orientation in the following sense. The reflection $s = s_{v^\perp}$ is easy to describe in the basis v, v' where v is orthogonal to v' . By definition $s(v) = -v$ while $s(v') = v'$ so the matrix with respect to this basis is $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ so the determinant is -1 . To connect with the above definition we would say that the bases (v, v') and $(s(v), s(v'))$ belong to different equivalence classes, i.e. orientations.

Since there are only two orientations for E and a single reflection reverses the orientation, all rotations preserve orientation. This is why they play such an important role.

Closely tied to the concept of rotation is the notion of angle. But what exactly *is* an angle? This is not as straightforward as it seems. In fact we will define four related notions of angle:

1. The oriented angle $\angle(u, v)$ between two vectors in $u, v \in E$.
2. The oriented angle of lines where we identify $\angle(u, v)$ with $\angle(u, -v)$.
3. The geometric angle which identifies $\angle(u, v)$ and $\angle(v, u)$.
4. Given an orientation of E , the measure of $\angle(u, v)$, a real number mod 2π .

What corresponds to the 'usual' notion of angles as real numbers modulo 2π is the measure of oriented angle. However it is NOT well defined in our Euclidean E vector space. The common definition of an angle θ between vectors u, v using the inner product as $\cos \theta = \frac{\langle u, v \rangle}{|u||v|}$ is also incomplete. The right hand side is symmetric in u, v so without additional assumptions it cannot possibly distinguish between θ and $-\theta$. Do we go from u to v or vice versa and which way do we go?

The issue is that there is no natural choice for orientation of E . Of course we can simply choose an arbitrary orientation but this muddles the theory and clutters the proofs. Much like affine space arises from our refusal to make an arbitrary choice of origin.

4.1 Angles

In this subsection we define four notions of angle the most important of which is called the oriented angle between two vectors. It is firmly rooted in our understanding of rotations.

Lemma 4.1. (*Audin prop 1.1, p.65*)

For any (ordered) pair of unit vectors in $(u, u') \in E$ there is a unique rotation that sends u to u' .

Proof. Suppose u, u' are unit vectors and choose a unit vector v such that u, v form an orthonormal basis. Then $u' = au + bv$ for some $a, b \in \mathbb{R}$ with $a^2 + b^2 = 1$. With respect to the basis u, v any orthonormal linear map sending u to u' and determinant 1 has matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, see also Lemma 3.2. \square

Our oriented angles now come out by considering pairs of vectors, modulo rotations.

Definition 4.3. (*Audin p.66*)

Define $\hat{\mathcal{A}}$ to be the set of pairs of unit vectors (u, v) in E . Two pairs $(p_1, p_2), (q_1, q_2) \in \hat{\mathcal{A}}$ are said to be equivalent if there exists a rotation that sends both p_1 to q_1 and p_2 to q_2 . The equivalence classes are called **oriented angles** and the set of oriented angles is denoted \mathcal{A} . The class of (u, v) is denoted² $\angle(u, v)$ and is called the oriented angle between u and v . The **flat angle** is $\angle(v, -v)$ for any unit vector v .

This definition is extended to any pair of non-zero vectors u', v' where we set $\angle(u', v') = \angle\left(\frac{u'}{|u'|}, \frac{v'}{|v'|}\right)$.

To clarify the relation between angles and rotations further we define a map from oriented angles to rotations $O^+(E)$ as follows.

Lemma 4.2. (*Audin lemma 1.2, p.66*)

The map $\hat{\Phi} : \hat{\mathcal{A}} \rightarrow O^+(E)$ is defined by setting $\hat{\Phi}(u, v)$ to be the unique rotation sending u to v . It defines a bijection $\Phi : \mathcal{A} \rightarrow O^+(E)$ by sending the class of $\angle(u, v)$ to $\hat{\Phi}(u, v)$.

¹Remember to write down the matrix of a linear transformation one just lists the images of the base vectors as columns of this matrix.

²Audin uses just (u, v) to denote both the pair and the oriented angle.

Proof. To show $\hat{\Phi}$ is surjective choose any $f \in O^+(E)$ and unit vector $u \in E$. We have $\hat{\Phi}(u, f(u)) = f$ since both sides send u to $f(u)$ and there is only one such rotation by Lemma 4.1.

Next we need to check that Φ is well-defined by showing that if (p_1, p_2) is equivalent in the sense of Definition 4.3 to (q_1, q_2) then $\hat{\Phi}$ maps them to the same rotation in $O^+(E)$. By Lemma 4.1 there is a unique rotation g taking p_1 to q_1 and by assumption it also takes p_2 to q_2 . Suppose $f = \hat{\Phi}(p_1, p_2)$ so $f(p_1) = p_2$ then plane rotations commute so $f(q_1) = f \circ g(p_1) = g \circ f(p_1) = g(p_2) = q_2$, showing $f = \hat{\Phi}(q_1, q_2)$.

Finally we should show that if $\hat{\Phi}(p_1, p_2) = \hat{\Phi}(q_1, q_2) = f$ then there is a single rotation that sends both p_1 to q_1 and p_2 to q_2 . By Lemma 4.1 there is a unique rotation g sending p_1 to q_1 . By assumption $f \circ g \circ f^{-1}$ sends p_2 to p_1 to q_1 to q_2 . Since the rotations form a commutative group $f \circ g \circ f^{-1} = g$. \square

The map Φ is bijective so with any rotation we may identify an oriented angle in \mathcal{A} . We showed there is a unique rotation that is the image of $\angle(u, v)$ so it rotates unit vector u to v . We call this the rotation over angle $\angle(u, v)$ and to any rotation there corresponds such an angle.

Moreover since the map Φ is a bijection we may use it to 'add' angles by composing their images in the group $O^+(E)$. More precisely we set

$$\angle(u, v) + \angle(w, z) = \Phi^{-1}(\Phi(\angle(u, v)) \circ \Phi(\angle(w, z)))$$

Unpacking the definitions this says that the sum of the angles is an angle $\angle(a, b)$ where $b = f(a)$ and the rotation f is the result of first rotating u to v and then w to z . At the cost of changing z we may always choose $v = w$ and then this kind of addition is just placing the angles next to each other (draw a picture!). This gives the important special case

$$\angle(u, v) + \angle(v, z) = \angle(u, z)$$

Now let us turn to measuring angles with respect to an orientation. This depends on the following lemma.

Lemma 4.3. *Suppose (p_1, p_2) and (q_1, q_2) are two orthonormal bases of E in the same orientation \mathcal{O} and $r \in O^+(E)$ is a rotation. The matrix of r with respect to (p_1, p_2) is the same as that with respect to (q_1, q_2) . The orientation \mathcal{O} thus determines an isomorphism $F_{\mathcal{O}} : O^+(E) \rightarrow O^+(2)$.*

Proof. The linear map ρ sending basis p to basis q must be in $O^+(E)$ since both bases are orthonormal and in the same orientation. Now rotations commute so if $rp_1 = ap_1 + bp_2$ and $rp_2 = cp_1 + dp_2$ then also $rq_1 = r\rho p_1 = \rho r p_1 = \rho(ap_1 + bp_2) = aq_1 + bq_2$ and similarly $rq_2 = r\rho p_2 = \rho r p_2 = \rho(cp_1 + dp_2) = cq_1 + dq_2$. The isomorphism $F_{\mathcal{O}}$ comes from choosing any basis in \mathcal{O} . This is an isomorphism by Lemma 3.2. \square

Since the matrix of a rotation only depends on the orientation of the basis we can use it to define our measure of angle.

Definition 4.4. *(Measure of an oriented angle)*

*Given an orientation \mathcal{O} of E and $a \in \mathcal{A}$. The matrix $F_{\mathcal{O}}(\Phi a) \in O^+(2)$ is of the form $\begin{pmatrix} \cos \mu & \sin \mu \\ -\sin \mu & \cos \mu \end{pmatrix}$. This determines a number $\mu = \mu(a) \in \mathbb{R}$ modulo 2π and we call it the **measure** $\mu(a)$ of oriented angle a .*

It should be clear from this definition that the measure of the angle really does depend on the choice of orientation. Also the definition agrees with $\cos \mu = \langle u, v \rangle$ where $\mu = \mu \angle(u, v)$ for unit vectors u, v . The measure of the flat angle is $\pi \pmod{2\pi}$. Since $F_{\mathcal{O}}$ is an isomorphism we have $\mu(a) + \mu(b) = \mu(a + b) \pmod{2\pi}$ as implied by the word measure.

Returning to the oriented angles $\angle(u, v)$ we next define the angle between two lines and the geometric angle between two vectors. Neither depends on a choice of orientation.

Definition 4.5. *(Angle between lines and geometric angle)*

*The **angle between two lines** spanned by u and v in E is defined to be $\angle(u, v) \pmod{\angle(u, -u)}$*

*The **geometric angle** between two vectors $u, v \in E$ is the equivalence class of $\angle(u, v)$ under the equivalence relation where we identify $\angle(x, y)$ with $\angle(y, x)$ for any $x, y \in E$.*

One should check that the definition of the angle between lines does not depend on the choice of spanning vectors u and v . To this end recall that $\angle(u, -u)$ represents the flat angle and flipping the sign of u or v amounts to adding or subtracting such a flat angle.

The point of defining the notion of geometric angle is that this is what is actually invariant under all isometries. By definition a rotation $\rho \in O^+(E)$ sends preserves oriented angles in the sense that $\angle(u, v) = \angle(\rho u, \rho v)$. However a reflection s does not.

Lemma 4.4. *(Audin prop 1.10, p. 70)*

For any unit vectors $u, v \in E$ and a reflection s we have $\angle(su, sv) = \angle(v, u)$.

Proof. The reflection $s' = s_{(u-v)^\perp}$ sends u to v and v to u . The composition $s \circ s'$ is a rotation so $\angle(v, u) = \angle(s \circ s'(v), s \circ s'(u)) = \angle(su, sv)$. \square

It follows from this discussion that geometric angles are preserved by all isometries, both rotations and reflections.

So far we considered angles in a two-dimensional Euclidean vector space E . These can be carried over to the affine Euclidean plane \mathcal{E} as follows.

Definition 4.6. (*Angles in the affine Euclidean plane*)

For points $O, A, B \in \mathcal{E}$ with $O \neq A$ and $O \neq B$, the oriented angle between the segments (1-simplices) $[O, A]$ and $[O, B]$ is $\angle(OAB)$, defined as $\angle OAB = \angle(\overrightarrow{OA}, \overrightarrow{OB})$.

For a pair of lines $\mathcal{L}_1, \mathcal{L}_2$ intersecting at O the angle between them is the angle between the lines $\Theta_O(\mathcal{L}_i)$ in E . The geometric angle and the measure of the angle $\angle OAB$ are defined as explained above.

4.2 Plane constructions involving angles

To illustrate what we just learned about angles let us prove a few elementary facts involving angles.

Lemma 4.5. (*Angle sum, Audin prop 1.14 p.73*)

If A, B, C are distinct points in an affine Euclidean plane then $\angle ABC + \angle BCA + \angle CAB$ is the flat angle and so is $\angle ACB + \angle BAC + \angle CBA$.

Proof. The rotation $h = h_{C, -1}$ sends \overrightarrow{CA} to \overrightarrow{AC} and \overrightarrow{CB} to \overrightarrow{BC} so $\angle(\overrightarrow{CA}, \overrightarrow{CB}) = \angle(h\overrightarrow{CA}, h\overrightarrow{CB}) = \angle(\overrightarrow{AC}, \overrightarrow{BC})$. Now $\angle(x, y) + \angle(y, z) = \angle(x, z)$ so

$$\angle(\overrightarrow{AB}, \overrightarrow{AC}) + \angle(\overrightarrow{BC}, \overrightarrow{BA}) + \angle(\overrightarrow{CA}, \overrightarrow{CB}) = \angle(\overrightarrow{AB}, \overrightarrow{AC}) + \angle(\overrightarrow{BC}, \overrightarrow{BA}) + \angle(\overrightarrow{AC}, \overrightarrow{BC}) = \angle(\overrightarrow{AB}, \overrightarrow{BA})$$

which is the flat angle. The second statement follows from $\angle(x, y) = -\angle(y, x)$. \square

Notice the cyclic permutation of the order of the vertices A, B, C in the angle sum: ABC, BCA, CAB .

In the presence of an orientation this becomes the familiar theorem about the angle sum in a triangle. If A, B, C are distinct points in an oriented affine Euclidean plane then $\theta_{\angle ABC} + \theta_{\angle BCA} + \theta_{\angle CAB} = \pi \pmod{2\pi}$. This follows from the previous statement since the measure of the sum is the sum of the measures modulo 2π .

More interestingly we have the classic proposition about the angles in the circumference.

Theorem 4.1. (*Audin prop 1.17, p.74*)

If A, B, C are distinct points in an affine Euclidean plane and O is the circumcenter of the triangle $[A, B, C]$ then

$$\angle OAB = 2\angle CAB$$

Proof. O is the center of the circle through A, C so it is fixed by the reflection in the perpendicular bisector \mathcal{H} of segment $[A, C]$. This is the affine line of all points of equal distance to both A and C . Since $\sigma_{\mathcal{H}}$ reverses orientation and permutes A and C we have $\angle COA = \angle ACO$. Lemma 4.5 tells us that $\angle COA + \angle ACO + \angle OAC = 2\angle COA + \angle OAC$ is flat. In the same way $2\angle CBO + \angle OCB$ is also flat. Adding two flat angles gives the angle 0 so

$$0 = 2\angle COA + \angle OAC + 2\angle CBO + \angle OCB = 2\angle CBA + \angle OAB$$

Since $\angle(x, y) = -\angle(y, x)$ we are done. \square

Returning briefly to our discussion of centers of triangles, for each pair of lines there is a pair of bisectors of the two angles defined by the lines. If the lines are directed by unit vectors u, v then these bisectors are directed by $u + v$ and $u - v$. They are perpendicular and cut the angles in two as reflection in either of the bisectors permutes the original two lines.

One can prove that given a triangle the six bisectors of the lines generated by the sides of the triangle meet in triples to define four points. These points are the centers of four circles tangent to the extended sides of the triangle. One of them is inside the circle and is called the inscribed circle. As was mentioned earlier these circles are all tangent to the nine-point circle, making it effectively a thirteen-point circle. Curiously, if instead we trisect each angle then the six trisectors inside the triangle meet in adjacent pairs to define a regular triangle. This is known as Morley's theorem.

Finally let us mention an interesting map that is not an isometry but as we will see later it reverses angles as if it were an ordinary reflection. In some sense inversions are reflections in a circular mirror.

Definition 4.7. (*Audin def 4.1, p.84*)

For $\lambda \in \mathbb{R}$ and O a point in affine Euclidean plane \mathcal{E} , consider the inversion map $I_{O, \lambda} : \mathcal{E} - \{O\} \rightarrow \mathcal{E} - \{O\}$ defined by $I_{O, \lambda}(M) = M'$ where M' is the point on line OM such that $\overrightarrow{OM'} = \frac{\lambda}{|\overrightarrow{OM}|^2} \overrightarrow{OM}$.

5 Three dimensional Euclidean geometry

In this section we briefly explore three dimensional Euclidean geometry. Our main focus is the classification of the regular polyhedra, also known as the Platonic solids. Traditionally they are a centerpiece of mathematics, for example Euclid's elements concludes with constructions of the polyhedra and an argument that there can only be five such. Even though the Platonic solids do not seem to have much significance in themselves, the tools needed to define, classify and construct them nicely illustrate this kind of geometry.

A new feature in dimension three is that in addition to usual angles between planes one now also has a notion of 'solid angle' or a piece of a sphere cut out by a cone. As such we are naturally led to consider the geometry of the sphere itself.

5.1 Rotations revisited

As in the plane we define a rotation in general to be a composition of two reflections in hyperplanes through the origin. According to Theorem 3.1 every linear isometry of 3-dimensional Euclidean space E is a composition of 0, 1, 2 or 3 reflections in planes through the origin. Looking at the determinant it is still true that $O^+(E)$ consists of only rotations (including the identity).

In three dimensions the planes P, P' we are reflecting in generally meet in a line L and points on this line are left fixed by the rotation $r = s_P \circ s_{P'}$. The plane L^\perp is also sent to itself (but not pointwise) because r preserves the inner product. Since $E = L \oplus L^\perp$ we may describe the rotation r by saying how it behaves in L^\perp . Viewed as a Euclidean vector space in its own right L^\perp is two dimensional and the restriction of r to L^\perp is a plane rotation as described in the previous sections. This is because the restriction of r is now the composition of reflections in the lines $P \cap L^\perp$ and $P' \cap L^\perp$ viewed as an element of $O^+(L^\perp)$. As such we may describe it using an oriented angle.

A more concrete approach is to choose an orientation \mathcal{O}_E for E usually referred to as the 'right hand rule' and also an orientation \mathcal{O}_L of line L usually thought of an arrow on L . In that case we obtain an orientation \mathcal{O}_{L^\perp} of L^\perp as follows. Choose a basis vector ℓ in \mathcal{O}_L . A basis (a, b) of L^\perp is in the orientation \mathcal{O}_{L^\perp} if the triple (a, b, ℓ) is in orientation \mathcal{O}_E . In other words it should agree with the right hand rule where a, b are index finger and middle finger and ℓ is the thumb.

With all these orientations in place we are in a position to use the measure of the oriented angle of the rotation in L^\perp . The rotation can then be described by giving an axis ℓ and an angle $\mu \in \mathbb{R} \bmod 2\pi$. Choosing the angle to be in $[0, 2\pi)$ we may encode the rotation into the single axis vector ℓ by requiring $|\ell| = \mu$. We will then use the notation r_ℓ to denote the rotation with axis ℓ and angle $|\ell| \in [0, 2\pi)$. By convention the case $\ell = 0$ refers to the identity.

Taking the standard basis (e_1, e_2, e_3) of $E = \mathbb{R}^3$ to represent our orientation the most commonly used rotations are $r_{\mu e_2}$ (rolling), $r_{\mu e_1}$ (pitching) and $r_{\mu e_3}$ (yawing)³. For pilots the coordinate system is often thought of being attached to the cockpit such that e_2 points forwards, e_1 is pointing to the right wing and e_3 points upwards. We prefer instead to fix the coordinate system once and for all.

This notation is sufficient in the sense that any other rotation can be written as a composition of rolls, pitches and yaws. Actually rolls and pitches suffice. We will briefly return to this point using quaternions later.

5.2 Convex polyhedra

Convexity is an important property of objects in affine space (no inner product/distance necessary!). When discussing polyhedra in space we (as did Euclid) often assume implicitly that the polyhedron is convex. Roughly it means the polyhedron looks like a ball without cavities, craters or holes. More formally,

Definition 5.1. A subset S of affine space \mathcal{E} is called **convex** if for all $A, B \in S$ the line segment (1-simplex) $[A, B] \subset S$.

In any dimension the ball $B = \{x \in E : |x| \leq r\}$ and the simplex are convex (Exercise!). The regular polygons, triangle, square, pentagon ... in the plane (equal angles and equal sides) are also convex.

Inspecting the definition we remark that the intersection of two convex subsets is again convex (Exercise!). This allows us to generate many examples of convex subsets by taking the convex hull and such objects are called polyhedra when the set is finite.

Definition 5.2. (Audin Def 5.5 p.29)

The **convex hull** of $S \subset \mathcal{E}$ is the intersection of all convex subsets of \mathcal{E} that contain S .

The convex hull of a sphere S^{n-1} is the solid ball B contained in it because any point in B is on a straight line between two elements of the sphere. This means that this point must be contained in any convex subset containing the sphere.

³rollen, stampen en gieren in goed Nederlands.

Definition 5.3. (Audin Def 4.1 p.122)

A **polyhedron** in $P \subset \mathcal{E}$ is the convex hull of finitely many points in \mathcal{E} . It is said to be non-degenerate if the affine subspace generated by these points is \mathcal{E} . A point $p \in P$ is called interior if there exists a ball around p that is also contained in P . The set of all interior points is called P° .

In other words a non-degenerate polyhedron always contains an affine frame and hence the simplex spanned by it. It also means that there exist interior points. For non-degenerate polyhedra we will give a careful definition of the vertices, edges and faces. This is not as straightforward as one might think.

Definition 5.4. Define the vertices of a polyhedron P to be the intersection of all $T \subset \mathcal{E}$ whose convex hull is P . Any subset of n of the vertices determines a hyperplane H . We say $H \cap P$ is a face of P if $P^\circ \cap H = \emptyset$. In case the intersection of two distinct faces contains two vertices we call it an edge. The sets of vertices, edges and faces are generally denoted by V, E, F .

Even though a polyhedron P may be the convex hull of ten points, it may have fewer vertices, for example take the eight corners of a regular cube and place two more points inside. The convex hull will still be that cube and the eight corners will be the vertices. It is true that if P is the convex hull of set S then V is a subset of S (exercise!).

Our terminology is mostly adapted for dimension three. In two dimensions the sides of a triangle should now be called faces and there are no edges. In three dimensions the edges are intervals connecting two vertices (exercise!) but in higher dimensions they will have higher dimension too and more should be said to describe the combinatorial structure of the polyhedron completely.

Now that we defined faces properly we can say what we mean by a regular polyhedron. Again the definition we give is mostly tailored to dimension three, in higher dimensions we can ask even more regularity.

Definition 5.5. A non-degenerate convex polyhedron in Euclidean affine space is called **regular** if all its faces are isometric to a single regular polygon and if for each pair of vertices a, b there is an isometry taking a to b that sends any edge adjacent to a to an edge adjacent to b .

Simple examples of regular polyhedra in any dimension are the orthoplexes which are the convex hull of the unit vectors $\pm e_i$ in \mathbb{R}^n (viewed as affine n space). Dually the n -cubes $[-1, 1]^n$ are also regular.

These two examples illustrate a general phenomenon of duality of convex polyhedra. Recall that for triangles we defined the centroid to be the center of gravity. This notion works for any finite subset $S \subset E$. It is just the average of the points in S . The centroid of a polyhedron is the centroid of its vertices.

Definition 5.6. The dual of a convex polyhedron P is the convex hull of the centroids of the faces of P .

5.3 Spherical geometry

Throughout this subsection we work in a three dimensional Euclidean vector space E . Our main object of study is the sphere $S^2 = S^2(0, 1)$ of radius 1 centered at the origin.

Definition 5.7. A great circle on S^2 is the intersection of S^2 with a two dimensional linear subspace of E . A half space defined by vector $v \in E$ is the set $H_v = \{x \in E \mid \langle v, x \rangle \geq 0\}$. A **convex spherical polygon** is the intersection of S^2 and finitely many half spaces.

The most important special cases of convex spherical n -gons are: the case $n = 1$ which is known as a hemisphere. The case $n = 2$ is called a time zone (lune) and the case $n = 3$ is called a spherical triangle.

We think of the great circles as the 'straight lines' (geodesics). Later in the course we will see that indeed among all paths that run on the surface of the sphere the shortest path between two points is always an arc of a great circle. By arc of great circle we just mean a connected part. Unlike Euclidean straight lines, any great circles intersects and not just in one point but actually in two points. This pair of points is always antipodal (opposite). There is a fundamental isometry called the antipodal isometry $a : E \rightarrow E$ defined by $a(x) = -x$ and we say it sends points x on the sphere to their antipode $-x$.

We make the assumption that there exists an area function $\text{Area} : \mathcal{P} \rightarrow \mathbb{R}$ defined on the set \mathcal{P} of all $P \subset S^2$ bounded by finitely many segments of great circles⁴. We assume Area satisfies the following properties:

1. $\text{Area}(S^2) = 4\pi$
2. If $P, Q \in \mathcal{P}$ and $P \cap Q$ consists of segments of great circles then $\text{Area}(P \cup Q) = \text{Area}(P) + \text{Area}(Q)$.
3. For any isometry ϕ of E and $P \in \mathcal{P}$ we have $\text{Area}(P) = \text{Area}(\phi(P))$.

⁴One should not include all subsets of the sphere because this will allow paradoxal constructions akin to the Banach-Tarski paradox.

The angle between two great circles is the geometric angle between the two corresponding planes.

Theorem 5.1. (*Audin Prop 3.1, p. 121*)

The sum of the angles of a spherical triangle is the area of the triangle plus π .

Proof. (Nicer than Audin!). Extend the sides of our spherical triangle ABC to three great circles. Each pair of great circles will meet in an additional point antipodal to the original, say A' is antipodal to A and B' the antipode of B and C' the antipode of C . The antipodal isometry sends triangle ABC to triangle $A'B'C'$ which means they have equal area. For each vertex of our triangle two of the three great circles intersect and these two circles bound two time zones ending at the vertex and its antipode. One of these time zones contains ABC the other contains $A'B'C'$. The time zones are related by a π rotation with axis $\overrightarrow{AA'}$. The measure of the geometric angle of both the time zones at A is called $\alpha \in [0, \pi)$, the angle at B is called β and the one at C is called γ .

The total of six time zones covers the sphere completely with a triple overlap precisely at ABC and $A'B'C'$. Since the area of a time zone with angle μ is 2μ and the area of the sphere is 4π finishes the proof:

$$4\pi = \text{Area}(S^2) = 4\alpha + 4\beta + 4\gamma - 2\text{Area}(ABC) - 2\text{Area}(A'B'C') = 4(\alpha + \beta + \gamma - \text{Area}(ABC))$$

□

5.4 Euler's formula

In the remainder of this subsection we assume the dimension of \mathcal{E} is 3.

Theorem 5.2. (*Euler's polyhedron formula*)

If V , E and F denote the number of vertices, edges and faces of a convex non-degenerate polyhedron in \mathcal{E} then

$$V - E + F = 2$$

Proof. Without loss of generality we may assume that all faces of our polyhedron $Q \subset \mathcal{E}$ are triangular. This is because we can always divide a face with more corners by adding a diagonal edge. This increases both E and F by one while not changing the number of vertices. Hence $V - E + F$ is unchanged. By non-degeneracy we may choose a point O in the interior of our polyhedron Q . Applying the map $\Theta_O : \mathcal{E} \rightarrow \mathbb{E}$ where we wrote the underlying vector space as \mathbb{E} to avoid conflict of notation we now consider a polyhedron $P = \Theta_O(Q)$ containing the origin in its interior.

The radial projection map $\rho : \mathbb{E} - \{0\} \rightarrow S^2$ defined by $\rho(x) = \frac{x}{|x|}$ sends the vertices, edges and faces of P to a tiling of the sphere S^2 by triangles. More precisely the images of the edges end at the images of the vertices and their complement on the sphere is a union of spherical triangles.

We calculate in two ways the sum of the (measures of the geometric) angles of all the triangles. Collecting the angles around each vertex we see that they add up to 2π at each vertex so we get $2\pi V$ in total. On the other hand we can add the angle sum of each of the triangular faces to get $\sum_{\text{faces } f} \text{Area}(f) + \pi = 4\pi + \pi F$ since the F faces cover S^2 precisely.

Putting it all together we found

$$2\pi V = 4\pi + \pi F$$

Since all faces are assumed to be triangles we have $3F = 2E$ so $-\frac{F}{2} = -E + F$. Dividing the above by 2π we find $2 = V - \frac{E}{2} = V - E + F$. □

Euler's polyhedron formula restricts the possibilities for regular polyhedra severely.

Lemma 5.1. *If in a convex polyhedron all faces have s vertices and around every vertex r edges meet then $\{r, s\} \subset \{\{3, 3\}, \{3, 4\}, \{3, 5\}\}$*

Proof. Any edge belongs to two faces so $sF = 2E$ and every edge has two end points so $rV = 2E$. Solving for E in Euler's formula we find $2 = V - E + F = \frac{2}{r}E - E + \frac{2}{s}E$ so dividing by $2E$ we find

$$\frac{1}{E} + \frac{1}{2} = \frac{1}{r} + \frac{1}{s}$$

The number of edges is positive so $\frac{1}{r} + \frac{1}{s} > \frac{1}{2}$. Since r, s are supposed to be integers ≥ 3 there are only finitely many possibilities to satisfy this equation. Five to be precise as one can see by listing all possibilities. □

So far we have shown that if regular polyhedra exist in 3-space then the possible values of s, r are as above. However we do not yet know if for each of these allowed values there exists such a polyhedron. This is the content of the final book of Euclid's Elements:

Theorem 5.3. *(the Platonic solids)*

Up to isometries and dilations there exist precisely five regular polyhedra in Euclidean affine 3-space. They are:

1. The tetrahedron $(r, s) = (3, 3)$
2. The cube $(3, 4)$
3. The octahedron $(4, 3)$
4. The dodecahedron $(3, 5)$
5. The icosahedron $(5, 3)$

The existence of tetrahedron, cube and octahedron is not hard to establish explicitly. The construction of the dodecahedron and icosahedron is more surprising.

The strategy is to start with a cube and place roofs on each of the faces of a cube in a symmetric fashion. The roof has a square horizontal base and a top line segment parallel to a pair of sides of the cube. The remaining faces are determined by connecting the vertices of the square to the endpoints of the top segment. The five edges not part of the square should have equal length. The two dihedral angles that the faces make with the square base should add up to $\pi/2$. The top angles of the triangles should be $2\pi/5$.

The vertices of a dodecahedron are $(\pm 1, \pm 1, \pm 1), (0, \pm\tau, \pm\tau^{-1}), (\pm\tau^{-1}, 0, \pm\tau), (\pm\tau, \pm\tau^{-1}, 0)$.

Higher dimensional analogues of the Platonic solids exist in any dimension. We have already met the simplices and the generalization of the cube is the hypercube $[-1, 1]^n$. Its dual is the analogue of the octahedron and also exists in any dimension. However dodecahedron and icosahedron are unique to three dimensions. In dimensions five and up the three regular solids: simplex, hypercube and its dual the orthoplex are the only ones. In dimension four there are three more that go by the name of 120-cell 600-cell (analogues of dodecahedron and icosahedron) and something new called the 24-cell. One way to understand their existence is as lifts of the symmetries of the three-dimensional Platonic solids. This uses the fact that $SO(3)$ is doubly covered by the 3-sphere S^3 .

5.5 Quaternions

No discussion of three-dimensions is complete without mention of the quaternions. Much like complex numbers are very effective in discussing the plane the four-dimensional space of quaternions illuminates the isometry group of three space. Following Hamilton we boldly construct these new numbers:

Definition 5.8. *The quaternions \mathbf{H} is a 4-dimensional vector space with basis $1, i, j, k$ and bilinear, associative product $\mathbf{H} \times \mathbf{H} \rightarrow \mathbf{H}$ determined by Hamilton's relations*

$$i^2 = j^2 = k^2 = -1 \quad ij = k \quad jk = i \quad ki = j$$

If $q = w + xi + yj + zk \in \mathbf{H}$ the conjugate is $\bar{q} = w - xi - yj - zk$. The norm is defined as $|q| = q\bar{q} = w^2 + x^2 + y^2 + z^2 \geq 0$.

Calculus books still use $\mathbf{i}, \mathbf{j}, \mathbf{k}$ to denote the unit vectors in \mathbb{R}^3 and this originated from Hamilton's work. If we identify vector $v = (v_1, v_2, v_3) \in \mathbb{R}^3$ with the imaginary quaternion $q_v = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k} \in \mathbf{H}$ then we have

$$q_v q_u = -v \cdot u + q_{v \times u}$$

This means that vectors $v \perp u$ if and only if $q_v q_u + q_u q_v = 0$ because $\bar{q}_v = -q_v$.

Now assume u is a unit vector so that $q_u \bar{q}_u = 1$. Then $q_u^2 = -1$ and $v \perp u$ if and only if $q_v = q_u q_v q_u$. Notice that the left hand side is again a pure quaternion. This means there is a map $s : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ given by $v \mapsto q_u q_v q_u = q_{s(v)}$. We claim $s = s_{u^\perp}$ the reflection in the plane orthogonal to unit vector u . Indeed s fixes vectors in the plane u^\perp by the above discussion and $q_{s(u)} = q_u^3 = -q_u = q_{-u}$ since $q_u^2 = -1$.

As in the plane a rotation in three-dimensional Euclidean space E is still a composition of two reflections and any element of $O^+(E)$ is a rotation since it is the composition of 0, 1, 2 or 3 reflections.

For two unit vectors u, t composing the reflections s_{u^\perp} and s_{t^\perp} we get a map sending q_v to $q_t q_u q_v q_u q_t$. Now $q_t q_u = (q_u q_t)^{-1}$ since both u, t are unit vectors. The product $q_t q_u$ is also a unit quaternion so all rotations are written in quaternions as conjugation by a unit quaternion.

6 Differential geometry

In the second half of this course we explore what remains of geometry when the space is curved and distances are distorted. Such spaces will be introduced as Riemannian⁵ charts⁶. Our main concern will be to formulate what we mean by a 'straight line' in such a curved space. The curves taking the place of straight lines are known as geodesics and we will show that they exist at least locally as the curve minimizing the distance from point A to point B . In doing so we will need a fair amount of techniques from differential calculus including the existence of solutions to ordinary differential equations and a little variational calculus.

6.1 Derivative

Definition 6.1. A C^1 function is a function whose partial derivatives exist and are continuous. Likewise C^2 means the partial derivatives of the partial derivatives are continuous and so on.

Imagine a C^1 function $P \xrightarrow{f} \mathbb{R}^m$ defined on an open subset $P \subset \mathbb{R}^n$. Set $f = \sum_{i=1}^m f_i e_i$ for some functions $f_i : P \rightarrow \mathbb{R}$. The derivative of f at $p \in P$ is the linear transformation $\mathbb{R}^n \xrightarrow{f'(p)} \mathbb{R}^m$ whose matrix with respect to the standard bases is $(\frac{\partial f_i}{\partial x_j}(p))_{i=1\dots m, j=1\dots n}$.

In case P is not open we say f is differentiable if there exists a differentiable function on an open set containing P that coincides with f on P .

The matrix of the derivative with respect to the standard bases is also known as the Jacobian matrix. Its (i, j) -th entry is the j -th partial derivative $\frac{\partial f_i}{\partial x_j}(p)$ of the i -th component of f . Even though it is often enough to work in the standard bases it is important to notice that our derivative is the linear transformation corresponding to the Jacobian matrix, not the Jacobian itself. In other words $f'(p)e_j = \partial_j f(p) = \sum_i \frac{\partial f_i}{\partial x_j}(p)$. In geometry it is often useful to be able to change to a basis adapted to the situation. Also, phrasing things in terms of linear transformations makes the formulas easier to manage.

It is traditional to call a function $X : P \rightarrow \mathbb{R}^n$ for $P \subset \mathbb{R}^n$ a **vector field** on P . We will take for granted the following properties of the derivative:

Lemma 6.1. (Properties of the derivative)

1. (Chain rule) For C^1 functions $P \xrightarrow{f} Q \xrightarrow{g} R$ we have $(g \circ f)'(p) = g'(f(p)) \circ f'(p)$.
2. For C^1 functions $P \xrightarrow{f,g} Q$ and $\alpha, \beta \in \mathbb{R}$ the function $\alpha f + \beta g$ is also C^1 .
3. For C^1 functions $P \xrightarrow{f,g} \mathbb{R}$ the product fg is also C^1 .
4. The function $P \xrightarrow{f} Q$ is C^1 if and only if its coefficient functions f_i are C^1 where $f(p) = \sum_i f_i(p)e_i$.

Definition 6.2. A C^1 function $\mathbb{R}^n \supset P \xrightarrow{f} Q \subset \mathbb{R}^n$ is said to be a diffeomorphism if it is a bijection whose inverse is also C^1 .

Diffeomorphisms are often important as symmetries or coordinate changes because by definition we do not lose any information applying them to some object. The inverse function theorem provides lots of theoretical examples of diffeomorphism. One way to state the theorem is that if a C^1 function $P \xrightarrow{f} Q \xrightarrow{g} R$ has invertible derivative $f'(p)$ for some $p \in P$ then there are open neighbourhoods $p \in A \subset P$ and $f(p) \in B \subset Q$ such that the restriction of f to A is a diffeomorphism with image B .

A geometric way to think about the derivative of a map $P \xrightarrow{f} Q$ between open subsets $P \subset \mathbb{R}^n$ and $Q \subset \mathbb{R}^m$ is the following. At every point $p \in P$ the derivative at p is a linear map $\mathbb{R}^n \xrightarrow{f'(p)} \mathbb{R}^m$ and to visualize $f'(p)v$ for some $v \in \mathbb{R}^n$ we use the chain rule. To make the vector v more geometric we interpret it as the velocity vector of some curve $\gamma : (-\epsilon, \epsilon) \rightarrow P$ for some small $\epsilon > 0$. So $\gamma(0) = p$ and $\dot{\gamma}(0) = v$. We cannot apply f directly to v but we can now consider what happens to γ as we apply f . This produces a new curve $\beta = f \circ \gamma$ with image in Q . The velocity vector w of this new curve β at $f(p) = \beta(0)$ is a natural geometric way to transport the vector v at p to a vector w at $f(p)$. In formulas

$$w = \dot{\beta}(0) = (f \circ \gamma)'(0) = f'(\gamma(0))\dot{\gamma}(0) = f'(p)v$$

In the third step we used the chain rule. We leave it as an exercise to the reader that this construction did not depend on the particular choice of γ .

⁵B. Riemann wrote his PhD thesis on this topic. His advisor was C. Gauss.

⁶Usually one considers Riemannian manifolds (see the course in the 3rd year) instead of charts but for our purposes charts are sufficient.

Another good way to think about the derivative $f'(p)v$ is as a directional derivative $\partial_v f(p)$:

$$\partial_v f(p) = \lim_{h \rightarrow 0} \frac{f(p + hv) - f(p)}{h}$$

The chain rule tells us that $\partial_v f(p) = f'(p)v$. This is left as an exercise to the reader.

6.2 Riemannian metrics

Definition 6.3. (*Riemannian chart*)

A **Riemannian chart** (P, g) is an open subset $P \subset \mathbb{R}^n$ together with a **Riemannian metric** g . The Riemannian metric is a choice of inner product $g(p)$ for each $p \in P$ in such a way that the functions $g_{ij} : P \rightarrow \mathbb{R}$ defined by $g_{ij} = g(p)(e_i, e_j)$ are C^1 .

Euclidean n -space is the fundamental example of a Riemannian chart. It looks like (\mathbb{R}^n, g_E) where the Riemannian metric g_E is given by $g_E(p)(v, w) = \langle v, w \rangle$, the usual inner product. Since the standard basis e_i is orthonormal we have $g_{ij}(p) = \delta_{ij}$ for all p which is surely a C^1 function.

Lengths and measures of angles can still be defined in a Riemannian chart (P, g) :

Definition 6.4. (*Length and angle*)

Suppose $\beta, \gamma : [a, b] \rightarrow P$ are curves in Riemannian chart (P, g) .

The **length** $L(\gamma)$ of curve γ is the integral $L(\gamma) = \int_a^b \sqrt{g(\gamma(t))(\gamma'(t), \gamma'(t))} dt$.

If $\beta(q) = \gamma(q) = p \in P$ the **angle** between β, γ at p is the angle between the vectors $\beta'(q), \gamma'(q)$ in the Euclidean space $(\mathbb{R}^n, g(p))$, provided that $\beta'(q), \gamma'(q) \neq 0$.

Metrics can also be transferred by pull-back as follows.

Definition 6.5. (**Pull-back metric**)

Given a C^2 differentiable map $P \xrightarrow{\varphi} Q$ between open $P \subset \mathbb{R}^m$ and Riemannian chart (Q, g) such that $\varphi'(p)$ is injective for all $p \in P$, we may define a metric φ^*g on P by $(\varphi^*g)(p)(v, w) = g(\varphi(p))(\varphi'(p)v, \varphi'(p)w)$.

The derivative of φ is required to be injective in order to ensure non-degeneracy of the pulled back inner product.

For example we may take the sphere and geographic coordinates $G = (0, \pi) \times (-\pi, \pi)$ and $G \ni (\mu, \lambda) \xrightarrow{\varphi} (\cos \lambda \sin \mu, \sin \lambda \sin \mu, \cos \mu) \in \mathbb{R}^3$. Here μ is the latitude coordinate and λ the longitude, for example Leiden is the point $\varphi(52.1601, 4.4970) \frac{\pi}{180}$ written in degrees. Explicitly the inner product φ^*g_{Eucl} is given by calculating

it at every point for the basis vectors e_1, e_2 . Since the matrix for $\varphi'(p)$ is $\begin{pmatrix} \cos \lambda \cos \mu & -\sin \lambda \sin \mu \\ \sin \lambda \cos \mu & \cos \lambda \sin \mu \\ -\sin \mu & 0 \end{pmatrix}$ we get $\varphi^*g_{Eucl}(p)(e_1, e_1) = 1, \varphi^*g_{Eucl}(p)(e_1, e_2) = 0, \varphi^*g_{Eucl}(p)(e_2, e_2) = \sin^2 \mu$.

Given a geometric object $Q \subset \mathbb{R}^n$ in Euclidean space there are two ways to study it: **intrinsically** or **extrinsically**. The intrinsic way of studying Q would be to find a differentiable map $P \xrightarrow{\varphi} Q$ if injective derivative and turn P into a Riemannian chart (P, g) as above by pulling back the Euclidean metric: $g = \varphi^*g_{Eucl}$. Properties of Q that can be described in this way are called **intrinsic** properties. The length of a curve $\gamma : [0, 1] \rightarrow Q$ is an example of an intrinsic property.

The other way is to work with Q from the 'outside' or **extrinsically** by studying how it sits in \mathbb{R}^n using the tools from Euclidean geometry and differential calculus. A typical extrinsic property of Q is the Euclidean distance between two points in Q .

Both the intrinsic and extrinsic are important but in modern geometry there is an emphasis on the intrinsic features. Often working intrinsically is more efficient in that one needs less coordinates and is less distracted by 'irrelevant' extrinsic features. Also sometimes our space Q is not presented to us as subset of some Euclidean space but still has a Riemannian metric. In that case the intrinsic approach is the only option. For example when Q is the universe or when $Q = O(E)$ the orthogonal group or the space of quaternions of unit length.

Finally another important example of a Riemannian chart is hyperbolic space $\mathbb{H}^n = \mathbb{R}^{n-1} \times \mathbb{R}_{>0}$ with metric $g_{hyp}(x, y)(v, w) = \frac{1}{y^2} g_E(v, w)$.

For example in the hyperbolic plane the length of the vertical line between $(0, a)$ and $(0, b)$ given by the curve γ defined by $\gamma(t) = (t(b-a) + a)e_2$ with $a < b$ is given by $L(\gamma) = \int_0^1 \frac{b-a}{t(b-a)+a} dt = \log b - \log a$. Also notice that angles in the hyperbolic plane are just the Euclidean angles.

Definition 6.6. (**Isometry**)

An **isometry** is a C^2 -diffeomorphism that preserves the metric in the following sense. $(P, g) \xrightarrow{\varphi} (Q, h)$ is an isometry if $g = \varphi^*h$.

Since all geometric properties derive from the metric, isometries can be thought of as those transformations that preserve shape. As the name suggests the Euclidean linear isometries $O(E)$ from a Euclidean vector space to itself are examples of the above Riemannian notion of isometry. In this case we have $(P, g) = (Q, h) = (\mathbb{R}^n, g_E)$, the standard inner product in every point.

Other examples of isometries arise when we describe the same object in Euclidean space using two different Riemannian charts. For example let us make a completely different chart describing (part of) the sphere in \mathbb{R}^3 . The stereographic map $\sigma : \mathbb{R}^2 \rightarrow S^2 \subset \mathbb{R}^3$ is defined as $\sigma(p) = \frac{(2p, |p|^2 - 1)}{|p|^2 + 1}$ where $p = (p_1, p_2)$. It has inverse $\sigma^{-1}(x, z) = \frac{x}{1-z}$ where $x = (x_1, x_2) \in \mathbb{R}^2$ defined on all of S^2 except the north pole where $z = 1$.

The stereographic map σ has injective derivative at every point and is a C^2 function. Therefore we can use it to turn $P = \mathbb{R}^2$ into a Riemannian chart (P, g) with pull-back metric $g = \sigma^*g_E$. Recall we also had the geographic Riemannian chart (G, φ^*g_E) of the S^2 . The map $\sigma^{-1} \circ \varphi$ when restricted to the correct domain is an example of an isometry between Riemannian charts.

Another simple example is an isometry of $t : \mathbb{H}^2 \rightarrow \mathbb{H}^2$ defined by $t(x, y) = (x + 1, y)$. Since $t'(p)$ is the identity at any point p we get $t^*g_{hyp} = g_{hyp}$. There are many more isometries of \mathbb{H}^2 and we will get back to them later.

6.3 Calculus of variations

The purpose of this subsection is to find a differential equation for the shortest curve between two points in a Riemannian chart (P, g) , where $P \subset \mathbb{R}^n$. Since the length of a curve γ is given as $\int_0^1 |\dot{\gamma}(t)| dt$ we are trying to minimize the function Length on some space of curves in \mathbb{R}^n . This is reminiscent of minimizing ordinary functions by setting their derivative to zero and is known as calculus of variations. It makes sense to generalize a little and treat more general functions S on the space of curves.

Consider the space \mathcal{C} of C^2 curves $\gamma : [0, 1] \rightarrow \mathbb{R}^n$. For a C^2 function $\mathcal{L} : \mathbb{R}^{2n+1} \rightarrow \mathbb{R}$ define $S : \mathcal{C} \rightarrow \mathbb{R}$ by $S(\gamma) = \int_0^1 \mathcal{L}(\gamma(t), \gamma'(t), t) dt$.

Problem: Suppose S obtains a minimum value, how do we find it?

Theorem 6.1. (Euler-Lagrange equation)

If $\gamma \in \mathcal{C}$ is a minimum of S then for all $i \in \{1, \dots, n\}$:

$$\partial_i \mathcal{L}(\gamma(t), \gamma'(t), t) = \frac{d}{dt} (\partial_{n+i} \mathcal{L}(\gamma(t), \gamma'(t), t))$$

The proof of this theorem depends on a simple lemma

Lemma 6.2. Suppose M is a C^2 function on $M : [0, 1] \rightarrow \mathbb{R}^n$. If M is such that for all C^2 functions $h : [0, 1] \rightarrow \mathbb{R}^n$ with $h(0) = h(1) = 0$ we have

$$\int_0^1 M(t)h(t) dt = 0$$

then $M(t) = 0$ for all $t \in [0, 1]$.

Proof. Define a function $h : [0, 1] \rightarrow \mathbb{R}$ by $h(t) = (1 - t)tM(t)$. The function h is C^2 because M is and $h(0) = h(1) = 0$ and $M(t)h(t) > 0$ for $t \in (0, 1)$. By assumption $0 = \int_0^1 M(t)h(t) dt$ but by the mean value theorem for integrals this integral equals $M(s)h(s)$ for some $s \in (0, 1)$. \square

Proof. (Of Theorem 6.1)

Let us assume curve $\gamma \in \mathcal{C}$ minimizes S in the sense that for any other curve $\beta \in \mathcal{C}$ $S(\gamma) \leq S(\beta)$.

Choose $i \in \{1, \dots, n\}$ and consider perturbing the curve γ by adding an arbitrary function h in the i direction. More precisely choose some fixed C^2 function $h : [0, 1] \rightarrow \mathbb{R}$ with $h(0) = h(1) = 0$. The function $V : (-v, v) \times [0, 1] \rightarrow P$ defined by $V(\epsilon, t) = \gamma(t) + \epsilon h(t)e_i$ represents a family of curves, one for each value of ϵ and v is chosen so that all the curves actually have image contained in P . So for any fixed $|\epsilon| < v$ we get a curve $V_\epsilon : [0, 1] \rightarrow P$ defined by $V_\epsilon(t) = V(\epsilon, t)$. For example $V_0 = \gamma$ and for small ϵ we get a curve V_ϵ deviating slightly from γ .

By assumption $S(V_\epsilon) \geq S(V_0)$ and this means that the function $H : (-v, v) \rightarrow \mathbb{R}$ defined by $H(\epsilon) = S(V_\epsilon)$ has a minimum at $\epsilon = 0$. Ordinary calculus tells us that $H'(0) = 0$. Unpacking the definition of H this means

$$0 = H'(0) = \frac{d}{d\epsilon} \Big|_{\epsilon=0} \int_0^1 \mathcal{L}(V(\epsilon, t), \partial_t V(\epsilon, t), t) dt = \int_0^1 \partial_\epsilon \Big|_{\epsilon=0} \mathcal{L}(V(\epsilon, t), \partial_t V(\epsilon, t), t) dt \quad (1)$$

To apply the chain rule in a clean way we introduce $W : (-v, v) \times \mathbb{R} \rightarrow P \times \mathbb{R}^{n+1}$ defined by $W(\epsilon, t) = (V(\epsilon, t), V'(\epsilon, t), t)$. Then the derivative with respect to ϵ inside the integral is equal to $(\mathcal{L} \circ W)'(0, t)e_1 =$

$\mathcal{L}'(W(0,t))W'(0,t)e_1$ by the chain rule⁷. Now $W'(0,t)e_1 = h(t)e_i + h'(t)e_{n+i}$ so our integrand from (1) is

$$(\mathcal{L} \circ W)'(0,t)e_1 = \mathcal{L}'(W(0,t))(h(t)e_i + h'(t)e_{n+i}) = (\partial_i \mathcal{L})(\gamma(t), \gamma'(t), t)h_i(t) + (\partial_{i+n} \mathcal{L})(\gamma(t), \gamma'(t), t)h'_i(t)$$

Partial integration gets rid of the h' in the second term (using $h(0) = h(1) = 0$) and turns our integral into

$$0 = H'(0) = \int_0^1 \partial_i \mathcal{L}(\gamma(t), \gamma'(t), t)h_i(t) - \partial_t((\partial_{i+n} \mathcal{L})(\gamma(t), \gamma'(t), t))h_i(t) dt = \int_0^1 M(t)h(t) dt = 0$$

$$M = \partial_i \mathcal{L}(\gamma(t), \gamma'(t), t) - \partial_t((\partial_{i+n} \mathcal{L})(\gamma(t), \gamma'(t), t))$$

Since this must hold for arbitrary choices of h Lemma 6.2 shows $M = 0$ finishing the proof. \square

6.4 Geodesics

Our main application of the calculus of variations is to minimize the length of curves in a Riemannian chart. Such curves will be shown to be solutions to a system of ordinary differential equations known as the geodesic equation (4).

Definition 6.7. Any C^2 curve in a Riemannian chart that solves the geodesic equation (4) is called a **geodesic (curve)**.

For technical reasons it is easier not to directly find curves that minimize the length but rather minimize the energy $S(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2 dt$. By reparametrizing we can assume our curve γ minimizing the length has unit speed, so $|\dot{\gamma}(t)| = 1$. Since $1^2 = 1$ we see γ must also minimize the easier S which uses the squared norm.

For example take a curve in flat 2-space $\gamma : [0, 1] \rightarrow \mathbb{R}^2$. We aim to minimize $S(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2 dt$ so $\mathcal{L}(x, y, t) = y_1^2 + y_2^2$.

The Euler-Lagrange equations we need to solve are: $\gamma''(t) = 0$ since

$$\frac{d}{dt}(\partial_{2+i} \mathcal{L}(\gamma(t), \gamma'(t), t)) = \frac{d}{dt}(2\gamma'_i(t)) = 2\gamma''_i(t) = 0 = \partial_i \mathcal{L}(\gamma(t), \gamma'(t), t)$$

It follows that $\gamma(t) = at + b$ for some $a, b \in \mathbb{R}^2$.

Moving on to the general case, consider a curve $\gamma : [0, 1] \rightarrow \mathbb{R}^n$. For the argument it is important to allow all curves, not just those in the Riemannian chart (P, g) with $P \subset \mathbb{R}^n$. Writing $\gamma(t) = \sum_{i=1}^n \gamma_i e_i$ we find

$$|\dot{\gamma}(t)|^2 = \sum_{i,j} g(\gamma(t))(\gamma'_i(t)e_i, \gamma'_j(t)e_j) = \sum_{i,j} g_{ij}(\gamma(t))\gamma'_i(t)\gamma'_j(t)$$

In this case we seek to minimize $S(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2 dt = \int_0^1 \sum_{i,j} g_{ij}(\gamma(t))\gamma'_i(t)\gamma'_j(t) dt$ so $\mathcal{L}(x, y, t) = \sum_{i,j=1}^n g_{ij}(x)y_i y_j$ is a function from \mathbb{R}^{2n+1} to \mathbb{R} .

The Euler-Lagrange equations now read for any fixed $k \in \{1, \dots, n\}$:

$$\partial_k \mathcal{L}(\gamma(t), \gamma'(t), t) = \frac{d}{dt}(\partial_{n+k} \mathcal{L}(\gamma(t), \gamma'(t), t)) \quad (2)$$

The left hand side is $\sum_{i,j=1}^n (\frac{\partial g_{ij}}{\partial x_k})y_i y_j$. For the right hand side we first compute

$$\partial_{n+k} \mathcal{L}(x, y, t) = \sum_{j=1}^n g_{kj}(x)y_j + \sum_{i=1}^n g_{ik}(x)y_i = 2 \sum_{i=1}^n g_{ik}(x)y_i$$

so the right hand side of (2) is

$$\frac{d}{dt}(\partial_{n+k} \mathcal{L}(x, y, t)) = 2 \frac{d}{dt} \sum_{i=1}^n g_{ik}(\gamma(t))\gamma'_i(t) = 2 \sum_{i,j} \frac{\partial g_{ik}}{\partial x_j}(\gamma(t))\gamma'_i(t)\gamma'_j(t) + 2 \sum_{i=1}^n g_{ik}(\gamma(t))\gamma''_i(t)$$

Using symmetry to interchange the dummy indices i, j we get $\sum_{i,j} \frac{\partial g_{ik}}{\partial x_j}(\gamma(t))\gamma'_i(t)\gamma'_j(t) = \sum_{i,j} \frac{\partial g_{jk}}{\partial x_i}(\gamma(t))\gamma'_i(t)\gamma'_j(t)$

Putting it all together we found a differential equation for γ :

$$0 = 2 \sum_{s=1}^n g_{ks}(\gamma(t))\gamma''_s(t) + \sum_{i,j=1}^n \left(\frac{\partial g_{jk}}{\partial x_i}(\gamma(t)) + \frac{\partial g_{ik}}{\partial x_j}(\gamma(t)) - \frac{\partial g_{ij}}{\partial x_k}(\gamma(t)) \right) \gamma'_i(t)\gamma'_j(t) \quad (3)$$

⁷Here e_1 refers to the first coordinate in $(-v, v) \times [0, 1]$, which is ϵ . Also $W'(0, t)e_1 = \partial_\epsilon W(\epsilon, t)|_{\epsilon=0}$.

It is convenient to isolate the second derivative term $\gamma_s''(t)$. This can be achieved since the matrix $(g_{ij}(\gamma(t)))$ describing the metric is positive definite and symmetric. It follows that it has only positive eigenvalues and hence positive determinant. To see this pick an eigenvector v with eigenvalue λ and write $v^T(g_{ij}(\gamma(t)))v = v^T \lambda v = \lambda|v|^2 > 0$. It thus has an inverse $(g_{rk}^{-1}(\gamma(t)))$ so $\sum_k g_{rk}^{-1}(\gamma(t))g_{ks}(\gamma(t)) = \delta_{rs}$ (Kronecker delta). Applying this to both sides of our differential equations we get the **geodesic equation**. For fixed $r \in \{1, \dots, n\}$ we have:

$$0 = \gamma_r''(t) + \sum_{i,j=1}^n \Gamma_{ij}^r(\gamma(t))\gamma_i'(t)\gamma_j'(t) \quad (4)$$

For any value of $i, j, r \in \{1, \dots, n\}$ the function $\Gamma_{ij}^r : P \rightarrow \mathbb{R}$ is known as the **Christoffel symbols** and is defined as:

$$\Gamma_{ij}^r(p) = \frac{1}{2} \sum_{k=1}^n g_{rk}^{-1}(p) \left(\frac{\partial g_{jk}}{\partial x_i}(p) + \frac{\partial g_{ik}}{\partial x_j}(p) - \frac{\partial g_{ij}}{\partial x_k}(p) \right) \quad (5)$$

For example the eight Christoffel symbols for the sphere with geographic coordinates are $\Gamma_{ij}^1 = 0$ except $\Gamma_{22}^1 = -\cos \mu \sin \mu$. $\Gamma_{11}^2 = \Gamma_{22}^2 = 0$ and $\Gamma_{12}^2 = \Gamma_{21}^2 = \cot \mu = \frac{\cos \mu}{\sin \mu}$. So the geodesic equations for $\gamma = (\gamma_1, \gamma_2)$ is $\ddot{\gamma}_1 - \cos \mu \sin \mu \dot{\gamma}_2 \dot{\gamma}_2 = 0 = \ddot{\gamma}_2 + 2 \cot \mu \dot{\gamma}_1 \dot{\gamma}_2$. While a little complicated at first sight it should be clear that the meridians $\gamma(t) = (t, c)$ for some constant $c \in \mathbb{R}$ satisfy this equation.

Returning to the general case: the fundamental theorem on the existence and uniqueness of ordinary differential equations assures us that geodesics always exist. Intuitively the next theorem states that in a Riemannian chart one can start walking 'straight' in any direction at any point of the chart.

Theorem 6.2. (Existence and uniqueness of geodesics)

In a Riemannian chart (P, g) any $p \in P \subset \mathbb{R}^n$ and $v \in \mathbb{R}^n$ determines a unique C^2 curve $\gamma : [C, D] \rightarrow P$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = v$ and for all m we have $\gamma_m''(t) + \sum_{i,j=1}^n \Gamma_{ij}^m \gamma_i'(t)\gamma_j'(t)$. The domain $[-D, D]$ is not quite unique but if we have another curve $\tilde{\gamma}$ with the same properties and domain $[A, B]$ then $\gamma = \tilde{\gamma}$ on $[A, B] \cap [C, D]$.

We arrived at geodesics by attempting to minimize the distance between two points but what we found is actually more like the curves that are 'straight'. Depending on global issues travelling a straight line may or may not actually minimize the distance travelled.

6.5 Surfaces in Euclidean space

In this subsection we examine the meaning of the geodesic equation from the *extrinsic* point of view. This means we assume our Riemannian chart (P, g) comes with an injective C^2 map $\phi : P \rightarrow \mathbb{R}^m$ with injective derivative and the metric is the pull-back of the Euclidean metric $g = \phi^* g_E$. We will assume $P \subset \mathbb{R}^n$ throughout. The reader should pay particular attention to the case $n = 2, m = 3$ that we see around us every day.

The tangent space to $\phi(P)$ at point $\phi(p)$ is spanned by the partial derivatives $\partial_i \phi(p) = \phi'(p)e_i$ or more generally the directional derivatives $\partial_v \phi(p)$. Let us denote the **tangent space** by $T_{\phi(p)} = \phi'(p)(\mathbb{R}^n)$. It is an n -dimensional linear subspace of \mathbb{R}^m , translated by vector $\phi(p)$ it actually becomes an affine subspace tangent to $\phi(P)$.

We will see that in this setting the geodesic equation simply states that the acceleration of a geodesic is perpendicular to the tangent space. Actually given a curve γ in P we also have its extrinsic counterpart $\phi \circ \gamma = \beta$ and we really mean $\ddot{\beta}(t) \perp T_{\beta(t)}$ is equivalent to γ being a geodesic in (P, g) . In other words, the acceleration is 0 just like we are used to for straight lines in the plane except some acceleration in the normal direction is needed to keep our curve on the surface $\phi(P)$.

In summary the theorem we would like to prove is the following:

Theorem 6.3. *With $\beta = \phi \circ \gamma$ we have $\ddot{\beta} \perp T_{\beta(t)}$ is equivalent to the geodesic equation for γ :*

$$0 = \ddot{\gamma}_r(t) + \sum_{i,j=1}^n \Gamma_{ij}^r(\gamma(t))\gamma_i'(t)\gamma_j'(t)$$

The key to the proof is the relation between the Christoffel symbols and the second order partial derivatives:

Lemma 6.3.

$$\langle \partial_i \partial_j \phi(p), \partial_k \phi(p) \rangle = \sum_{r=1}^n \Gamma_{ij}^r(p) g_{kr}(p) = \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij})$$

Proof. Dropping the evaluation at p for brevity the right hand side can be expanded using (5):

$$\frac{1}{2} \sum_{r,\ell} g_{kr} g_{r\ell}^{-1} (\partial_i g_{j\ell} + \partial_j g_{\ell i} - \partial_\ell g_{ij}) = \frac{1}{2} \sum_{\ell} \delta_{k\ell} (\partial_i g_{j\ell} + \partial_j g_{\ell i} - \partial_\ell g_{ij}) = \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij})$$

Here we used the defining equation for the inverse of g , namely $\sum_r g_{kr}g_{r\ell}^{-1} = \delta_{k\ell}$.

Writing the dot product with a \cdot and using the product rule for the dot product we have: $\partial_i g_{jk} = \partial_i(\partial_j\phi \cdot \partial_k\phi) = (\partial_i\partial_j\phi) \cdot \partial_k\phi + \partial_j\phi \cdot (\partial_i\partial_k\phi)$. write this equation three times while cyclically permuting the indices i, j, k :

$$\partial_i g_{jk} = (\partial_i\partial_j\phi) \cdot \partial_k\phi + \partial_j\phi \cdot (\partial_i\partial_k\phi) \quad (6)$$

$$\partial_j g_{ki} = (\partial_j\partial_k\phi) \cdot \partial_i\phi + \partial_k\phi \cdot (\partial_j\partial_i\phi) \quad (7)$$

$$\partial_k g_{ij} = (\partial_k\partial_i\phi) \cdot \partial_j\phi + \partial_i\phi \cdot (\partial_k\partial_j\phi) \quad (8)$$

It follows that $\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij} = 2(\partial_i\partial_j\phi) \cdot \partial_k\phi$. □

Proof. (of the Theorem)

Since the vectors $\partial_k\phi(p)$ span the tangent space $T_{\phi(p)}$ as k runs from 1 to n it suffices to prove

$$\forall k : \langle \ddot{\beta}(t), \partial_k\phi(\gamma(t)) \rangle = 0$$

By the chain rule $\dot{\beta}(t) = (\phi \circ \gamma)'(t) = \phi'(\gamma(t))\dot{\gamma}(t) = \sum_{i=1}^n \partial_i\phi(\gamma(t))\dot{\gamma}_i(t)$. Differentiating once more using both product rule and chain rule we find

$$\ddot{\beta}(t) = \sum_{i,j=1}^n \partial_j\partial_i\phi(\gamma(t))\dot{\gamma}_i(t)\dot{\gamma}_j(t) + \sum_{i=1}^n \partial_i\phi(\gamma(t))\ddot{\gamma}_i(t)$$

Taking the inner product with $\partial_k\phi(\gamma(t))$ on both sides and abbreviating $p = \gamma(t)$ we can make use of our lemma as follows

$$\begin{aligned} \langle \ddot{\beta}(t), \partial_k\phi(\gamma(t)) \rangle &= \sum_{i,j=1}^n \langle \partial_j\partial_i\phi(p), \partial_k\phi(p) \rangle \dot{\gamma}_i(t)\dot{\gamma}_j(t) + \sum_{s=1}^n \langle \partial_s\phi(p), \partial_k\phi(p) \rangle \ddot{\gamma}_s(t) = \\ &= \sum_{i,j=1}^n \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}) \dot{\gamma}_i(t)\dot{\gamma}_j(t) + \sum_{s=1}^n g_{sk}(p) \ddot{\gamma}_s(t) = 0 \end{aligned}$$

Up to a factor 2 the final line is precisely equation (3) which is equivalent to the geodesic equation. □

Lemma 6.3 actually gives us more intuition about the Christoffel symbols. They express the projection of the second derivative in terms of the basis $\partial_k\phi(p)$ of the tangent space $T_{\phi(p)}$. More precisely, for any i, j we have $\partial_i\partial_j\phi(p) = \sum_{k=1}^n \Gamma_{ij}^k(p)\partial_k\phi(p) + N_{ij}(p)$ for some vectors $N_{ij}(p)$ perpendicular to $T_{\phi(p)}$. If we denote by π_q the orthogonal projection of \mathbb{R}^m onto the linear subspace T_q then we may reformulate our lemma as

$$\pi_{\phi(p)}\partial_i\partial_j\phi(p) = \sum_{k=1}^n \Gamma_{ij}^k(p)\partial_k\phi(p) \quad (9)$$

6.6 Covariant derivative and curvature

Intuitively the meaning of the geodesic equation is that one should set the acceleration of the curve to 0 except for some Christoffel symbols that correct for the fact that space we are studying is 'curved'. In this subsection we make the notion of being curved painfully precise by introducing the Riemann curvature. It is built on a generalization of the arguments in the previous subsections. A basic geometric question that curvature answers is why all maps of the earth have to be distorted in some way. We cannot flatten the surface of a sphere onto the plane without stretching or tearing.

The new feature we want to emphasize is the invariance of our constructions under isometries. To this end we pass from differentiation with respect to the arbitrary coordinates in our chart to differentiation in any direction. Partial derivatives should be replaced by the directional derivative $\partial_v(p)F = F'(p)v$.

Since the partial derivatives of the previous subsection were viewed as functions on P , their collection really forms a bunch of vector fields. So instead of merely the directional derivative we would like to differentiate one vector field in the direction of another. At every point. This produces a new vector field as follows.

Definition 6.8. For C^1 vector fields $X, Y : P \rightarrow \mathbb{R}^n$ on open set $P \subset \mathbb{R}^n$ we define a new vector field $\partial_X Y : P \rightarrow \mathbb{R}^n$ by $(\partial_X Y)(p) = (\partial_{X(p)} Y)(p) = Y'(p)X(p)$. The commutator is $[X, Y] = \partial_X Y - \partial_Y X$.

If $P \xrightarrow{f} Q$ is injective and has injective derivative and X a vector field on P then we denote by $f'X$ the vector field on $f(P) \subset Q$ defined by $f'X(f(p)) = f'(p)X(p)$.

For example in $P = \mathbb{R}^2$ with $X(x, y) = (2x + y^2, x)$ and $Y(x, y) = (3y + x^3, y^2)$ we get $\partial_X(Y)(x, y) = \partial_{(2x+y^2, x)}Y(x, y) = Y'(x, y)(2x + y^2, x) = \begin{pmatrix} 3x^2 & 3 \\ 0 & 2y \end{pmatrix} \begin{pmatrix} 2x+y^2 \\ x \end{pmatrix} = \begin{pmatrix} 6x^3+3x^2y^2+3x \\ 2xy \end{pmatrix}$. More generally if $X = \sum_{i=1}^n X_i e_i$ and $Y = \sum_{j=1}^n Y_j e_j$ then $(\partial_X Y)(p) = \sum_{i,j} X_i(p)(\partial_i Y_j)(p)e_j$. Another way to think about $(\partial_X Y)(p)$ is that we find an integral curve γ for X so $\gamma(0) = p$ and $\dot{\gamma}(0) = X(p)$. Then $(\partial_X Y)(p) = \frac{d}{dt}Y(\gamma(t))$.

Our new operation on vector fields can be used to give an important criterion for space being 'flat'.

Lemma 6.4. *For any three vector fields X, Y, Z on \mathbb{R}^n we have*

$$\partial_X(\partial_Y Z) - \partial_Y(\partial_X Z) - \partial_{[X, Y]}Z = 0$$

Proof. To see why we simply expand the definitions, for brevity we drop the evaluation at p from our notation:

$$\partial_X(\partial_Y Z) = \sum_{i,j,k} X_k \partial_k (Y_i (\partial_i Z_j)) e_j = \sum_{i,j,k} X_k (\partial_k Y_i + Y_i \partial_k) \partial_i Z_j e_j$$

Therefore interchanging X and Y and also the indices i, k we find

$$\partial_X(\partial_Y Z) - \partial_Y(\partial_X Z) = \sum_{i,j,k} (X_i \partial_i Y_k \partial_k - Y_i \partial_i X_k \partial_k) Z_j e_j$$

On the other hand $[X, Y] = \partial_X Y - \partial_Y X = \sum_{i,k} (X_i \partial_i Y_k - Y_i \partial_i X_k) e_k$ proving the lemma. \square

As we saw in the geodesic equation the proper generalization of $\partial_X Y$ to an arbitrary Riemannian chart is called the **covariant derivative** also known as the Levi-Civita connection $\nabla_X Y$. It turns out to be related to the Christoffel symbols by $\nabla_{e_i} e_j = \sum_{i,j} \Gamma_{ij}^k e_k$. We will see that in this notation the geodesic equation would be $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$.

Definition 6.9. (Levi-Civita connection)

Denote the set of all C^2 vector fields on P by $\text{Vec}(P)$. An **LC-connection** on a Riemannian chart (P, g) is a map $\nabla : \text{Vec}(P) \times \text{Vec}(P) \rightarrow \text{Vec}(P)$ with the following properties. For any $X, Y, Z \in \text{Vec}(P)$ scalar $a \in \mathbb{R}$ and C^1 -function $f : P \rightarrow \mathbb{R}$ we have:

1. $\nabla_X(fY + Z) = f\nabla_X Y + (\partial_X f)Y + \nabla_X Z$
2. $\nabla_{fX+Z}(Y) = f\nabla_X Y + \nabla_Z Y$
3. $g(p)(\nabla_X Y(p), Z(p)) + g(p)(Y(p), \nabla_X Z(p)) = \partial_{X(p)}G(p)$, where $G(s) = g(s)(Y(s), Z(s))$
4. $\nabla_X Y - \nabla_Y X = [X, Y]$

In (\mathbb{R}^n, g_E) the formula $\nabla_X Y = \partial_X Y$ defines an LC-connection. In checking the first two axioms the hardest bit is $\nabla_X(fY) = \sum_{i=1}^n X_i \partial_i(fY) = \sum_{i=1}^n X_i (\partial_i f)Y + X_i f \partial_i Y = (\partial_X f)Y + f\nabla_X Y$. The third axiom is true because the partial derivative satisfies a product rule with respect to the standard dot product. The last axiom was our definition of the commutator so there is nothing to check.

While a little abstract this definition is encoding exactly what we were doing before. In the theorem below we will see the Christoffel symbol comes out as the coefficients of $\nabla_{e_i} e_j$ so $\nabla_{e_i} e_j = \sum_{i,j,k} \Gamma_{ij}^k e_k$. Assuming this for a moment the geodesic equation can be rewritten as $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$ with the caveat that we need to extend $\dot{\gamma}$ to a vector field V such that $V(\gamma(t)) = \dot{\gamma}(t)$ for all t . Indeed using $V = \sum_{j=1}^n V_j e_j$ the properties of the connection give us

$$\nabla_{\dot{\gamma}} \dot{\gamma} = \sum_{i,j} \dot{\gamma}_i \nabla_{e_i} V_j e_j = \sum_{i,j} \dot{\gamma}_i (\partial_i V_j) e_j + \dot{\gamma}_i V_j \nabla_{e_i} e_j = \sum_j \frac{d}{dt} V_j(\gamma(t)) e_j + \sum_{i,j,k} \Gamma_{ij}^k \dot{\gamma}_i V_j e_k = \left(\ddot{\gamma}_k(t) + \sum_{i,j,k} \Gamma_{ij}^k \dot{\gamma}_i V_j \right) e_k$$

What is important is that the Riemannian metric g determines the LC-connection uniquely, much like our discussion in the previous subsection.

Theorem 6.4. (Fundamental lemma of Riemannian geometry)

There exists a unique Levi-Civita connection ∇ on Riemannian chart (P, g) .

Also $\nabla_{e_i} e_j = \sum_{i,j,k} \Gamma_{ij}^k e_k$.

Proof. For simplicity we write the Riemannian metric as $\langle \cdot, \cdot \rangle$ and do not explicitly write its dependence on the point. Condition 3) is then abbreviated as

$$\partial_X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_Z X \rangle + \langle Y, [X, Z] \rangle \quad (10)$$

We start by proving uniqueness of the LC-connection by proving it is determined entirely by g much like the proof of Lemma 6.3. For vector fields X, Y, Z we write Equation (10) three times cycling around X, Y, Z :

$$\begin{aligned}\partial_X \langle Y, Z \rangle &= \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_Z X \rangle + \langle Y, [X, Z] \rangle \\ \partial_Y \langle Z, X \rangle &= \langle \nabla_Y Z, X \rangle + \langle Z, \nabla_X Y \rangle + \langle Z, [Y, X] \rangle \\ \partial_Z \langle X, Y \rangle &= \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Y Z \rangle + \langle X, [Z, Y] \rangle\end{aligned}$$

Subtracting the third equation from the sum of the first two we can solve for $\langle \nabla_X Y, Z \rangle$ to get

$$\langle \nabla_X Y, Z \rangle = \frac{1}{2} \left(\partial_X \langle Y, Z \rangle + \partial_Y \langle Z, X \rangle + \partial_Z \langle X, Y \rangle - \langle Y, [X, Z] \rangle - \langle Z, [Y, X] \rangle + \langle X, [Z, Y] \rangle \right) \quad (11)$$

Uniqueness follows from this equation since if we had two LC-connections $\nabla, \tilde{\nabla}$ then for all Z we would find $\langle \nabla_X Y, Z \rangle = \langle \tilde{\nabla}_X Y, Z \rangle$ which is only possible if $\nabla_X Y = \tilde{\nabla}_X Y$. Existence follows from the same equation (11) because taking $X, Y, Z = e_i, e_j, e_k$ we find

$$\langle \nabla_{e_i} e_j, e_k \rangle = \frac{1}{2} \left(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij} \right) = \sum_{r=1}^n \Gamma_{ij}^r(p) g_{kr}(p)$$

using $[e_a, e_b] = 0$ and the definition of the Christoffel symbols (5). Using the inverse g^{-1} we thus find $\nabla_{e_i} e_j = \sum_{k=1}^n \Gamma_{ij}^k e_k$ as desired. This proves that an LC-connection exists because by properties 1) and 2) we have

$$\nabla_X Y = \sum_{i,j} X_i \nabla_{e_i} (Y_j e_j) = \sum_{i,j} X_i \partial_i Y_j e_j + X_i Y_j \nabla_{e_i} e_j = \sum_{i,j} X_i \partial_i Y_j e_j + \sum_k X_i Y_j \Gamma_{ij}^k$$

□

When our Riemannian metric comes from pulling back the Euclidean metric along a suitable map $\phi : P \rightarrow \mathbb{R}^m$ as in the previous subsection then the LC-connection is easier to understand. It is basically using the LC-connection in (\mathbb{R}^m, g_E) given by Definition 6.8 and then projecting the result onto the tangent space in \mathbb{R}^m . More precisely set $g = \phi^* g_E$ to be a metric on $P \subset \mathbb{R}^n$ and $\phi : P \rightarrow \mathbb{R}^m$ as in the previous subsection. We claim that the LC-connection in that case is given by $\phi' \nabla_X Y = \pi \nabla_{\tilde{X}} \tilde{Y}$ where $\pi(p) : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is the linear map that projects orthogonally onto the tangent space $\text{Image}(\phi'(p))$ and \tilde{X} and \tilde{Y} are vector fields on \mathbb{R}^m such that $\tilde{X} \circ \phi = \phi' X$ and likewise $\tilde{Y} \circ \phi = \phi' Y$. Also recall that by $\phi' X$ we mean the vector field on $\text{Image}(\phi)$ defined by $\phi' X(\phi(p)) = \phi'(p) X(p)$. This means that on the image of ϕ we get

$$\nabla_{\tilde{X}} \tilde{Y}(\phi(p)) = (\tilde{Y}(\phi(p)))' \tilde{X}(\phi(p)) = (\phi' Y)'(p) X(p) = \partial_X(\phi' Y)(p)$$

When $X = e_i$ and $Y = e_j$ we get $\phi' Y = \partial_j \phi$ so that by equation (9) $\phi' \nabla_{e_i} e_j = \pi \partial_i \partial_j \phi = \sum_{k=1}^n \Gamma_{ij}^k(p) \partial_k \phi$. Since $\phi' e_k = \partial_k \phi$ we conclude $\nabla_{e_i} e_j = \sum_{k=1}^n \Gamma_{ij}^k(p) e_k$.

Lemma 6.5. (Isometry invariance)

For any isometry $P \xrightarrow{f} Q$ between Riemannian charts we have $f'(\nabla_X Y) = \nabla_{f'X} f'Y$. Also γ is a geodesic if and only if $f \circ \gamma$ is a geodesic.

Proof. Define a map $\Delta : \text{Vec}(P) \times \text{Vec}(P) \rightarrow \text{Vec}(P)$ by $\Delta_X Y = (f^{-1})' \nabla_{f'X} f'Y$. If we can show that Δ is an LC-connection on P then by uniqueness $\nabla = \Delta$ so $\nabla_X Y = (f^{-1})' \nabla_{f'X} f'Y$ proving $f'(\nabla_X Y) = \nabla_{f'X} f'Y$.

Next, if γ is a geodesic in P and $\beta = f \circ \gamma$ then $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$. Applying f' to both sides yields $0 = f' \nabla_{\dot{\gamma}} \dot{\gamma} = \nabla_{f' \dot{\gamma}} f' \dot{\gamma} = \nabla_{\dot{\beta}} \dot{\beta}$, showing β is a geodesic too. Conversely if β is a geodesic then so is γ because f^{-1} is an isometry too. □

We we have seen that for the LC-connection on (\mathbb{R}^n, g_E) we had $\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z = 0$. For other Riemannian charts this is not at all the case.

Definition 6.10. (Riemannian Curvature)

The Riemann curvature of n -dimensional Riemannian chart (P, g) is the map $R : \text{Vec}(P) \times \text{Vec}(P) \times \text{Vec}(P) \rightarrow \text{Vec}(P)$ defined by

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

It follows that curvature is invariant under isometries in the following sense

Lemma 6.6. (Isometry invariance of curvature)

For any isometry $P \xrightarrow{f} Q$ between Riemannian charts we have $f'(R(X, Y)Z) = R(f'X, f'Y)f'Z$

Proof. Using Lemma 6.5

$$f'(R(X, Y)Z) = \nabla_{f'X} \nabla_{f'Y} f'Z - \nabla_{f'Y} \nabla_{f'X} f'Z - \nabla_{[f'X, f'Y]} f'Z = R(f'X, f'Y)f'Z$$

□

Concretely the coefficients of the Riemann curvature can be computed in terms of Christoffel symbols.

Lemma 6.7. *The curvature is linear in X, Y, Z and for C^2 -functions $f : P \rightarrow \mathbb{R}$ we have $fR(X, Y)Z = R(X, Y)fZ = R(X, fY)Z = R(fX, Y)Z$. If we set $R(e_i, e_j)e_k = \sum_{\ell} R_{i,j,k}^{\ell} e_{\ell}$ then*

$$R_{i,j,k}^{\ell} = \partial_i \Gamma_{jk}^{\ell} - \partial_j \Gamma_{ik}^{\ell} + \sum_r \Gamma_{jk}^r \Gamma_{ir}^{\ell} - \Gamma_{ik}^r \Gamma_{jr}^{\ell}$$

Proof. Linearity in X, Y, Z follows directly from the properties of the LC-connection. The property with function f is left to the reader as an exercise.

Notice $[e_i, e_j] = 0$ so $R(e_i, e_j)e_k = \nabla_{e_i} \nabla_{e_j} e_k - \nabla_{e_j} \nabla_{e_i} e_k$. Now

$$\nabla_{e_i} \nabla_{e_j} e_k = \nabla_{e_i} \sum_r \Gamma_{jk}^r e_r = \sum_r \partial_i \Gamma_{jk}^r e_r + \sum_{r,s} \Gamma_{jk}^r \Gamma_{ir}^s e_s = \sum_{\ell} \left(\partial_i \Gamma_{jk}^{\ell} + \sum_r \Gamma_{jk}^r \Gamma_{ir}^{\ell} \right) e_{\ell}$$

Therefore swapping i and j and subtracting we obtain the desired result. □

As promised we can now understand how the geometry of the sphere is intrinsically different from that of the plane.

Theorem 6.5. *Suppose (P, g) is a Riemannian chart such that $g = \phi^* g_E$ and $\phi : P \rightarrow S^2 \subset \mathbb{R}^3$ is an injective C^2 map with injective derivative. Also take (Q, h) to be a part of the plane with the Euclidean metric so $Q \subset \mathbb{R}^2$ and $h = g_E$. There is no isometry between (P, g) and (Q, h) .*

Proof. Restricting the domains P, Q if necessary, Lemma 6.6 tells us we can assume without loss of generality that $P = (0, \pi) \times (-\pi, \pi)$ and that ϕ is the usual spherical coordinate map $\phi(\mu, \lambda) = (\cos \lambda \sin \mu, \sin \lambda \sin \mu, \cos \mu)$. By computing the Christoffel symbols in this case we see R is not zero in P . For example $R_{1,2,2}^1 = \sin^2 \mu$ is non-zero.

In Q the curvature maps any triple of vector fields to 0 by Lemma 6.4. □

Of course the Riemann curvature does much more than distinguishing the sphere from the plane but we leave this for further study except for a few general comments. Since the Riemann curvature coefficients contain a lot of information it is common to package them into simpler quantities known as the Ricci curvature coefficients $R_{ij} = \sum_{k=1}^n R_{kij}^k$ and the scalar curvature $S = \sum_{i,j} g_{ij}^{-1} R_{ij}$. For surfaces in 3-space all the information of the Riemann curvature is already contained in S and is known to be twice the Gauss curvature which describes how the normal vector turns as one walks around a point. In general a simple interpretation of the scalar curvature is possible in any Riemannian chart by comparing the circumference of a small (geodesic) circle with radius r to $2\pi r$. Similar expressions are available in terms of area and volume. For future reference we state a definition for volume in a Riemannian chart.

Definition 6.11. (Volume)

For a closed subset D of a Riemannian chart (P, g) we define the determinant of the metric $\det g : P \rightarrow \mathbb{R}$ by $\det g(p) = \det(g_{ij}(p))_{i,j=1,\dots,n}$, the determinant of the matrix formed by the functions g_{ij} . The volume of D is defined to be

$$\text{vol}(D) = \int_D \sqrt{\det g} dx_1, \dots, dx_n$$