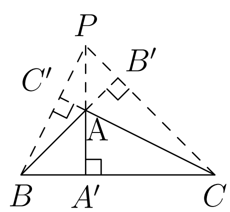
Affine Geometry

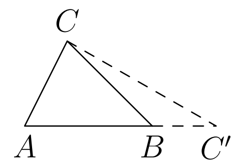
0.1 Oriented Lengths, Areas, Volumes

A careful reader might have noted that we have cheated slightly in lecture 1: we formulated Ceva's theorem only for the case that three Cevians meet inside the triangle. Then we applied it in a situation when Cevians don't necessarily meet inside the triangle. For instance this is the case for three heights in an obtuse triangle.



Later, when we introduced the notion of ratio of oriented lengths of segments, we had refined the formulation of Ceva's theorem to include that case as well. However we never gave a proof of Ceva's theorem that covers this case. (In fact the proof with center of masses does work if we allow negative masses, but we didn't emphasize that enough). Let us see what the problem can be. In the proof of Ceva's theorem we used multiple times an identity of the following kind: if C' is a point on side AB in triangle ABC, then $\text{Area}ABC = \text{Area}AC'C + \text{Area}CC'B$. It is correct if the point C' is inside the segment AB, but it becomes wrong once the point C' goes beyond the point A or B along the line AB. Indeed, in the picture below the following identity holds: $\text{Area}ABC = \text{Area}AC'C - \text{Area}CC'B$.

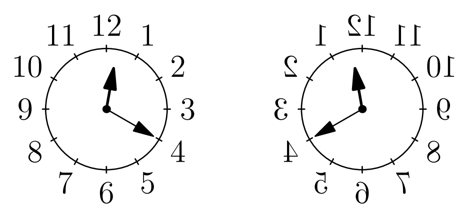
Note that the area of triangle $CC'B$ must be taken with minus sign. In this section we are going to revisit notions of length, area and volume to make them better behaved and easier to deal with.



We will start with recollecting what oriented length is on an oriented line. An oriented line is just a line with a choice of preferred direction on it (one-sided road is an example). On such a line we can measure the oriented distance from point A on the line to point B on the same line. It's the distance between A and B taken with the sign "+" if the direction from A to B agrees with the chosen direction along the line, and sign "-" otherwise (this will make perfect sense to a driver on the one-sided road — he will have to drive negative distances in reverse). One thing we used several times in proofs, but didn't emphasize is that it is always the case that while the equality of lengths $|AB| = |AC| + |CB|$ holds only for points C inside the segment AB, the equality of signed lengths $AB = AC + CB$ is true for any choice of point C on the line.

Directed area

The notion of directed area makes sense on a plane with a choice of one of the two possible orientations: clockwise and counterclockwise.



We say that oriented area of triangle ABC is the area of triangle ABC taken with sign "+" if in going from A to B to C to A one rotates in the direction that was chosen on the plane and sign "-" otherwise. The concept can be understood better by looking at a picture contrasting positive oriented area and negative one.

Note that the sign of oriented area (as is also in the case of oriented length) depends on the order in which the vertices of the triangle are traversed: $\text{Area}ABC = \text{Area}BCA = \text{Area}CAB = -\text{Area}ACB = -\text{Area}CBA = -\text{Area}BAC$. For oriented areas it is always true that if ABC is a triangle and P is any point in the plane, then $\text{Area}ABC = \text{Area}ABP + \text{Area}BCP + \text{Area}CAP$.

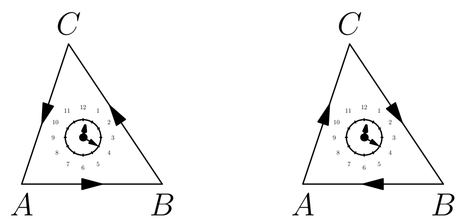
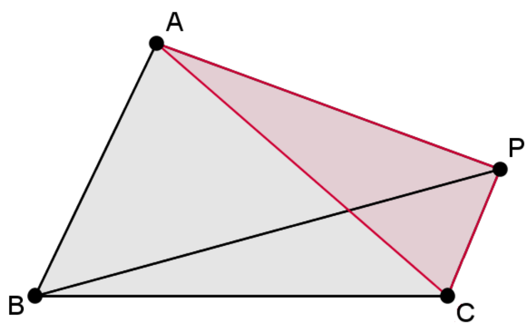
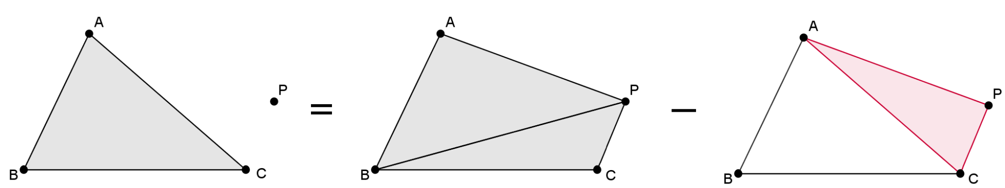


Figure 1: With the clockwise orientation of the plane, the oriented area of ABC is negative, while that of ACB is positive



For non-oriented areas this is the case only when the point P is inside the triangle ABC. In other cases we should change the signs in the equation above.



(Thanks to Chenchen Li for the illustration)

To define the oriented area of any closed polygon $A_1A_2\dots A_n$ we can choose a point P anywhere in the plane and define the area of $A_1A_2\dots A_n$ to be the sum $\mathrm{Area}A_1A_2P + \mathrm{Area}A_2A_3P + ... + \mathrm{Area}A_nA_1P$. The reader is advised to check that this definition does not depend on the choice of the point P in the plane.

In space we will consider one of the two orientations - right-handed and left-handed. Suppose we've chosen the right-handed orientation.

Space for figure.

Then to determine the sign of oriented volume of a tetrahedron $ABCD$ we can curve the index finger of our right hand along the direction of rotation from A to B to C and then look at our thumb: if it points in the direction of the half-space containing the vertex D, then the volume is positive. Otherwise it will be negative.

Space for figure.

The identity $Vol(ABCD) = Vol(ABCP) + Vol(PBCD) + Vol(CDAP) + Vol(PDAB)$ holds for any point P in the space if the volumes are interpreted as oriented volumes.

Space for figure.

To discuss further properties of oriented volumes and their generalizations, we will have to recall what linear functions are. We will remind the definitions in the next section. If the reader feels free with the material on linear functions, he may safely skip it.

0.2 Vector spaces and linear functions

Recall that if we choose an origin in a plane or space, we can identify points with the corresponding vectors from the origin. Once we do so, we get the possibility to add these vectors using the parallelogram rule and multiply them by real numbers (i.e. scale them). People have crystallized the essence of this situation into the definition of an abstract vector space.

According to the definition, a vector space over a field of scalars K is a set with operations of addition (one can add any two vectors and get a vector) and multiplication by scalars (one can take any vector and multiply it by a scalar from the field K to get another vector). These operations should satisfy some familiar properties that addition and scaling of two- or three-dimensional vectors satisfy (like $v + w = w + v, \lambda(v + w) = \lambda v + \lambda w$ etc). In this lecture the field of scalars will be just the field of real numbers R, but in later lectures the fields C (complex numbers) and Q (rational numbers) will be also useful.

The main examples of vector spaces are the spaces R^n: for every natural number n the space R^n consists of n-tuples of real numbers $(a_1, ..., a_n)$ with operations of addition and multiplication by scalar defined coordinatewise $((a_1, \ldots, a_n) + \lambda(b_1, \ldots, b_n) = (a_1 + \lambda b_1, \ldots, a_n + \lambda b_n)$.

The space of vectors in the plane can be identified with R^2 once a basis is chosen, i.e. when we choose what two non-collinear vectors e_1 and

e_2 will correspond to $(1,0)$ and $(0,1)$. Once these are chosen, any other vector can be represented as a linear combination $a_1e_1 + a_2e_2$ of these two (a_1, a_2 are some real numbers). Such a vector will be identified with the pair (a_1, a_2). Similarly we can identify the space of spacial vectors with R^3 by choosing three non-coplanar vectors e_1, e_2, e_3 that will be identified with $(1,0,0), (0,1,0), (0,0,1)$.

Thus the vector space R^n is the natural n-dimensional analogue of our notions of two- and three-dimensional spaces.

One notion that is important for the discussion of vector spaces is that of linear mapping: a mapping between vector spaces that preserves linear structure. More precisely, a function $f : V- > W$ between K-vector spaces V, W is a linear mapping if $f(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 f(v_1) + \lambda_2 f(v_2)$ for any $v_1, v_2 \in V, \lambda_1, \lambda_2$ in the field K.

The linear maps from the vector space V to the field K (considered as one-dimensional vector space) are called linear functionals.

Let's examine some examples of linear maps:

1. Rotations around the origin are linear maps. If we have a rectangular coordinate system in our space, then the rotations have the form $(x, y)- > (cos(\phi)x + sin(\phi)y, -sin(\phi)x + cos(\phi)y)$. Space for figure.

2. Reflections in an axis passing through the origin. For instance the reflection in the x-axis is given in coordinates as $(x, y)- > (x, -y)$. Refer to exercise for representation in coordinates of a more general reflection.

3. Shearing in direction of x-axis given in coordinates by $(x, y)- > (x + \lambda y, y)$.

4. Scaling (aka homothety) with center at the origin. Scaling by factor λ is given by the map $v- > \lambda v$.

5. Stretching by factor λ_1 in x-direction and factor λ_2 in y-direction given in coordinates by $(x, y)- > (\lambda_1 x, \lambda_2 y)$.

6. Parallel projection along some direction onto an axis that passes through origin. For instance the projection onto the line $x = y$ along the direction of x-axis is given by $(x, y)- > (x, x)$.

Note that rotations preserve distances, angles and oriented areas, reflections preserve distances, but reverse the signs of angles and oriented areas, shearings preserve areas, but not distances or angles, homotheties preserve angles, but not distances or areas and finally the last mapping doesn't preserve any of the three.

One can represent linear maps between finite-dimensional spaces by means of matrices. let $f : V- > W$ be any linear map and let $v_1, ..., v_n$ be a basis for the vector space V. Then any vector $v \in V$ can be uniquely represented as $v = \lambda_1 v_1 + ... + \lambda_n v_n$ for some scalars $\lambda_1, ..., \lambda_n \in K$. Hence for such v we have $f(v) = \lambda_1 f(v_1) + ... + \lambda_n f(v_n)$. Thus the value of the function f on any vector is determined by its value on the basis vectors v_j. Now if $w_1, ..., w_m$ is a basis for the vector space W, then each vector $f(v_j)$ can be expressed as a linear combination of the w_i: $f(v_j) = a_{1,j}w_1 + a_{2,j}w_2 + ... + a_{m,j}w_m$.

The scalars $a_{i,j}$ determine the linear function f uniquely. Thus the linear mapping f can be identified with the mapping $v- > A \cdot v$, where A is the matrix

$$\begin{pmatrix} a_{1,1} & ... & a_{1,n} \\ ... & & ... \\ a_{m,1} & ... & a_{m,n} \end{pmatrix}$$

of the function f relative to the bases $v_1, ..., v_n$ of V and $w_1, ..., w_m$ of W.

0.3 Volume as multilinear function, determinants, Cramer's rule

Recall that if we choose an orientation on the line R, we can measure the oriented length of any vector OA from the origin to point A. This length is a linear function from R to R. We want to explore what linearity properties the oriented areas and volumes have.

Let's start with oriented areas in R^2. For this suppose we've chosen orientation of R^2, e.g. counterclockwise one. Let $V(v_1, v_2)$ denote the oriented area of the parallelogram with vertices $0, v_1, v_1 + v_2, v_2$. Let $v_0^\perp$ denote a unit length vector orthogonal to v such that the basis $v_1, v_1^\perp$ is oriented positively. Then the area $V(v_1, v_2)$ is equal to the product of length (unoriented) of vector v and the oriented length of the projection of w to the line spanned by $v_0^\perp$ (the orientation on this line being given by demanding that the oriented length of $v_1^\perp$ is $+1$). This length is equal to the scalar product $< v_1^\perp, w >$ and as such it is linear function in w.

What follows from this line of arguments is that the function $V(v_1, v_2)$ is linear in the second parameter (if the first one is being held fixed, i.e. $V(v_1, v_2 + \lambda v_2') = V(v_1, v_2) + \lambda V(v_1, v_2')$. By a similar argument (or by using the identity $V(v_1, v_2) = -V(v_2, v_1)$) we can conclude that $V(v_1, v_2)$ is linear in the first parameter as well. Thus the oriented area is linear in each of its parameters. This seemingly simple observation allows us to express the area of parallelogram explicitly in terms of coordinates of its vertices. Namely let

e_1, e_2 be two basis vectors. Let $v_1 = a_{11}e_1 + a_{21}e_2, v_2 = a_{12}e_1 + a_{22}e_2$. Then

$$V(v_1, v_2) = V(a_{11}e_1 + a_{21}e_2, v_2) = a_{11}V(e_1, v_2) + a_{21}V(e_2, v_2)$$

$$= a_{11}V(e_1, a_{12}e_1 + a_{22}e_2) + a_{21}V(e_2a_{12}e_1 + a_{22}e_2)$$

$$= a_{11}a_{12}V(e_1, e_1) + a_{11}a_{22}V(e1, e2) + a_{21}V(e_2, e_1) + a_{21}a_{22}V(e_2, e_2)$$

Since $V(e_1, e_1) = 0, V(e_2, e_2) = 0$ (the parallelogram degenerates to a segment in these cases) and $V(e_2, e_1) = -V(e_1, e_2)$, we get that $V(a_{11}e_1 + a_{21}e_2, a_{12}e_1 + a_{22}e_2) = (a_{11}a22 - a_{12}a_{21})V(e_1, e_2)$.

The first factor in this expression is the determinant of the matrix

$$\begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix}$$

(one can think of this more invariantly as the determinant of the transformation sending the basis vectors e_1, e_2 to vectors v_1, v_2).

The second factor is the volume of the parallelogram spanned by the vectors e_1, e_2. If we choose the vectors e_1, e_2 to be the vectors $(1, 0)$ and $(0, 1)$ in R^2, this area is just $+1$. So we established a close relation between oriented area of parallelograms and determinants of 2x2 matrices.

Let's see whether the same applies to volumes. Choose some orientation of the space. Let $V(v_1, v_2, v_3)$ denote the oriented volume of parallelepiped spanned by the vectors v_1, v_2, v_3.

Space for figure

Let $v_{12}^{\perp}$ denote a unit vector perpendicular to the plane of v_1, v_2 in the direction such that the basis $v_1, v_2, v_{12}^{\perp}$ is positively oriented. Then the volume $V(v_1, v_2, v_3)$ is equal to the product of the unoriented area of parallelogram spanned by v_1, v_2 and the oriented length of the projection of v_3 to $v_{12}^{\perp}$. This length is equal to $< v_{12}^{\perp}, v_3 >$ and thus it is linear in the parameter v_3. Hence the volume itself is also linear in the parameter v_3. By using similar argument (or using $V(v_1, v_2, v_3) = V(v_2, v_3, v_1) = V(v_3, v_1, v_2)$) we get that the volume $V(v_1, v_2, v_3)$ is linear in each of its parameters. In the same manner we did it for areas, we can now write what the volume of parallelepiped spanned by vectors v_1, v_2, v_3 is in terms of the coordinates of the vectors v_1, v_2, v_3 with respect to a basis e_1, e_2, e_3. Indeed, if $e_j = a_{1,j}e_1 + a_{2,j}e_2 + a_{3,j}e_3$ for $j = 1..3$, then $V(v_1, v_2, v_3) = V(\sum_i a_{i,1}e_i, \sum_j a_{j,2}e_j, \sum_k a_{k,3}e_k) = \sum_{i,j,k} a_{i_1}a_{j,2}a_{k,3}V(e_i, e_j, e_k)$. Now if at least two indices among i, j, k are equal, then $V(e_i, e_j, e_k) = 0$ (since the parallelepiped degenerates to a parallelogram in this case). The remaining six volumes are related by $V(e_1, e_2, e_3) = V(e_2, e_3, e_1) = V(e_3, e_1, e_2) = -V(e_2, e_1, e_3) = -V(e_3, e_2, e_1) = -V(e_1, e_3, e_2)$. Thus $V(v_1, v_2, v_3) = (a_{1,1}a_{2,2}a_{3,3} + a_{2,1}a_{3,2}a_{1,3} + a_{3,1}a_{1,2}a_{2,3} - a_{2,1}a_{1,2}a_{3,3} + a_{3,1}a_{2,2}a_{1,3} - a_{1,1}a_{3,2}a_{2,3})V(e_1, e_2, e_3)$, or $V(v_1, v_2, v_3) = det(A)V(e_1, e_2, e_3)$ where the matrix A is the matrix

$$\begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{matrix}$$

(one can think instead of the matrix of the transformation of R^3 to itself sending the basis vectors e_1, e_2, e_3 to the vectors v_1, v_2, v_3 - the matrix of this transformation with respect to the basis e_1, e_2, e_3 is A).

The discussion above suggests two ways of extension of the notion of volume to higher dimensional spaces: one is an inductive definition formulated in geometric terms, namely the volume of the n-dimensional parallelepiped spanned by vectors $v_1, ..., v_n$ is the $n-1$-dimensional volume of the "base" parallelepiped spanned by $v_1, ..., v_{n-1}$ multiplied by the "height" - the length of projection of v_n onto the line orthogonal to hyperplane spanned by $v_1, ..., v_{n-1}$ and taken with appropriate sign to account for orientations. The other way is to define the volume as the algebraic expression of the determinant of the matrix, whose columns are the coordinates of the vectors $v_1, ..., v_n$ with respect to some fixed basis $e_1, ..., e_n$. As we've seen in dimensions 2 and 3 the two definitions agree: determinants have meaning of volumes and volumes can be computed as determinants.

0.4 Cramer's rule

Let v be a vector in R^3 (we will write our computations in R^3, but the case of R^n for any positive integer n is no different). Let e_1, e_2, e_3 be a basis of R^3. Then we know that we can express v as a linear combination of e_1, e_2, e_3: $v = \lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3$. How can we find these λ's?

One way is by using the multilinearity property of volumes we've seen in previous section: we can expand the volume $V(v, e_2, e_3)$ as $V(\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3, e_2, e_3) = \lambda_1 V(e_1, e_2, e_3)$. Then we get that $\lambda_1 = \frac{V(v,e_2,e_3)}{V(e_1,e_2,e_3)}$. Similarly $\lambda_2 = \frac{V(e_1,v,e_3)}{V(e_1,e_2,e_3)}$ and $\lambda_3 = \frac{V(e_1,e_2,v)}{V(e_1,e_2,e_3)}$. This is the geometrical meaning of the familiar Cramer's rule that expresses the solution of linear system of equations $A \cdot \lambda = v$ using determinants. One just has to take the e_i's from our previous discussion to be the columns of the matrix A.

0.5 Barycentric coordinates

We will now combine the ideas of center of mass and linearity of volumes together to get convenient coordinated in plane naturally associated with a given triangle in it.

So let ABC be a (non-degenerate) triangle. We've seen that the vector sum $\lambda_1 OA + \lambda_2 OB + \lambda_3 OC$ doesn't depend on the choice of origin O provided that $\lambda_1 + \lambda_2 + \lambda_3 = 1$. We will write such a sum as $\lambda_1 A + \lambda_2 B + \lambda_3 C$ (a sort of linear combination of points A,B,C). Such a sum is always in the plane of triangle ABC (because one can choose the origin O to be in the plane A, B, C). We now ask the question: can an arbitrary point P in the plane of ABC be represented as a combination $\lambda_1 A + \lambda_2 B + \lambda_3 C$ for some real $\lambda_1, \lambda_2, \lambda_3$ with $\lambda_1 + \lambda_2 + \lambda_3 = 1$? If so, is the representation unique and what is the meaning of coefficients $\lambda_1, \lambda_2, \lambda_3$ in it?

So let P denote a point in plane ABC. Choose the origin O to lie outside the plane ABC, so that OA, OB, OC are now three linearly independent vectors in three-dimensional space. Since three linearly independent vectors in R^3 form a basis, the vector OP can be written uniquely as a linear combination of the vectors OA, OB, OC: $OP = \lambda_1 OA + \lambda_2 OB + \lambda_3 OC$. To see that in this representation $\lambda_1 + \lambda_2 + \lambda_3 = 1$ we can project all the vectors onto a line passing through points O, P, the projection being in direction parallel to the plane ABC. Since P, A, B, C lie in the plane along which we are projecting, they all map to the same point, say K. The vector OK is non-zero, since O doesn't lie in the plane ABC. Now we can apply our projection to the equality $OP = \lambda_1 OA + \lambda_2 OB + \lambda_3 OC$ and use that projection is a linear map to get that $OK = \lambda_1 OK + \lambda_2 OK + \lambda_3 OK$, or $1 = \lambda_1 + \lambda_2 + \lambda_3$.

Moreover, we already know how to find the coefficients $\lambda_1, \lambda_2, \lambda_3$ using Cramer's rule: for instance λ_1 is equal to $\frac{V(OP,OB,OC)}{V(OA,OB,OC)}$. Now the volume of the parallelepiped spanned by the vectors OP, OB, OC is equal to 6 times the volume of tetrahedron $OPBC$ and similarly the volume $V(OA, OB, OC)$ is 6 times the volume of the tetrahedron $OABC$. These two tetrahedra share the same height to the face lying in the plane ABC, hence the ratio of their volumes is the same as the ratio of areas of their bases lying in the plane ABC. Finally we get $\lambda_1 = \frac{\text{Area}PBC}{\text{Area}ABC}$. Similarly $\lambda_2 = \frac{\text{Area}APC}{\text{Area}ABC}$ and $\lambda_3 = \frac{\text{Area}APC}{\text{Area}ABC}$.

exercise: put O at B and reprove the formulas.

exercise: verify that the answer above satisfies $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

Notice that when $\lambda_1 + \lambda_2 + \lambda_3 = 1$, the expression $\lambda_1 A + \lambda_2 B + \lambda_3 C$ is just the center of mass of point masses $\lambda_1, \lambda_2, \lambda_3$ put at points A, B, C respectively. To summarize, we've proved the following: for any point P in the plane ABC we can find unique set of three masses $\lambda_1, \lambda_2, \lambda_3$ with total mass $\lambda_1 + \lambda_2 + \lambda_3$ equal to 1, so that when we put them at points A, B, C the center of mass $\lambda_1 A + \lambda_2 B + \lambda_3 C$ coincides with P. These masses can be expressed as $\lambda_1 = \frac{\text{Area}PBC}{\text{Area}ABC}, \lambda_2 = \frac{\text{Area}APC}{\text{Area}ABC}, \lambda_3 = \frac{\text{Area}APC}{\text{Area}ABC}$. These masses $\lambda_1, \lambda_2, \lambda_3$ are called the barycentric coordinates of the point P (with respect to triangle ABC).

0.6 Affine geometry

We are now going to discuss a kind of geometry where we can start with bare minimum - knowing what points are collinear and what lines intersect. We shall see that even though notions of length and angles are not defined in this geometry, some remarkable theorems can still be proved.

The space that we will work with will be the usual space R^n. Consider the following group of transformations of R^n to itself - the group of transformations $L_{A,b}$ given by $L_{A,b}(x) = A \cdot x + b$, where A is some invertible matrix and b is some vector. We will be interested in the question "what geometrical notions are invariant under all the transformations from this group?".

Let us see for example that the notion of a line is invariant. Every line can be parametrized as the set of points $\{P + \lambda v | \lambda \in R\}$, where P is some point on the line and v is direction along this line. After applying the transformation $L_{A,b}$ to this set, we get the set $\{L_{A_b}(P + \lambda v) | \lambda \in R\}$, or, equivalently, $\{(A \cdot P + b) + \lambda(A \cdot v) | \lambda \in R\}$, which is of course the line passing through the point $A \cdot P + b$ with direction vector $A \cdot v$.

Furthermore, the notion of parallel lines is preserved as well: all lines with direction vector v go after application of transformations $L_{A,b}$ to lines in direction $A \cdot v$.

Another notion that is preserved is that of quotient of directed lengths: indeed, suppose that B_1, B_2, B_3 are three collinear points on the line $\{P + \lambda v\}$. Let $B_1 = P + \lambda_1 v$, $B_2 = P + \lambda_2 v$, $B_3 = P + \lambda_3 v$. Then the quotient $\frac{B_1 B_2}{B_2 B_3}$ is equal to $\frac{(P+\lambda_2 v)-(P+\lambda_1 v)}{(P+\lambda_3 v)-(P+\lambda_2 v)} = \frac{(\lambda_2 - \lambda_1)v}{(\lambda_3 - \lambda_1)v} = \frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_1}$. If we apply $L_{A,b}$ to the points B_1, B_2, B_3 we get $L_{A,b}(B_i) = (A \cdot P + b) + \lambda_i A \cdot v$ and so the quotient of lengths $\frac{B_1B_2}{B_2B_3}$ gets mapped to $\frac{L_{A,b}(B_1)L_{A,b}(B_2)}{L_{A,b}(B_2)L_{A,b}(B_3)} = \frac{\lambda_2 A\cdot v - \lambda_1 A\cdot v}{\lambda_3 A\cdot v - \lambda_1 A\cdot v} = \frac{\lambda_2-\lambda_1}{\lambda_3-\lambda_1}$.

Note however that many standard Euclidean notions, like lengths and angles, are not preserved by these transformations.

For instance we can map any non-degenerate planar triangle ABC to any other non-degenerate triangle $A'B'C'$ by means of one of the maps $L_{M,b}$. Indeed, there is a unique matrix M mapping the two linearly independent vectors AB and AC to the two vectors $A'B'$ and $A'C'$. If we choose now b as the vector by which we have to translate MA to get A', the transformation $L_{M,b}$ will be the one we are looking for (exercise: verify that $L_{M,b}(B) = B'$).

Now it is time to discuss what geometry is in general.

0.7 What is geometry?

The question in the title is not as easy as it might sound and involves answering subquestions like: what spaces do we allow to study? What figures have geometric meaning? What is a geometric theorem? And so on and so on. The answers should be general enough to include at least all the meaningful examples we considered so far, but restrictive enough to exclude too pathological examples. One way to make precise what we mean by "geometry" is to follows F. Klein's ideas in Erlangen program: let X be any set. This will be our space. Let G be a group of maps from X to itself*. This will be our group of transformations. Now we can define a geometric figure: consider all subsets of X up to the following equivalence relation. Two subsets are called equivalent (sometimes simply referred to as "equal") if there is a transformation from group G mapping onto another. The term "group" has a precise abstract definition. Here however we only need the following definition - a group of transformations of space X is a subset of all invertible maps of X to itself closed under composition. Equivalent classes of such subsets will be called "geometrical figures".

We, admittedly, made a definition too general to be useful - we allowed all kinds of "spaces" X and all kinds of groups G. A meaningful definition would restrict these notions. We, instead, will consider some examples. Example 1 - Euclidean geometry. The space X from the definition of geometry will be R^n and the group G will be the group of distance-preserving maps of R^n to itself. In this geometry all points are of course equivalent (one can use a translation to map any point in R^n to any other), but two pairs of points are equivalent if and only if the distance between the two points of the first pair is the same as the distance between the two points of the other pair.

Example 2 - Affine geometry. The space X is again R^n, but the group of transformations is now the group of affine transformations $\{x- > M\cdot x+b|M$ is invertible matrix, b is a vector in $R^n\}$ (exercise: prove it is a group of transformations). In this geometry not only all pairs of distinct points are equivalent, but we just proved that all non-degenerate triangles are equivalent as well. However if one considers triples A, B, C of collinear points, then the ratio of lengths $\frac{AB}{BC}$ is invariant under affine transformations, so not all triples are equivalent. So we are lead to believe that in affine geometry the notion of ratio of lengths of collinear segments is a meaningful geometric object.

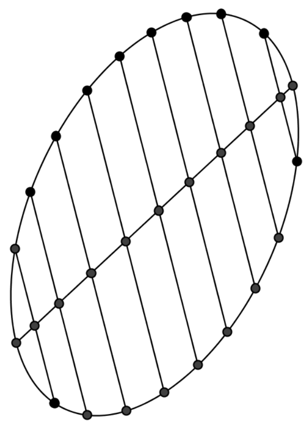
0.8 A couple of theorems from affine geometry

Some of the notions of Euclidean geometry, like "line", "collinearity", "concurrency" are also meaningful in affine geometry (i.e. are invariant under affine transformations). Many others, like "length", "volume", "angle", "circle" do not have any meaning in affine geometry. Instead one can deal with "quotient of oriented lengths", "quotient of oriented volumes", "ellipsi" etc. These notions make sense in affine geometry. Later we will see another example of geometry - projective geometry - where even these do not make sense and instead we have to consider still different notions: "double ratios", "quadrics" etc.

Some theorems we have already seen are in fact of affine nature - their proper formulations use only affine notions and, moreover, affine proofs can be given. This is the case for Ceva's and Menelaus's theorems - they can be formulated in such a way that only the notions of collinearity/concurrency and quotients of oriented lengths of segments are used.

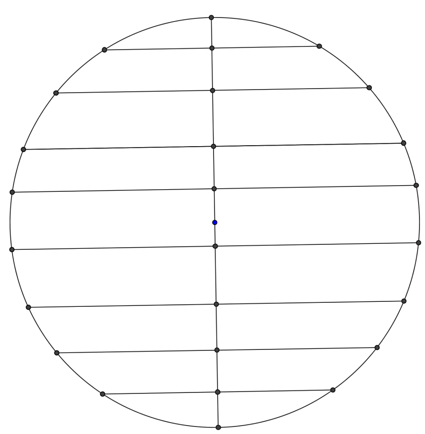
Note, however, that some of the corollaries of Ceva's and Manelaus' theorems we've seen in lecture 1 do not have affine analogues. Indeed, while the notion of medians in triangle is affine, the notions of heights or angle bisectors simply do not exist in affine geometry.

Let's now prove another theorem of affine geometry. Let E be an ellipse and consider a family of parallel lines that intersect E. For each line l in the family consider the mid-point of the segment of l inside the ellipse. Then all these midpoints lie on the same line.



Proof: The notions of collinearity, midpoints of segments and parallel lines are all invariant under affine transformations. So instead of proving the original theorem, we can first apply an affine transformation and prove the theorem for the image we get under this transformation. We will now choose an affine transformation that maps the ellipse to a circle (such a transformation exists: we can for instance stretch the ellipse in the direction of one of its major axes).

Now we have to prove the theorem in the case E is a circle. But it is obvious now: all the midpoints lie on the diameter of the circle that is perpendicular to the family of parallel lines.



Exercise: Prove that if an ellipse is inscribed into triangle, then the lines connecting the vertices to points of tangency of the ellipse with the sides of the triangle are concurrent.

0.9 Why affine geometry is natural?

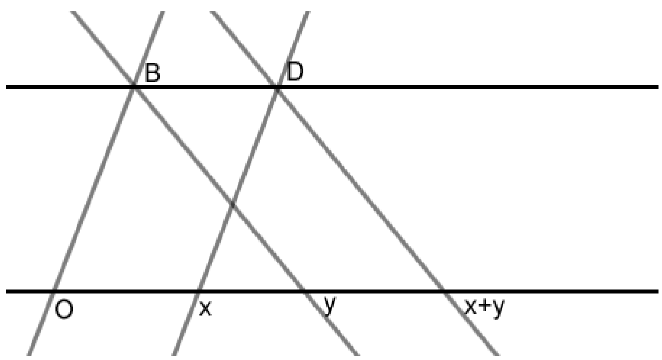
In previous sections we've introduced the group of affine transformations of the form $x- > Mx + b$ for some invertible matrix M and vector b. While we've seen that considering such a group is rewarding in terms of theorems we can prove, the choice of this group seems to be quite artificial. In this section we will prove the following theorem that explains that if one is interested in properties related only to collinearity, he is lead to work with affine geometry. Theorem: Let f be an invertible function of R^n to itself. Suppose that f maps lines to lines (i.e. if l is a line in R^n then $f(l)$ is also a line) and that $n \geq 2$. Then f is an affine transformation, i.e. there exists an invertible matrix M and vector b so that $f(x) = Mx + b\ \forall x \in R^n$.

Proof: We will prove this theorem for $n = 2$ and refer the reader to exercise ... for the general case.

Let O, A, B be three non-collinear points. Then the points $f(O), f(A), f(B)$ are also non-collinear (indeed, suppose the opposite is true. Then the lines OA, OB must map to the line through $f(O), f(A), f(B)$. Then any two points A', B' lying on the lines OA, OB should be mapped to points on the line through $f(O), f(A), f(B)$. But then the line connecting them should also be mapped to the line through $f(O), f(A), f(B)$ and every point in the plane lies on such a line for some choice of points A', B'. This contradicts the assumption that f is invertible).

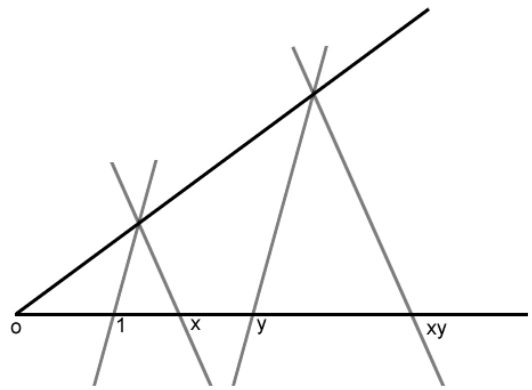
If so, we can find an affine transformation L mapping $f(O)$ to O, $f(A)$ to A and $f(B)$ to B. Since affine transformations map lines to lines and are invertible, the composite mapping $\phi = L \circ f$ maps lines to lines, is invertible and fixes points O, A, B. We will prove below that any such map must be the identity map. Once we prove it, it will follow that $f = L^{-1}$, so f is affine. Suppose now that function $\phi : R^2- > R^2$ is an invertible mapping that maps line to lines and fixes three non-collinear points O, A, B. We will identify the line through O and A with R in such a way that O is identified with $0 \in R$ and A - with $1 \in R$. Then when we restrict ϕ to this line, we get a map (we will call it ϕ as well) from R to R fixing 0 and 1. Let's prove a couple of properties of this mapping ϕ: Property 1: $\phi(X + Y) = \phi(X) + \phi(Y)$ for any $X, Y \in R$.

Proof: Parallel lines get mapped by ϕ to parallel lines (because in Euclidean geometry parallel lines can be defined simply as lines without any common points, so invertible functions that send lines to lines preserve this notion). Hence parallelograms get mapped to parallelograms. Now we can construct the point $x + y$ from x and y by the following procedure that involves only drawing parallelograms: draw parallelogram $OXDB$ spanned by vectors OB and OX (the point D is just the name for the other vertex of this parallelogram). Now draw the parallelogram spanned by vectors BD and BY. Its fourth vertex is the point $X + Y$.



Now under ϕ the point $X + Y$ gets mapped to $\phi(X + Y)$ and the construction shows that it must get mapped to $\phi(X)+\phi(Y)$. Hence $\phi(X+Y) = \phi(X) + \phi(Y)$. Property 2: $\phi(X \cdot Y) = \phi(X) \cdot \phi(Y)$ for any $X, Y \in R$.

Proof: We can construct the point $X \cdot Y$ by the following procedure involving only construction of parallel lines through points: construct line through X that is parallel to the line AB. Let D be the point of intersection of this line with line OB. Construct the line through D that is parallel to BY. The point of intersection of this line with OA is $X \cdot Y$. Under the mapping ϕ the point $X \cdot Y$ gets mapped to $\phi(X \cdot Y)$ and, following the construction, to $\phi(X) \cdot \phi(Y)$. So $\phi(X \cdot Y) = \phi(X) \cdot \phi(Y)$.



(Thanks to Yeonsuk Oh for the pictures)

Now properties 1 and 2 suffice to prove that $\phi : R- > R$ is the identity. Indeed, for any natural number p, $\phi(p) = \phi(1 + 1 + ... + 1) = p\phi(1) = p$. Hence for any pair of natural numbers p, q, $q\phi(p/q) = \phi(p/q) + ... + \phi(p/q) = \phi(q \cdot p/q) = \phi(p) = p$, i.e. $\phi(p/q) = p/q$. Also $\phi(-p/q) + \phi(p/q) = \phi(0) = 0$, so $\phi(-p/q) = -p/q$. So ϕ fixes all rational numbers (note that so far we used only property 1). To prove that ϕ fixes all real numbers, we should show some continuity. We first claim that if $x \in R$ is positive, then $\phi(x)$ is also positive. Indeed, if $x \geq 0$, then there is some $a \in R$ such that $x = a^2$. Then $\phi(x) = \phi(a \cdot a) = \phi(a)^2 \geq 0$.

It follows that if $y \geq z$, the $\phi(y) \geq \phi(z)$: if $y \geq z$, then $y = z + x$ for some positive x, and then $\phi(y) = \phi(z) + \phi(x) \geq \phi(z)$. Combining this property and the fact that ϕ fixes all rational numbers, we get that if y lies between two rational numbers, then $\phi(y)$ lies between the same rational numbers. This is enough to guarantee that $\phi(y) = y$. By now we have showed that ϕ fixes any point on OA. Similarly it fixes any point on OB. Hence it fixes all points on lines through a point on OA and a point on OB. But every point in the plane lies on such a line, so all points are fixed by ϕ.

AFFINE TRANSFORMATION

0.9.1 Problem 1

- a) Prove that for every trapezoid there exists an affine transformation that transforms it into isosceles trapezoid.

- b) Prove that in any trapezoid the middles of the bases, the point of intersection of the diagonals and the intersection of the extensions of the sides lie on one line.

0.9.2 Problem 2

- Suppose that in a pentagon $ABCDE$ all the diagonals are parallel to the opposite sides. Suppose that $AC \cap BE = P$ and $AD \cap BE = Q$. Calculate PQ/BE.

0.9.3 Problem 3

- a) Prove, that the image of an ellipse in any affine transformation is an ellipse.

- b) Prove, that for any ellipse there exits an affine transformation that transfers it into a circle.

- c) Prove that an affine transformation transfers the center of symmetry of a ellipse into the center of symmetry of an ellipse.

- d) Suppose A is a point, which lies outside an ellipse. AB_1 and AB_2 are tengent to the ellipse, B_1 and B_2 lie on the ellipse. Prove that point A, the center of the ellipse and the center of the section B_1B_2 lie on one line.

- e) E is an ellipse, points A and B lie on it. Prove that there is a one and only one affine transformation that leaves the ellipse E on its place and transfers A into B.

0.9.4 Problem 4

- a) Prove that the image of a parabola is a parabola.

- b) Prove that any parabola can be transferred into any other parabola.

- c) Prove that the axis of a parabola transfers into a line parallel to the axis of the image of the parabola.

- d) P is an parabola, points A and B lie on it. Prove that there is a one and only one affine transformation that leaves the parabola on its place and transfers A into B.
- e) Suppose A is a point, which lies outside a parabola. AB_1 and AB_2 are tangent to the parabola, B_1 and B_2 lie on the parabola. Prove that point A and the center of the section B_1B_2 lie on a line parallel to the axis of the parabola.

0.9.5 Problem 5

Prove similar statements for a hyperbola.

0.9.6 Problem 6

Suppose point A lies on an ellipse, A' is a point symmetric to A in relation to the center of the ellipse, l is a line, that goes through the center of the ellipse and parallel to the tangent lines of the ellipse, go through A and A'. B and B' are the points of intersection of l and the ellipse. Prove that the tangent lines to the ellipse, that go through points B and B' are parallel to line AA'.

0.9.7 Problem 7

Was incorrect

0.9.8 Problem 8

- a) Prove that the medians of any triangle intersect in one point (the center of gravity of the triangle) and that point divides the medians in ratio of $1:2$.
- b) Prove that the center of gravity of the triangle in any affine transformation is transformed into the center of gravity of the image of the given triangle.
- c) Median of a tetrahedron is a section that join the vertex of a tetrahedron and the center of gravity of the counter face. Prove that the medians of a tetrahedron intersect in one point (center of gravity of

tetrahedron) and divide each other in ratio of $3:1$. Prove that in any affine transformation the image of the center of gravity of the tetrahedron is the center of gravity of the image of the given tetrahedron.

- d) Median of a n-dimensional simplex is a section that joins its vertex with the center of gravity of the counter face (which is a $(n-1)$-dimensional simplex). Prove that all of the $n+1$ medians of the simplex intersect in one point (called the center of gravity of the simplex) which divides then in ratio of $n:1$. In any affine transformation the center of gravity of a simplex transfers into the center of gravity of the image of the simplex.

0.9.9 Problem 9

Prove that parallel affine subspaces $L_1, L_2 \subset \mathbb{R}^n$ ether do not intersect or coincide.

0.9.10 Problem 10

Linear transformation $B : \mathbb{R}^2 \to \mathbb{R}^2$ leaves in place

- a) line $y = 1$,
- b) line $x = y$.

Write the conditions for a matrix $\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ of this transformation. Check that the set of these matrixes form a group and build an isomorphism if this group and the group of affine transformations $F : \mathbb{R}^1 \to \mathbb{R}^1$.

Convex geometry

0.10 Dehn-Sommerville relations

In this section we will talk about convex polytopes. One possible definition of a convex polytope uses the notion of a half-space:

A closed half-space in $\mathbf{R}^n$ is the set of points lying on the same side of some hyperplane (this hyperplane is included in the set as well to make it "closed"). A little bit more precisely it is the set of points $x = (x_1, ..., x_n) \in R^n$ satisfying the inequality $a_1x_1 + ...a_n \cdot x_n \leq c$ for given numbers $a_1, ..., a_n, c \in R$.

A closed convex figure is an intersection of finitely or infinitely many closed half-spaces. Thus a closed convex figure is the set of solutions of a system of inequalities of the type $a_1x_1 + ...a_n \cdot x_n \leq c$.

Later in this chapter we will discuss another way to define a convex figure — it can be defined as a closed set (in the sense of topology of subsets of $\mathbf{R}^n$) which along with any two points in it contains the whole segment connecting them.

A polytope is a closed convex figure that can be represented as intersection of finitely many closed half-spaces. Note that in general a polytope may be unbounded.

A convex polytope is called bounded if the distance between any two points of this polytope is bounded by some a priori chosen number.

In this section the word "polytope" will be reserved only for closed bounded polytopes.

Space for figure

We will single out a special class of polytopes - that of simple polytopes. A polytope is called simple if the vectors along the edges of every vertex form a basis of $\mathbf{R}^n$. In particular only n edges meet at every vertex.

Exercise: prove that any face of a simple polytope is a simple polytope (in the affine space spanned by it).

With any polytope we will associate an n-tuple of numbers $(f_0, ..., f_n)$ that describe some of its combinatorics. Namely f_0 will be the number of vertices of the polytope, f_1 - the number of edges, f_2 - the number of 2-dimensional faces, f_k - the number of k-dimensional faces. For example f_n is always just 1 — a polytope has only one n-dimensional face.

At first sight it might seem that these numbers $f_0, ..., f_n$ might be more or less arbitrary. However there are some relations that they always satisfy. For example Euler has noticed that $f_0 - f_1 + f_2 - ... + (-1)^n f_n$ is always equal to one.

In fact we will see below that in case of a simple polytope, the numbers $f_0, ..., f_n$ possess a hidden and marvelous symmetry. Let's now build the tools to uncover it.

It is convenient to package the numbers $f_0, ..., f_n$ into one polynomial $f(t) = f_0 + f_1 t + ... + f_n t^n$. We will also need the polynomial h defined by $h(t) = f(t-1)$.

It turns out that the beautiful properties of the numbers $f_0, ..., f_n$ are most easily expressed in terms of the coefficients $h_0, ..., h_n$ of this polynomial $h(t) = h_0 + h_1 t + ... + h_n t^n$. For instance Euler's observation $f_0 - f_1 + f_2 - ... + (-1)^n f_n = 1$ translates into the statement that $h_0 = f(-1) = 1$.

Now we will state the main result presented in this section, the Dehn-Sommerville relations.

Theorem 1. *The numbers $h_0, ..., h_n$ associated to a simple polytope satisfy*
$h_i = h_{n-i}$ *(symmetry)*
$h_i \geq 1$ *(positivity)*

Add exercises and comments on non-simple polytopes.
The proof of this theorem will occupy us to the end of this section.
The following definitions will be crucial for the proof.

1. A linear functional L on $\mathbf{R}^n$ is a linear function $L : \mathbf{R}^n \to \mathbf{R}$. A linear functional always has the form $L(x_1, ..., x_n) = a_1 x_1 + ... + a_n x_n$ for some real numbers $a_1, ..., a_n$. We will abbreviate this by writing $L(x) = \langle a, x \rangle$.

   We say that a functional L is generic with respect to a given polytope if its values at different vertices are all different.

   It is clear that a functional chosen at random will be generic with probability one. Indeed, if $v_1, ..., v_k$ are the vertices of the polytope, then the functional $L(x) = \langle a, x\rangle$ is non-generic with respect to the polytope if and only if $\langle a, v_i\rangle = \langle a, v_j\rangle$for some pair of indices i, j. Equivalently a belongs to the hyperplane $\langle a, v_i - v_j\rangle = 0$. Since the number of vectors of the form $v_i - v_j$ is finite, the generic functionals are abundant — except for vectors a in a finite union of hyperplanes, all vectors a define generic functionals.

2. A vertex v has index i with respect to the functional L if there are exactly i edges going out of v along which L is decreasing.

Space for figure.

The main statement that is needed for the proof is the following:

Lemma 2. *For every functional l generic with respect to the polytope P, the number $\tilde{h}_i$ of vertices of index i is equal to the number h_i — the i-th coefficient of the h-polynomial of the polytope P.*

Proof. We will call a vertex v "maximal" for the face F if the the maximum of the restriction of functional L to the face F is attained at it. The assumption that the functional L is generic implies that every face of P has exactly one maximal vertex.

Thus if we count the number of pairs (F, v) such that v is the maximal vertex of the m-dimensional face F, we get just the number f_m of m-dimensional faces of the polytope P.

We can compute the same number in a different way — counting the number of relevant faces for each vertex first and then summing up the results over all vertices.

Suppose that a vertex v has index i. Then the number of m-dimensional faces for which v is maximal is $\binom{i}{m}$ — there are i edges going out of v along which L is decreasing and the choice of an m-dimensional face for which v is maximal is equivalent to a choice of m of these edges.

Note that here we strongly use the assumption that P is simple — if P is not simple, then different choices of m edges may lead to the same choice of m-dimensional face.

Space for figure.

So to find the number of pairs (F, v) of m-dimensional faces F for which the vertex v is maximal, we should sum the numbers $\binom{i}{m}$ over all vertices of index i and over all indices i. The number of such pairs is equal to $\sum_{i\geq m}\binom{i}{m}\tilde{h}_i$ (recall that $\tilde{h}_i$ denotes the number of vertices of index i).

Comparing the two ways of counting we can conclude that $f_i = \sum_{i\geq m}\binom{i}{m}\tilde{h}_i$. But this means that

$$f(t) = \sum_{i=0}^{n} f_i t^i = \sum_{i=0}^{n}\sum_{i\geq m}\binom{i}{m}\tilde{h}_i t^i = \sum_{0\leq m\leq i\leq n}\binom{i}{m}\tilde{h}_i t^i = \sum_{i=0}^{n}(\sum_{m=0}^{i}\binom{i}{m}t^m)\tilde{h}_i = \sum_{i=0}^{n}\tilde{h}_i(1+t)^i$$

Thus $f(t-1) = \sum_{i=0}^{n}\tilde{h}_i t^i$, or $h_i = \tilde{h}_i$.

□

To conclude, we just showed that $\tilde{h}_i$, the number of vertices of index i with respect to the linear functional L is equal to h_i, a number that can be expressed in terms of combinatorics of the polytope alone! In particular the number of vertices of index i does not depend on the choice of the generic linear functional L.

Now we will exploit this last result to prove Dehn-Sommerville relations. If v has index i with respect to the functional L, then its index with respect to the functional $-L$ is $n-i$ (here we use simplicity of the polytope P once again). Thus the number h_i of vertices of index i with respect to functional L is equal to the number h_{n-i} of the vertices of index $n-i$ for the functional $-L$. This proves the first part of the theorem: $h_i = h_{n-i}$. For the second part, choose a vertex of the polyhedron and choose the generic linear functional L so that this vertex will have index i with respect to it. Then the number h_i of index i vertices will be at least 1. This proves the final assertion of Dehn-Sommerville theorem.

Exercise: find all h and f polynomials for 3-dimensional polytopes (usually referred to as polyhedra). For each one of the polynomials find an example of polyhedron that realizes this polynomial.

0.11 Convex closed sets

In this section we will focus on properties of more general convex sets - not necessarily bounded polytopes.

Large parts of the theory of convex sets can be developed in a general context of topological vector spaces, and such generality is indeed useful for purposes of functional analysis. We, however, will focus on more geometric parts of the story and restrict ourselves to studying convex set in Euclidean spaces.

A subset Δ of an Euclidean space is called **convex** if along with any two points a and b in δ, the whole line segment joining a and b is contained in δ.

Space for figure.

Note that an intersection of an arbitrary collection of convex sets is convex again.

Also note that a half-space $\{x | \langle a, x \rangle \leq c\}$ is an example of a convex set. Intersections of half-spaces provide us with a lot of additional examples of convex sets.

One should note however, that all examples obtained in this way have to be closed[1].

Our first goal will be to show that in fact every closed convex set can be obtained in this way - a set is closed and convex if and only if it can be represented as the intersection of some collection of half-spaces. Before we prove this result, we will formulate a very closely related result:

Theorem 3 (Separation theorem)**.** *Let p be a point located outside a closed convex set Δ. Then there exists a hyperplane that separates them, i.e. a hyperplane with the property that the point p and the set Δ lie on different sides of it.*

Proof. First we will prove that there exists a point $q \in \Delta$ such that the distance from p to q is the same as the distance between p and the set Δ[2]. Since the distance between the point p and the set Δ is the infimum of distances between p and points in Δ, we can find a sequence of point $q_i \in \Delta$ such that $\lim_{i \to \infty} dist(p, q_i) = dist(p, \Delta)$ ($dist$ denotes the Euclidean distance). The sequence q_i is bounded (starting from some index i_0 all q_i's satisfy $dist(p, q_i) < dist(p, \Delta) + 1$, hence $|q_i| < |p| + dist(p, \delta) + 1$) and hence there exists a converging subsequence of it. Let q denote the limit of this subsequence. By the assumption that the set Δ is closed, the point q belongs to it and by continuity of the distance, $dist(p, q) = dist(p, \Delta)$.

Now let H be the hyperplane orthogonal to the segment pq and such that p and q lie on different sides of it. We claim that H separates p from Δ. Indeed, if r is any point in Δ, then the segment from q to r must be contained

[1]Recall that a set is called closed if for any converging sequence of points in the set, the limit of the sequence also belongs to the set. One can easily check that an arbitrary intersection of closed sets is a closed set.

[2]The distance between a point p and a set Δ is defined as the infimum of distances between p and points in Δ. One can easily check that if the set Δ is not closed, this distance might be strictly smaller than all the distances between p and points of Δ

in Δ. Since the distance from p to any point on this segment must be greater that the distance from p to q, the angle pqr must be obtuse (greater than 90 degrees). This implies that r is on the same side of H as q is.

$\square$

Now we can prove that a convex closed set is an intersection of half-spaces: let Δ be any closed convex set and $\tilde{\Delta}$ be the convex closed set obtained as the intersection of all half-spaces that contain Δ. Clearly the set Δ is contained in $\tilde{\Delta}$. If a point p does not belong to Δ, then there is a hyperplane that separates p from Δ. Thus p does not belong to the half-space determined by this hyperplane and containing Δ. Thus p doesn't belong to $\tilde{\Delta}$ — the intersection of all half-spaces that contain Δ. This way we showed that $\Delta = \tilde{\Delta}$.

0.12 Convex Hull

We will study now the minimal convex set that contains a given one.

Definition 1. *The convex hull $\Delta(S)$ of the set S is the smallest convex closed set that contains S. It can be described also as the intersection of all convex closed sets that contain S, or, as we showed in the last section, as the intersection of all the closed half-spaces that contain S.*

For example the convex hull of a finite collection of points in a plane is the smallest convex polygon containing all of them.

We will now give an alternative description of the convex hull of the set S:

Claim 1. *A point p belongs to the convex hull of the set S if and only if there exist points $x_1, ..., x_m \in S$ and numbers $\lambda_1, ..., \lambda_m \geq 0$ such that $\lambda_1 + ... + \lambda_m = 1$ and $p = \lambda_1 x_1 + ... + \lambda_m x_m$.*

Proof. First we show that if $x_1, ..., x_n$ belong to S and $\lambda_1, ..., \lambda_m$ are positive numbers such that $\lambda_1 + ... + \lambda_m = 1$, then $\lambda_1 x_1 + ... + \lambda_m x_m$ belongs to the convex hull of S. We show this by induction on m. If $m = 1$, then $\lambda_1 = 1$ and the claim is obvious. Suppose we proved the claim for m and want to show it for $m+1$. The point $\lambda_1 x_1 + ... + \lambda_m x_m + \lambda_{m+1} x_{m+1}$ is the center of mass of the system of positive masses $\lambda_1, ..., \lambda_{m+1}$ located at points $x_1, ..., x_{m+1}$. By the properties of center of mass, it must be located on the segment connecting the point x_{m+1} to the center of mass of the system C of masses $\lambda_1, ..., \lambda_m$ located at the points $x_1, ..., x_m$. By inductive assumption, the center of mass of C lies in the convex hull. Clearly x_{m+1} lies in the convex hull as well. Since the convex hull of any set is convex, the entire segment connecting x_{m+1} to C must lie in the convex hull. In particular it contains the point $\lambda_1 x_1 + ... + \lambda_m x_m + \lambda_{m+1} x_{m+1}$.

For the other direction let $\tilde{\Delta}$ be the set of linear combinations $\lambda_1 x_1 + ... + \lambda_m x_m$ of points in S such that $\lambda_i \geq 0$ and $\lambda_1 + ... + \lambda_m = 1$. This set is convex. Indeed, if p is the center of mass of points $x_1, ..., x_m$ with positive masses $\lambda_1, ..., \lambda_m$ and q is the center of masses of points $y_1, ..., y_k$ with positive masses $\mu_1, ..., \mu_k$, then any point $t \cdot p + (1-t) \cdot q$ on the segment connecting p to q is the center of mass of the system of masses $t\lambda_1, ..., t\lambda_m, (1-t)\mu_1, ..., (1-t)\mu_k$. Hence $\tilde{Delta}$ contains the convex hull of S. But the first part of the proof shows that the convex hull is contained in $\tilde{\Delta}$, hence they coincide.

□

Now we will prove an explicit bound on the number of points one needs in order to represent any point in the convex hull of the set S as a convex linear combination. The bound is known under the name of Caratheodory theorem.

Theorem 4 (Caratheodory). *If S is a subset of $\mathbf{R}^d$, then any point in the convex hull of S can be written as $\lambda_1 x_1 + ... + \lambda_{d+1} x_{d+1}$ for some points $x_1, ..., x_{d+1}$ in S and numbers $\lambda_i \geq 0$, $\lambda_1 + ... + \lambda_{d+1} = 1$.*

Alternatively we can formulate the theorem as saying that any point in the convex hull of the set S is in the convex hull of at most $d+1$ points of S.

Proof. We already proved that any point x can be written as the center of mass of some collection $\lambda_1, ..., \lambda_n$ of positive masses with total mass 1 of points $x_1, ..., x_n$. Now we will show that if $n > d+1$, then we can do without one of the masses. By repeating this argument, we will end up with a system of at most $d+1$ positive masses with the same center of mass.

To show we can get rid of one of the masses in case $n > d+1$ we will "slide" the masses without changing the center of mass and make one of the masses in the end of the process equal to zero.

Here's how we do it: the $n-1$ vectors $x_2 - x_1, ..., x_n - x_1$ must be linearly dependent (since $n-1 > d$). Hence there are some numbers $\mu_2, ..., \mu_n$, not all zero, such that $\mu_2(x_2 - x_1) + ... + \mu_n(x_n - x_1) = 0$. Let $\mu_1 = -(\mu_2 + ... + \mu_n)$, so that $\mu_1 + ... + \mu_n = 0$ and $\mu_1 x_1 + ... + \mu_n x_n = 0$. These two equalities show that if we will add to the mass λ_i at point x_i the mass $k\mu_i$ for some $k \geq 0$ (the same k for all points x_i), then the total mass of the system won't change and the location of the center of mass won't change as well. Since λ_i's are all positive and at least one of the μ_i's is negative, there is some maximal k so that $\lambda_i + k\mu_i$ is still non-negative for all i. For such k one of the masses $\lambda_i + k\mu_i$ must vanish (otherwise, if all these masses are strictly positive, we can increase k a little bit more).

□

0.13 Helly's theorem

In this section we will prove the following remarkable theorem: if in a collection of convex subsets of R^d every $d + 1$ sets intersect, then all the sets of the collection intersect. Because the theorem is extremely beautiful, we will repeat it once again for the case $d = 2$: if in a collection of convex sets in the plane every three have a point in common, then all the sets in the collection have some point in common.

To appreciate the theorem we should see that the assumptions are really necessary.

Space for figure - three convex planar sets, each two intersect, but the three don't

Space for figure - four non-convex planar sets, each three intersect, but the four don't.

To prove this theorem we will rely on a lemma by Radon.

The lemma claims that if S is a finite set of at least $d + 2$ points in R^d, then the set S can be partitioned as a union of two subsets, A and B such that the convex hulls of A and B intersect non-trivially. We will prove both Radon's and Helly's theorem on line and in plane and leave the more general case as a series of exercises for the reader.

Radon's lemma for the line: Suppose we have a set S that contains at least 3 points on the real line R. We want to show that S can be partitioned as a union of two subsets A and B such that their convex hulls intersect. Let A consist of the smallest and largest points in S (we order the points as if they are just numbers in R). Let B consist of all other points. Clearly the convex hull of A coincides with the convex hull of S, hence in particular intersects (in fact contains) the convex hull of B.

Radon's lemma for the plane: Suppose we have a set S that contains at least 4 points on the plane R^2. We want to show that S can be partitioned as a union of two subsets A and B such that their convex hulls intersect. We can assume that no three points in S are collinear, for if they were, we could use Radon't theorem for the line that contains at least three points and we would be done.

So take any 3 points p, q, r in S. They form a non-degenerate triangle. We will refer to the half-spaces defined by the lines pq,qr and rs and containing the triangle pqr as H_1, H_2 and H_3. Take now any other point s from S (i.e. a point different from p, q and r). The point s can lie in one of the seven regions to which the lines pq,qr,pr subdivide the plane. There are three possible cases: one case is that the point s lies inside the triangle pqr, i.e. inside all three half-planes H_1, H_2, H_3. Then we will take A to be the set of three points p, q, r and let B be the compliment of A in S. With this choice the point s will lie in the intersection of the convex hulls of A and B. Another possibility is that the point s lies inside one of the half-planes and in the outside of the other two. For instance we will consider the case that it lies inside H_1, the half-plane defined by pq, and outside the other two half-spaces. Then the point r must lie inside the triangle pqs and we can take A to consist of the points p, q and s and B be the compliment of A. The intersection of the convex hulls of A and B contains the point r. The last option is that the point s lies inside two half-planes, say H_1 and H_2, and outside the other one. Then the segments pq (the one that defines the half-space to which s doesn't belong) and rs must intersect. We will take in this case A to consist of p and q and B be the compliment. The intersection of the convex hulls of A and B will contain the intersection point of pq and rs.

Remark: there is no fourth case, where the point s doesn't belong to any of the half-planes: their union covers all the plane.

Space for figure describing the three cases.

Radon's lemma, while it is quite simple, allows us to prove a much deeper result, Helly's theorem, which otherwise is not very easy to prove. We will now recall what Helly's theorem (in the plane) tells us and prove it.

Theorem 5. *If in a finite collection of convex sets every three intersect, then the intersection of all the sets in the collection is not empty.*

Proof. We will prove the theorem by induction on the number n of sets in the collection. The case $n = 3$ is clear. Even though we don't have to, we will verify the case $n = 4$ separately, to give the reader a feeling for the general argument.

Let G_1, G_2, G_3 and G_4 be the four convex sets. Suppose that the intersection of every triple of them is non-empty. Let A_1, A_2, A_3, A_4 be points in these intersections, i.e. $A_1 \in G_2 \cap G_3 \cap G_4$, $A_2 \in G_1 \cap G_3 \cap G_4$, $A_3 \in G_1 \cap G_2 \cap G_4$ and $A_4 \in G_1 \cap G_2 \cap G_3$. Radon's theorem tells us that we can subdivide the collection A_1, A_2, A_3 and A_4 to two subsets, whose convex hulls intersect. If one of the four points, say A_1 belongs to the convex hull of the other three, then A_1 must belong to G_1 - all three of the points A_2, A_3, A_4 belong to G_1 by their definition, hence any point in their convex hull belongs to G_1 as well. Since by its definition A_1 belongs to the other three convex sets, it belongs to the intersection of all four.

Similar argument applies if the Radon's theorem gives us that the two sets, whose convex hulls intersect, consist of two points each. Suppose for instance that the segments $[A_1, A_2]$ and $[A_3, A_4]$ intersect at a point P. Since both A_1 and A_2 belong to G_3 and G_4, every point in their convex hull also belong both to G_3 and G_4. In particular the point P lies in $G_3 \cap G_4$. A similar argument about the points A_3 and A_4 and the sets G_1 and G_2 implies that P belongs to $G_1 \cap G_2$ as well. Hence P belongs to the intersection of the four sets G_1, G_2, G_3 and G_4.

Now suppose we have proved the theorem for every collection of n planar sets and want to prove it for a collection of $n+1$ sets. By induction hypothesis the intersection of every n-tuple of sets in our collection is non-empty. Let A_i be a point in the intersection of all the sets, except the i-th one. Radon's theorem guarantees us that we can subdivide the set $A_1, ..., A_{n+1}$ to two subsets, whose convex hulls intersect. By renaming the sets if necessary,we can assume that the two subsets are $\{A_1, ..., A_k\}$ and $\{A_{k+1} ..., A_{n+1}\}$. Let P be a point in the intersection of their convex hulls. Since all the points $A_1, ..., A_k$ belong to $G_{k+1} \cap ... \cap G_{n+1}$, all points in their convex hull also lie in this intersection. In particular the point P lies there. Similarly, considering the points $A_{k+1}, ..., A_{n+1}$, we deduce that the point P must lie in the intersection of the first k sets, $G_1 \cap ... \cap G_k$. This shows that the point P lies in the intersection of all the sets in the collection we started with.

$\square$

Let's see how we can apply Helly's theorem to a problem from geometry.

Problem: prove that if every three points of a finite collection of points in a plane lie inside some circle of radius one, then all the points of this collection lie inside some circle of radius one.

Solution: Consider the set of circles of radius one with centers at the points from our collection of points. Every three circles in this set must intersect. Indeed, saying that the three circles of radius one centered at points p_1, p_2, p_3 do not intersect is the same as saying that there is no point in the plane, whose distance from the three points p_1, p_2, p_3 is smaller than one. But this means also that there is no circle of radius one containing the three points p_1, p_2, p_3.

Now Helly's theorem implies that the intersection of all the circles in the set we considered is non-empty, i.e. there is some point P, whose distance from all the points in our collection is smaller than one. This is exactly what we wanted to prove.

Exercises about Helly's and Radon's theorems:

In the d-dimensional space R^d Radon's theorem tells that if S is a set of n points in R^d and $n \geq d+2$, then we can subdivide S to two subsets so that their convex hulls intersect. Prove the theorem following the outlined steps:

1. If there is a collection of $d+1$ points from S that lie on a $d-1$-dimensional hyperplane, reduce the problem to $d-1$-dimensional Radon theorem.

2. Show it is enough to prove the theorem for $n = d+2$.

3. Suppose that S consists of $d+2$ points $p_1, ..., p_{d+2}$ and every collection of $d+1$ of them doesn't lie on one hyperplane. Show that we can find non-zero numbers $\lambda_1, ..., \lambda_{n+1}$ with sum 1, so that $p_{d+2} = \lambda_1 p_1 + ... + \lambda_{d+1} p_{d+1}$.

4. Suppose the numbers $\lambda_1, ..., \lambda_k$ in the above representation $p_{d+2} = \lambda_1 p_1 + ... + \lambda_{d+1} p_{d+1}$ are positive and the other λ's are negative. Show that there exist positive numbers $\mu_1, ..., \mu_{d+2}$ such that $\mu_1 + ... + \mu_k = \mu_{k+1} + ... + \mu_{d+2}$ and

$$\frac{\mu_1 p_1 + ... + \mu_k p_k}{\mu_1 + ... + \mu_k} = \frac{\mu_{k+1} p_{k+1} + ... + \mu_{d+2} p_{d+2}}{\mu_{k+1} + ... + \mu_{d+2}}$$

.

Hint: why the choice $\mu_1 = \lambda_1, ..., \mu_k = \lambda_k, \mu_{k+1} = -\lambda_{k+1}, ..., \mu_{d+1} = -\lambda_{d+1}, \mu_{d+2} = 1$ works?

5. Deduce Radon's theorem from the result of step 4.

Exercise: deduce Helly's theorem in d-dimensional space R^d from Radon's theorem proved in the exercise above. Namely show that if every $d+1$ sets in a finite collection of convex sets intersect, then all the sets in this collection intersect.

Hint: the arguments are very close to those used in the deduction of planar Helly's theorem from planar Radon's theorem.

CONVEX SET

0.13.1 Problem 1

Prove that section A$[a, b] \subset \mathbb{R}^n$ is a convex set.

0.13.2 Problem 2

Specify all the nonzero convex subsets of $\mathbb{R}$.

0.13.3 Problem 3

A convex polygon lies on a plane. It does not lie on one line and lines that join to point of the polygon is a side of the polygon. Prove that the polygon is a triangle.

0.13.4 Problem 4

Point a lies on the face F of a convex set X and is a convex combination of points $a_1, \dots, a_k \in X$ with nonzero coefficients. Prove that $a_1, \dots, a_k \in F$.

0.13.5 Problem 5

Prove that a convex polygon in $\mathbb{R}^3$ any two vertexes of which are joined by an edge is a tetrahedron.

0.13.6 Problem 6 (Radon's theorem on plane)

On a plain there are given $m \geq 4$ points $a_1, \dots, a_m$. Prove that they can be divided into two groups, convex hulls of which intersect.

0.13.7 Problem 7

Suppose that $a_1, \dots, a_m \in \mathbb{R}^n$, $m \geq n + 2$. Prove that there exist $\lambda_1 \dots, \lambda_m$ not all of them zero that $\lambda_1 a_1 + \cdots + \lambda_m a_m = 0$, $\lambda_1 + \cdots + \lambda_m = 0$.

0.13.8 Problem 8 (Radon's theorem)

Suppose that $a_1, \ldots, a_m \in \mathbb{R}^n$, $m >\geq n+2$. Prove that they can be divided into two groups, convex hulls of which intersect.

0.13.9 Problem 9 (Helly's theorem)

Suppose $M_1, \ldots, M_m \subset \mathbb{R}^n$ - convex sets, $m \geq n+1$. It is given that the intersection of any $n+1$ set M_i is nonzero. Prove that the intersection of all of the M_i is nonzero.

0.13.10 Problem 10

Suppose that $T_1, \ldots, T_m$ – parallel sections on a plane, $m \geq 3$. It is given that for any three sections there exits a line that intersects all of them. Prove that there exits a line that intersects all of the sections.