Preliminaries on convex sets and convex functions

Given $x,y \in R^n$, a **linear combination** of x and y is a point $\alpha x+\beta y$, where $\alpha,\beta \in R$.

A set $C \subseteq R^n$ is a **subspace** if it contains all the linear combinations of any two points in C.

For example, $C = \{x \in R^n : Ax = 0\}$, where A is a m×n matrix, all the solution of this equation belongs to the subspace. The dimension of the space depends on the rank of the matrix, on how many equations are linearly independent.
Other examples:

- {0}, only the origin, with dimension 0
- any line which passes through zero is a subspace with dimension 1
- the solution set of a homogeneous system of linear equations

An **affine combination** of x and y is a point $\alpha x+\beta y$, where $\alpha+\beta=1$, a linear combination with a constraint (they can be positive or negative).

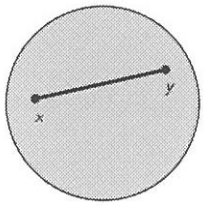
A set $C \subseteq R^n$ is an **affine set** if it contains all the affine combinations of any two points in C.
Examples:

- any single point {x}, not only the origin, with dimension 0
- any line, passing through the origin or not, with dimension 1
- the solution set of a system of linear equations

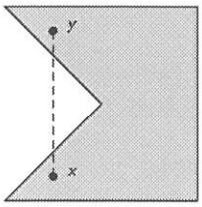
  $C = \{x \in R^n : Ax = b\}$, where A is a m × n matrix and $b \in R^m$
  (not homogeneous, we have b)
- any subspace

A **convex combination** of two given points x and y is a special affine combination, a point $\alpha x +\beta y$, where $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$. We have three constraints.

A set $C \subseteq R^n$ is **convex** if it contains all the convex combinations of any two points in C.



convex set



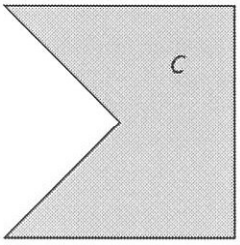
non-convex set

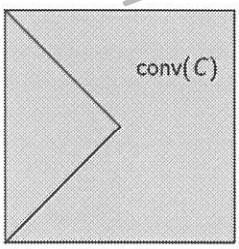
If given any two points, the segment between them is contained in the same set, we have a convex set.
The same property holds for a convex combination of k points in the set, not only two points
If C is convex, then for any x1,...,xk ∈ C and α1,...,αk ∈(0,1) s.t.

$$\sum_{i=1}^{k} \alpha_i = 1, \text{ one has } \sum_{i=1}^{k} \alpha_i x^i \in C$$

It can be proven by induction.

If the set is not convex, we can find the smallest possible subset which is convex containing that set.
The **convex hull** conv(C) of a set C is the smallest convex set containing C.





conv(C) = {all convex combinations of points in C}.
C is convex if and only if C = conv(C).

Examples of convex sets:

- subspace
- affine set
- line segment
- halfspace {x ∈ R^n : a^Tx ≤ b}
- polyhedron P = {x ∈ R^n : Ax ≤ b} solution set of a system of linear inequalities

We can replace the equality with an inequality.
The square, which is a polyhedron, is not affine but it is convex.

We define a **ball** B(x,r)={y ∈ R^n :‖y−x‖ ≤ r}, where ‖·‖ is any norm.
The ball is another kind of convex set.
the set of vectors such that holds.
This definition depends on the norm, if we choose an Euclidian we get a sphere, a quadratic function.

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \text{ (Euclidean norm)}$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \text{ (Manhattan distance)}$$

$$\|x\|_\infty = \max_{i=1,\dots,n} |x_i| \text{ (Chebyshev norm)}$$

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}, \text{ with } 1 \le p < +\infty$$

$$\|x\|_A = \sqrt{x^{\mathsf{T}} A x}, \text{ where } A \text{ is a symmetric and positive definite matrix, i.e.,}$$

$$x^{\mathsf{T}} A x > 0 \qquad \forall\, x \neq 0.$$

We have a different shape of the ball with respect to the norm, sphere for (1), and so on..
Find the unit ball B(0, 1) w.r.t. ‖ · ‖1, ‖ · ‖∞ and ‖ · ‖A, where A= (2 0, 0 1).
The one norm does not provide a quadratic function, we have a linear function, and it is a square with the vertices on the different axis, in (-1,0), (1,0), etc.
With the max norm, we have a square with vertices in (-1,-1), (1,1), etc.
With these two norms we typically use linear functions, and so we can use all the theorem of linear programming.
||x||A = sqrt(2x1^2 + x2^2) <= 1
2x1^2 + x2^2 <= 1 and this is an hehlypsis.
Any ball with respect to any norm is a convex set.

Now let's see some **operation that preserve convexity** of a set.

Sum and difference: if C_1 and C_2 are convex, then $C_1 + C_2 := \{x + y : x \in C_1, y \in C_2\}$ is convex.
If C_1 and C_2 are convex, then $C_1 - C_2 := \{x - y : x \in C_1, y \in C_2\}$ is convex.

Intersection: if C_1 and C_2 are convex, then $C_1 \cap C_2$ is convex. If we have a segment in C_1, if it's also in C_2, we will have it also in the intersection.
If $\{C_i\}_{i \in I}$ is any (possibly infinite) family of convex sets, then is convex. An intersection of any number of convex sets is convex.

Union: if C_1 and C_2 are convex, then $C_1 \cup C_2$ is convex? No. For example, we can think at two spheres.
Closure and interior: if C is convex, then cl(C) is convex. If C is convex, then int(C) is convex.
Interior: set of points such that the Euclidian ball is contained in the set.
Closure: interior + borders.

Affine functions:

Let $f : R^n \to R^m$ be affine, i.e. f(x)=Ax+b, with $A \in R^{m \times n}$, $b \in R^m$.

- If $C \subseteq R^n$ is convex, then f(C) = {f(x): x∈C} is convex.
  The image of a convex set with an affine function is a convex set in the image space, independently of A or b.
  We can use the convexity to prove it.
- If $C \subseteq R^m$ is convex, then $f^{-1}(C)=\{x \in R^n : f(x) \in C\}$ is convex.

The same holds for the counter-image.

Examples of affine functions:

- scaling, e.g. f(x)=αx, with α>0
- translation, e.g. f(x)=x+b, with b ∈ R^n, so A = I
- rotation, e.g. f(x) = (cos θ -sin θ, sin θ cos θ), with θ ∈ (0,2π), independently of the rotation angle

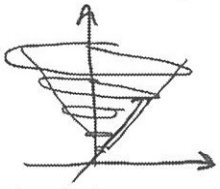
A set C ⊆ R^n is a **cone** if the following implication holds: x∈C, α≥0 ⇒ αx∈C. This implication has to be satisfied for any scalar greater or equal than 0. Any cone contains 0 surely, but if contains points different than the origin, it contains all the half-lines starting from the origin and pointing to that point.

Examples:

- R^n_+ is a convex cone.
- $\{x \in R^2 : x_1x_2 = 0\}$ is a non-convex cone. Not all cones are convex. The solutions set here are the two axes.
- Given a polyhedron P = {x : Ax ≤ b}, the recession cone of P is defined as
  rec(P):={d: x+αd ∈ P for any x ∈ P, α≥0}.
  It is the set of directions such that the half-line starting from any point of the polyhedron and along d considering any positive scalar, so in the first plane, this line is contained in the polyhedron.



  We cannot consider a negative alpha or it's not considered.
  The recession cone is a polyhedron for the system of linear inequality.
  It is easy to prove rec(P) = {x : Ax ≤ 0}, if we consider 0, thus it is a polyhedral cone.
  Some cones are also polyhedral.
- $\{x \in R^3 : x_3 \geq x_1^2 + x_2\}$ is a non-polyhedral cone



  Not all the cones are polyhedral, not all cones are convex.

Exercises

1) Write the vector (1, 1) as the convex combination of the vectors (0, 0), (3, 0), (0, 2), (3, 2).

The convex hull of the point is the rectangle, so it can be written as a convex combination, it's inside the rectangle, the solution must exist, but it's not unique.

We want to be written in the form (1,1) = a1 (0,0) + a2 (3,0) + a3 (0,2) + a4 (3,2)

with a1,a2,a3,a4 scalars non-negative with sum equal to 1.

1 = 3a2 + 3a4

1 = 2a3 + 2a4

a1 + a2 + a3 + a4 = 1

We can express a2 and a3 in terms of a4 and then a1 in terms of a4 and then solve the equation. Then we want them to be non-negative.

Ex. a4 = 0, and then derive a1, a2, a3.

a4 has to vary from 0 and 1/3, as constraint for having them >= 0.

Dependently of what we fix = 0 we are considering a different triangle.

We have an infinite number of solutions.

2) Which of the following sets are polyhedron?

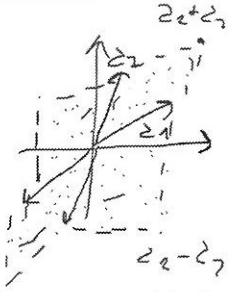
$\{y_1 a_1 + y_2 a_2 : -1 \leq y_1 \leq 1, -1 \leq y_2 \leq 1\}$, where $a_1, a_2 \in \mathbb{R}^n$.

$\left\{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^{n} x_i = 1, \sum_{i=1}^{n} a_i x_i = b_1, \sum_{i=1}^{n} a_i^2 x_i = b_2\right\}$, where $b_1, b_2, a_1, \ldots, a_n \in \mathbb{R}$.

$\{x \in \mathbb{R}^n : x \geq 0, a^T x \leq 1 \text{ for all } a \text{ with } \|a\|_2 = 1\}$.

$\{x \in \mathbb{R}^n : x \geq 0, a^T x \leq 1 \text{ for all } a \text{ with } \|a\|_1 = 1\}$.

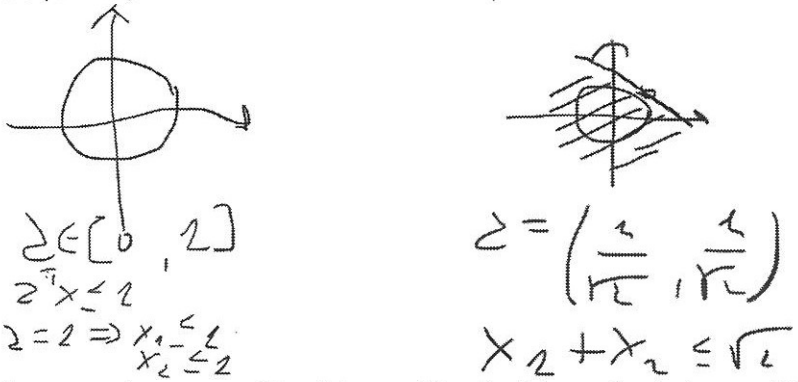
2.1) a1 and a2 are given. It's a polyhedron it's given by the intersection of a finite number of half-planes. The half-planes are the one below each line, each border.



We have 4 inequalities to create this polygon, we need to find A and b.
We need to write the equations of the four lines and get the system from it.
One line pass from a1 + a2 and a1 – a2 for example.

2.2) The coefficients are given. We have three linear constraints (it's ai^2 not xi^2). We have a system of linear equations. We have a system of n + 6 inequalities. This is a polyhedron; we have linear inequalities, and the shape will depend. The important things is that they are linear, the variables are inside the relation with power 1.

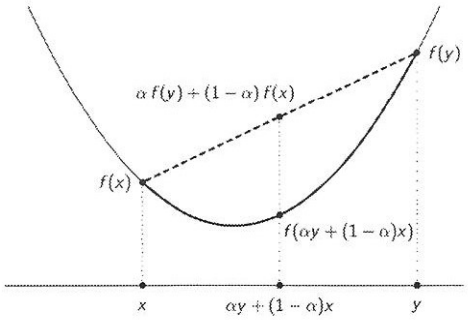
2.3) We have linear inequalities with respect to x, one for each vector of norm 1, so an infinite number of inequalities, we don't have a finite set of inequalities. We need to understand what the shape is.



For any vector, we consider all inequalities, the intersection between all this number of halfplanes is? For any, we are intersecting an infinite number of halfplanes tangent to the circle, and the intersection is the circle itself.
We need to consider only the part of the circle in the first plane, a fourth of a circle, but we cannot write it as a finite number of linear inequalities. It is convex because it is an intersection of halfplanes, but we need to consider an infinite number of halfplanes, so it's not a polyhedron, we cannot find a finite number of rows for A.

Given a convex set C ⊆ R^n, a function f : C → R is **convex** if

$$f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha)f(x) \qquad \forall\, x, y \in C, x \neq y, \forall\, \alpha \in (0,1)$$



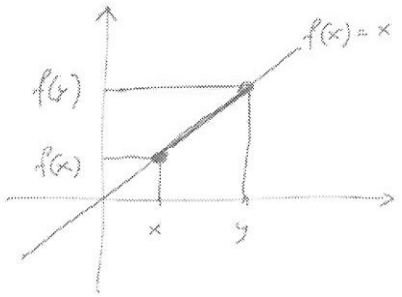
If this is satisfied for any pair of x,y and any scalar in that range.
The content of f is a convex combination.
The value of f at that point is less or equal of the convex combination of f with the same coefficients.
From a geometric point of view, the graph of f is below or equal to the values that are in the segments between f(x) and f(y), for any pair of points x and y.
When it's exactly along the segment, the function is linear.

f is said **concave** if −f is convex, it is the opposite of a convex function.
From a geometric point of view, the situation is symmetrical.

Ex. Prove that if f is convex, then for any x1, . . . , xk ∈ C and
$\alpha_1, \ldots, \alpha_k \in (0,1)$ s.t. $\sum_{i=1}^{k} \alpha_i = 1$, one has $f\left(\sum_{i=1}^{k} \alpha_i x^i\right) \leq \sum_{i=1}^{k} \alpha_i f(x^i)$.
When a function is convex, the function combination of a set C is less or equal of the convex combination of f with the same coefficients.
It can be proven in a similar way than the previous.

Given a convex set C ⊆ R^n, a function f : C → R is **strictly convex** if
$$f(\alpha y + (1-\alpha)x) < \alpha f(y) + (1-\alpha) f(x) \qquad \forall\, x, y \in C, x \neq y, \forall\, \alpha \in (0,1)$$
The definition is the same with the strict equality.
We cannot have a linear function, we don't have the equality, a linear function cannot be strictly convex.

Given a convex set C ⊆ R^n, a function f : C → R is **strongly convex** if there exists tau > 0 s.t.
$$f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha) f(x) - \frac{\tau}{2}\alpha(1-\alpha)\|y-x\|_2^2$$
$$\forall\, x, y \in C, x \neq y, \forall\, \alpha \in (0,1)$$
We have the convex combination of two points, this is less or equal than the convex combination of f(x) and f(y) with the same coefficients minus that value.
The additional term is positive because tau, alpha, 1 – alpha and the square norm are positive.
We want that the value is below the convex combination minus the term, that will depend on alpha, tau and the distance of x and y.
The set of convex functions contains the set of strictly convex functions that contains the set of strongly convex functions. The opposite verse is not valid, e.g. linear function.



$$f(\alpha x + (1-\alpha)y) = \alpha f(x) + (1-\alpha) f(y)$$

convex ⟹ strictly convex? No.
Thereom: f is strongly convex if and only if
$\exists\, \tau > 0$ s.t. $f(x) - \frac{\tau}{2}\|x\|_2^2$
is convex.
strictly convex !⇒ strongly convex
We can prove this by using the derivative when it exists.
When we have derivatives, and they are continuous we have a theorem that characterize these three notions.

First order conditions
Assume that C ⊆ R^n is open and convex, f : C → R is continuously differentiable.
We consider the first order approximation of f(x).
We take the tangent plane of the graph at point f(x).
Locally f is approximately by that value, when y is close to x, we can replace the function with a linear function.