# Optimization & Operational Research : First Part

Antoine Gourru

Slides built by Guillaume Metzler, updated by Ievgen Redko

January 2024 - Semester II

# Topic of the course

**Headline**

- Mathematical background : Convex sets and derivatives.
- Convex function and their properties.
- What is a convex optimization problem ?
- Algorithms for convex optimization.

# Some references

## Linear Algebra

- K.B Petersen, M.S Pedersen, *The Matrix Cookbook*,2012. Available at : http ://matrixcookbook.com

## Convex Optimization

- Stephen Boyd & Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2014

# Mathematical Background

# Norms

Given $x, y \in \mathbb{R}^n$, the inner product is given by :

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i.$$

The inner product of $x$ with itself is called the square of the norm of $x$

$$\langle x, x \rangle = \|x\|^2.$$

## Definition

Let $E$ be a $\mathbb{R}$-vector space, then the application $\|.\|$ is said to be a norm if for all $u, v \in E$ and $\lambda \in \mathbb{R}$

1. (positive) $\|u\| \geq 0$ ,
2. (definite) $\|u\| = 0 \iff u = 0$,
3. (scalability) $\|\lambda u\| = |\lambda| \|u\|$,
4. (triangle inequality) $\|u + v\| \leq \|u\| + \|v\|$.

# Norms

Given $x, y \in \mathbb{R}^n$, the inner product is given by :

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i.$$

The inner product of $x$ with itself is called the square of the norm of $x$

$$\langle x, x \rangle = \|x\|^2.$$

## Definition

Let $E$ be a $\mathbb{R}$-vector space, then the application $\|.\|$ is said to be a norm if for all $u, v \in E$ and $\lambda \in \mathbb{R}$

1. (positive) $\|u\| \geq 0$ ,
2. (definite) $\|u\| = 0 \iff u = 0$,
3. (scalability) $\|\lambda u\| = |\lambda| \|u\|$,
4. (triangle inequality) $\|u + v\| \leq \|u\| + \|v\|$.

# Norms

Given $x, y \in \mathbb{R}^n$, the inner product is given by :

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i.$$

The inner product of $x$ with itself is called the square of the norm of $x$

$$\langle x, x \rangle = \|x\|^2.$$

## Definition

Let $E$ be a $\mathbb{R}$-vector space, then the application $\|.\|$ is said to be a norm if for all $u, v \in E$ and $\lambda \in \mathbb{R}$

1. (positive) $\|u\| \geq 0$ ,
2. (definite) $\|u\| = 0 \iff u = 0$,
3. (scalability) $\|\lambda u\| = |\lambda| \|u\|$,
4. (triangle inequality) $\|u + v\| \leq \|u\| + \|v\|$.

# Norms

Given $x, y \in \mathbb{R}^n$, the inner product is given by :

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i.$$

The inner product of $x$ with itself is called the square of the norm of $x$

$$\langle x, x \rangle = \|x\|^2.$$

## Definition

Let $E$ be a $\mathbb{R}$-vector space, then the application $\|.\|$ is said to be a norm if for all $u, v \in E$ and $\lambda \in \mathbb{R}$

1. (positive) $\|u\| \geq 0$ ,
2. (definite) $\|u\| = 0 \iff u = 0$,
3. (scalability) $\|\lambda u\| = |\lambda| \|u\|$,
4. (triangle inequality) $\|u + v\| \leq \|u\| + \|v\|$.

# Norms

The norm can be seen as distance between two vectors $x, y$ in the same vector space

$$\text{dist}(x, y) = \|x - y\|.$$

Example of usual norms :

- $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ (Manhattan)
- $\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$ (Euclidean)
- $\|x\|_\infty = \max\left(|x_1|, \ldots, |x_n|\right)$
- More generally we define the norm $\|.\|_p$ for all integers $p$ as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

# Norms

The norm can be seen as distance between two vectors $x, y$ in the same vector space

$$\text{dist}(x, y) = \|x - y\|.$$

Example of usual norms :

- $\|x\|_1 = \sum_{i=1}^n |x_i|$ (Manhattan)
- $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$ (Euclidean)
- $\|x\|_\infty = \max(|x_1|, \ldots, |x_n|)$
- More generally we define the norm $\|.\|_p$ for all integers $p$ as

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}.$$

# Norms

The norm can be seen as distance between two vectors $x, y$ in the same vector space

$$\text{dist}(x, y) = \|x - y\|.$$

Example of usual norms :

- $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ (Manhattan)
- $\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$ (Euclidean)
- $\|x\|_\infty = \max(|x_1|, \ldots, |x_n|)$
- More generally we define the norm $\|.\|_p$ for all integers $p$ as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

# Norms

The norm can be seen as distance between two vectors $x, y$ in the same vector space

$$\text{dist}(x, y) = \|x - y\|.$$

Example of usual norms :

- $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ (Manhattan)
- $\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$ (Euclidean)
- $\|x\|_\infty = \max\left(|x_1|, \ldots, |x_n|\right)$
- More generally we define the norm $\|.\|_p$ for all integers $p$ as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

# Example 1/2

We will show that the Euclidean norm is a true norm. Let $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ then

1. It is obvious that $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$ is positive.
2. As $|x_i|^2 \geq 0$ then $\sum_{i=1}^n |x_i|^2 = 0$ if and only if $\forall i,\ x_i = 0$
3. Finally,

$$\begin{aligned} \|\lambda x\|_2 &= \sqrt{\sum_{i=1}^n |\lambda x_i|^2} \\ &= \sqrt{\sum_{i=1}^n |\lambda|^2 |x_i|^2} \\ &= |\lambda| \sqrt{\sum_{i=1}^n |x_i|^2}. \end{aligned}$$

# Example 2/2

To prove the last point we will use the Cauchy-Schwartz Inequality :

$$\langle x, y\rangle \leq \|x\|\|y\|.$$

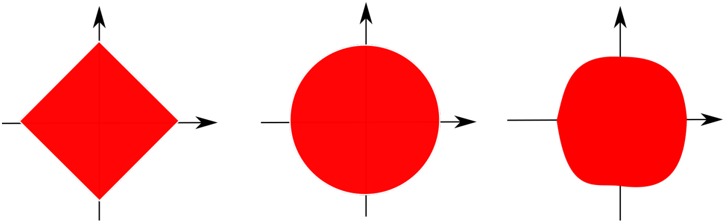
We have,

$$\begin{aligned} \|x+y\|_2^2 &= \|x\|_2^2 + 2\langle x, y\rangle + \|y\|_2^2 \\ &\leq \|x\|_2^2 + 2\|x\|_2\|y\|_2 + \|y\|_2^2 \\ &\leq \left(\|x\|_2 + \|y\|_2\right)^2. \end{aligned}$$

By taking the square root, which is an increasing function, we get the result.

# Norms and Unit Ball

Unit ball for the norms $\|\|_p$ for $p = 1, 2$ and $p > 2$



## Exercise

1. Represent the unit ball for the norm $\|.\|_\infty$.
2. Show that $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ is a norm.

# Correction

- The Unit Ball using the $\|.\|_\infty$ is a full square.
- We have to check the four points of the definition.

1. $\|x\|_1 = \sum_{i=1}^n |x_i| \geq 0$ by definition of the absolute value.
2. $\|x\|_1 = \sum_{i=1}^n |x_i| \geq 0 \implies x = 0$ because the sum of positive numbers is equal to zero if and only if all the terms are equal to zero.
3. $\|\lambda x\|_1 = \sum_{i=1}^n |\lambda x_i| = |\lambda| \sum_{i=1}^n |x_i| = |\lambda| \|x\|_1$.
4. $\|x + y\|_1 = \sum_{i=1}^n |x_i + y_i| \leq \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| = \|x\|_1 + \|y\|_1$.

# Norms on matrices

It is also to define an inner product and a norms on matrices :

1. Given two matrices $X, Y \in \mathbb{R}^{m \times n}$ the **inner product** is defined by :

$$\langle X, Y \rangle = Tr\left(X^T Y\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} y_{ij}.$$

2. A classical norm used with matrices is the Frobenius norm :

$$\|X\|_F = \sqrt{Tr\left(X^T X\right)} = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}^2 \right)^{1/2}.$$

What is the inner product of the symmetric matrices $X, Y \in \mathcal{S}^n(\mathbb{R})$ ?

# Convex Sets

**Definition**

A set $C$ is said to be convex if, for every $(u, v) \in C$ and for all $t \in [0, 1]$ we have :

$$tu + (1 - t)v \in C.$$

In other words, C is said to be convex if **every point on the segment connecting $u$ and $v$ is in the set.**

# Convex Sets

## Definition

A set $C$ is said to be convex if, for every $(u, v) \in C$ and for all $t \in [0, 1]$ we have :

$$tu + (1 - t)v \in C.$$

In other words, C is said to be convex if **every point on the segment connecting $u$ and $v$ is in the set.**

## Proposition

Let $(u_1, u_2, \ldots, u_n)$ be a set of $n$ points belonging to a convex set $C$. Then for every reel numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that $\sum_{i=1}^{n} \lambda_i = 1$ :

$$v = \sum_{i=1}^{n} \lambda_i u_i \in C.$$

# Convex Sets

**Definition**

A set $C$ is said to be convex if, for every $(u, v) \in C$ and for all $t \in [0, 1]$ we have :

$$tu + (1 - t)v \in C.$$

In other words, C is said to be convex if **every point on the segment connecting $u$ and $v$ is in the set.**

**Proposition**

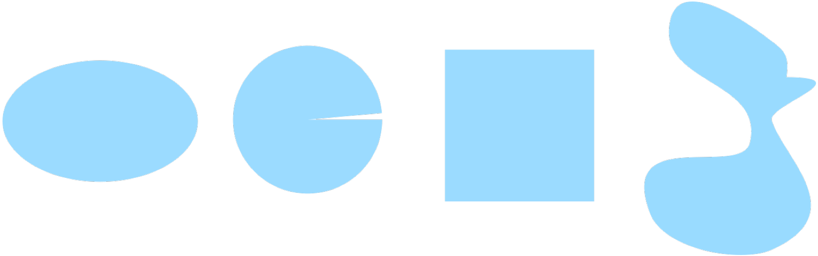
Let $(u_1, u_2, \ldots, u_n)$ be a set of $n$ points belonging to a convex set $C$. Then for every reel numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that $\sum_{i=1}^{n} \lambda_i = 1$ :

$$v = \sum_{i=1}^{n} \lambda_i u_i \in C.$$

**Every convex combination** of points in a convex set is **in the convex set**.

# Convex Sets

Which of the sets are convex ?



# Examples of Convex Sets

1. $\mathcal{B} = \{u \in \mathbb{R}^n \mid \|u\| \leq 1\}$ is convex.
2. Every segment in $\mathbb{R}$ is convex.
3. Every hyperplane $\{x \in \mathbb{R}^n \mid a^T x = b\}$ is convex.
4. If $C_1$ and $C_2$ are two convex sets, then the intersection $C = C_1 \cap C_2$ is also convex.

# Examples of Convex Sets

1. $\mathcal{B} = \{u \in \mathbb{R}^n \mid \|u\| \leq 1\}$ is convex.
2. Every segment in $\mathbb{R}$ is convex.
3. Every hyperplane $\{x \in \mathbb{R}^n \mid a^T x = b\}$ is convex.
4. If $C_1$ and $C_2$ are two convex sets, then the intersection $C = C_1 \cap C_2$ is also convex.

**Exercise**

1. Prove that the Euclidean Unit Ball is convex.
2. (At home) Prove that a set A is convex if and only if its intersection with any line is convex.

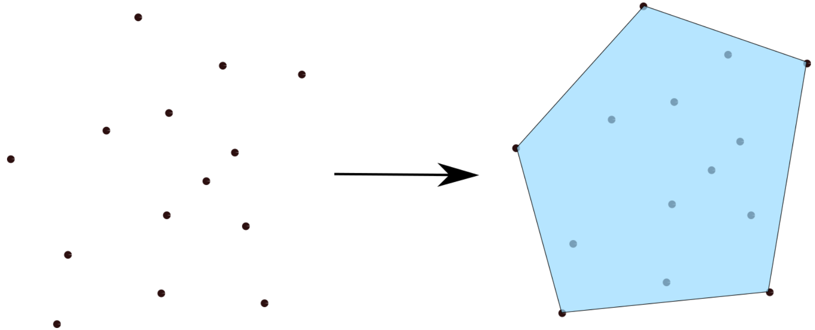
# Correction

- For the first point, consider $\lambda \in [0,1]$ and $u, v$ two vectors in the unit ball. Then set $z = \lambda u + (1-\lambda)v$. (i) take the norm of $z$, (ii) apply the triangle inequality and (iii) the scalability of the norm.
- Use the definition of convexity

# Build a convex set

For a convex set and a set of point $x_1, \ldots x_n$, it is possible to build a convex set. This new set is called the convex hull $\mathcal{H}$ of a set of points

$$\mathcal{H} = \{y = \sum_{i=1}^{n} \lambda_i x_i \mid \sum_{i=1}^{n} \lambda_i = 1\}.$$



# Derivative for real functions

**Recall**

Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and $x_0 \in \mathbb{R}$. We say that $f$ is differentiable at $x_0$ if the limit :

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

exists and is finite.

If $f$ is continuously differentiable at $x_0$, so for $h \simeq 0$ we have

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \varepsilon(h).$$

This formula (Taylor's Formula) can be generalized to a function $g$ n-times continuously differentiable :

$$f(x_0 + h) = f(x_0) + \sum_{i=1}^{n} \frac{h^{(i)}}{i!} f^{(i)}(x_0) + \varepsilon(h^n).$$

# First order derivative

## Definition

Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a $C^0$ application and $x \in \mathbb{R}^m$. Then $f$ is differentiable at $x_0$ if it exists $J \in \mathbb{R}^{m \times n}$ such that :

$$\lim_{x \to x_0} \quad \frac{\|f(x) - f(x_0) - Jf(x_0)(x - x_0)\|}{\|x - x_0\|} = 0.$$

D is called the Jacobian of the application $f$.

For all $i, j$ the elements of the matrix $J$ are given by :

$$J_{ij}f(x_0) = \left.\frac{\partial f_i(x)}{\partial x_j}\right|_{x=x_0}$$

# First order derivative

**Remark**

Usually $f : \mathbb{R}^m \to \mathbb{R}$ so the Jacobian of the application $f$ (also called the gradient) will be a **vector** $\nabla f(x_0)$

The gradient gives the possibility to approximate the function near the point its gradient is calculated. For all $x \in V(x_0)$ we have

$$f(x) \simeq f(x_0) + \nabla f(x_0)(x - x_0)$$

**This affine approximation of the function $f$ will help us to characterize convex functions.**

# First order derivative : example

Let us consider a function $f : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$f(x, y, z) = 3x^2 + 2xyz + 6z + 5yz + 9xz.$$

We want to calculate the Jacobian of this function. To do so, we need to calculate : $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$. The Jacobian of $f$ at $(x, y, z)$ is given by :

$$J_{f(x,y,z)} = \begin{pmatrix} 6x + 2yz + 9z, & 2xz + 5z & 2xy + 6 + 5y + 9x \end{pmatrix}$$

# First order derivative

## Exercise

1. Let $x, y, z \in \mathbb{R}^n$. Calculate the Jacobian of the function
$$f(x,y,z) = exp(xyz) + x^2 + y + log(z).$$
2. Linear Regression. Let $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times d}$ and $\beta \in \mathbb{R}^d$. Calculate the derivative of the function
$$f(\beta) = \|Y - X\beta\|_2^2$$
3. Log-Sum-Exp. Let $x, b \in \mathbb{R}^n$. Calculate the derivative of the function
$$f(x) = \log \sum_{i=1}^{n} \exp\left(x_i + b_i\right)$$

# Correction

- You simply have to apply the definition as it we have done in the previous example and you will have :

$$\nabla f(x,y,z) = \left( yz\exp(xyz) + 2x, xz\exp(xyz) + 1, xy\exp(xyz) + \frac{1}{z} \right).$$

- Here, you have to use the face that : $\|x\|^2 = \langle x, x \rangle$. Then you compute the derivative using the fact that $f$ is defined as a product of two functions of $\beta$.

$$\nabla f(\beta) = -X^T(Y - X\beta) + ((Y - X\beta)^T(-X))^T = -2X^T(Y - X\beta).$$

- Remember that the Jacobian $\nabla_f = J_f$ is a vector where each entry $i$ is equal to :

$$\nabla f(x)_i = \frac{\exp(x_i + b_i)}{\sum_{i=1}^n \exp(x_i + b_i)}.$$

# Second order derivative

**Definition**

Let $f : \mathbb{R}^m \to \mathbb{R}$ be a real function. Provided that this function is twice diffentiable, the second derivative $H$, (also called the Hessian)of $f$ at $x_0$ is given by :

$$H_{ij}f(x_0) = \left.\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right|_{x=x_0},$$

and $H \in \mathbb{R}^{m \times m}$

**Hessian is useful** to prove that a function $f$ is **convex** or not and also to build efficient algorithms.

# Second order derivative : example

Let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = 4x^2 + 6y^2 + 3xy + 2\left(\cos(x) + \sin(y)\right)$$

and calculate the Hessian of this function. We first have to calculate the Jacobian of the matrix and then the Hessian.

$$J_{f(x,y)} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 8x + 3y - 2\sin(x) & 12y + 3x + 2\cos(y) \end{pmatrix}$$

$$H_{f(x,y)} = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial^2 y} \end{pmatrix} = \begin{pmatrix} 8 - 2\cos(x) & 3 \\ 3 & 12 - 2\sin(y) \end{pmatrix}$$

# Second order derivative : example

**Exercise**

Calculate the second order derivative of the following functions :

- $f(x, y) = \log(x + y) + x^2 + 2y + 4$
- $f(x, y, z) = \dfrac{6x}{1 + y} + \exp(xy) + z$

# Correction

The process is similar as in the previous example, so I only give the results.

$$H_{f(x,y)} = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial^2 y} \end{pmatrix} = \begin{pmatrix} 2 - \frac{1}{(x+y)^2} & -\frac{1}{(x+y)^2} \\ -\frac{1}{(x+y)^2} & -\frac{1}{(x+y)^2} \end{pmatrix}$$

$$H_{f(x,y)} = \begin{pmatrix} y^2 \exp(xy) & -\frac{6}{(1+y)^2} + (xy+1)\exp(xy) \\ -\frac{6}{(1+y)^2} + (xy+1)\exp(xy) & \frac{12x}{(1+y)^3} + x^2 \exp(xy) \\ 0 & 0 \end{pmatrix}$$