# UE Optimisation:

## Introduction à l’optimisation

“La nature agit toujours par les voies les plus courtes et les plus simples”

Pierre de Fermat, 1657

# Plan

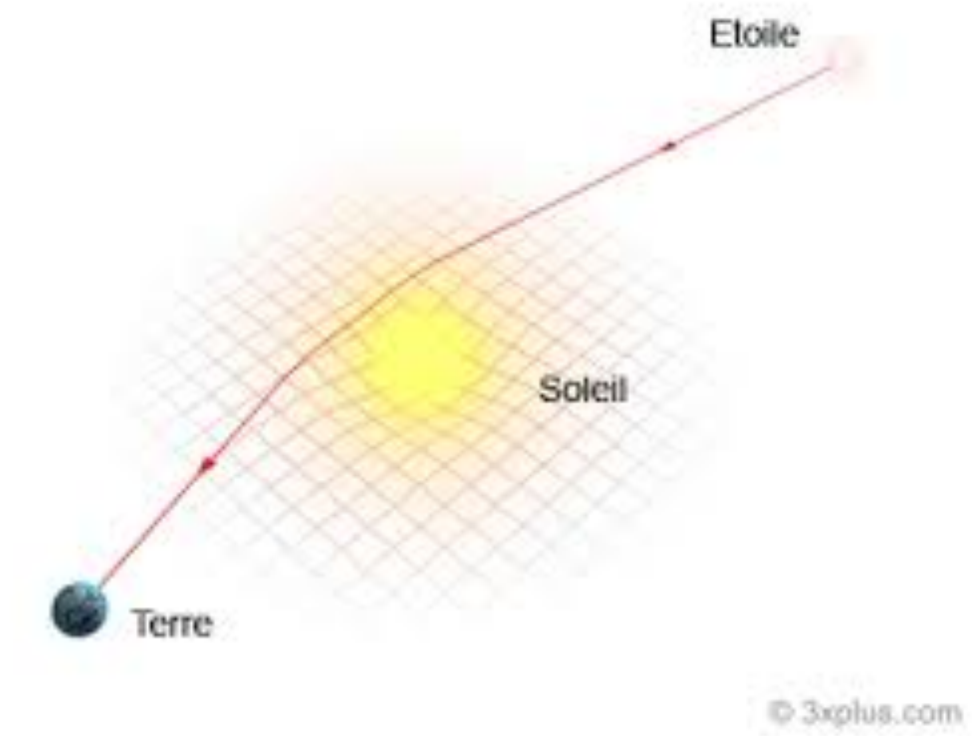
- 1. Introduction

# Introduction

- **Principe de moindre action**
  - En physique, le comportement observé d’un système correspond à la **minimisation** (ou **à la maximisation**) d’une certaine grandeur relative à ce phénomène.
  - Principe d’abord phénoménologique mis en évidence au 17$^{e}$ siècle en optique (Fermat).

# Introduction

- **Principe de moindre action**



# Introduction

- **Principe de moindre action**
  - La lumière se propage d'un point à un autre sur des trajectoires telles que la durée du parcours est stationnaire.
  - **Il s'agit d'un problème d'optimisation :** la durée du parcours doit être extrémale, en générale minimale, par rapport à une petite variation du trajet.

# Introduction

- **Principe de moindre action**
  - A l’aide de ce principe, on démontre immédiatement la loi de propagation rectiligne de la lumière dans un milieu homogène…
  - Et les lois de Snell-Descartes (réfraction, réflexion) qui en découlent.

# Introduction

- **Le problème de la Reine Didon : l'isopérimétrie**
  - L'isopérimétrie a débuté avec le problème auquel a été confrontée la Reine Didon, qui a du trouver la forme de la frontière à poser au sol (en utilisant des bandes de peau de boeuf) **pour encercler une surface d'aire maximale**.
  - Si l'on suppose la côte droite, alors la réponse, qui a manifestement été trouvée par la Reine Didon, est de poser ces bandes en forme de demi-cercle.
  - C'est ainsi que Carthage serait né ☺

http://math.arizona.edu/~dido/lord-elin.html

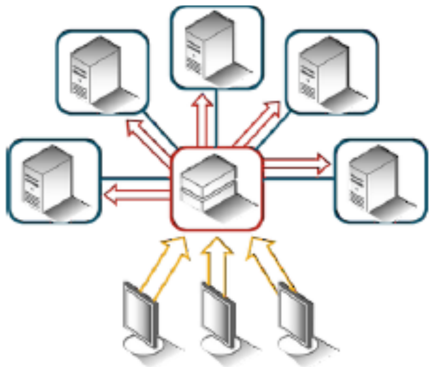
# Introduction

- **Optimisation : une définition**
  - En mathématique, un problème d'optimisation consiste à trouver, parmi un ensemble donné, un élément* **minimisant ou maximisant une fonction donnée de cet ensemble sur IR.**

**dans le cadre de ce cours, il s'agira souvent d'un vecteur de $IR^n$ mais ce peut être aussi un vecteur d'entiers, une fonction…*

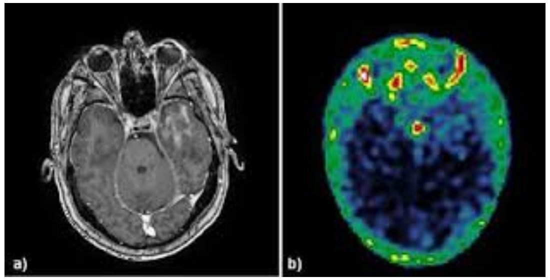
# Introduction


Chaîne logistique

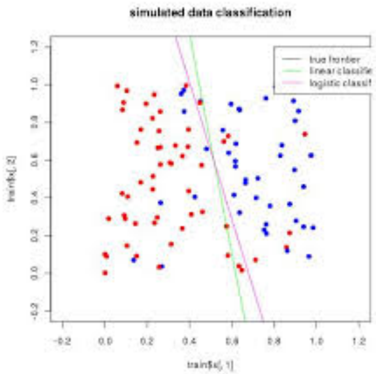


Equilibrage de charge sur un réseau *(électrique, telecom)*

Traitement du signal, de l'image *(recalage)*




Planification *(Scheduling)*




Apprentissage statistique *(SVM)*

# Introduction

- **Optimisation : en résumé**
  - L'art de formuler les problèmes *(de décision)*
  - Une théorie mathématique *(un mal nécessaire ☺)*
  - Des techniques algorithmiques

***Entrons dans le vif du sujet***

# Plan

- 1. Introduction
- 2. **Formalisation**

# Formalisation

- **En écriture mathématique, optimiser c'est :**

$$J(x^*) = \min_{x \in C} J(x) \longleftarrow x = (x_1, x_2 \ldots, x_n)^T \in \Re^n$$

Fonction de coût

$J : \Re^n \mapsto \Re$

Ensemble des contraintes

$C \subset \Re^n$

Variables recherchées

$x^* \in C$

# Formalisation

- **Exemple : Le portefeuille de Markovitz**
  - *En français :* l'objectif est d'optimiser les sommes d'argent dédiées à divers placement bancaires (caractérisés par un rendement)
    - En minimisant les risques de perte...
    - Et en ne plaçant pas plus que ce qu'il y a sur notre compte en banque
    - *How to become rich?*

# Formalisation

- **Exemple : Le portefeuille de Markovitz**
  - *En langage « optimisation » :*

$$\min_{x \in \Re^n} \left( -\sum_1^n r_i x_i + \frac{1}{2} \alpha x^T \Sigma x \right); \quad x \geq 0, \sum_1^n x_i = 1$$

  - *$r=(r_1, r_2, \ldots, r_n)$ sont les rendements des parts investies*
  - *α est un coefficient d'aversion au risque*
  - *Σ la matrice de variance/covariance des montants placés*

# Formalisation

$$\min_{x\in\Re^n}\left(-\sum_1^n r_i x_i+\frac{1}{2}\alpha x^T\Sigma x\right);\quad x\geq 0,\ \sum_1^n x_i=1$$

On maximise l'objectif de rendement

On minimise les risques de placement au regard d'un critère donné

On ne place pas de l'argent qu'on a pas !

**Les contraintes définissent un domaine de faisabilité**

# Formalisation

- **Cas d'une fonction à une seule variable**
  - *Minimisation*

$$J(x^*) = \min_{x \in C} J(x)$$

  - *Illustration : Le problème du maître nageur*

# Formalisation

- **Le problème du maître nageur**
  - Telle la lumière qui se propage moins vite dans l'eau que dans l'air, un(e) maître nageur(se) court plus vite qu'il (elle) ne nage.
  - Il (elle) se trouve au point **A** (plage) lorsqu'il aperçoit un(e) jolie fille (garçon) qui se noie en **B** (océan).
  - **Question :** Comment arriver en **B** le plus vite possible ?

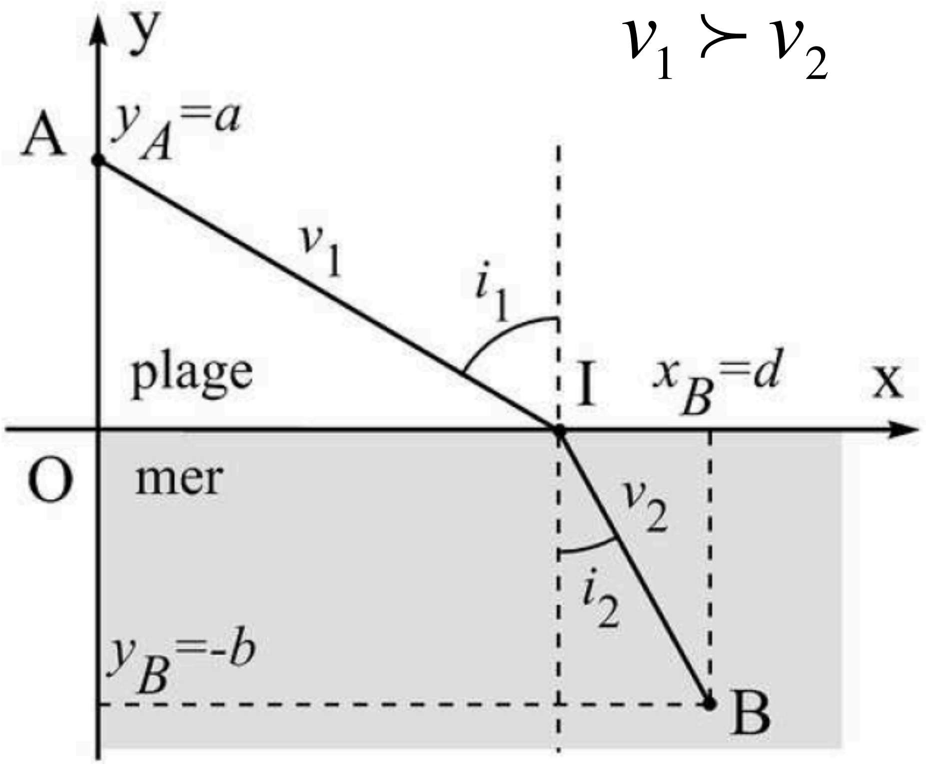
# Formalisation

- **Le problème du maître nageur**
  - **Réponse :**
  - Il faut trouver un compromis entre la ligne droite et le « chemin » qui **rend minimal le temps de parcours.**
  - La réponse mathématique a été donnée par Maupertuis (bien avant David Hasselhoff !) en 1744.

# Formalisation

- **Illustration :**

$$v_1 \succ v_2$$

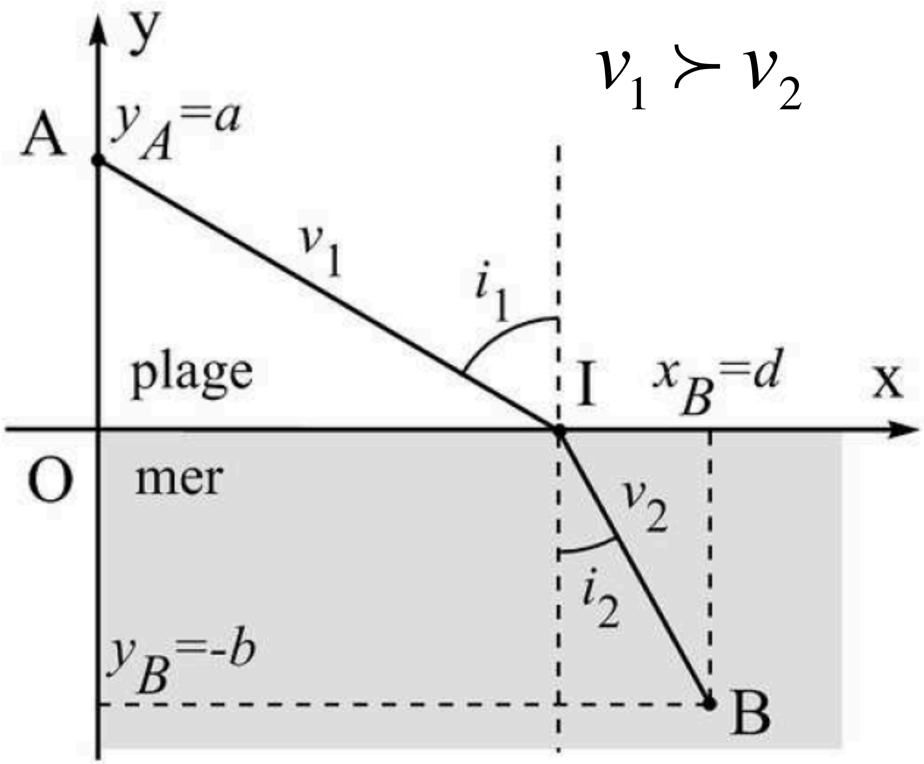



La trajectoire du maître nageur est constituée de deux droites **AI** et **IB**, où *I(x, 0)* est le point où le maître nageur plonge.

A priori la distance **AI** sera plus grande que la distance **IB** car il court plus vite qu'il ne nage.

# Formalisation

- **Illustration :**



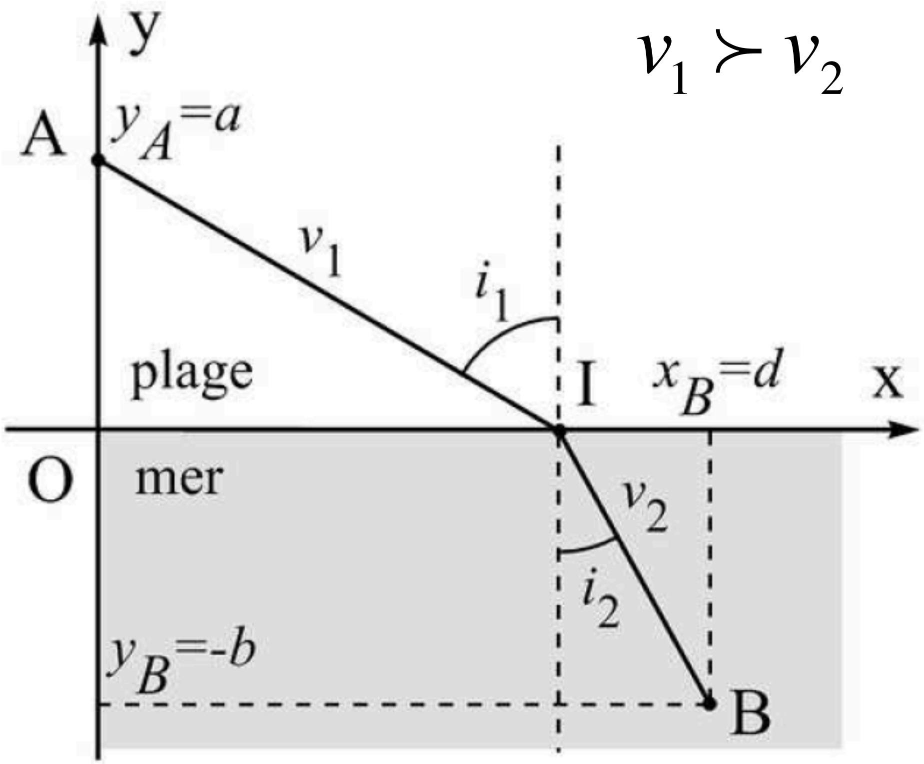

$$v_1 \succ v_2$$

La fonction temporelle *t* mis par le maître nageur pour aller de A en B est :

$$t(x) = \frac{AI}{v_1} + \frac{IB}{v_2}$$

# Formalisation

- **Illustration :**




$$v_1 \succ v_2$$

$$t(x) = \frac{AI}{v_1} + \frac{IB}{v_2}$$

**Problème de minimisation :**

$$x^* = \arg\min_{x \in \Re^+} t(x)$$

# Formalisation

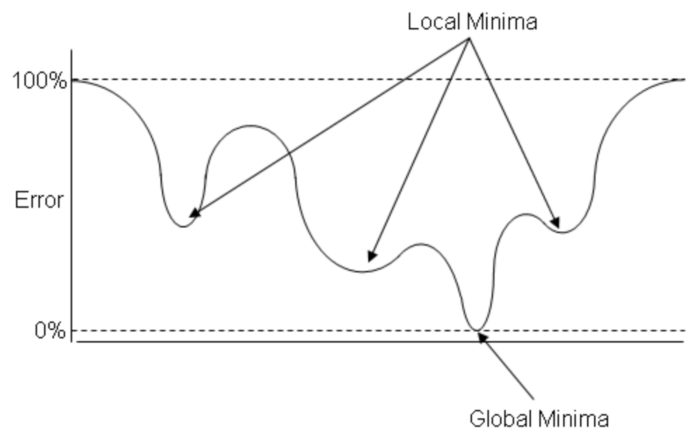
- **Solution (à défaut d'avoir un hors-bord) :**
  - La fonction *t* est minimale quand sa dérivée première est nulle

$$\frac{dt(x)}{dx} = 0 = \frac{1}{v_1}\frac{x}{\sqrt{x^2 + a^2}} - \frac{1}{v_2}\frac{d - x}{\sqrt{(d-x)^2 + b^2}}$$

  - **NB** Ou encore : $\frac{1}{v_1}\sin i_1 = \frac{1}{v_2}\sin i_2$

# Formalisation

- **Quelques remarques**
  - Dans cet exemple, nous avons fait les hypothèses implicites suivantes :
    - Le minimum de la fonction *t* existe
    - La fonction *t* est dérivable (et même $C^1$)
  - Autres questions possibles
    - Le minimum est-il unique ?
    - Et si non, le minimum est-il
      - global ?
      - local ?



# Formalisation

- **Quelques remarques**
  - **Sur la solution**
    - Dans ce cas simple, on peut la déterminer analytiquement en résolvant $t'(x)=0$,

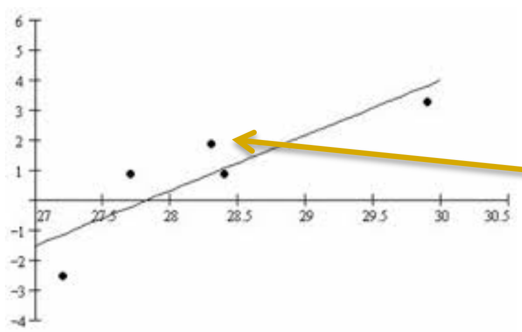
      ***Il s'agit alors d'une solution exacte***

    - Quand on ne peut résoudre analytiquement le problème, on cherche une solution approchée

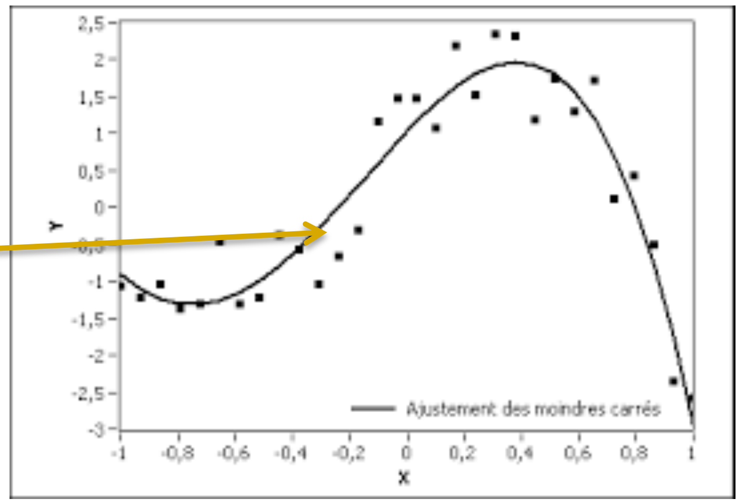
      ***On construit alors une suite $x^k$ convergente (descente de gradient par exemple)***

# Formalisation

- **Problème multidimensionnel**
  - **Même questionnement qu'à une variable mais en « un peu plus » compliqué**
  - *Exemple :* Algorithme des moindres carrés




On cherche à minimiser la somme des carrés des écarts des valeurs observées à la fonction « référence » utilisée pour la régression



# Formalisation

- **Problème multidimensionnel**
  - **Même questionnement qu'à une variable mais en « un peu plus » compliqué**
  - *Exemple :* Algorithme des moindres carrés

$$f(\theta^*) = \min_{\theta \in \Re^n} \sum_1^p \left(y_i - g(x_i : \theta)\right)^2$$

Résidu de l'erreur de distance de la fonction de paramètres aux données d'observations $(x_i, y_i)$

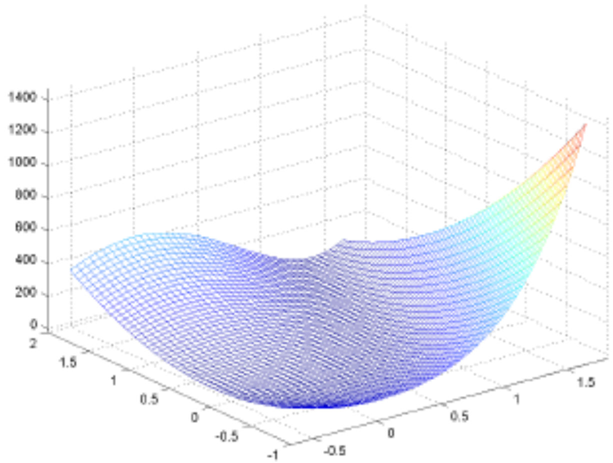
Résolution : $$\nabla_\theta f(\theta^*) = 0$$

# Formalisation

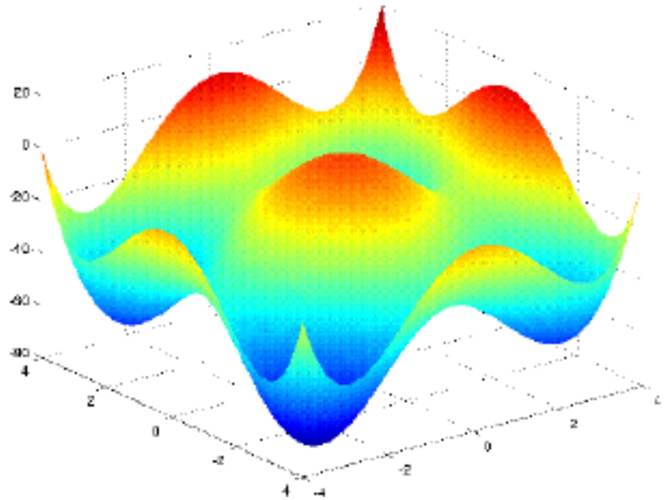
- **Problème multidimensionnel**
  - **Minimum global ou local**

$$x^l = \arg \min_{x \in V_\varepsilon(x^l)} J(x)$$

$V_\varepsilon(x)$ Un voisinage de $x$



$$x^* = \arg \min_{x \in \Re^2} J(x)$$



# Formalisation

- **Synthèse**
  - *Un problème d'optimisation c'est :*
    - *Une fonction de coût à minimiser*
    - *Très souvent des contraintes à respecter*
  - *Les questions associées*
    - *Dimensionnalité de la variable ?*
    - *Propriété de la fonction de coût ? (continuité, dérivabilité, topologie, linéarité)*
    - *Unicité de la solution ?*
    - *Méthodes de résolution correspondantes ?*

# Plan

- 1. Introduction
- 2. Formalisation
- 3. Rappels de mathématiques

# "Rappels" de mathématiques

- **Quelques définitions fondamentales**
  - *Définition 1 : <u>produit scalaire</u>*

Pour tout $x, y \in \Re^n$, on note $\langle x, y \rangle$ le produit scalaire de $x$ et $y$ qui est donné par

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

# "Rappels" de mathématiques

- **Quelques définitions fondamentales**
  - *Définition 2 : <u>norme euclidienne</u>*

Pour tout $x \in \Re^n$, on note $\|x\| \geq 0$ la norme euclidienne de $x$ qui est donnée par

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\langle x, x \rangle}$$

# “Rappels” de mathématiques

- **Proposition 1**
  - *Inégalité de Cauchy-Schwarz*

$$\text{Pour tout } x, y \in \Re^n$$

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

# “Rappels” de mathématiques

## Aparte sur la notion de norme

- *Soit E un espace vectoriel sur R.*
- *Une norme* N *est une application de E dans R telle que :*

$$\boxed{\begin{array}{l} 1.\ \forall x \in E, \quad \mathrm{N}(x) = 0 \Leftrightarrow x = 0 \\ 2.\ \forall x \in E, \quad \mathrm{N}(x) \geq 0 \\ 3.\ \forall x, y \in E, \quad \mathrm{N}(x+y) \leq \mathrm{N}(x) + \mathrm{N}(y) \\ 4.\ \forall x \in E, \forall \lambda \in \Re, \quad \mathrm{N}(\lambda x) = |\lambda| \mathrm{N}(x) \end{array}}$$

**Rq:** Un espace vectoriel muni d’une norme est appelé espace vectoriel normé (EVN)

# “Rappels” de mathématiques

- **Aparte sur la notion de norme**
  - *Quelques exemples classiques sur $R^n$*
    - *La classique norme $L^2$ donc ou distance euclidienne*

$$\forall x \in \Re^n, \quad \|x\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$

    - *La norme $L^1$*

$$\forall x \in \Re^n, \quad \|x\|_1 = |x_1| + |x_2| + \ldots + |x_n|$$

# “Rappels” de mathématiques

- **Aparte sur la notion de norme**
  - *Quelques exemples classiques sur $R^n$*
    - *Et la norme $L^\infty$*

$$\forall x \in \Re^n, \quad \|x\|_\infty = \max\left(x_1, x_2, \ldots, x_n\right)$$

# “Rappels” de mathématiques

- **Quelques définitions fondamentales**
  - *Définition 3 : <u>fonction de classe $C^m$</u>*

$$\boxed{\begin{array}{c}\text{Soit } \Omega \subset \Re^n,\ f \text{ est de classe } C^m \text{ sur } \Omega \\ \text{si toutes les dérivées partielles jusqu'à} \\ \text{l'ordre } m \text{ de } f \text{ existent et sont continues}\end{array}}$$

# “Rappels” de mathématiques

- **Quelques définitions fondamentales**
  - *Définition 4 : <u>Dérivées partielles</u>*

Quand elle existe, on définit la dérivée partielle de la fonction $f$ dans la direction $x_i$ de la manière suivante :

$$\frac{\partial f(x)}{\partial x_i} = \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}$$

avec $e_i$ le symbole de Kronecker

$(e_i)_j = 1$ si $i = j$

0 sinon

# “Rappels” de mathématiques

- **Quelques définitions fondamentales**
  - *Définition 5 : Gradient*

Quand il existe, on note

$$\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)^T$$

le gradient de la fonction $f$

# “Rappels” de mathématiques

- **Quelques définitions fondamentales**
  - *Définition 6 : <u>Dérivée de f dans la direction h</u>*

D'une manière générale, la dérivée de

de $f$ dans la direction $d \in \Re^n$ est donnée par

$$\frac{\partial f(x)}{\partial d} = \left\langle \nabla f(x), d \right\rangle = \nabla f(x)^T d$$

# “Rappels” de mathématiques

- **Quelques définitions fondamentales**
  - *Définition 7 : <u>Matrice hessienne de f</u>*

Le Hessien de la fonction $f$ quand il existe est la matrice telle que :

$$\left(\nabla^2 f(x)\right)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

**Rq :** Si la fonction *f* est $C^2$ le hessien est une matrice symétrique (Théorème de Schwarz)

# “Rappels” de mathématiques

- **Proposition 2**
  - *Formules de Taylor-Young*

$$\text{Soit } a \in \Omega \ (\subset \Re^n) \text{ et } h \in \Re^n$$

$$\text{Si } f \text{ est } C^1 \text{ sur } \Omega$$

$$f(a+h) = f(a) + \left\langle \nabla f(a), h \right\rangle + \circ\left(\|h\|\right)$$

$$\text{Si } f \text{ est } C^2 \text{ sur } \Omega$$

$$f(a+h) = f(a) + \left\langle \nabla f(a), h \right\rangle + \frac{1}{2}\left\langle \nabla^2 f(a)h, h \right\rangle + \circ\left(\|h\|^2\right)$$

- **Définition 8**
  - *Minimum absolu*

Soit $U \subset \Re^n$, $u^* \in U$ et $f : U \mapsto \Re$

On dit que $u^*$ est un point de minimum **absolu** (ou global)

de $f$ sur $U$ si

$$f(u) \geq f(u^*),\ \forall u \in U$$

# “Rappels” de mathématiques

- **Définition 9**
  - *Minimum relatif*

$$\text{Soit } U \subset \Re^n,\ u^* \in U \text{ et } f : U \mapsto \Re$$

On dit que $u^*$ est un point de minimum **relatif** (ou local) de $f$ sur $U$ si $\exists V$ un voisinage de $u^*$ dans $\Re^n$ tel que

$$f(u) \geq f(u^*),\ \forall u \in U \cap V$$

# “Rappels” de mathématiques

- **Théorème 1**
  - *Condition nécessaire d’optimalité*

Soit $U \subset \Re^n$, $u^* \in U$ et $f : U \mapsto \Re$ de classe $C^2$

**Si** $u^*$ est un minimum local de $f$ sur $U$) **alors**

1. $\nabla f(u^*) = 0$

2. $\nabla^2 f(u^*)$ est une matrice définie positive

# “Rappels” de mathématiques

- **Théorème 2**
  - *Condition suffisante d’optimalité*

Soit $U \subset \Re^n$, $u^* \in U$ et $f : U \mapsto \Re$ de classe $C^2$

**Si** $f$ vérifie en $x^*$

1. $\nabla f(u^*) = 0$

2. $\nabla^2 f(u^*)$ est une matrice définie positive

**alors** $f$ admet un minimum local en $u^*$

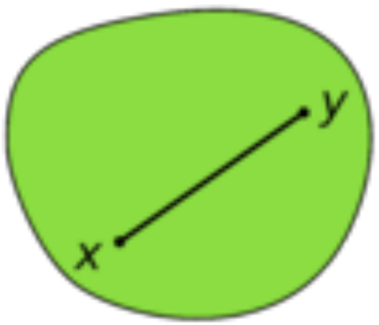
# Plan

- 1. Introduction
- 2. Formalisation
- 3. Rappels de mathématiques
- 4. Optimisation et convexité
- 5. Classification

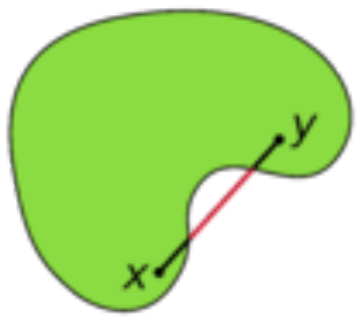
# Optimisation et convexité

- **Convexité**
  - *Ensemble convexe*

$$\boxed{\begin{array}{l}\text{Soit } U \subset \Re^n,\ U \text{ est convexe si} \\ \quad \forall x, y \in R^n, \quad [x, y] \subset U\end{array}}$$




Convexe




Non-Convexe

# Optimisation et convexité

- **Convexité**
  - *Ensemble convexe: notation équivalente*

$$\boxed{\begin{array}{c}\text{Soit } U \subset \Re^n, \text{ et } t \in [0,1], \ U \text{ est convexe si} \\ \forall x, y \in R^n, \quad tx + (1-t)y \subset U\end{array}}$$


Convexe Non-Convexe

# Optimisation et convexité

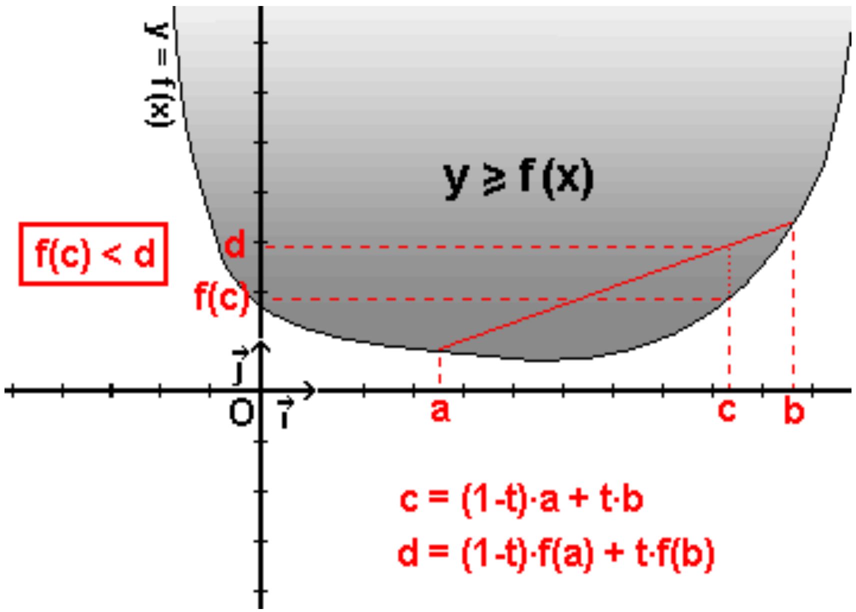
- **Convexité**
  - *Fonction convexe*

$$\text{Soit } U \subset \Re^n,\ \ t \in [0,1],\ \text{et}\, f : U \to \Re,$$
$$f \ \text{ est convexe sur } U \ \text{ ssi}$$
$$\forall x, y \in R^n, \quad f\left(tx + (1-t)y\right) \leq t.f(x) + (1-t).f(y)$$

**Rq :** On parle de stricte convexité si l'inégalité ci-dessus est stricte

# Optimisation et convexité

- **Convexité**
  - *Fonction convexe : illustration*



# Optimisation et convexité

- **Convexité**
  - *Fonction convexe : en pratique*
    - La condition précédente n'est en pratique pas simple à vérifier
    - En pratique donc, on préfère utiliser des conditions sur le gradient ou le hessien.

# Optimisation et convexité

- **Convexité**
  - *Fonction convexe : en pratique*
    - La condition précédente n'est en pratique pas simple à vérifier
    - En pratique donc, on préfère utiliser des conditions sur le **gradient (si f est de classe *C*$^1$)** ou le hessien:

$$\boxed{\begin{array}{c}\text{Soit } U \subset \Re^n, \quad f : U \to \Re, \\ f \text{ est convexe sur } U \textbf{ ssi} \\ \forall x, y \in U, \quad \left\langle \nabla f(y) - \nabla f(x), x - y \right\rangle \geq 0\end{array}}$$

**Rq:**
Le gradient de *f* est dit monotone

# Optimisation et convexité

- **Convexité**
  - *Fonction convexe : en pratique*
    - La condition précédente n'est en pratique pas simple à vérifier
    - En pratique donc, on préfère utiliser des conditions sur le gradient ou le **hessien (si f est de classe *C*²)** :

$$\boxed{\begin{array}{c}\text{Soit } U \subset \Re^n, \quad f : U \to \Re, \\ f \text{ est convexe sur } U \text{ ssi} \\ \forall x, y \in U, \quad \left\langle \nabla^2 f(x)(x-y), x-y \right\rangle \geq 0\end{array}}$$

# Optimisation et convexité

- **Convexité**
  - *Fonction convexe : exemple*
    - Montrer la convexité (et même la stricte convexité) de la fonction définie de *R* dans *R*

$$f(x) = x^2$$

# Optimisation et convexité

- **Convexité**
  - *Fonction convexe : exemple*
    - Montrer la convexité (et même la stricte convexité) de la fonction définie de *R* dans *R*

$$f(x) = x^2$$

    - Les normes sont des fonctions convexes

# Optimisation et convexité

- **Concavité**
  - *Fonction concave*

$$\boxed{\begin{array}{c}\text{Soit } U \subset \Re^n, \ \text{ et } f: U \to \Re, \\ f \text{ est concave sur } U \text{ ssi} \\ \ldots\end{array}}$$

# Optimisation et convexité

- **Concavité**
  - *Fonction concave*

# Optimisation et convexité

- **Concavité**
  - *Fonction concave*

$$\text{Soit } U \subset \Re^n, \ \text{ et } f : U \to \Re,$$
$$f \text{ est concave sur } U \text{ ssi}$$
$$\textbf{-f est convexe}$$

# Optimisation et convexité

- **Théorème 3**
  - *Optimalité et convexité*

$$\boxed{\begin{array}{c}\text{Soit } U \subset \Re^n,\ u^* \in U \ \text{ et } f : U \to \Re, \\ \textbf{Si } f \text{ est convexe sur } U \text{ et admet} \\ \text{un minimum local en } u^* \ \textbf{alors} \\ f \text{ admet aussi un minimum global en } u^*\end{array}}$$

# Optimisation et convexité

- **Théorème 4**
  - *Optimalité et convexité*

Soit $U \subset \Re^n$, $u^* \in U$ et $f : U \to \Re$,

**Si** $f$ est convexe sur $U$ et est au moins $C^1$ sur $U$

**alors**

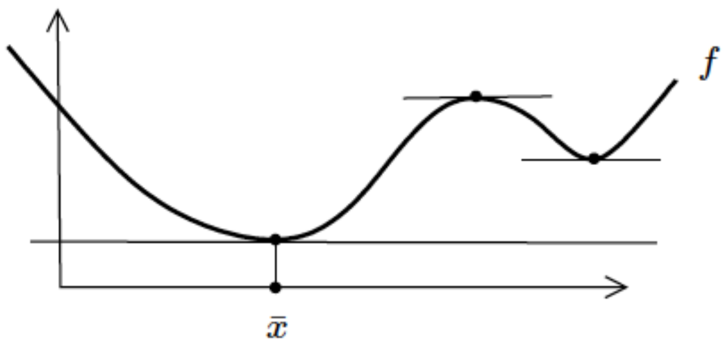
$f$ admet un minimum global en $u^*$ **ssi**

$$\nabla f(u^*) = 0$$

# Optimisation et convexité

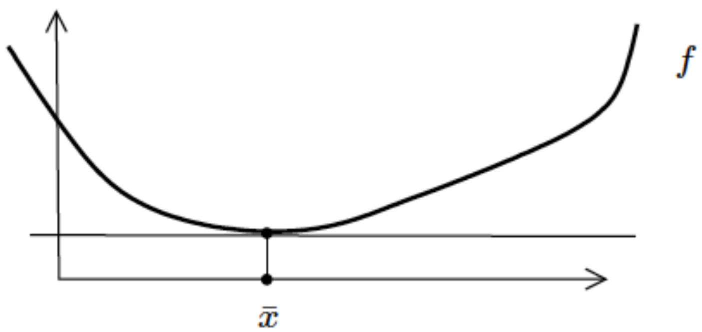
- **Illustration**
  - *Optimalité et convexité*

$$x^* = \arg\min(f(x)) \Rightarrow \nabla f(x^*) = 0$$




*f* est ici différentiable

$$x^* = \arg\min(f(x)) \Leftrightarrow \nabla f(x^*) = 0$$




*f* est ici différentiable
et convexe

# Plan

- 1. Introduction
- 2. Formalisation
- 3. Rappels de mathématiques
- 4. Optimisation et convexité
- 5. Classification

# Classification

- **Comment classer les problèmes d'optimisation**
  - M. Minoux 1983
    - Classer les problèmes d'optimisation en fonction des caractéristiques mathématiques de la fonction de coût et de celles des contraintes du problème posé (s'il y en a)
    - On distinguera alors :
      - *L'optimisation linéaire, quadratique, non-linéaire, convexe*
    - Toute classification a ses imperfections… par exemple pour d'autres (dont Guy Cohen) l'optimisation linéaire fait partie de la branche des pb d'optimisation convexe

# Classification

- ***1. Optimisation linéaire (ou LP)***
  - La fonction de coût et les contraintes sont des fonctions linéaires de la variable *x* considérée (dans $R^n$)
  - *Exemple en dimension 2:*
    - *On cherche à minimiser la fonction coût J telle que :*

$$\begin{cases} J(x_1, x_2) = ax_1 + bx_2 \\ x_1 \geq 0, \, x_2 \geq 0 \\ x_1 + x_2 \leq c \end{cases} \qquad a, b, c \in \Re$$

# Classification

- ***1. Optimisation linéaire (ou LP)***
  - *Généralisation*
    - *On cherche à minimiser la fonction coût J telle que :*

$$\begin{cases} J(x) = \sum_{i=1}^{n} c_i x_i \\ \text{avec } \sum_{i=1}^{n} a_{ij} x_i \le b_j, \quad j = 1, m \\ \text{et } x_j \ge 0 \end{cases}$$

Contraintes d'inégalités

Contraintes de signe

# Classification

- ***1. Optimisation linéaire (ou LP)***
  - *Généralisation*
    - *Forme canonique pure:*

$$\begin{cases} J(x) = c^T x \\ Ax \leq b \\ x \geq 0 \end{cases} \qquad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

# Classification

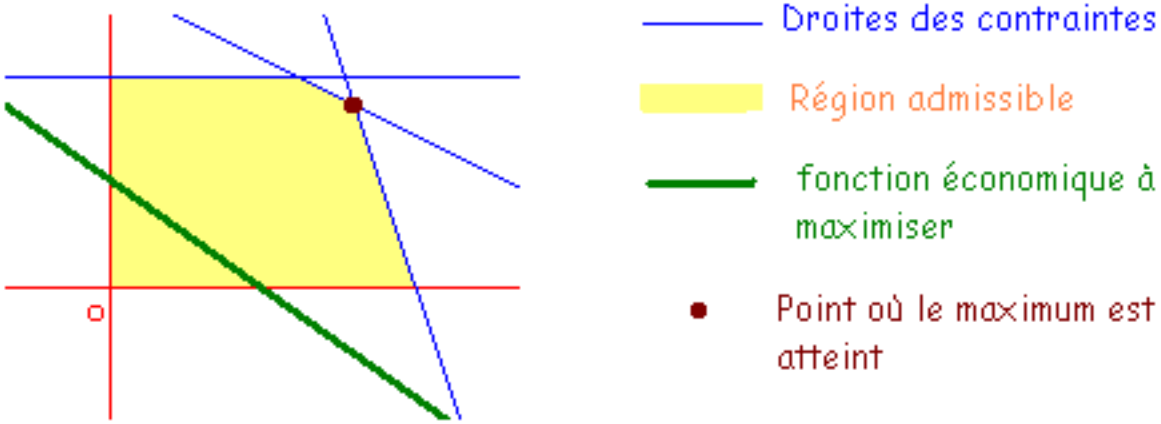
- ***1. Optimisation linéaire (ou LP)***
  - ***Avantages***
    - *Les fonctions linéaires sont convexes*
    - *Un minimum local est donc également globale*
    - *Le minimum est unique si la stricte convexité peut-être montrée (souvent le cas)*
    - *« Facile » à résoudre en programmation linéaire (méthode du simplexe utilisée depuis les années 40)*

# Classification

- ***1. Optimisation linéaire (ou LP)***
  - ***Algorithme du simplexe en dimension 2*** *(maximisation)*

$$J(x,y) = 5x + 7y$$

$$\begin{cases} x + 2y \le 60 \\ 2y \le 42 \\ 3x + y \le 90 \end{cases}$$




Un problème de programmation linéaire

# Classification

- ***1. Optimisation linéaire (LP)***
  - *Un peu d'histoire*
    - ***1900*** *Début de l'étude mathématique de la convexité (H. Minkowski)*
    - ***1947*** *Algorithme du simplexe pour l'optimisation linéaire (G. Dantzig)*
    - ***1948*** *Premières applications militaires "recherche opérationnelle"*
    - ***1970*** *Analyse convexe (W. Fenchel, J.-J. Moreau, T. Rockafellar)*
    - ***1994*** *Algorithme de points intérieurs (Y. Nesterov & A.Nemirovski)*

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *La fonction à minimiser* ***q*** *est définie en $x=(x_1, x_2, x_3, \ldots, x_n)$ de $R^n$ par la relation matricielle :*

$$q(x) = g^T x + \frac{1}{2} x^T H x + c$$

avec $g$ un vecteur de $\Re^n$, $c \in \Re$, et

$H$ une matrice réelle symmétrique

de $\Re^n \times \Re^n$

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *La fonction à minimiser **q** est définie en $x=(x_1, x_2, x_3,\ldots,x_n)$ de $R^n$ par la relation matricielle :*

$$q(x) = g^T x + \frac{1}{2} x^T H x + c$$

  - La constante *c* peut ne pas être prise en compte de par le fait qu'elle n'influe pas sur le problème de minimisation

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *Ou encore sous forme développer en omettant c :*

$$q(x) = \sum_{1}^{n} g_i x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} H_{ij} x_i x_j$$

# Classification

- ***2. Optimisation Quadratique (QP)***
  - ***Convexité :***
  - Une fonction quadratique est convexe **ssi** *H est semi-définie positive*
  - Une fonction quadratique est **strictement** convexe **ssi** *H est définie positive*

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *Rappels au cas où ☺ :*
    - Une matrice carrée symétrique est définie positive si toutes ses valeurs propres sont strictement positives.
    - Une matrice carrée symétrique est semi-définie positive si toutes ses valeurs propres sont positives ou nulles.

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *Rappels (toujours au cas où ☺)*:
    - Une matrice $H$ symétrique de $R^n \times R^n$ possèdent $n$ valeurs propres $\lambda_i$
    - On détermine les valeurs propres en résolvant :

$$\det\left(H - \lambda I\right) = 0$$

$I$ étant la matrice identité et $\lambda=(\lambda_1, \lambda_2, \ldots, \lambda_n)$

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *Les contraintes du problème d'optimisation*
    - Les *m (m<n)* contraintes sont représentées par des fonctions affines linéaires de la variable *x de* $R^n$

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *Les contraintes du problème d'optimisation*
    - Les $m$ $(m<n)$ contraintes sont représentées par des fonctions affines linéaires de la variable $x$ de $R^n$
    - **Contraintes d'égalité**

$$Ax = b$$

$A$ est une matrice de $R^m \times R^n$ et $b$ un vecteur réel de $R^m$

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *Les contraintes du problème d'optimisation*
    - Les *m* ($m<n$) contraintes sont représentées par des fonctions affines linéaires de la variable *x de* $R^n$
    - **Contraintes d'inégalité**

$$b_l \leq Ax \leq b_h$$

$A$ est une matrice de $R^m \times R^n$ et $b_l$ et $b_h$ deux vecteur réels de $R^m$

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *En résumé*

$$J(x^{*}) = \min\left(g^{T}x + \frac{1}{2}x^{T}Hx\right)$$

$$\text{sous contraintes que}$$

$$b_{l} \leq Ax \leq b_{h} \text{ ou } Ax = b$$

# Classification

- ***2. Optimisation Quadratique (QP)***
  - *Résolution*
    - *Si les contraintes sont des égalités, on retombe sur un problème d'optimisation linéaire*
    - *Pour des contraintes de type inégalité, le problème devient NP-complexe… ☹*
      - *Algorithme du gradient conjugué (des extensions),*
      - *Algorithme du gradient projeté,*
      - *Algorithme du simplexe (des extensions).*
      - *Algorithmes de points intérieurs.*

# Classification

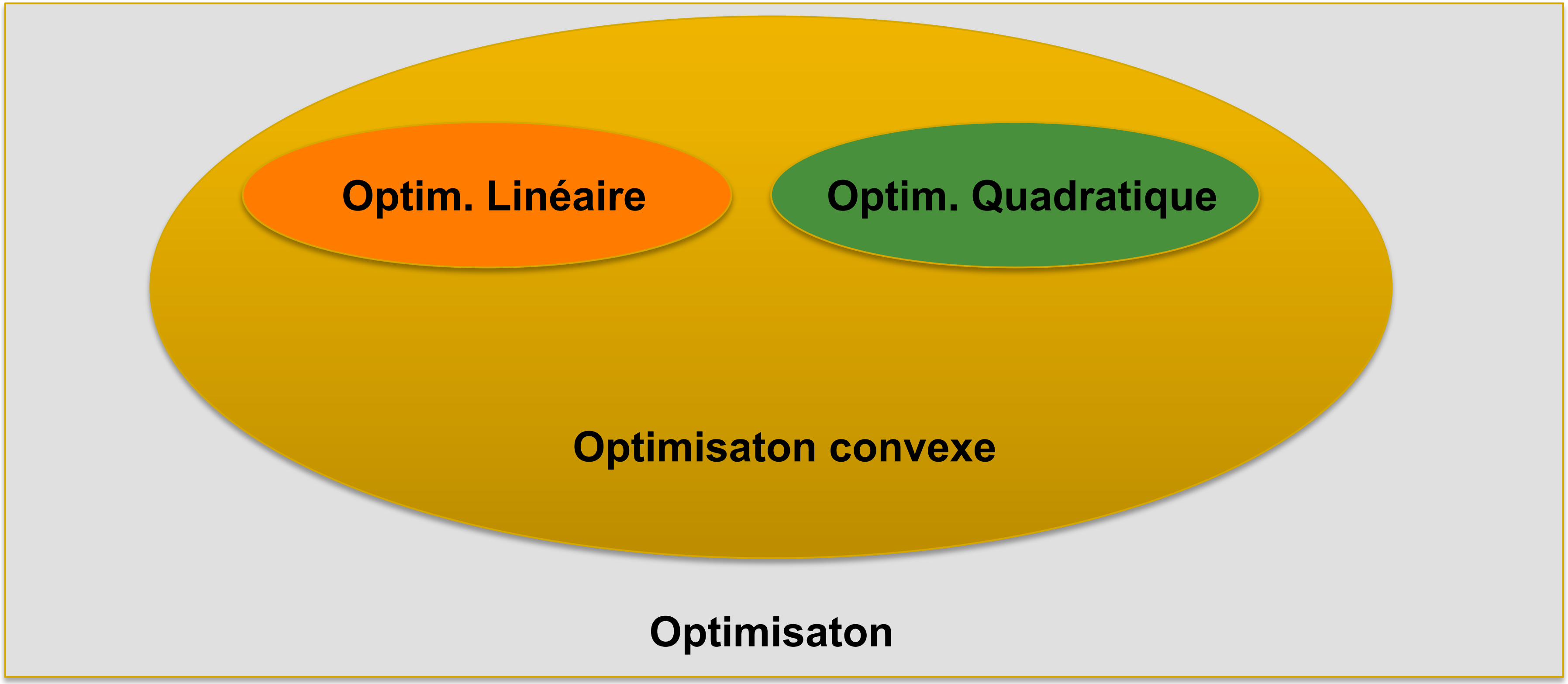
- ***2. Optimisation Quadratique (QP)***
  - *Cas particulier* ***d'un pb QP sans contraintes***

    - *Il s'agit en fait d'un simple problème de moindre carré !*

    - *La solution peut se calculer analytiquement (voir section exemple)*

# Classification

- ***3. Optimisation Convexe***



# Classification

- ***3. Optimisation Convexe***
  - ***Cas le meilleur !***
  - ***Un formalisme mathématique riche assurant l'existence et l'unicité de la solutions aux problèmes d'optimisation***
  - ***Des algorithmes efficaces (voir chapitre suivant)***

# Plan

- 1. Introduction
- 2. Formalisation
- 3. Rappels de mathématiques
- 4. Optimisation et convexité
- 5. Classification
- 6. Exemple : les moindres carrés

# Exemple

- ***Les moindres carrés***
  - *On se place dans $R^n$*
  - *On dispose d'une variable expliquée (ou réponse)* ***b de $R^n$***
  - *On dispose d'un ensemble de mesures pouvant se mettre sous la forme d'une matrice de taille m*n notée A (semi-)définie positive*
  - ***On cherche la variable explicative x permettant d'approximer au mieux b au sens de la norme euclidienne (erreur) connaissant A***

# Exemple

- ***Les moindres carrés***
  - *Problème d'optimisation quadratique sans contrainte*

$$J(x^*) = \min_{x \in \Re^n} \|Ax - b\|^2$$

  - ***Rq:*** *les normes étant des fonctions strictement convexes, on est dans un cas « idéal »*

# Exemple

- ***Les moindres carrés***
  - *Solution analytique du problème*
    - *J étant strictement convexe, trouver le min de J est équivalent à résoudre :*

$$\nabla J(x^*) = 0$$

    - *Résolution ?*
      - *Il faut dériver une fonction quadratique du type*

$$f(x) = x^T A x$$

# Exemple

- ***Les moindres carrés***
  - *Solution analytique du problème*
    - *Dérivée d'une fonction quadratique de matrice A (forme standard)*

$$\frac{\partial f(x)}{\partial x} = (A + A^T)x$$

    - *Développement de J*

$$J(x) = \left\| Ax - b \right\|^2 = \left\| Ax \right\|^2 - 2\left\langle Ax, b \right\rangle + \left\| b \right\|^2$$

# Exemple

- ***Les moindres carrés***
  - *Solution analytique du problème*
    - *Développement de J*

$$J(x) = \|Ax - b\|^2 = \|Ax\|^2 - 2\langle Ax, b\rangle + \|b\|^2$$

    - *Ou encore*

$$J(x) = \|Ax - b\|^2 = x^T A^T Ax - 2x^T A^T b + b^T b$$

# Exemple

- ***Les moindres carrés***
  - *Solution analytique du problème*
    - *Calcul du gradient de J*

$$\nabla J(x) = \left(A^T A + AA^T\right)x - 2A^T b + 0 = 2A^T Ax - 2A^T b$$

- *Minimum atteint quand le gradient s'annule, soit pour :*

$$A^T Ax = A^T b$$

**Rq :** En statistique, si on fait une hypothèse de gaussianité de l'erreur autour des valeurs réelles, *Ax* est l'estimateur sans biais du maximum de vraisemblance de *b*