Analyse 1: Optimisation sans contrainte

Joseph Salmon

Septembre 2014

Exemples de problèmes d'optimisation en apprentissage / statisiques

- les moyennes (pondérées ou non)
- les moindres carrées (contraints ou non)
- la plus grande valeur propre d'une matrice symétrique : lien avec l'analyse en composante principale (PCA)
- les classifieurs classiques : le perceptron, les Support Vector Machine (SVM), la régression logistique
- ...

L'optimisation : minimisation

Minimisation

Tout problème d'optimisation peut se ramener à l'étude de la **minimisation** d'une fonction.

- Si l'on a un problème de maximisation d'une fonction g, on passe à un problème de minimisation de fonction en considérant $f = -g$.
- Dans la suite on ne parlera donc que de problème de **minimisation**.
- On appelle **minimum** de f un point qui atteint la plus petite valeur de f, et l'on écrit le problème de la manière suivante :

$$\boxed{x^* \in \arg\min_{x \in K} f(x)}$$

où $K \subset \mathbb{R}^d$ est un espace de contrainte. Quand un tel x^* est unique on note $x^* = \arg\min_{x \in K} f(x)$

Optimisation, apprentissage et statistique

La moyenne comme problème d'optimisation

Prenons $x_1, \dots, x_n$ qui sont n vecteurs de $\mathbb{R}^d$. Le vecteur moyen noté $\bar{x}_n$ est la solution du problème d'optimisation :

$$\bar{x}_n = \underset{x \in \mathbb{R}^d}{\arg\min} \left(\sum_{i=1}^{n} \|x_i - x\|^2 \right)$$

Rem: Dans ce cas on connaît la solution ; on verra que l'on peut l'exprimer de manière explicite (on retrouve la définition usuelle) :

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Résolution de systèmes matricielle

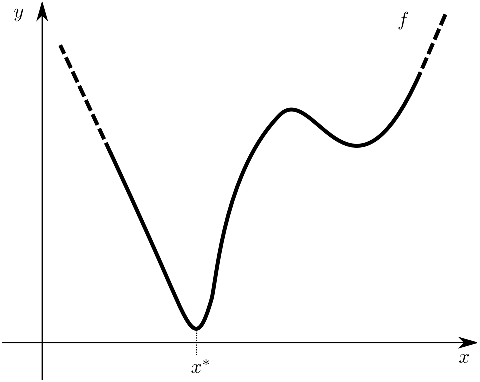
- Méthodes classiques pour la résolution d'un système linéaire du type $Ax = b$, avec $A \in \mathbb{R}^{n \times d}$ et $x \in \mathbb{R}^d$: le pivot de Gauss (*cf.* par exemple Golub et van Loan (1996))
- Point de vue itératif : résolution du système linéaire $Ax = b$ équivalent à trouver le minimum de la fonction $x \mapsto \|Ax - b\|$ ou plus simplement de $x \mapsto \|Ax - b\|^2$

<u>Rem</u>: On n'inverse rarement une matrice en pratique ; on résout plutôt le système associé : $Ax = b$

Condition d'existence d'un minimum I

Théorème de coercivité

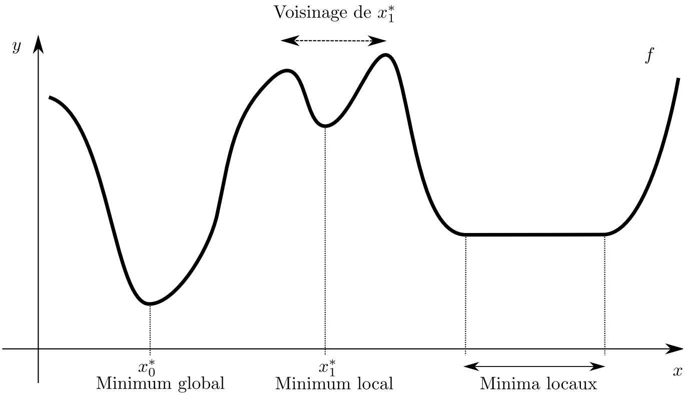
Si une fonction $f : \mathbb{R}^d \mapsto \mathbb{R}$ est continue et vérifie $\lim_{\|x\| \to \infty} f(x) = +\infty$ (*i.e.*, **coercive**) alors il existe un point x^* qui atteint le minimum : $x^* \in \underset{x \in \mathbb{R}^d}{\arg\min} f(x)$



Minimum local, minimum global

Définition : minimum local

On dit qu'une fonction $f : \mathbb{R}^d \mapsto \mathbb{R}$ admet un **minimum local** en un point x^* si x^* est le minimum de la restriction de f sur un voisinage de x^*

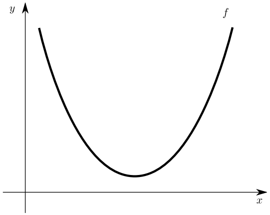


Rem: : un minimum global est donc aussi un minimum local

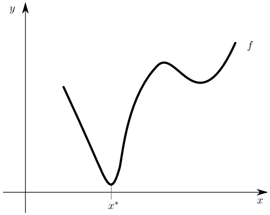
Cas convexe : local = global

Théorème : équivalence local/global dans le cas convexe

Si une fonction $f : \mathbb{R}^d \mapsto \mathbb{R}$ est une fonction convexe, alors tout minimum local de f est aussi un minimum global de f.



Convexe : 1 minimum global

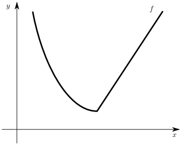


Non-convexe : 1 min. local + 1 global

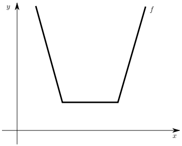
Convexité et minimum

Divers type de comportement pour des fonctions convexes :

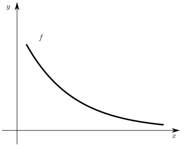
- un minimum global *e.g.*, quadratique, etc.
- plusieurs minima *e.g.*, affine par morceaux
- Sans minimum, borne inférieure finie *e.g.*, exponentielle
- Sans minimum, borne inférieure infinie *e.g.*, affine ou $-log(\cdot)$



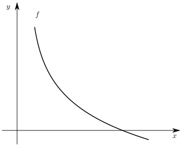
minimum global



intervalle de minima



borné inf.



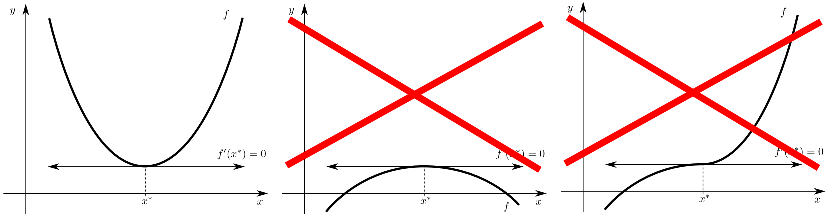
non bornée inf.

Condition du premier ordre pour un minimum local (CNO)

Théorème : règle de Fermat

Si f est différentiable en un minimum local x^* alors le gradient de f est nulle en x^*, *i.e.*, $\nabla f(x^*) = 0$

<u>Rem</u>: Attention ce n'est pas une condition suffisante



minimum local | maximum local | point selle

Retour sur les exemples classiques

La moyenne comme problème d'optimisation

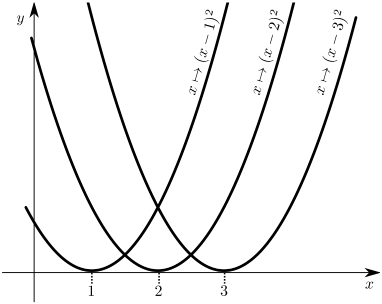
Prenons $x_1, \ldots, x_n$ qui sont n vecteurs de $\mathbb{R}^d$. Le vecteur moyen noté $\bar{x}_n$ est la solution du problème d'optimisation :

$$\bar{x}_n = \arg\min_{x \in \mathbb{R}^d} \frac{1}{2} \left(\sum_{i=1}^{n} \|x_i - x\|^2 \right)$$

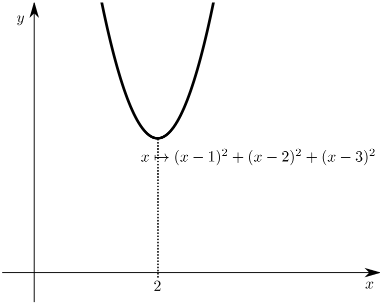
En utilisant le gradient de la fonction coercive (strictement convexe) $f_i : x \mapsto \|x - x_i\|^2/2$, $\nabla f_i(x) = (x - x_i)$ la CNO devient

$$\sum_{i=1}^{n} (\bar{x}_n - x_i) = 0 \Leftrightarrow \bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Retour sur la moyenne : visualisation



Objectifs individuels



Objectif global

Références I

- G. H. Golub and C. F. van Loan.
  Matrix computations.
  Johns Hopkins University Press, Baltimore, MD, third edition, 1996.