Introduction à l'optimisation non linéaire sans contraintes: Aspects théoriques

Djaffar Boussaa

CNRS/LMA

Contact : boussaa@lma.cnrs-mrs.fr

Plan

Formulation du problème

Optimisation sans contraintes : conditions d'optimalité

Convexité

Plan

Formulation du problème

Optimisation sans contraintes : conditions d'optimalité

Convexité

Le problème

Trouver $x^\star \in X$ tel que pour tout $x \in X$ on ait $f(x^\star) \leq f(x)$

où

- $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction à n variables, continue et suffisamment différentiable
- X est $\mathbb{R}^n$ ou une partie de $\mathbb{R}^n$ ayant des propriétés de "continuité"

Lien entre maximisation et minimisation

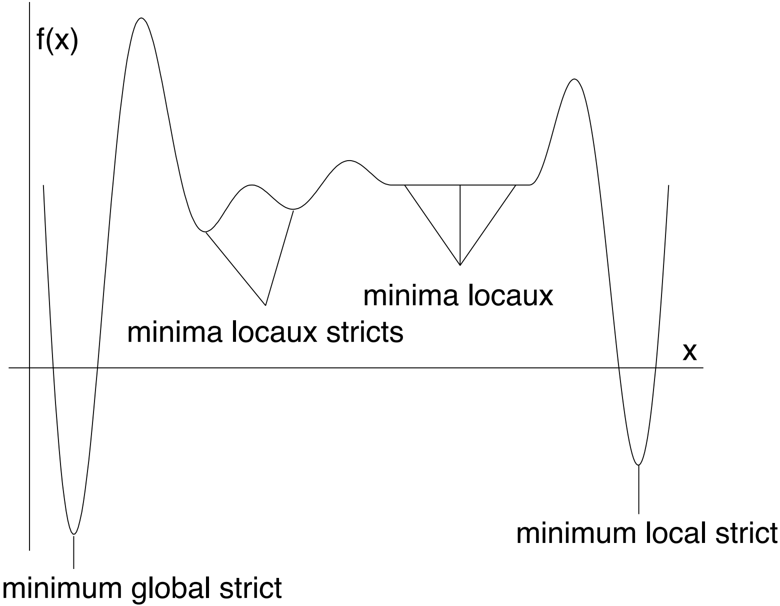
$$\boxed{\max f(x) = -\min -f(x)}$$

Premières définitions

- f critère (fonction-coût, fonction économique, fonction objectif, etc.)
- X ensemble admissible
- x variables de décision (variable de commande, variables d'état, paramètres, variables de design, etc.)
- $x^\star$ solution du problème ou minimum du problème. On parle de solution ou **minimum global** pour le distinguer du minimum local.
- **Minimum local** Un point $x^\star$ est une solution locale ou minimum local s'il existe un voisinage V de $x^\star$ tel que

$$x^\star \in X \quad \text{et} \quad f(x^\star) \leq f(x) \text{ pour tout } x \in X \cap V$$

- **Minimum local strict** Un minimum local est dit **strict** si $f(x^\star) < f(x)$ pour tout $x \in X \cap V \setminus \{x^\star\}$


f(x)
minima locaux
minima locaux stricts
x
minimum local strict
minimum global strict

Notations usuelles

Le problème général

$$\left\{ \begin{array}{c} \min f(x) \\ x \in X \end{array} \right. \qquad \min\{f(x) : x \in X\} \qquad \min_{x \in X} f(x)$$

Forme générale du problème considéré dans le cours

$$X = \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1, \ldots, m; g_j(x) \leq 0, \quad j = 1, \ldots, p\}$$

$$\left\{ \begin{array}{ll} \min f(x) & x \in \mathbb{R}^n \\ h_i(x) = 0 & i = 1, \ldots, m \\ g_j(x) \leq 0 & j = 1, \ldots, p \end{array} \right.$$

h_i contrainte d'égalité (fonction de $\mathbb{R}^n$ dans $\mathbb{R}$)
g_j contrainte d'inégalité (fonction de $\mathbb{R}^n$ dans $\mathbb{R}$)

Existence de solution

Soit le problème

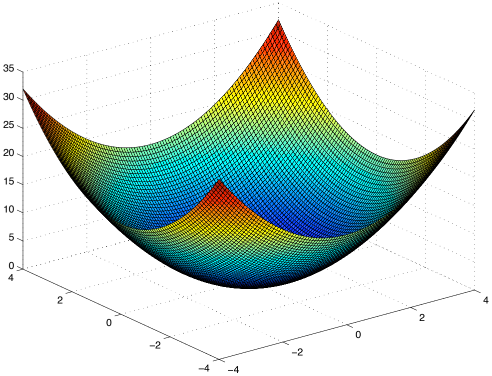
$$\min \{f(x) : x \in X\}$$

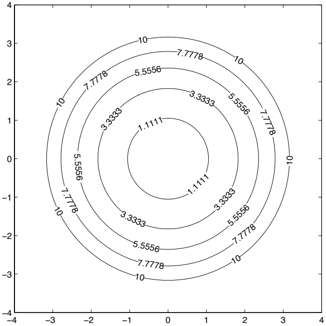
Deux possibilités

1. L'ensemble $\{f(x) : x \in X\}$ n'est pas borné inférieurement.
   Le problème n'admet pas de solution
2. L'ensemble $\{f(x) : x \in X\}$ est borné inférieurement.
   2.1 Le problème admet un minimum *global* si f est continue et X est compact
   2.2 Le problème admet un minimum *global* si X est fermé et f est coercive (ie $\lim f(x) \to \infty$ lorsque $\|x\| \to \infty$)

Exemple 1

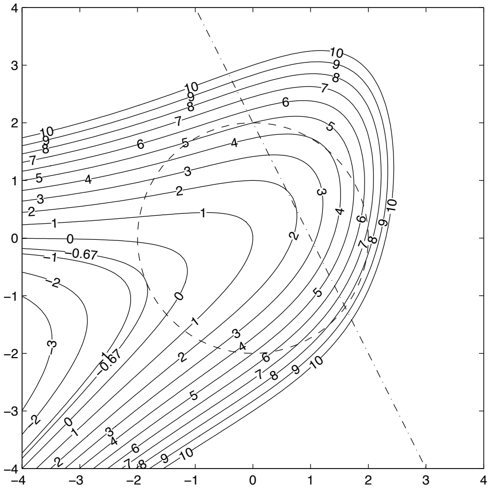
$$\min \; x_1^2 + x_2^2 \qquad (x_1, x_2) \in \mathbb{R}^2$$





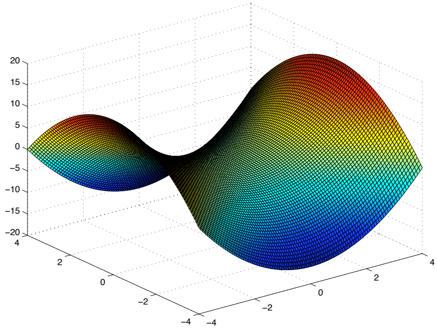
Exemple 2

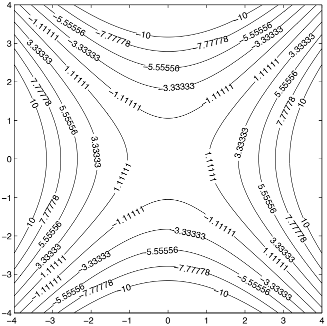
$$\begin{cases} \min \;\; e^{x_1} - x_1 x_2 + x_2^2 \\ 2x_1 + x_2 - 2 \leq 0 \\ x_1^2 + x_2^2 - 4 = 0 \\ (x_1, x_2) \in \mathbb{R}^2 \end{cases}$$



Exemple 3

$$\min \; x_1^2 - x_2^2 \qquad (x_1, x_2) \in \mathbb{R}^2$$





Un peu de jargon

- **Minimisation sans contrainte** Si $X = \mathbb{R}^n$, le problème est dit problème de minimisation sans contrainte
- **Point acceptable** Si $x \in X$, x est dit acceptable (on dit aussi admissible ou réalisable)
- **Optimisation linéaire** Si f est linéaire et X polyédrique (f, les h_i et les g_j sont linéaires), le problème d'optimisation est dit linéaire. Sinon, on parle d'optimisation non linéaire.

Le cours traitera de l'optimisation non linéaire.

Gradient et matrice hessienne

Gradient

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Matrice Hessienne

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Rappel

Si f est deux fois continûment différentiable, alors la matrice hessienne est symétrique.

Gradient et matrice hessienne

Exemples de calcul

$f(x)$	$\nabla f(x)$	$\nabla^2 f(x)$
$c'x + d$	c	0
$\frac{1}{2} x'Qx + b'x + c$	$\frac{1}{2}(Q + Q')x + b$	$\frac{1}{2}(Q + Q')$
$\frac{1}{2}\sum_{i=1}^{m}(r_i(x))^2$	$\sum_{i=1}^{m} r_i \nabla r_i$	$\sum_{i=1}^{m}\left(\nabla r_i \nabla r_i' + r_i \nabla^2 r_i\right)$

Exercice

Déterminer le gradient et la hessienne de la fonction de Rosenbrock généralisée

$$f(x) = \sum_{i=1}^{n/2}\left[\alpha\left(x_{2i} - x_{2i-1}^2\right)^2 + (1 - x_{2i-1})^2\right]$$

Matrice définie positive, matrice semidéfinie positive

Définitions

Soit Q une matrice $n \times n$ symétrique. Q est dite

- définie positive si $x'Qx > 0$ pour tout $x \in \mathbb{R}^n$, $x \neq 0$
- semidéfinie positive si $x'Qx \geq 0$ pour tout $x \in \mathbb{R}^n$

De même, Q est dite

- définie négative si $x'Qx < 0$ pour tout $x \in \mathbb{R}^n$, $x \neq 0$
- semidéfinie négative si $x'Qx \leq 0$ pour tout $x \in \mathbb{R}^n$

Reconnaître une matrice définie positive

Proposition

Soit Q une matrice $n \times n$ symétrique et Q_k la sous matrice de Q suivante :

$$Q_k = \begin{pmatrix} q_{11} & \cdots & q_{1k} \\ \vdots & \ddots & \vdots \\ q_{k1} & \dots & q_{kk} \end{pmatrix}$$

Alors Q est définie positive si et seulement si

- les valeurs propres de Q sont toutes strictement positives,
- $\det Q_k > 0$ pour tout $k \in \{1, \dots, n\}$,
- il existe une matrice triangulaire inférieure inversible L dont tous les termes diagonaux sont strictement positifs telle que $Q = LL^{'}$. ($Q = LL'$ est la factorisation de Cholesky de Q et L le facteur de Cholesky de Q).

Formules de Taylor

Soit $f : D \subset \mathbb{R}^n \to \mathbb{R}$, où D est un ouvert de $\mathbb{R}^n$. Si f est continûment différentiable, alors pour tout $x, y \in D$, on peut écrire

$$f(y) = f(x) + \nabla f(x)'(y-x) + R_1(x,y)$$

où le reste R_1 vérifie

$$\lim_{x \to y} \frac{R_1(x,y)}{\|x-y\|} = 0.$$

Si f est deux fois continûment différentiables, alors on peut écrire

$$f(y) \quad = \quad f(x) + (y-x)'\nabla f(x) + \frac{1}{2}(y-x)'\nabla^2 f(x)(y-x) + R_2(x,y)$$

où le reste R_2 vérifie

$$\lim_{x \to y} \frac{R_2(x,y)}{\|x-y\|^2} = 0.$$

Formules de Taylor (suite)

Il existe $y \in (x, x+p)$ tel que

$$f(x+p) = f(x) + \nabla f(y)' p$$

Il existe $z \in (x, x+p)$ tel que

$$f(x+p) = f(x) + \nabla f(x)' p + \frac{1}{2} p' \nabla^2 f(z) p$$

Plan

Formulation du problème

Optimisation sans contraintes : conditions d'optimalité

Convexité

Condition nécessaire du premier ordre

Le problème

$$\min_{x\in\mathbb{R}^n} f(x)$$

Proposition

Si $x^\star$ est un minimum local de f et si f est différentiable en $x^\star$, alors $\nabla f(x^\star) = 0$.

Définition

Un point $x^\star$ tel que $\nabla f(x^\star) = 0$ est dit stationnaire.

Condition nécessaire du second ordre

Proposition

Si $x^\star$ est un minimum local de f et si f est deux fois différentiable en $x^\star$, alors $\nabla f(x^\star) = 0$ et $\nabla^2 f(x^\star)$ est semidéfinie positive

Ces conditions ne pas suffisantes Soit $f(x) = x^3$. Le point $x = 0$ satisfait les conditions nécessaires du premier et second ordres, mais il n'est pas un minimum de f.

Preuves

- **Condition du premier ordre** Soit $p \in \mathbb{R}^n$ une direction fixe quelconque. Comme $x^\star$ est un minimum local,

$$0 \leq \lim_{t \downarrow 0} \frac{f(x^\star + tp) - f(x^\star)}{t} = p' \nabla f(x^\star)$$

on conclut que $\nabla f\left(x^\star\right) = 0$ CQFD.

- **Condition du second ordre** En vertu de Taylor au 2nd ordre

$$f(x^\star + tp) - f(x^\star) = tp' \nabla f(x^\star) + \frac{t^2}{2} p' \nabla^2 f(x^\star) p + o(t^2)$$

Comme $x^\star$ est un minimum local, $\nabla f\left(x^\star\right) = 0$ et il existe $\bar{t}$ suffisamment petit tel que pour tout $t \in]0, \bar{t}[$

$$0 \leq \frac{f(x^\star + tp) - f(x^\star)}{t^2} = \frac{1}{2} p' \nabla^2 f(x^\star) p + \frac{o(t^2)}{t^2}$$

En passant à la limite sur t $(t \to 0)$, on aboutit au résultat souhaité.

Condition suffisante de minimalité locale

Proposition Si un point $x^\star$ est tel que $\nabla f(x^\star) = 0$ et $\nabla^2 f(x^\star)$ est définie positive alors $x^\star$ est minimum local strict de f

Preuve Soit a la plus petite valeur propre de $\nabla^2 f(x^\star)$. Comme $\nabla^2 f(x^\star)$ est définie positive, $a > 0$. En vertu de Taylor

$$\begin{aligned} f(x^\star + p) - f(x^\star) &= \frac{1}{2} p' \nabla^2 f(x^\star) p + o(\|p\|^2) \\ &\geq \frac{a}{2} \|p\|^2 + o(\|p\|^2) \\ &= \left(\frac{a}{2} + \frac{o(\|p\|^2)}{\|p\|^2} \right) \|p\|^2 \end{aligned}$$

Pour $\|p\|$ suffisamment petit, $o(\|p\|^2) / \|p\|^2$ est négligeable devant $a/2$. Si bien que ce second membre est strictement positif pour $\|p\|$ suffisamment petit.

Un exemple

$$\min\left\{\tfrac{1}{3}x_1^3 + \tfrac{1}{2}x_1^2 + 2x_1x_2 + \tfrac{1}{2}x_2^2 - x_2 + 9 : (x_1, x_2) \in \mathbb{R}^2\right\}$$

$$\nabla f(x) = \begin{pmatrix} x_1^2 + x_1 + 2x_2 \\ 2x_1 + x_2 - 1 \end{pmatrix} \qquad \nabla^2 f(x) = \begin{pmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{pmatrix}$$

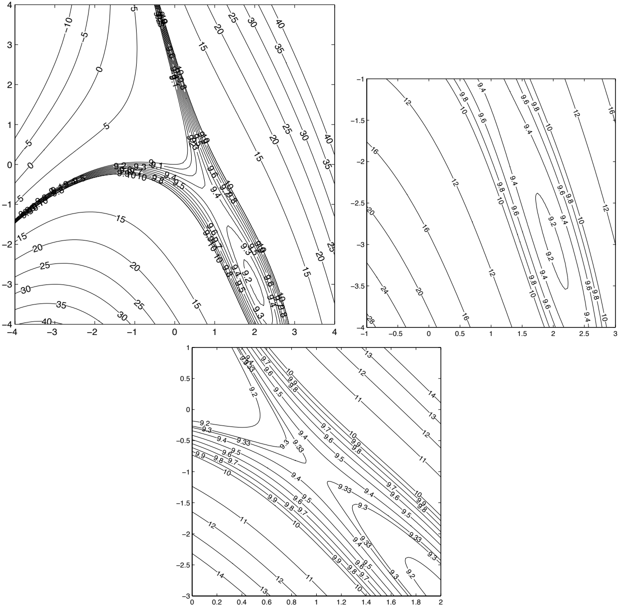
Les points stationnaires sont $x_a = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ et $x_b = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$

La hessienne en ces points est respectivement

$\nabla^2 f(x_a) = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$ et $\nabla^2 f(x_b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$

En x_a, hessienne indéfinie, x_a ni un minimum ni un maximum.

En x_b, hessienne définie positive, x_b un minimum.



−4 −3 −2 −1 0 1 2 3 4
−1 −0.5 0 0.5 1 1.5 2 2.5 3
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2

Cas quadratique

Le problème

$$\min_{x\in\mathbb{R}^n} f(x) \text{ avec } f(x) = \frac{1}{2}x'Qx - c'x$$

où Q matrice $n \times n$ symétrique et $c \in \mathbb{R}^n$

Conditions nécessaires

$$\nabla f(x^\star) = Qx^\star - c = 0 \qquad \nabla^2 f(x^\star) = Q \geq 0$$

Discussion

- $Q > 0$: $x^\star = Q^{-1}c$ est l'unique minimum global.
- $Q \geq 0$ non inversible : Soit pas de solution, soit infinité de solutions
- $Q \not\geq 0$: f n'a pas de minimum local

Plan

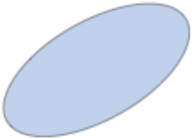
Formulation du problème

Optimisation sans contraintes : conditions d'optimalité

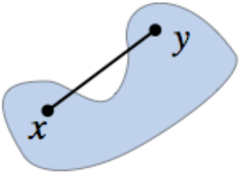
Convexité

Ensemble convexe

Un ensemble $C \subset \mathbb{R}^n$ est dit convexe si pour tout $x,\ y \in X$ le segment $[x, y] = \{(1-t)x + ty : t \in [0,1]\}$ est contenu dans X.



convexe

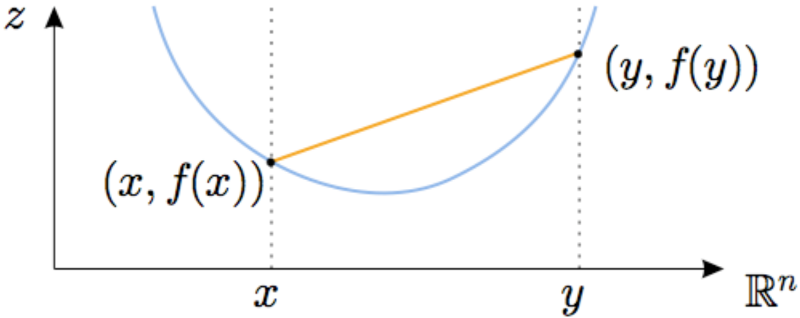


non convexe

Fonction convexe

Soit C un ensemble convexe. La fonction $f : C \subset \mathbb{R}^n \to \mathbb{R}$ est dite convexe si pour tout $x,\ y \in \mathbb{R}^n$ et pour tout $t \in [0,1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$



Exemples

- Les fonctions linéaires et affines
- Les normes
- Les fonctions quadratiques $f(x) = \frac{1}{2}x'Qx + q'x + r$ avec $Q \geq 0$
- $f(x) = x^2$
- $f(x) = -\ln x \qquad (x > 0)$
- $f(x) = 1/x$
- $e^{\alpha x}$
- $|x|$
- $\max(0, x)$
- $\max(0, -x)$

Reconnaître une fonction convexe par ses dérivées

Proposition

Soit f une fonction continûment différentiable. Alors f est convexe sur un ensemble convexe C si et seulement si

$$f(x) \geq f(y) + (x - y)' \nabla f(y) \qquad \text{pour tout } x, y \in C$$

Proposition

Soit f une fonction deux fois continûment différentiable. Alors f est convexe sur un ensemble convexe C si et seulement si $\nabla^2 f(x)$ est semidéfinie positive pour tout $x \in C$.

Problème d'optimisation convexe

$\min_{x \in C} f(x)$ avec f une fonction convexe et C un ensemble convexe

Convexité et minimalité

Proposition

Soient $C \subseteq \mathbb{R}^n$ un ensemble convexe et $f : C \to \mathbb{R}^n$ une fonction convexe, alors tout minimum local de f est aussi un minimum global.

Preuve

Soit x un minimum local de f qui ne soit pas un minimum global. Alors il existe un $y \neq x$ tel que $f(y) < f(x)$. L'inégalité de convexité permet d'écrire
$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) < f(y)$ pour tout $t \in [0, 1[$, ce qui contredit l'hypothèse que x est un minimum local.