Optimisation

Dr. Ibtissem DIDI

01 Mars 2020

Plan

Historique

L'***optimisation*** est une branche des mathématiques cherchant à modéliser, à analyser et à résoudre analytiquement ou numériquement les problèmes qui consistent à minimiser ou maximiser une fonction sur un ensemble. L'optimisation joue un rôle important en :

- la recherche opérationnelle (domaine à la frontière entre l'informatique, les mathématiques et l'économie),
- mathématiques appliquées (fondamentales pour l'industrie et l'ingénierie),
- analyse et en analyse numérique,
- statistique pour l'estimation du maximum de vraisemblance d'une distribution,
- la recherche de stratégies dans le cadre de la théorie des jeux,
- théorie du contrôle et de la commande.

Historique

Les premiers problèmes d'optimisation auraient été formulés par **Euclide**, au III^e siècle avant notre ère, dans son ouvrage historique *Éléments*.

Figure – Mécanicien et un mathématicien grec

Historique

300 ans plus tard, **Héron d'Alexandrie** dans Catoptrica énonce le « principe du plus court chemin » dans le contexte de l'optique.

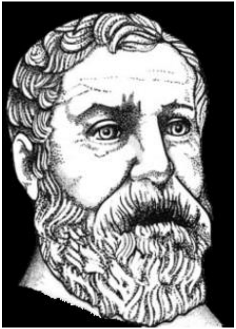


Figure – Ingénieur

Historique

Au XVIIe siècle, l'apparition du calcul différentiel entraîne l'invention de techniques d'optimisation, ou du moins en fait ressentir la nécessité. **Newton** met au point une méthode itérative permettant de trouver les extremums locaux d'une fonction en faisant intervenir la notion de dérivée.

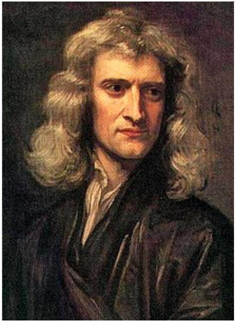


Figure – Philosophe, mathématicien, physicien, alchimiste, astronome et théologien anglais, puis britannique

Historique

Cette nouvelle notion issue de ses travaux avec **Leibniz** permet de grandes avancées dans l'optimisation de fonctions car le problème est ramené à la recherche des racines de la dérivée.

Figure – Philosophe, scientifique, mathématicien, logicien, diplomate, juriste, bibliothécaire et philologue allemand

Historique

Durant le XVIIIe siècle, les travaux des mathématiciens **Euler** et **Lagrange** mènent au calcul des variations, une branche de l'analyse fonctionnelle regroupant plusieurs méthodes d'optimisation.

Figure – Mathématicien et physicien suisse

Historique

Lagrange invente une technique d'optimisation sous contraintes : les multiplicateurs de Lagrange.

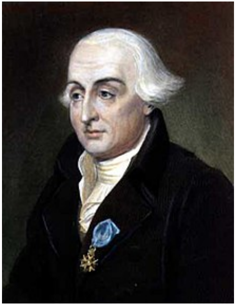


Figure – Mathématicien, mécanicien et astronome italien naturalisé français

Historique

Le XIXe siècle est marqué par l'intérêt croissant des économistes pour les mathématiques. Ceux-ci mettent en place des modèles économiques qu'il convient d'optimiser, ce qui accélère le développement des mathématiques. Depuis cette période, l'optimisation est devenue un pilier des mathématiques appliquées.

Historique

On peut tout de même évoquer l'invention de plusieurs méthodes itératives utilisant le gradient de la fonction, ainsi que l'utilisation du terme « programmation mathématique », pour désigner des problèmes d'optimisation. Historiquement, le premier terme introduit fut celui de « programmation linéaire », inventé par **George Dantzig** vers 1947.

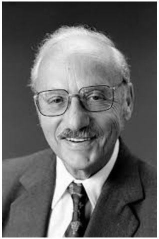


Figure – Mathématicien américain, notamment inventeur de l'algorithme du simplexe en optimisation linéaire

Problème d'optimisation

Un problème d'optimisation $(\mathbb{P})$ s'exprime de la manière suivante :

Définition

Etant donné une fonction $f : \mathcal{R} \longrightarrow \mathbb{R}$ trouver un élément $\bar{x}$ de $\mathcal{R}$ tel que $f(\bar{x}) \leq f(x)$ pour tout $x \in \mathcal{R}$. On dit que l'on cherche à minimiser la fonction f sur l'ensemble $\mathcal{R}$.

- La fonction f peut s'appeler : *fonction-coût*, ou simplement *coût*, *fonction-objectif*, ou simplement *objectif*, *critère*, etc.
- L'ensemble $\mathcal{R}$ est appelé : *ensemble admissible*, et les points de $\mathcal{R}$ sont appelés *les points admissibles* de $(\mathbb{P})$.
- Le point $\bar{x}$ est appelé *solution* ou *minimum* ou bien *minimiseur* de $(\mathbb{P})$.

Problème d'optimisation

Le problème d'optimisation $(\mathbb{P})$ peut s'écrire de plusieurs manières :

$$\min_{x \in \mathcal{R}} f(x)$$

$$\min \{f(x) | x \in \mathcal{R}\}$$

$$\min f(\mathcal{R})$$

ou

$$\begin{cases} \min f(x) \\ x \in \mathcal{R} \end{cases}$$

Problème d'optimisation

Dans la pratique, il est souvent difficile de déterminer une fonction-objectif à minimiser, et les diverses fonctions-objectifs choisies donnent lieu à autant de problèmes d'optimisation différents. X est l'espace des paramètres (disant que X est un espace vectoriel de dimension finie). La partie $\mathcal{R}$ de X est l'ensemble des ***contraintes***. $\mathcal{R}$ est représenté de diverses manières, par des égalités $a_i(x), i = \overline{1,p}$ ou des inégalités $c_j(x), j = \overline{1,q}$ à respecter, des bornes à ne pas dépasser. Si $\mathcal{R}$ est l'espace X tout entier, c'est-à-dire si on laisse toute liberté dans le choix de x, on parlera de problème d'optimisation sans contraintes.

Problème d'optimisation

Moyennant le changement de fonction f en $-f$, on peut convertir une formulation "maximiser" en une formulation "minimiser" et vice versa.

Maximiser f revient à minimiser $-f$.

Explication

$$f(x) \leq f(\bar{x}), \forall x \in \mathcal{R} \Longleftrightarrow -f(x) \geq -f(\bar{x}), \forall x \in \mathcal{R}$$

Problème d'optimisation

Toutefois il ne faut pas s'y fier : si certaines propriétés de f sont maintenues par passage de f en $-f$ (continuité, différentiabilité ou linéarité), d'autres, cruciales, sont détruites par cette transformation ; par exemple "minimiser f sur $\mathcal{R}$ avec f et $\mathcal{R}$ convexes" peut être considérer comme un problème facile, alors que "maximiser f sur $\mathcal{R}$ avec f et $\mathcal{R}$ convexes" est de fait un problème d'optimisation très difficile.
La convexité est un exemple de propriété tournée davantage vers la minimisation que vers la maximisation.

Région admissible

N'importe quel point x satisfaisant les contraintes d'égalités et d'inégalités est dit *point admissible* de $\mathbb{P}$. L'ensemble de point satisfaisant ces contraintes est dit *région admissible* de $f(x)$. Ainsi la région admissible peut être définie par l'ensemble

$$\mathcal{R} = \{x/a_i(x) = 0 \text{ pour } i = \overline{1,p} \text{ et } c_j(x) \geq 0 \text{ pour } j = \overline{1,q}\}$$

N'importe quel point x qui n'appartient pas à $\mathcal{R}$ est dit *point non admissible*

Région admissible

- Si les contraintes du problème $\mathbb{P}$ étaient toutes des inégalités, les contraintes se divisent en 3 parties :
  1. Point intérieur est un point pour lequel $c_j(x) > 0$ pour tout j. Ceci est un point admissible.
  2. Point au bord est un point pour lequel au moins un $c_j(x) = 0$. Ceci peut être ou ne pas être un point admissible.
  3. Point extérieurs est un point pour lequel au moins un $c_j(x) < 0$. Ceci est un point non admissible.
- Si une contrainte $c_m(x)$ est nulle durant une itération spécifique, elle est dite *active*.
- Si $c_m(\bar{x})$ est nulle la convergence est atteinte, l'optimum $\bar{x}$ est localisé au bord. Dans ce cas là, le point optimal est dit *contraint*.

Exemple 1

En utilisant la méthode graphique, résoudre le problème d'optimisation suivant :

$$\begin{aligned} \min f(x) \quad &= x_1^2 + x_2^2 - 4x_1 + 4 \\ sc : c_1(x) \quad &= x_1 - 2x_2 + 6 \geq 0 \\ c_2(x) \quad &= -x_1^2 + x_2 - 1 \geq 0 \\ c_3(x) \quad &= x_1 \geq 0 \\ c_4(x) \quad &= x_2 \geq 0 \end{aligned}$$

Solution

La fonction objectif est exprimée comme suit :

$$(x_1 - 2)^2 + x_2^2 = C$$

Ainsi les contours de $f(x)$ dans le plan (x_1, x_2) sont des cercles de centre $(2, 0)$ et de rayon $r = \sqrt{C}$. Les contraintes $c_1(x)$ et $c_2(x)$ impliquent

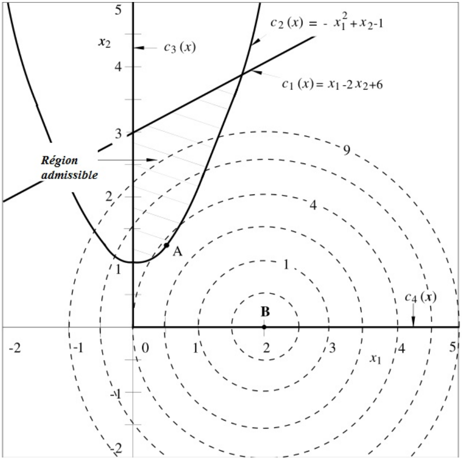
$$x_2 \leq \frac{1}{2}x_1 + 3$$

et

$$x_2 \geq x_1^2 + 1$$

respectivement, tant que $c_3(x)$ et $c_4(x)$ impliquent la positivité de x_1 et x_2.

Solution



Solution

La solution est localisée au point $\mathcal{R}$ sur le bord de $c_2(x)$. Le point optimal est donc contraint car $c_2(\bar{x}) = 0$. Par conséquent, si le problème était résolu par une méthode de programmation, la contrainte $c_2(x)$ deviendra active dés que la solution sera atteinte.
Si les deux contraintes $c_1(x)$ et $c_2(x)$ n'existent pas, le minimum sera le point B.

Exemple 2

En utilisant la méthode graphique, résoudre le problème d'optimisation suivant :

$$\begin{aligned} \min f(x) &= x_1^2 + x_2^2 + 2x_2 \\ sc : a_1(x) &= x_1^2 + x_2^2 - 1 = 0 \\ c_1(x) &= x_1 + x_2 - 0.5 \geq 0 \\ c_2(x) &= x_1 \geq 0 \\ c_3(x) &= x_2 \geq 0 \end{aligned}$$

Solution

La fonction objectif est exprimée comme suit :

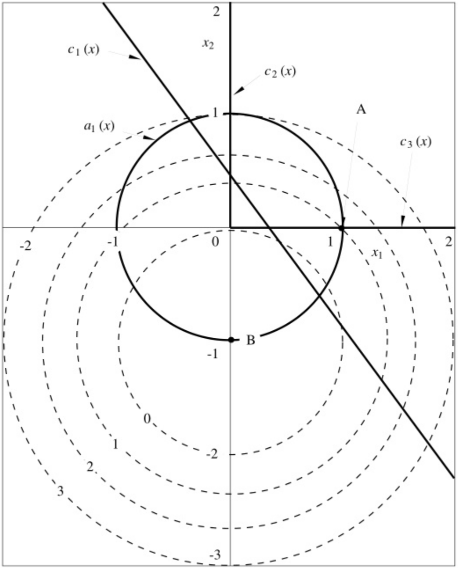
$$x_1^2 + (x_2 + 1)^2 = C + 1$$

Ainsi les contours de $f(x)$ dans le plan (x_1, x_2) sont des cercles de centre $(0, -1)$ et de rayon $r = \sqrt{C+1}$. La contrainte $a_1(x)$ est un cercle de centre $(0, 0)$ et de rayon $r = 1$. La contrainte $c_1(x)$ est une droite

$$x_2 \leq 0.5 - x_1$$

Les contraintes $c_2(x)$ et $c_3(x)$ impliquent la positivité de x_1 et x_2.

Solution



Solution

Dans ce cas là, la région admissible est l'arc de cercle $a_1(x) = 0$ dans le quadrant positif du plan (x_1, x_2). La solution est localisée au point $\mathcal{R}$. Le point optimal $\bar{x}(1, 0)$ est donc contraint car $a_1(\bar{x}) = 0$ et $c_3(\bar{x}) = 0$. Par conséquent, si le problème était résolu par une méthode de programmation, la contrainte $c_2(x)$ deviendra active dés que la solution sera atteinte.
Si les deux contraintes $c_1(x)$ et $c_2(x)$ n'existent pas, le minimum sera le point B.

Introduction

Dans ce chapitre, on va donner des définitions du vecteur gradient, la matrice hessienne et les différents types d'extremum (maximum et minimum). Les conditions nécessaires et suffisantes d'optimalité. Par conséquent, une classe de fonctions convexes et concaves sera introduite. A travers ce chapitre ,on va se focaliser sur des problèmes d'optimisation non linéaire

$$\begin{array}{ll} \min & f = f(x) \\ sc: & x \in \mathcal{R} \end{array}$$

Définition d'une forme quadratique

Définition

Soit H une matrice symétrique $n \times n$ et $g \in \mathbb{R}^n$. On appelle *forme quadratique* la fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par

$$f(x) = \frac{1}{2}x^T H x - g^T x$$

Lorsque la matrice H possède certaines propriétés, la fonction f peut prendre un nom particulier. La propriété à laquelle nous allons nous intéresser est la **positivité**.

Définition d'une forme quadratique

Définition

Soit H une matrice symétrique $n \times n$ et $g \in \mathbb{R}^n$. On dit que H est semi-définie positive (SDP) et on note $H \geq 0$, quand

$$x^T H x \geq 0, \forall x \in \mathbb{R}^n$$

On dit que H est définie positive (DP) et on note $H > 0$, quand

$$x^T H x > 0, \forall x \in \mathbb{R}^n, x \neq 0$$

Définition d'une forme quadratique

Cette définition peut être reliée aux valeurs propres de la matrice H :

Propriété

Soit H une matrice symétrique $n \times n$. On note $\lambda_i, i = \overline{1, n}$ ses valeurs propres (réelles). On a les équivalences suivantes :

$$H \geq 0 \Longleftrightarrow \lambda_i \geq 0, i = \overline{1, n}$$

$$H > 0 \Longleftrightarrow \lambda_i > 0, i = \overline{1, n}$$

Lorsque la matrice H est définie positive (resp. semi-définie positive), on dira que $f(x)$ est une forme quadratique définie positive (resp. semi-définie positive).

Définition de la différentiabilité

Dans $\mathbb{R}^n$ on note x le vecteur colonne

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

et la notation $\|.\|$ désignera, la norme euclidienne

$$\|x\| = \left(\sum_{k=1}^{n} x_k^2 \right)^{\frac{1}{2}}$$

Définition de la différentiabilité

Définition

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ représentée dans la base canonique de $\mathbb{R}^m$ par le vecteur

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

continue en $a \in \mathbb{R}^n$. On dit que f est différentiable en a s'il existe une application linéaire, notée $f'(a)$, telle que pour tout $d \in \mathbb{R}^n$ on ait

$$f(a+d) = f(a) + f'(a)d + \|d\| \, \epsilon(d)$$

où $\epsilon(.)$ est une fonction continue en 0 vérifiant $\lim_{d \to 0} \epsilon(d) = 0$. On appel $f'(a)$ dérivée de f au point a.

Calcul de la dérivée première

On peut d'ores et déjà donner un résultat "pratique" permettant de calculer directement la dérivée à partir du développement.

Proposition

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ différentiable en a, alors

$$\lim_{t\to 0} \frac{f(a+td)-f(a)}{t} = f'(a)d$$

Cette méthode d'estimation du gradient est souvent appelée *différences finies*. La quantité $f'(a)d$ est la dérivée directionnelle de f au point a dans la direction d. La proposition suivante fait le lien entre la matrice de $f'(a)$ et les dérivées partielles de f au point a :

Calcul de la dérivée première

Proposition

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ différentiable en a, alors on peut représenter $f'(a)$ par sa matrice dans les bases canoniques de $\mathbb{R}^n$ et de R^m et on a

$$[f'(a)]_{ij} = \frac{\partial f_i}{\partial x_j}(a)$$

On appelle souvent $f'(a)$ la *matrice jacobienne* de f au point a. Lorsque $m = 1$ on adopte une notation et un nom particuliers : *le gradient* est le vecteur noté $\nabla f(a)$ et défini par

$$f'(a) = \nabla f(a)^T$$

et on a

$$f(a+d) = f(a) + \nabla f(a)^T d + \|d\| \,\epsilon(d)$$

Dérivée seconde

On se place maintenant dans le cas $m = 1$, soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.

Définition

L'application $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est dite deux fois différentiable s'il existe une matrice symétrique $\nabla^2 f(a)$ telle que

$$f(a+d) = f(a) + \nabla f(a)^T d + d^T \nabla^2 f(a) d + \|d\|^2 \epsilon(d)$$

On appelle $\nabla^2 f(a)$ *matrice hessienne* de f au point a. Comme l'énonce le théorème suivant, cette matrice s'obtient à partir des dérivées secondes de f :

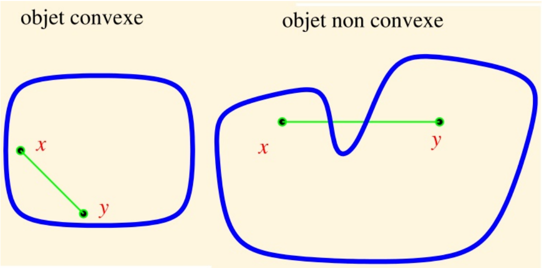
Théorème

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ deux fois différentiable en un point a. Si on note $g(x) = \nabla f(x)$ alors la matrice hessienne est définie par $\nabla^2 f(a) = g'(a)$, soit

$$[\nabla^2 f(a)]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Introduction

La convexité est à la base une propriété géométrique, assez intuitive d'ailleurs, qui permet de caractériser certains objets. On voit assez bien ce qu'est un objet convexe dans un espace à deux ou trois dimensions. Nous allons maintenant montrer comment cette propriété peut aussi s'appliquer aux fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$.



Ensemble convexe

Définition (ensemble convexe)

Soient x_1 et x_2 deux vecteurs de $\mathbb{R}^n$. Le segment de droite joignant l'extrémité de ces vecteurs, l'ensemble des point :

$$D = \{x \in \mathbb{R}^n / x = \alpha x_1 + (1-\alpha)x_2, 0 \leq \alpha \leq 1\}$$

Exemple. Considérons les deux vecteurs $x_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$ et $x_2 = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$.

Soit $[P_1, P_2]$ le segment de droite reliant les extrémités de x_1 et x_2.

$$[P_1, P_2] = \left\{x \in R^2 / x = \alpha \begin{pmatrix} 4 \\ 1 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 2 \\ 5 \end{pmatrix}, 0 \leq \alpha \leq 1\right\}$$

En faisant varier la valeur de α entre 0 et 1, on obtient les points de segment.

Exemple

$$\alpha = 0 \quad x = \begin{pmatrix} 2 \\ 5 \end{pmatrix} = P_2$$

$$\alpha = \frac{1}{4} \quad x = \frac{1}{4}\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \frac{3}{4}\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} \frac{5}{2} \\ 4 \end{pmatrix}$$

$$\alpha = \frac{1}{2} \quad x = \frac{1}{2}\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$$\alpha = \frac{3}{4} \quad x = \frac{3}{4}\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} \frac{7}{2} \\ 2 \end{pmatrix}$$

$$\alpha = 1 \quad x = \begin{pmatrix} 4 \\ 1 \end{pmatrix} = P_1$$

Exemple

Ces points sont représentés sur la figure suivante.

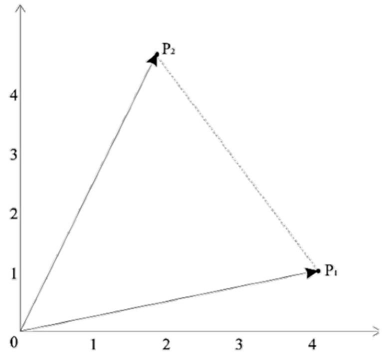


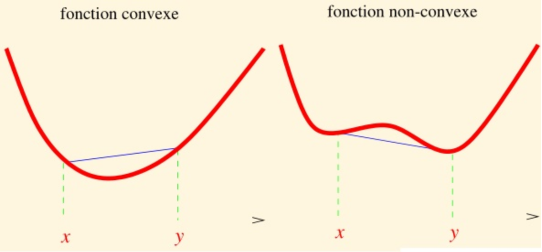
Figure – Segment de droite reliant P_1 à P_2.

Fonction convexe

Définition (fonction convexe)

Soit S un ens. convexe de $\mathbb{R}^n$; $f : S \longrightarrow \mathbb{R}$ est dite *convexe* sur S si :
$\forall (x,y) \in S^2, \forall \alpha \in [0,1], f\,(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$

On dira que f est *strictement convexe* sur S si pour $x \neq y$:
$\forall (x,y) \in S^2, \forall \alpha \in \,]0,1[\,, f\,(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y$



Caractérisation de la convexité en termes du hessien

Propriété

Si $f : \mathbb{R} \to \mathbb{R}$ est 2 fois continûment dérivable sur S convexe alors f est convexe ssi $f''(x) \geq 0$, $\forall x \in S$ et strictement convexe ssi $f''(x) > 0$, $\forall x \in S$.

Ce résultat se généralise pour $n > 1$: le résultat suivant fait le lien entre le hessien et la propriété de convexité :

Théorème

Si $f : \mathbb{R}^n \to \mathbb{R}$ est 2 fois continûment dérivable sur S convexe alors f est convexe ssi $\nabla^2 f(x) \geq 0$, $\forall x \in S$ et strictement convexe ssi $\nabla^2 f(x) > 0$, $\forall x \in S$.

Caractérisation de la convexité en termes du hessien

Le corollaire suivant est immédiat :

Corollaire

Soit f une forme quadratique définie par

$$f(x) = \frac{1}{2}x^T Hx - g^T x$$

alors f est convexe ssi $H \geq 0$, et strictement convexe ssi $H > 0$.

Cela provient du fait que $\nabla^2 f(x) = H$

Caractérisation de la convexité en termes du gradient

Dans le cas où la fonction f n'est supposée qu'une fois différentiable, on a le résultat suivant :

Théorème

Soit $f : S \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction **une fois différentiable**, alors f est convexe ssi

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \forall (x, y) \in S^2$$

La fonction f est strictement convexe ssi

$$f(y) > f(x) + \nabla f(x)^T (y - x), \forall (x, y) \in S^2, x \neq y$$

Types d'extremum

Définition 1

Un point $\bar{x} \in \mathcal{R}$, est dit un *minimiseur local* de $f(x)$ s'il existe $\epsilon > 0$ tel que

$$f(\bar{x}) \leq f(x)$$

si

$$x \in \mathcal{R} \text{ et } \|x - \bar{x}\| < \epsilon$$

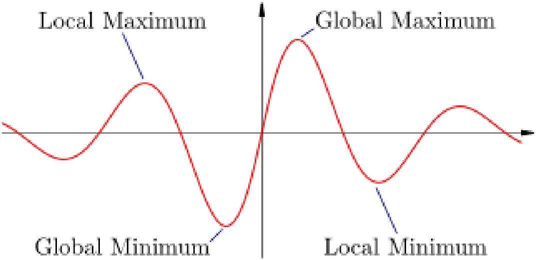
Définition 2

Un point $\bar{x} \in \mathcal{R}$, est dit un *minimiseur global* de $f(x)$ tel que $\forall x \in \mathcal{R}$

$$f(\bar{x}) \leq f(x)$$

Si la définition 2 est satisfaite en $\bar{x}$ alors la définition 1 est aussi satisfaite en $\bar{x}$ et donc un minimiseur global est aussi un minimiseur local.

Types d'extremum



Introduction

Le gradient $g(x)$ et le Hessien $H(x)$ doivent satisfaire certaines conditions en $\bar{x}$ (minimiseur local). Deux conditions vont être présentées :

1. Conditions qui vont être satisfaite en $\bar{x}$. Ce sont des conditions nécessaires.
2. Conditions qui vont garantir que $\bar{x}$ est un minimiseur local. Ce sont des conditions suffisantes.

Définition (Direction)

Soit $\delta = \alpha d$ une variation de x où $\alpha > 0$ et d est le vecteur direction. Si $\mathcal{R}$ est la région admissible et une constante $\hat{\alpha} > 0$ existe telle que

$$x + \alpha d \in \mathcal{R}$$

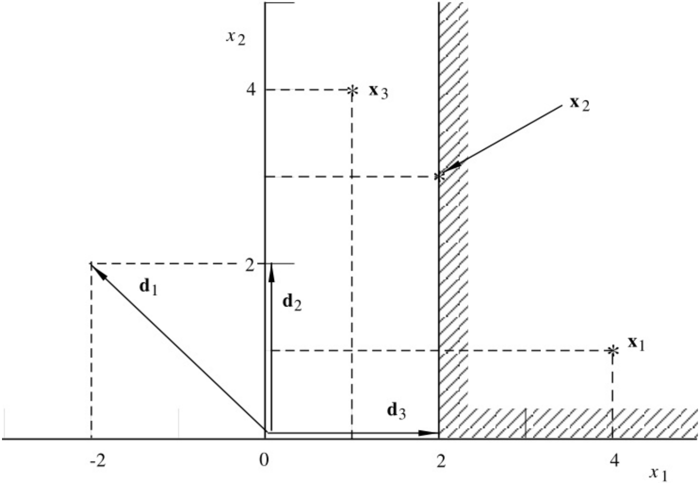
Pour tout α avec $0 \leq \alpha \leq \hat{\alpha}$, alors d est dite *direction admissible* au point x.

Exemple. La région admissible d'un problème d'optimisation est donnée par :

$$\mathcal{R} = \{x / x_1 \geq 2, x_2 \geq 0\}$$

Lequel des vecteurs $d_1 = [-2\ 2]^T$, $d_2 = [0\ 2]^T$ et $d_3 = [2\ 0]^T$ est une direction admissible au points $x_1 = [4\ 1]^T$, $x_2 = [2\ 3]^T$ et $x_3 = [1\ 4]^T$?

Exemple



x_2
4
$\mathbf{x}_3$
$\mathbf{x}_2$
2
$\mathbf{d}_1$
$\mathbf{d}_2$
$\mathbf{x}_1$
$\mathbf{d}_3$
-2
0
2
4
x_1

Exemple

- d_1 est une direction admissible au point x_1 pour tout $0 \leq \alpha \leq \hat{\alpha}$ pour $\hat{\alpha} = 1$.
- d_2 et d_3 sont des directions admissibles au point x_1 pour tout $0 \leq \alpha \leq \hat{\alpha}$ avec $\hat{\alpha} > 0$.
- d_1 n'est pas une direction admissible au point x_1.
- $\exists \hat{\alpha} > 0$ pour lequel d_2 et d_3 sont des directions admissibles au point x_2.
- d_1, d_2 et d_3 ne sont pas des directions admissibles au point x_3, car x_3 n'appartient pas à la région admissible.

Conditions nécessaire du premier ordre

La fonction objectif doit satisfaire deux types de conditions dans le but d'avoir un minimum, dites, conditions du premier et du second ordre. Les conditions du premier ordre utilisent la première dérivée, i.e., le gradient.

Théorème

1. Si $f(x) \in \mathcal{C}^{1}$ et $\bar{x}$ est un minimiseur local, alors

$$g(\bar{x})^T d \geq 0$$

pour toute direction admissible d au point $\bar{x}$.

2. Si $\bar{x}$ est localiser à l'intérieur de $\mathcal{R}$ alors

$$g(\bar{x}) = 0$$

Conditions nécessaire du second ordre

Les conditions nécessaire du second ordre utilisent non seulement les dérivées premières mais aussi les secondes dérivées, équivalent, le gradient et le Hessien.

Soit d est une direction arbitraire au point x.
Rappelons que,la forme quadratique $d^T H(x)d$ est dite **D.P**, **S.D.P**, **S.D.N** et **D.N** si $d^T H(x)d > 0, \geq 0, \leq 0$ et < 0 respectivement, pour tout $d \neq 0$ au point x. Si $d^T H(x)d$ admet des valeurs positives et négatives est dite **indéfinie**.

Conditions nécessaire du second ordre

Théorème

1. Si $f(x) \in \mathcal{C}^{\,2}$ et $\bar{x}$ est un minimiseur local, alors pour toute direction admissible d au point $\bar{x}$,
   1. $g(\bar{x})^T d \geq 0$
   2. Si $g(\bar{x})^T d = 0$, alors $d^T H(\bar{x}) d \geq 0$
2. Si $\bar{x}$ est un minimiseur local à l'intérieur de $\mathcal{R}$ alors,
   1. $g(\bar{x}) = 0$
   2. $d^T H(x) d \geq 0$, $\forall d \neq 0$.

Exemple 1. Soit $\bar{x} = \left[\frac{1}{2} \;\; 0\right]^T$ un minimiseur local du problème

$$\min f(x) = x_1^2 - x_1 + x_2 + x_1 x_2$$

$$\text{sc} : x_1 \geq 0, x_2 \geq 0$$

Montrer que les C.N du second ordre sont vérifiées.

Conditions nécessaire du second ordre

Solution. Les dérivées partielles premières de f sont :

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2 - 1,\ \frac{\partial f}{\partial x_2} = x_1 + 1$$

Si $d = [d_1\ d_2]^T$ est une direction admissible, on obtient :

$$g(x)^T d = (2x_1 + x_2 - 1)d_1 + (x_1 + 1)d_2$$

au point $x = \bar{x}$

$$g(\bar{x})^T d = \frac{3}{2} d_2$$

et pour $d_2 \geq 0$, on a :

$$g(\bar{x})^T d \geq 0$$

Ainsi, les C.N du premier ordre sont satisfaites.

Conditions nécessaire du second ordre

Maintenant, si $d_2 = 0$,

$$g(\bar{x})^T d = 0$$

le Hessien est :

$$H(\bar{x}) = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$$

et

$$d^T H(\bar{x}) d = 2d_1^2 \geq 0$$

Pour toute valeur de d_1. Ainsi ; les C.N du second ordre sont satisfaites.

Conditions nécessaire du second ordre

Exemple 2. Les points $p_1 = [0\ 0]^T$ et $p_2 = [6\ 9]^T$ sont des minimiseurs probables du problème :

$$\min f(x) = x_1^3 - x_1^2 x_2 + 2x_2^2$$

$$\text{sc} : x_1 \geq 0, x_2 \geq 0$$

Solution. Les dérivées partielles premières sont :

$$\frac{\partial f}{\partial x_1} = 3x_1^2 - 2x_1 x_2,\ \frac{\partial f}{\partial x_2} = 4x_2 - x_1^2$$

Si $d = [d_1\ d_2]^T$ est une direction admissible, on obtient :

$$g(x)^T d = (3x_1^2 - 2x_1 x_2)d_1 + (4x_2 - x_1^2)d_2$$

aux points p_1 et p_1

$$g(p_i)^T d = 0$$

Ainsi, les C.N du premier ordre sont satisfaites.

Conditions nécessaire du second ordre

Le Hessien

$$H(x) = \begin{pmatrix} 6x_1 - 2x_2 & -2x_1 \\ -2x_1 & 4 \end{pmatrix}$$

si $x = p_1$ alors

$$H(p_1) = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$$

Donc

$$d^T H(p_1) d = 4d_2^2 \geq 0$$

Ainsi, les C.N du second ordre au point p_1 sont satisfaites.

Conditions nécessaire du second ordre

si $x = p_2$ alors

$$H(p_2) = \begin{pmatrix} 18 & -12 \\ -12 & 4 \end{pmatrix}$$

Donc

$$d^T H(p_2) d = 18d_1^2 - 24d_1 d_2 + 4d_2^2$$

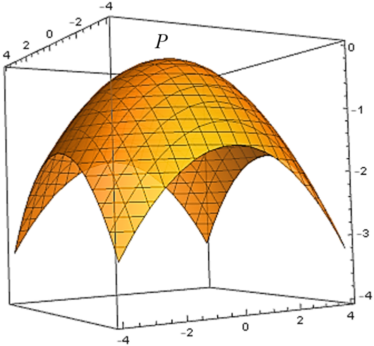
qui n'est pas définie. Ainsi, les C.N du second ordre au point p_2 ne sont pas satisfaites.

Classification des points stationnaires

Si les points extremums appelés minimiseurs et maximiseurs, sont localisés à l'intérieur de la région admissible, ils sont appelés **points stationnaires** lorsque $g(x) = 0$ en ces points.
Un autre type de points stationnaires est le **point selle** ou **point col**.
Illustrons toutes les définitions précédentes dans le cas d'une fonction à deux variables.

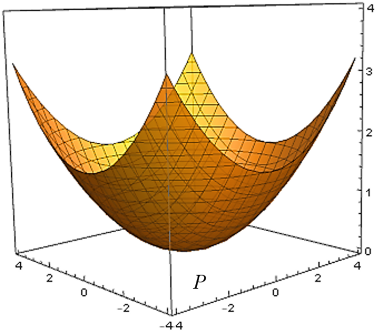
Classification des points stationnaires

Pour ce type de fonction, $z = f(x, y)$, le graphe est une surface dans l'espace à trois dimensions. Une telle fonction présente un maximum au point $P(x_0; y_0; f(x_0, y_0))$, si $f(x_0, y_0)$ atteint une valeur supérieure à toutes celles que prend $f(x, y)$ au voisinage de $x = x_0$ et $y = y_0$, comme indiqué sur la figure.



Classification des points stationnaires

De même, $f(x, y)$ possède un minimum au point $P(x_0; y_0; f(x_0, y_0))$, si $f(x_0, y_0)$ atteint une valeur inférieure à toutes celles que prend $f(x, y)$ au voisinage de $x = x_0$ et $y = y_0$; ce cas est illustré par la figure.



Classification des points stationnaires

Il en résulte qu'au point $P(x_0; y_0; f(x_0, y_0))$, il existe un plan tangent horizontal. Ce plan tangent est engendré par deux tangentes, elles-mêmes déterminées par :

$$\frac{\partial f}{\partial x} \; et \; \frac{\partial f}{\partial y}$$

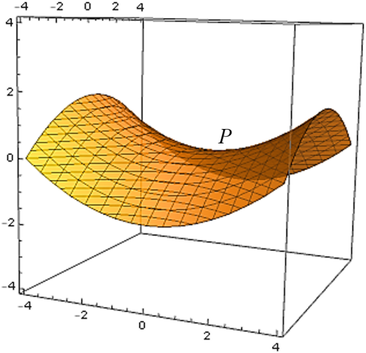
Ainsi, la condition nécessaire à l'existence d'un extremum est la suivante

$$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (0, 0)$$

Cette condition est nécessaire mais pas suffisante. En effet, il existe des fonctions pour lesquelles $\dfrac{\partial f}{\partial x} = \dfrac{\partial f}{\partial y} = 0$ sans qu'il existe un extremum en ce point. Dans ce cas, on parle de point-selle.

Classification des points stationnaires

Bien que les deux tangentes soient horizontales, il est toujours possible de trouver un point situé au-dessus du point-selle et un autre au-dessous, ceci quelque soit le voisinage du point-selle considéré. Notons encore, qu'en un point-selle la fonction présente un minimum pour l'une des variables et un maximum pour l'autre variable.



Classification des points stationnaires

Il faut donc remplir une condition suffisante qui est la suivante :

$$D = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 > 0$$

Ainsi on obtient le résultat suivant :

Résultat. Soit $P(x_0, y_0, f(x_0, y_0))$ le point en lequel :

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$$

Alors si en ce point :

1. $\frac{\partial^2 f}{\partial x^2} > 0$ et $D > 0$, f possède un minimum au point P.
2. $\frac{\partial^2 f}{\partial x^2} < 0$ et $D > 0$, f possède un maximum au point P.
3. Si $D < 0$, f ne possède ni minimum ni maximum au point P, mais un point-selle.
4. Si $D = 0$, on ne peut rien conclure.

Classification des points stationnaires

Exemple 1. Soit la fonction $z = f(x_1, x_2) = -x_1^2 - x_2^2$. Les points stationnaires s'obtiennent en résolvant le système d'équations :

$$\begin{cases} \dfrac{\partial f}{\partial x_1} &= 0 \\ \\ \dfrac{\partial f}{\partial x_2} &= 0 \end{cases}$$

c'est-à-dire :

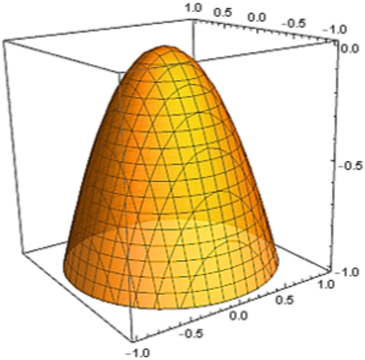
$$\begin{cases} -2x_1 = & 0 \\ -2x_2 = & 0 \end{cases} \Longleftrightarrow \begin{cases} x_1 = & 0 \\ x_2 = & 0 \end{cases}$$

Donc, il existe un point stationnaire qui $P = (0, 0, 0)$. Pour connaître la nature de ce point, il suffit d'appliquer le résultat précédent.

Ainsi on trouve :

$$\frac{\partial^2 f}{\partial x_1^2} = -2 < 0 \; et \; D = 4 > 0$$

Donc, $P = (0, 0, 0)$ est un maximum, comme le montre le graphe de la fonction f sur la figure.



Exemple 2. Soit $z = f(x_1, x_2) = 4x_1^2 - x_1x_2 + x_2^2 - x_1^3$
On résout le système suivant :

$$\begin{cases} 8x_1 - x_2 - 3x_1^2 & = 0 \\ -x_1 + 2x_2 & = 0 \end{cases}$$

Ainsi, on obtient deux points stationnaires : $P_1 = (0, 0, 0)$ et $P_2 = \left(\frac{5}{2}, \frac{5}{4}, \frac{125}{16}\right)$

1. Pour le point $P_1 = (0, 0, 0)$, nous avons :

$$\frac{\partial^2 f}{\partial x_1^2} = 8 > 0 \ et \ D = 4 > 0$$

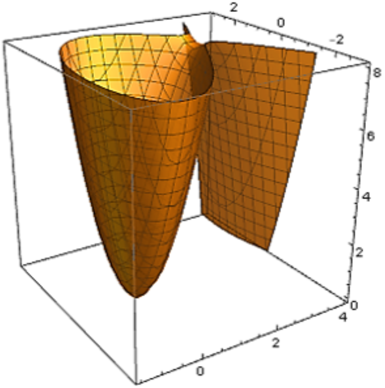
Donc le point $P_1 = (0, 0, 0)$ est un minimum.

2. Pour le point $P_2 = \left(\frac{5}{2}, \frac{5}{4}, \frac{125}{16}\right)$, nous avons :

$$\frac{\partial^2 f}{\partial x_1^2} = -7 < 0 \ et \ D = -15 < 0$$

Classification des points stationnaires

Il ne s'agit ni d'un minimum ni d'un maximum, mais d'un point-selle de la fonction. Comme le montre la figure.



Le point selle : Soit $x = \bar{x} + \alpha d \in \mathcal{R}$ au voisinage du point selle $\bar{x}$, la série de Taylor est donnée

$$f(x) = f(\bar{x}) + \frac{1}{2}\alpha^2 d^T H(\bar{x}) d + o(\alpha^2 \|d\|^2)$$

puisque $g(\bar{x}) = 0$. D'après la définition d'un point selle, il existe deux directions d_1 et d_2 telles que

$$f(\bar{x} + \alpha d_1) < f(\bar{x}) \text{ et } f(\bar{x} + \alpha d_2) > f(\bar{x})$$

Ainsi, on a

$$d_1^T H(\bar{x}) d_1 < 0 \text{ et } d_2^T H(\bar{x}) d_2 > 0$$

Du coup la matrice $H(\bar{x})$ doit être indéfinie.

Classification des points stationnaires

Exemple 3. Trouver la nature des points stationnaires de

$$f(x) = (x_1 - 2)^3 + (x_2 - 3)^3$$

Solution. Les dérivées partielles premières de $f(x)$ sont

$$\frac{\partial f}{\partial x_1} = 3(x_1 - 2)^2 = 0 \text{ et } \frac{\partial f}{\partial x_2} = 3(x_2 - 3)^2 = 0$$

Ainsi, $\bar{x} = (2, 3)$. Donc le Hessien est donné par

$$H(x) = \begin{bmatrix} 6(x_1 - 2) & 0 \\ 0 & 6(x_2 - 3) \end{bmatrix}$$

En remplaçant, $\bar{x}$ en $H(x)$, on obtient $H(\bar{x}) = 0$.

Classification des points stationnaires

H est semi-définie, donc on passe aux troisièmes dérivées.
Toutes les dérivées troisièmes sont nulles sauf $\dfrac{\partial^3 f}{\partial x_1^3}$ est $\dfrac{\partial^3 f}{\partial x_2^3}$ qu'elles sont égal toute les deux à 6.
Pour le point $\bar{x}+\delta$, le quatrième terme de la série de Taylor est donnée par

$$\frac{1}{3!}\left(\delta_1^3\frac{\partial^3 f}{\partial x_1^3}+\delta_2^3\frac{\partial^3 f}{\partial x_2^3}\right)=\delta_1^3+\delta_2^3$$

qui est positive pour $\delta_1,\delta_2>0$ et négative pour $\delta_1,\delta_2<0$. Ainsi

$$f(\bar{x}+\delta)>f(\bar{x}) \text{ pour } \delta_1,\delta_2>0$$

et

$$f(\bar{x}+\delta)<f(\bar{x}) \text{ pour } \delta_1,\delta_2<0$$

Donc $\bar{x}$ n'est ni un minimum, ni un maximum, c'est un point selle.